Fluctuations and Scaling in Biology

T. Vicsek
editor

2001

Contents

Chapter 1

Basic concepts

The complexity of biological systems is manifested in many ways. Here we shall consider those aspects of life which involve random fluctuations and a hierarchical underlying structure resulting in a power law dependence of the various quantities characterising these systems. As will be shown, these two features – fluctuations and scaling – are frequently and intimately related, although in some cases they appear independently of each other.

Are fluctuations an important, inherent ingredient of life? What is their origin and impact? These are important questions which have to be addressed before we lead the reader to the more advanced parts of the text.

It is almost a trivial statement that nearly all processes in biology involve randomness beyond a negligible level. On the other hand, except some artificial situations, all phenomena in nature have the elements of stochasticity in them. There is no such a thing as a noiseless or absolutely deterministic system and this is true from the behaviour of galaxies down to elementary particles. In the intermediate range the presence of temperature fluctuations, the various nonlinearities/instabilities represent the main source of randomness. This is known to be so and is understood in many cases for non-living systems.

Biological systems are not exceptions from the same rule. Perhaps the most typical feature of a biological object or signal is that they simultaneously possess some specific pattern and deviations around these patterns. If we take as an example an animal, let us say a dog, we can consider the following illustrations of the above statement: i) no two dogs are exactly the same, but they are similar in many ways, ii) they have typical reactions, but never react in a perfectly identical way, iii) their heart beats almost regularly, but a closer analysis shows specific fluctuations around the average heart beat rate, iv) if we study the electric signals (with the help of an electrode) of their brain we see an almost randomly looking series of spikes, v)

if we happened to see the motion of the individual cells or even large molecules in their body (e.g., blood cells, sperms, RNA molecules) we would observe an erratic behaviour around an average tendency. The list of such examples could be continued for long.

As will be discussed later, in many cases the above mentioned stochastic changes or fluctuations are not completely random and can be associated with power laws or scaling.

In short, if a system is made of many interacting units, specific statistical features involving fluctuations and scaling emerge.

There are a few important aspects of fluctuations when the origin of an apparently random behaviour is relatively well understood.

I)

Microscopic objects are subject to the so called thermal fluctuations. It is a fundamental feature of all systems that if they have a well defined temperature T, than each microscopic particle (atom or molecule) in them has in average an amount of kinetic energy $\frac{1}{2}kT$ (k is the Boltzmann constant) for each degree of freedom (mode of motion). A more complex molecule has many degrees of freedom, and is also subject to "kicks" from the neighbouring molecules moving with a velocity corresponding to their kinetic energy. As a result, an individual molecule interacting with many others follows a more or less random trajectory due to the many subsequent, randomly occurring "kicks" from their neighbours. It is a beautiful subject to understand how such fluctuations are tamed by the specific processes (diffusion, enzymatic interactions) inside living cells and result in an organised (with some fluctuations) behaviour.

II)

When many similar, but not necessarily microscopic objects (biological or non-living) are present in a system there are further reasons to consider the random aspect of the behaviour.

In general, if many moving objects interact, the motion of the individual objects is bound to be random-like (like in the above mentioned microscopic case), even if the motion or the interaction is deterministic and we consider macroscopic objects. The trajectories are subject to many changes of varying degree and instead of looking at such processes as deterministic it is conceptually more useful to assume that the motion is stochastic and it is the statistical features of the assembly of objects which should be determined when describing the system. It is not only the direction of motion which can be changed during interaction. For objects with a directedness their direction can also be modified due to some direction-dependent interaction.

Now, as a result of these interactions, various spontaneous processes may take in the system: for attractive forces, aggregates are likely to form, or groups with the same directedness of their members may appear. Schools of fish is a common

example for aggregation. However, due to the complex nature of a system with many objects moving both randomly, and orderly (in a group, but with some perturbations) another kind of fluctuations, a random distribution of group sizes is produced. Many times the distribution of group sizes follows a *power law*, or in other words, a *scaling distribution*.

III)

Non-linearities are known to lead to a very complex behaviour which – especially in the presence of thermal fluctuations – can be considered as random. A common spatial example is the formation of branching patterns under conditions leading to the unstable growth of more advanced branches. In these cases the smallest perturbations are likely to result in a new, quickly growing side-branch and the structure attains the well known branching morphology so common in biology (e.g., trees or blood vessel networks). Although such networks possess some specific features typical for the given biological object, they are also irregular. In addition, as will be shown they have a specific hierarchical structure best described by fractal geometry.

In the following parts of this introductory chapter I briefly discuss the basic concepts related to fluctuations and scaling in biology and summarise the most important findings obtained in the related investigations.

1.1 Fluctuations

1.1.1 Noise versus fluctuations

The origin of fluctuations can widely vary. In most of the cases, however, they are due to the above mentioned "thermal noise" or erratic motion of microscopic particles. Typically, noise is not correlated which means that the value of the fluctuating quantity F at the location $\mathbf{r}$ at time t does not depend on its value at a different location and at an earlier moment of time. Symbolically we write that the correlation function has the form of a delta function

$$c(\mathbf{r}) = \langle F(\mathbf{r},t)F(\mathbf{r}',t')\rangle - \langle F(\mathbf{r},t)\rangle\langle F(\mathbf{r},t)\rangle = C\delta(\mathbf{r}-\mathbf{r}',t-t')$$

where the δ function is equal to zero for any non-zero values of its arguments and C is some constant. The averaging (denoted by $\langle ...\rangle$ is made over all values of the arguments. This expression holds for the uncorrelated white or shot noise, while for correlated fluctuations $c(\mathbf{r})$ has more complex forms.

Fluctuations can be more complex than just white noise. Many times they represent an inherent, characteristic feature (reaction to white noise) of the system itself.

For example, due to the thermal (white) noise the fluctuations in the total magnetisation of a ferromagnet near its critical point can strongly increase and exhibit specific correlations.

Noise and fluctuations play a central role in ordering phenomena. A system of many interacting units with an interaction "trying" to force the units to behave in the same way in the presence of strong external fluctuations (noise) may not be able to order. On the other hand, as the magnitude of noise decreases (e.g., the temperature is lowered), the objects in the system may already assume their common or collective pattern of behaviour, for example, they spontaneously develop a common direction of motion, or find their right place for a crystalline structure. These noise driven transitions will be discussed briefly in the scaling part of this introductory chapter.

In the following when only their stochastic nature is relevant we use noise, fluctuations and random perturbations as synonyms. However, random changes appearing as an inherent behaviour (response) of the system itself will always be called fluctuations.

A further aspect of fluctuations involves transport processes in the presence of noise. Interestingly, even uncorrelated fluctuations (these are random changes without any tendencies) may result in a behaviour with a well defined tendency. This happens in the case of *molecular motors*, where white noise assists the motor proteins to proceed along specific intracellular tracks. On the other hand, white noise alone cannot produce currents, this would contradict the second law of thermodynamics. Here we mention the basic findings about this fascinating new direction in biological physics.

1.1.2 Molecular motors driven by noise and fluctuations

In the inorganic world transport always takes place along a macroscopic gradient of a potential (or an analogous quantity). Things fall down due to the gravitational force which can be obtained from the derivative of the gravitational potential. Even the global transport of microscopic objects such as molecules takes place along the extended gradient of the so called chemical potential. For example, particles tend to diffuse from a denser region to a less dense one making a macroscopic overall distance if the gradient of the density extends over that distance. Electrons move in a wire following the gradient of the electric potential which is larger at one end of the wire than at the other end.

This is not how transport is realized in most of the biological systems. The above mechanism tends to bring a system into a motionless state: as the objects move along the potential gradient they simultaneously decrease the value of the overall gradient. For example, difference in the concentration (driving the diffusional transport) de-

creases in time as the particles diffuse to the spots of smaller concentrations (and increase the density at these spots). Instead, life is about generating differences, building structures from a less patterned environment.

One mechanism for doing this is *motion along periodic, but locally asymmetric structures*. Imagine a sawtooth-like (or ratchet) potential: it is piece-wise linear, with two different gradients (slopes). We also have a particle in this potential, most of the time "sitting" in one of the minima ("valleys"). Now, if we pull this particle periodically back and forth in this sawtooth-like potential, the following cases are possible: i) the force is large enough for the particle to be pulled out from a minimum both to left and right ii) the force is strong enough only to pull out the particle in the direction of the smaller gradient (steepness), iii) the particle is not pulled out from the valley because the force is too weak.

Obviously, case ii) establishes a situation in which a transport is possible without any global potential difference: the particle moves in the direction of the smaller slope although the force acting on it is zero in average (acts back and forth). In this way particles can be collected at one end of a track with such a potential, thus, a concentration difference, i.e., structure can be built up.

However, life is not so simple. Such processes occur on the molecular level, where *fluctuations are very strong* for two reasons: a) The particle (called motor protein) is kicked by the other molecules in the system randomly, in a noisy, uncorrelated manner all the time, b) the periodic, deterministic back and forth driving cannot be established in a microscopic environment as well: instead, the energy supplied by the so called ATP molecules and providing the conformational changes of the molecular motor resulting in its tendency to move back and forth also arrives stochastically (the ATP molecules are "consumed" at times following a Poisson distribution).

The picture emerging is the following: A possible, simplified representation of biological motion is that of the motion of a Brownian particle in an asymmetric periodic potential. The corresponding equation have to account for the stochastic nature of the process: both for the white noise coming from the environment and for the irregular nature of the energy input. *The related Langevin equation approach is discussed in details in chapter 5.*

There is one interesting point here: in the case of motion along asymmetric periodic potentials noise may play a role *enhancing the transport*. This is a bit paradoxical, we are more used to the notion that random perturbations typically destroy tendencies. In the case of molecular motors, however, it may happen, that adding white noise results in stronger current. The easiest way to show this is the following: Imagine case iii) when the external force pulling the particle back and forth is not strong enough for the particle to escape in any of the two directions. Now, if we add noise (a randomly changing small amount of extra force) in some cases the

overall force may exceed the critical value necessary to pull out the particle from a valley. This will happen more frequently in the direction of the smaller slope, since the critical amount of force is smaller in that direction, and there will be an overall current in the direction of the smaller slope. This is why such systems are also called *thermal ratchets*.

Another interesting variant of this situation is when we consider one single force (instead of the sum of a deterministic periodic back and forth acting force and an uncorrelated white noise one) changing stochastically. Now, if this single fluctuating force is completely uncorrelated, or in other words, thermal or white noise, then no global transport is possible. If the fluctuations are thermal noise-like, than the system is in equilibrium and no transport is possible in thermal equilibrium.

On the other hand, if the fluctuating force (noise) *is correlated*, transport already becomes possible. The probability to leave in the less steep direction will be still larger than in the opposite direction. In this way, our ratchet "rectifies" the fluctuations, *it is able to make use of its non-white part.*

1.2 Scaling

A quantity F *scales* as a function of its argument x if changing the argument by a factor (e.g., changing x for Ax) does not change the form of the functional dependence of F on x (apart from a constant factor). This is trivially so for a function of the form of a power law, but is not, as a rule, true for other functions. For example, $F = x^2$ scales because $F' = F(Ax) = A^2x^2 = A^2F(x)$, while $F = \log(x + B)$ does not scale because in this case $F' = F(Ax + B)$ cannot be reduced to a form containing $F(x)$ and a constant factor only.

The scaling quantities we shall consider are mostly of stochastic nature. Thus, the specific functional dependences will be valid for the average of the given quantities as a function of their arguments. Each realization of some stochastic process has a fluctuating outcome, but making an average over several processes, or over a single process having sequences fluctuating around an average can provide the proper estimate of the quantity of interest.

1.2.1 Critical behaviour

Perhaps the most typical collective phenomena exhibited by an assembly of many interacting particles are the so called *phase transitions*, when, as a function of an external parameter (like temperature), the particles collectively change their behaviour.

For example, during freezing all of the molecules of a fluid move to a specific position so that the resulting structure becomes a crystal with regular microscopic structure.

In the vicinity of such transitions interesting spatial and temporal correlations can be observed in the systems and these features will be relevant from the point of the majority of the topics discussed in this book.

In particular, during second order phase transitions the so called "critical state" (or critical phenomena) can be observed in which the ordinarily exponential functional dependences are replaced by algebraic (power law) dependence of the relevant quantities on their parameters. A *power law dependence* of the quantity $n(s)$ (e.g., the number of schools containing s fish) is of the following form $n(s) \sim s^{-\tau}$, where $\sim$ expresses proportionality, and τ is some exponent. The power law dependence is very special: for example, a power law decay of the number of clusters (schools of fish) as a function of their size (number of fish in a school) means that very large clusters may occur with a probability which is not negligible (this probability would be extremely small if the number of clusters would decrease exponentially with growing cluster size, as it does for regular states). If a quantity changes according to a power law when the parameter it depends on is growing linearly, we say it *scales*, and the corresponding exponent is called *critical exponent*.

Why are such states called critical? Because they are extremely sensitive to small changes or perturbations. If a human being is in a critical state it means his or her state can get worse very easily. In the case of lattice models a small change in the temperature may lead to the quick collapse or birth of very large clusters. During the last two decades statistical physicists have worked out delicate theories and methods to interpret the behaviour of such critical transitions and states and in the following we shall consider the application of the related concepts to biological phenomena involving many similar units.

The important point is that scaling typically involves *universality*: instead of particles we can imagine similar organisms. If the interaction among these organisms is relatively simple, and is analogous to those which produce scaling or phase transitions in non-living systems, than we can expect the same type of behaviour in such systems of living entities as well.

1.2.2 Scaling of event sizes: Avalanches

As mentioned above, scaling can be observed during equilibrium phase transition, but in the following we shall argue that a power law dependence of the various important quantities can emerge in the non-equilibrium world (of life) as well. In fact, it is in the *non-equilibrium state* when structures can emerge spontaneously from an originally homogeneous medium.

A particular and important departure from equilibrium is when the system is "slowly driven" to a stationary state. Slow driving may mean the gradual addition of some quantity (energy) to a a system which may also loose this energy due to interactions. If the interaction between two parts of the system is such that a change exceeding a critical value of the given quantity in one unit results in a similar exceeding of the same critical value in the neighbouring unit, than *large, avalanche-like series of changes* may take place in the system when it is close to a (critical, balanced) state. As this state is both spontaneously emerging and critical, the associated phenomenon is called In this state the system is very sensitive to fluctuations, since a small perturbation may lead to a large avalanche. In this sense the system is in a critical state. Not all of the avalanches are large, the majority of them are small, but the probability of having a large avalanche does not go to zero very quickly with the avalanche size. In many such slowly driven systems scaling (power law) of the avalanche size distribution can be observed. Avalanches are sometimes very large (as we know from the news on skiing areas, but in this context an earthquake is also an avalanche) and they are the so called big events in the theory of slowly driven systems.

The simplest example is that of a *sandpile*. Imagine that we add grains of sand to a growing pile: as the grains are dropped, most of the time the surface of the pile becomes only slightly rearranged, however, time to time the new grain triggers a long series of events: grains rolling down the slope drag many more grains with them. A *simple model* along a line segment would contain columns of particles. A new particle is dropped at a random position. If the height difference between two columns becomes larger than 2, than from the higher column two particles are removed and added to the two neighbouring column. Particles at the edge of the segment are dropped out from the system completely.

In this book two examples for biological SOC are discussed. The structure of the lung is such that it can be brought into analogy with the sandpile model. The air entering the lung has to go through a sequence of airways each opening if the pressure exceeds a critical value. By forcing the air gradually enter the experimentally investigated lung, large jumps in the terminal airway resistance have been observed. These jumps corresponded to avalanches: to the subsequent opening of a large set of airways in a short time. The distribution of jumps followed a power law. The observation of strongly fluctuating extinction rates and the corresponding SOC related theory is also discussed in chapter 3.

1.2.3 Scaling of patterns and sequences: Fractals

Nature is full of beautiful complex shapes which are far more intricate than the idealised forms proposed by Euclid more than two thousand years ago. This is particularly true for the living world where complicated structures are generated during embriogenesis. Many of these patterns are random branching networks; examples include trees, the network of blood vessels, airways in the lung, neural nets, etc. These highly hierarchical patterns can be best interpreted in terms of fractal geometry.

Fractals are fascinating geometrical objects characterised by a non-trivial fractional dimension. Imagine a growing pattern whose mass M (the number of particles it contains) increases slower than the d-th power of its radius R, where d is the dimension of the space in which the pattern is developing. This is clearly different from the case of homogeneous structures that we are used to. For fractals $M(R) \sim R^D$, where D is called the fractal dimension since in many cases it is *not an integer*, but a fractional number less than d. If the above relation is true for a pattern, it is bound to be *self-similar* in the sense that a small part of it looks the same as the whole structure after it is expanded isotropically. For deterministic mathematical fractals the blown up piece looks exactly the same as the whole object. For random patterns self-similarity is satisfied in a stochastic manner. The fractal dimension can also be defined through the expression $c(r) = \frac{1}{N}\sum_{\mathbf{r}'} \rho(\mathbf{r})\rho(\mathbf{r}+\mathbf{r}') \sim r^{D-d}$ where $c(r)$ describes the density-density correlations within the pattern and $\rho(\mathbf{r})$ is equal to unity if there is a particle at the position $\mathbf{r}$, and it is equal to zero otherwise. For isotropic structures the correlation function $c(r)$ is equivalent to the probability that one finds a particle belonging to the cluster at a distance $r = |\mathbf{r} - \mathbf{r}'|$ from a fixed point on the cluster. In this case an averaging can be made over the directions as well.

The meaning of the above statements is that fractals can be looked at as structures exhibiting *scaling in space* since their mass as the function of size or their density as a function of distance behave as a power law.

Self-affine structures represent another type of fractals. For such objects a small part of the fractal must be enlarged in an anisotropic way to match the entire pattern. For example, if the fractal is embedded into two dimensions, for self-affine fractals one achieves matching by rescaling the size horizontally and vertically by different factors.

Fractal Bacterial Colonies

Perhaps the best defined biological systems exhibiting fractal growth are bacterial colonies. By a careful control of the experimental conditions it has been possible to obtain well reproducible results on the development of complex branching patterns made of many millions of bacteria as they multiply on the surface of an agar (gel) layer in a Petri dish. The related beautiful results are discussed in chapter 4.

Typically, bacterial colonies are grown on substrates with a high nutrient level and intermediate agar concentration. Under such "friendly" conditions, the colonies develop simple (almost structureless) compact patterns with relatively smooth envelope. This behaviour fits well the contemporary view of the bacterial colonies as a collection of independent unicellular organisms. However, in nature, bacterial colonies regularly must cope with hostile environmental conditions. What happens if we create hostile conditions in a Petri dish by using, for example, a very low level of nutrients or a hard surface (high concentration of agar), or both? The bacterial reproduction rate, which determines the growth rate of the colony, is limited by the level of nutrients concentration available for the bacteria. The latter is limited by the diffusion of nutrients towards the colony. Hence, the growth of the colony resembles *diffusion-limited growth* in inorganic systems leading to fractal patterns.

Diffusion-limited growth leads to random branching patterns because of the following *instability*: if a given part of the growing surface is a bit more advanced than the surrounding region, this part will advance faster than the neighbouring parts of the colony, because it will be closer to the source of the nutrient diffusing from the outer regions of the Petri dish. In turn, parts lagging behind tend not to grow any more since in those regions no nutrient will be available as the nutrient diffusing towards the colony will be consumed by the advanced parts of the colony. This is positive (negative) feedback: a protrusion grows faster (and produces a branch) the screened fjords stop to grow completely. The resulting pattern has a radially growing tree-like structure.

In reality the situation is somewhat more complex since the bacteria can communicate through *chemotaxis*. They are able to pass on information about their environment and increase/decrease the growth rate at other points in the colony. The communication enables each bacterium to be both actor and spectator (using Bohr's expressions) during the complex patterning. The bacteria developed a particle-field duality: each of the bacterium is a localised (moving) particle which can produce a chemical and physical field around itself. For researchers in the pattern formation field, the above communication regulation and control mechanism opens a new class of tantalising complex models exhibiting a much richer spectrum of patterns than the models for inorganic systems.

All this can be investigated by constructing suitable *computer models*. In the corresponding calculations a number of factors are taken into account (see chapter 4): good agreement with the experimental observations can be achieved by assuming a nutrient dependent multiplication rate, diffusional motion of the bacteria on the agar surface, chemotactic signalling, etc. These simulations are different from the common approaches in physics and biology. Physicists prefer to build simple models ignoring many of the details and look for universal behaviour. Biologists mostly

use specific models reflecting the biological details of the system under investigation. The models used to mimic bacterial colony development in the computer interpolate between these approaches and are *aiming at finding universal behaviour taking into account most of the biologically relevant details.*

Correlations in the genetic code

One possible representation of the vast information stored in the extremely long sequences of DNA data is a random walk built to correspond to such sequences. In this approach DNA is mapped onto a process which can be regarded as a walk: each of the four "letters" of a DNA sequence is identified with a step in a given direction. Then, the specific features of this walk can be analysed using methods borrowed from statistical physics.

Given the walk one can look for correlations. Two series of data (X and Y) are correlated if there is a relationship between the corresponding elements of the series. When searching for correlations within a single sequence of data we can ask how the value X_i is related to the value X_{i+j}. By comparing the two values with the average of X one can get information about the question whether two values in the data set separated by j elements are correlated.

An ordinary random walk has no long-range correlations. One of the most relevant questions one can raise in the context of DNA sequences is the location of coding and non-coding parts in the genome. In the case these two kinds of sub-sequences have different kinds of correlations we may be able to differentiate between coding and non-coding part without any prior knowledge about the sequences. Indeed, it has been shown that the random walks corresponding to *non-coding parts have long-range correlations* in contrast to the coding parts (which have short-range correlations only).

As an alternative to the DNA walk, the symbol sequence corresponding to a DNA molecule can be regarded as a written text composed by using four letters. Since we do not know the "language" of the text we have to to apply methods developed for analysing written (natural) texts of unknown origin. In particular, one can ask the question whether two texts were written in the same language or not. We expect larger correlations between text of the same origin (language). Here two sequences are correlated if the scalar product of the two appropriately defined vectors (corresponding to them) has a value different from that it would have for two uncorrelated sequences. In the first approximation his method is language insensitive.

When applying the vector space technique to DNA sequences in a way we look at DNA as an encoded text written in an unknown language, still, we expect to locate correlations between parts of the sequences due to similarities in their underlying structure. In case the language of the coding parts is different from the non-coding ones we get a higher value for the corresponding scalar product (as it has been demonstrated in some cases).

1.2.4 Scaling in group motion: Flocks

Group motion (flocking) is a beautiful phenomenon many times capturing our eyes. Here flocking is understood in a general sense of the word, including herding of quadrupeds, schooling of fish, etc). In the last chapter of this book we address the question whether there are some global, perhaps universal transitions in flocking when many organisms are involved and such parameters as the level of perturbations or the mean distance of the organisms is changed.

Everyone has experienced how an initially randomly directed group of birds feeding on the ground *is spontaneously ordered into a well organised flock* when they leave because of some external perturbation. This ordering is a highly non-trivial question since in a huge flock of several hundred or thousand birds usually there is no "leader" bird (we do not consider here the V shaped or other structured flight of some large bodied birds) and not even all birds can visually interact. Still, the whole flock selects a well defined direction. Such ordering is familiar from equilibrium phase transition of magnetic systems and the corresponding findings may provide clues to the understanding of the more complex far-from equilibrium ordering of moving organisms.

Self-propulsion is an essential feature of most living systems. In addition, the motion of the organisms is usually controlled by interactions with other organisms in their neighbourhood and randomness plays an important role as well. It is possible to design simple computer models which simulate the collective motion and take into account the most relevant ingredients of the phenomenon.

A *simple model of collective motion* consists of particles moving in one, two or three dimensions. The particles are characterised by their (off-lattice) location $\mathbf{x}_i$ and velocity $\mathbf{v}_i$ pointing in the direction ϑ_i. To account for the self-propelled nature of the particles the magnitude of the velocity is fixed to v_0. A simple local interaction is defined in the model: at each time step a given particle assumes the average direction of motion of the particles in its local neighbourhood with some uncertainty. Such a model is a transport related, non-equilibrium analogue of the ferromagnetic models. The analogy is as follows: the function tending to align the spins in the same direction in the case of equilibrium ferromagnets is replaced by the rule of aligning the direction of motion of particles, and the amplitude of the random perturbations can be considered proportional to the temperature.

In addition, collective motion can be described by continuum equations as well. The collection of "birds" is then looked at as particles in a fluid subject to fluctuations and satisfying the condition of trying to move with a given velocity.

Both theoretical approaches led to the conclusion that there are interesting, in

cases unexpected (compared to equilibrium systems) transitions in collective motion. For example, if the noise (level of perturbations, corresponding to temperature in the case of ferromagnets) is decreased, the *originally disordered flock becomes ordered in analogy with second order phase transitions*. The level of global order, the fluctuations around this order and several related quantities all scale, i.e., behave according to power law as a function of the distance from a critical level of perturbations.

Pedestrian simulations

A special kind of flocks is a group of people. In the last chapter interesting applications of pedestrian simulations are discussed. Just as in the case of other organisms, people can be represented by particles "dressed" by the appropriate interactions. Simulations of humans moving in confined places leads to a number of interesting effects reproducing related observations.

Freezing by heating is an effect observed when a particles are driven in opposite directions. The related simulations demonstrated that more nervous or hectic changes (heating) of the direction of motion can cause a breakdown of an efficient pattern of cooperative interactions and finally produce a deadlock (freezing). In particular, this may be relevant for panicking pedestrians in a smoky environment, who tend to build up fatal blockings. The system described in the last chapter consists of a mesoscopic number of driven particles with repulsive hard-core interactions moving into opposite directions under the influence of fluctuations. Example for such systems is pedestrians walking in a passage.

In short, "freezing by heating" means a transition from a fluid state (with self-organised lanes of uniform direction of motion to a solid, crystallised ("frozen") state just by increasing the noise amplitude ("temperature"). This is in contrast to, for example, melting (where increasing the temperature increases the energy and order is destroyed) and to noise-induced ordering in glasses or granular media, where increasing the temperature drives the system from a disordered metastable state (corresponding to a local energy minimum) to an ordered stable state (corresponding to the global energy minimum). Instead, "freezing by heating" shows an increase in the order at increasing temperature, although the total energy increases at the same time. The crystallised state can also be destroyed by *ongoing* fluctuations with extreme noise amplitudes giving rise to a third, disordered ("gaseous") state with randomly distributed particles. Thus, with increasing "temperature" θ, we have the atypical sequence of transitions *fluid* $\longrightarrow$ *solid* $\longrightarrow$ *gaseous*.

Further variants of pedestrian simulations allow *the quantitative investigation of trail formation, optimisation of passage geometries,* etc.

In this chapter I have attempted to present in a simplified manner a *selection of concepts, topics and results discussed in much more detail in the main body of the book*. For details necessary for a deeper understanding of the concepts and findings

related to fluctuations and scaling in biology I advise the reader to read other parts of our book as well (where the related references are also given).

Chapter 2

Introduction to complex patterns, fluctuations and scaling

In this book we mostly consider models of reality, since "reality" in the case of biology, is far too complex to allow complete theoretical treatment. On the other hand, whenever it is possible, we are trying to present models as realistic as possible in order to reflect the essential features of the specific phenomenon occurring in nature. In many cases we are concerned with *systems consisting of many similar objects* and this feature has particular implications on the kinds of models we consider.

Various models allowing exact or numerical treatment have been playing an important role in the studies of biological processes. Because of the complexity of the phenomena it is usually a difficult task to decide which of the many factors influencing the processes are the most significant. In a real system the number of all possible factors can be too large; this number is decreased to a few by appropriate model systems. Thus, the investigation of these models provides a possibility to detect the most relevant factors, and demonstrate their effects in the absence of any disturbance.

Systems consisting of many similar units can be successfully described in terms of *particles*, where the word particle can stand for a molecule as well as for more complex objects, including organisms. Then, the particular nature of the model is given by the features of the individual particles and by the ways these particles interact with each other.

The spatial arrangement of these particles is frequently of major interest. Structures consisting of connected particles are usually called *clusters* or *aggregates*. In most of the cases the particles are assumed to "exist" on a *lattice* for computational convenience, and two particles are regarded as connected if they occupy nearest neighbour sites of the lattice. However, for studying universality and related questions, *off-lattice* or further neighbour versions of clustering processes can also be investi-

gated. A lattice site with a particle assigned to it is called occupied or filled. An important additional feature included into the majority of models to be described is stochasticity which is typical for biological phenomena.

In this chapter we discuss the basic features of the complex patters produced by a wide variety of biological growth processes. In many cases biological growth leads to random fractal structures characterised by a non-integer dimension defined below. In the second section of this chapter the principles behind the motion in the presence of fluctuations will be presented. Since biological motion is produced by motor molecules acting on a microscopic scale, thermal noise and other stochastic perturbations are essential. Finally, we discuss continuous phase transitions where the concept of scaling plays a central role. Scaling means a power law dependence of a quantity on its argument and, as will be demonstrated, is a feature showing up in an unexpectedly large selection of biological systems. We know from the early studies of scaling in physics that it is a fundamental characteristics of a system. The power law dependence of a quantity usually involves a similar behaviour of the other quantities in a system; in addition, the exponent (corresponding to the power law) is typically not sensitive to the details universal) of the processes considered.

2.1 Fractal geometry

During the last decade it has widely been recognised by researchers working in diverse areas of science that many of the structures commonly observed possess a rather special kind of geometrical complexity. This awareness is largely due to the activity of Benoit Mandelbrot [1] who called attention to the particular geometrical properties of such objects as the shore of continents, the branches of trees, or the surface of clouds. He coined the name *fractal* for these complex shapes to express that they can be characterised by a *non-integer* (fractal) *dimensionality*. With the development of research in this direction the list of examples of fractals has become very long, and includes structures from microscopic aggregates to the clusters of galaxies. Objects of biological origin are many times fractal-like.

Before starting a more detailed description of fractal geometry let us first consider a simple example. Fig. 2.1 shows a cluster of particles which can be used for demonstrating the main features of fractals. This object was proposed to describe diffusion-limited growth [2] and has a loopless branching structure reminiscent of many shapes of biological origin. Imagine concentric circles of radii R centered at the middle of the cluster. For such an object it can be shown that the number of particles in a circle of radius R scales as

$$N(R) \sim R^D, \tag{2.1}$$

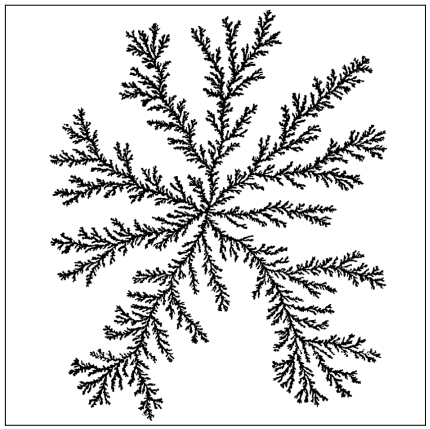


Figure 2.1: A typical stochastic fractal generated in a computer using the diffusion-limited aggregation model.

where $D < d$ is a non-integer number called the *fractal dimension*. Naturally, for a real object the above scaling holds only for length scales between a lower and an upper cutoff. Obviously, for a regular object embedded into a d dimensional Euclidean space Eq. 2.1 would have the form $N(R) \sim R^d$ expressing the fact that the volume of a d dimensional object grows with its linear size R as R^d. Clusters having a non-trivial D are typically *self-similar. This property means that a larger part of the cluster after being reduced "looks the same" as a smaller part of the cluster before reduction.* This remarkable feature of fractals can be visually examined on Fig. 2.1, where parts of different sizes (included into rectangular boxes) can be compared from this point of view.

2.1.1 Fractals as mathematical and biological objects

In addition to self-similarity mentioned above, a characteristic property of fractals is related to their volume with respect to their linear size. To demonstrate this we first need to introduce a few notions. We call *embedding dimension* the Euclidean dimension d of the space the fractal can be embedded in. Furthermore, d has to be the smallest such dimension. Obviously, the volume of a fractal (or any object),

$V(l)$, can be measured by covering it with d dimensional balls of radius l. Then the expression

$$V(l) = N(l)l^d \tag{2.2}$$

gives an estimate of the volume, where $N(l)$ is the *number of balls needed to cover the object completely* and l is much smaller than the linear size L of the whole structure. The structure is regarded to be covered if the region occupied by the balls includes it entirely. The phrase "number of balls needed to cover" corresponds to the requirement that $N(l)$ should be the smallest number of balls with which the covering can be achieved. For ordinary objects $V(l)$ quickly attains a constant value, while for fractals typically $V(l) \to 0$ as $l \to 0$. On the other hand, the surface of fractals may be anomalously large with respect to L.

There is an alternative way to determine $N(l)$ which is equivalent to the definition given above. Consider a d-dimensional hypercubic lattice of lattice spacing l which occupies the same region of space where the object is located. Then the number of boxes (mesh units) of volume l^d which overlap with the structure can be used as a definition for $N(l)$ as well. This approach is called *box counting*.

Returning to the cluster shown in Fig. 2.1 we can say that it can be embedded into a plane ($d = 2$). Measuring the total length of its branches (corresponding to the surface in a two-dimensional space) we would find that it tends to grow almost indefinitely with the decreasing length l of the measuring sticks. At the same time, the measured "area" of the cluster (volume in $d = 2$) goes to zero if we determine it by using discs of decreasing radius. The reason for this is rooted in the extremely complicated, self-similar character of the cluster. Therefore, such a collection of branches to be definitely much "longer" than a line but having infinitely small area: it is neither a one- nor a two-dimensional object.

Thus, the volume of a finite geometrical structure measured according to Eq. 2.2 may go to zero with the decreasing size of the covering balls while, simultaneously, its measured surface may diverge following a *power law* instead of the better behaving exponential convergence. In general, *we call a physical object fractal, if measuring its volume, surface or length with d, $d-1$ etc. dimensional hyperballs it is not possible to obtain a well converging finite measure for these quantities when changing l over several orders of magnitude.*

It is possible to construct mathematical objects which satisfy the criterion of self-similarity exactly, and their measured volume depends on l even if l or (l/L) becomes smaller than any finite value. Fig. 2.2 gives examples how one can construct such fractals using an iteration procedure. Usually one starts with a simple initial configuration of units (Fig. 2.2a) or with a geometrical object (Fig. 2.2b). Then, in the growing case this simple seed configuration (Fig. 2.2a, $k = 2$) is repeatedly added

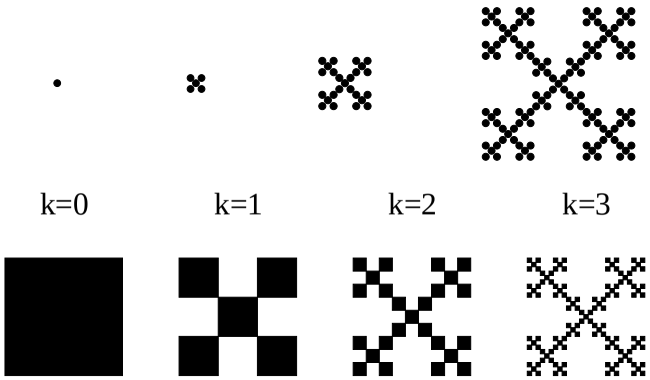


Figure 2.2: Fig. 2.2a demonstrates how one can generate a growing fractal using an iteration procedure. In Fig. 2.2b an analogous structure is constructed by subsequent divisions of the original square. Both procedures lead to fractals for $k \to \infty$ with the same dimension $D \simeq 1.465$ [3].

to itself in such a way that the seed configuration is regarded as a unit and in the new structure these units are arranged with respect to each other according to the same symmetry as the original units in the seed configuration. In the next stage the previous configuration is always looked at as the seed. The construction of Fig. 2.2b is based on division of the original object and it can be well followed how the subsequent replacement of the squares with five smaller squares leads to a self-similar, scale invariant structure.

One can generate many possible patterns by this technique; the fractal shown in Fig. 2.2 was chosen just because it has an open branching structure analogous to many observed biological fractals [3]. Only the first couple of steps (up to $k = 3$) of the construction are shown. *Mathematical fractals* are produced after *infinite number of such iterations*. In this $k \to \infty$ limit the fractal displayed in Fig. 2.2a becomes infinitely large, while the details of Fig. 2.2b become so fine that the picture seems to "evaporate" and can not be seen any more. Our example shows a connected construction, but disconnected objects distributed in a nontrivial way in space can also form a fractal.

In any real system there is always a lower cutoff of the length scale; in our case this is represented by the size of the particles. In addition, a real object has a finite

linear size which inevitably introduces an upper cutoff of the scale on which fractal scaling can be observed. This leads us to the conclusion that, in contrast to the mathematical fractals, for fractals observed *in natural phenomena (including biology) the anomalous scaling of the volume can be observed only between two well defined length scales*.

Then, a possible definition for a biological fractal can be based on the requirement that a power law scaling of $N(l)$ has to hold over *at least two orders of magnitude*.

2.1.2 Definitions

Because of the two main types of fractals demonstrated in Fig. 2.2, to define and determine the fractal dimension D one typically uses two related approaches.

For fractals having fixed L and details on very small length scale D is defined through the scaling of $N(l)$ as a function of decreasing l, where $N(l)$ is the number of d dimensional balls of diameter l needed to cover the structure.

In the case of growing fractals, where there exists a smallest typical size a, one cuts out d-dimensional regions of linear size L from the object and the volume, $V(L)$, of the fractal within these regions is considered as a function of the linear size L of the object. When determining $V(L)$, the structure is covered by balls or boxes of unit volume ($l = a = 1$ is usually assumed), therefore $V(L) = N(L)$, where $N(L)$ is the number of such balls.

The fact that an object is a mathematical fractal then means that $N(l)$ diverges as $l \to 0$ or $L \to \infty$, respectively, according to a non-integer exponent.

Correspondingly, for fractals having a finite size and infinitely small ramifications we have

$$N(l) \sim l^{-D} \tag{2.3}$$

with

$$D = \lim_{l \to 0} \frac{\ln N(l)}{\ln(1/l)}, \tag{2.4}$$

while

$$N(L) \sim L^{D} \tag{2.5}$$

and

$$D = \lim_{L \to \infty} \frac{\ln N(L)}{\ln(L)}. \tag{2.6}$$

for the *growing case*, where $l = 1$. Here, as well as in the following expressions the symbol $\sim$ means that the proportionality factor, not written out in 2.3, is independent of l.

Obviously, the above definitions for non-fractal objects give a trivial value for D coinciding with the embedding Euclidean dimension d. For example, the area (corresponding to the volume $V(L)$ in $d = 2$) of a circle grows as its squared radius which according to 2.6 results in $D = 2$.

Now we are in the position to calculate the dimension of the objects shown in Fig. 2.2. It is evident from the figure that for the growing case

$$N(L) = 5^k \qquad \text{with} \qquad L = 3^k, \tag{2.7}$$

where k is the number of iterations completed. From here using 2.6 we get the value $D = \ln 5/\ln 3 = 1.465...$ which is a number between $d = 1$ and $d = 2$ just as we expected.

2.1.3 Useful rules

In this section we mention a few rules which can be useful in predicting various properties related to the fractal structure of an object. Of course, because of the great variety of self-similar geometries the number of possible exceptions is not small and the rules listed below should be regarded, at least in part, as starting points for more accurate conclusions.

a) Many times it is the *projection* of a fractal which is of interest or can be experimentally studied (e.g. a picture of a fractal embedded into $d = 3$). In general, projecting a $D < d_s$ dimensional fractal onto a d_s dimensional surface results in a structure with the same fractal dimension $D_p = D$. For $D \geq d_s$ the projection fills the surface, $D_p = d_s$.

b) It follows from a) that for $D < d_s$ the density correlations $c(r)$ (see next section) within the projected image decay as a power law with an exponent $d_s - D$ instead of $d - D$ which is the exponent characterising the algebraic decay of $c(r)$ in d.

c) Cutting out a d_s dimensional slice (*cross-section*) of a D dimensional fractal embedded into a d dimensional space usually leads to a $D + d_s - d$ dimensional object. This seems to be true for self-affine fractals (next section) as well, with D being their local dimension

d) Consider two sets A and B having fractal dimensions D_A and D_B, respectively. *Multiplying* them together results in a fractal with $D = D_A + D_B$. As a simple example, imagine a fractal which is made of parallel sticks arranged in such a way that its cross-section is the fractal shown in Fig. 2.2b. The dimension of this object is $D = 1 + \ln 5/\ln 3$.

e) The *union* of two fractal sets A and B with $D_A > D_B$ has the dimension $D = D_A$.

f) The fractal dimension of the *intersection* of two fractals with D_A and D_B is given by $D_{A\cap B} = D_A + D_B - d$. To see this, consider a box of linear size L within the overlapping region of two growing stochastic fractals. The density of A and B particles is respectively proportional to L^{D_A}/L^d and L^{D_B}/L^d. The number of overlapping sites $N \sim L^{D_{A\cap B}}$ is proportional to these densities and to the volume of the box which leads to the above given relation. The rule concerning intersections of fractals with smooth hypersurfaces (rule c) is a special case of the present one.

g) The distribution of empty regions (holes) in a fractal of dimension D scales as a function of their linear size with an exponent $-D - 1$.

Self-similarity can be directly checked for a deterministic fractal constructed by iteration, but in the case of random structures one needs other methods to detect the fractal character of a given object. In fact, *random fractals are self-similar only in a statistical sense* (not exactly) and to describe them it is more appropriate to use the term scale invariance than self-similarity. Naturally, for demonstrating the presence of fractal scaling one can use the definition based on covering the given structure with balls of varying radii, however, this would be a rather troublesome procedure. It is more effective to calculate the so called *density-density or pair correlation function*

$$c(\vec{\mathbf{r}}) = \frac{1}{V}\sum_{\vec{\mathbf{r}}'} \rho(\vec{\mathbf{r}} + \vec{\mathbf{r}}')\ \rho(\vec{\mathbf{r}}') \tag{2.8}$$

which is the expectation value of the event that two points separated by $\vec{\mathbf{r}}$ belong to the structure. For growing fractals the volume of the object is $V = N$, where N is the number of particles in the cluster, and 2.9 gives the probability of finding a particle at the position $\vec{\mathbf{r}} + \vec{\mathbf{r}}'$, if there is one at $\vec{\mathbf{r}}'$. In 2.9 ρ is the local density, i.e., $\rho(\vec{\mathbf{r}}) = 1$ if the point $\vec{\mathbf{r}}$ belongs to the object, otherwise it is equal to zero. Ordinary fractals are typically isotropic (the correlations are not dependent on the direction) which means that the density correlations depend only on the distance r so that $c(\vec{\mathbf{r}}) = c(r)$.

Now we can use the pair correlation function introduced above as a criterion for fractal geometry. An object is non-trivially scale invariant if its correlation function determined according to 2.9 is unchanged up to a constant under rescaling of lengths by an arbitrary factor b:

$$c(br) \sim b^{-\alpha} c(r) \tag{2.9}$$

with α a non-integer number larger than zero and less than d. It can be shown that the only function which satisfies 2.9 is the power law dependence of $c(r)$ on r

$$c(r) \sim r^{-\alpha} \tag{2.10}$$

corresponding to an algebraic decay of the local density within a random fractal, since the pair correlation function is proportional to the density distribution around a given point. Let us calculate the number of particles $N(L)$ within a sphere of radius L from their density distribution

$$N(L) \sim \int_0^L c(r) d^d r \sim L^{d-\alpha}, \tag{2.11}$$

where the summation in 2.8 has been replaced by integration. Comparing 2.11 with 2.5 we arrive at the relation

$$D = d - \alpha \tag{2.12}$$

which is a result widely used for the determination of D from the density correlations within a random fractal.

2.1.4 Self-similar and self-affine fractals

There are three major types of fractals as concerning their scaling behaviour. Self-similar fractals are invariant under isotropic rescaling of the coordinates, while for self-affine fractals scale invariance holds for affine (anisotropic) transformation. Until this point mainly the former case has been discussed.

The random motion of a particle represents a particularly simple example of stochastic processes leading to growing fractal structures. A widely studied case is when the particle undergoes a random walk (Brownian or diffusional motion) making steps of length distributed according to a Gaussian in randomly selected directions. Such processes can be described in terms of the mean squared distance $R^2 = \langle R^2(t) \rangle$ made by the particles during a given time interval t. For random walks $R^2 \sim t$ independently of d which means that the Brownian trajectory is a random fractal in spaces with $d > 2$. Indeed, measuring the volume of the trajectory by the total number of places visited by the particle making t steps, $(N(R) \sim t)$, the above expression is equivalent to

$$N(R) \sim R^2 \tag{2.13}$$

and comparing 2.13 with 2.5 we conclude that for random walks $D = 2 < d$ if $d > 2$. In this case, rather unusually, the fractal dimension is an integer number. However, the fact that it is definitely smaller than the embedding dimension indicates that the object must be non-trivially scale invariant.

In many physically relevant cases the structure of the objects is such that it is invariant under dilation transformation only if the lengths are rescaled by direction dependent factors. These anisotropic fractals are called self-affine [4, 5, 6].

Single-valued, nowhere-differentiable functions represent a simple and typical form in which self-affine fractals appear. If such a function $F(x)$ has the property

$$F(x) \simeq b^{-H} F(bx) \tag{2.14}$$

it is self-affine, where $H > 0$ is some exponent. Eq. 2.14 expresses the fact that the function is invariant under the following rescaling: shrinking along the x axis by a factor $1/b$, followed by rescaling of values of the function (measured in a direction perpendicular to the direction in which the argumentum is changed) by a different factor equal to b^{-H}. In other words, by shrinking the function using the appropriate direction-dependent factors, it is rescaled onto itself. For some deterministic self-affine functions this can be done exactly, while for random functions the above considerations are valid in a stochastic sense (expressed by using the sign $\simeq$).

A definition of self-affinity equivalent to 2.14 is given by the expression for the *height correlation function* $c(\Delta x)$

$$c(\Delta x) = \langle [F(x + \Delta x) - F(x)]^2 \rangle \sim \Delta x^{2H} \tag{2.15}$$

which can be easily used for the determination of the exponent H. In addition to functions satisfying 2.14 and 2.15, there are also self-affine fractals different from single-valued functions.

Let us first construct a deterministic self-affine model, in order to have an object which can be treated exactly.

An actual construction of such a bounded self-affine function on the unit interval is demonstrated in Fig. 2.3. The object is generated by a recursive procedure by replacing the intervals of the previous configuration with the generator having the form of an asymmetric letter z made of four intervals. However, the replacement this time should be done in a manner different from the earlier practice. Here every interval is regarded as a diagonal of a rectangle becoming increasingly elongated during the iteration. The basis of the rectangle is divided into four equal parts and the z-shaped generator replaces the diagonal in such a way that its turnovers are always at analogous positions (at the first quarter and the middle of the basis). The function becomes self-affine in the $k \to \infty$ limit.

Such a random function is, for example, the plot of the distances $X(t)$ measured from the origin as a function of time t, of a Brownian particle diffusing in one dimension. It is obvious that a so called *fractional Brownian plot* for which $\langle X_H^2(t) \rangle \sim t^{2H}$ satisfies 2.15.

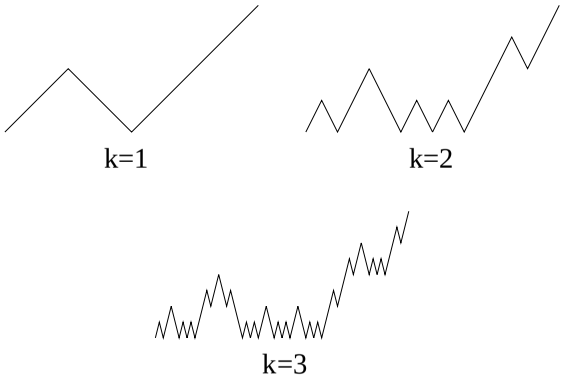


Figure 2.3: Self-affine functions can be generated by iteration procedures. This The single-valued character of the function is preserved by an appropriate distortion of the z-shaped generator ($k = 1$) of the structure [3].

Next we give a further basic feature of fractional Brownian motion.

Calculating the Fourier spectrum of a fractional Brownian function one finds that the coefficients of the series, $A(f)$, are independent Gaussian random variables and their absolute value scales with the frequency f according to a power law

$$|A(f)| \sim f^{-H-\frac{1}{2}}. \tag{2.16}$$

2.1.5 Multifractals

In the previous sections complex geometrical structures were discussed which could be interpreted in terms of a single fractal dimension. The present section is mainly concerned with the development of a formalism for the description of the situation when a singular distribution is defined on a fractal [7, 8].

It is typical for a large class of phenomena in nature that the behaviour of a system is determined by the spatial distribution of a scalar quantity, e.g., concentration, electric potential, probability, etc.. For simpler geometries this distribution function and its derivatives are relatively smooth, and they usually contain only a few (or none) singularities, where the word singular corresponds to a local power law behaviour of the function. (In other words, we call a function singular in the region surrounding point $\vec{\mathbf{x}}$ if its local integral diverges or vanishes with a non-integer exponent when

the region of integration goes to zero). In the case of fractals the situation is quite different: a process in nature involving a fractal may lead to a spatial distribution of the relevant quantities which possesses infinitely many singularities.

As an example, consider an isolated, charged object. If this object has sharp tips, the electric field around these tips becomes very large in accord with the behaviour of the solution of the Laplace equation for the potential. In the case of charging the branching fractals produced in the $k \to \infty$ limit of constructions shown in Fig. 2.1 or 2.2 one has infinite number of tips and corresponding singularities of the electric field. Moreover, tips being at different positions, in general have different local environments (configuration of the object in the region surrounding the given tip) which affect the strength of singularity associated with that position.

The above discussed time independent distributions defined on a fractal substrate are called *fractal measures*. In general, a fractal measure possesses an infinite number of singularities of infinitely many types. The term *"multifractality"* expresses the fact that points corresponding to a given type of singularity typically form a fractal subset whose dimension depends on the type of singularity. The description of the multifractal formalism goes beyond the scope of the present section, but can be found, for example, in Ref. 3.

2.1.6 Methods for determining fractal dimensions

When one tries to determine the fractal dimension of biological structures in practice, it usually turns out that the direct application of definitions for D given in the previous sections is ineffective or can not be accomplished. Instead, one is led to measure or calculate quantities which can be shown to be related to the fractal dimension of the objects. Three main approaches are used for the determination of these quantities: experimental, computer and theoretical.

Experimental methods for measuring fractal dimensions

A number of experimental techniques have been used to measure the fractal dimension of scale invariant structures grown in various experiments. The most widely applied methods can be divided into the following categories: (a) digital image processing of two-dimensional pictures, (b) scattering experiments and (c) direct measurement of dimension-dependent physical properties.

(a) *Digitising the image* of a fractal object is a standard way of obtaining quantitative data about geometrical shapes. The information is picked up by a scanner or an ordinary video camera and transmitted into the memory of a computer (typically a

PC). The data are stored in the form of a two-dimensional array of pixels whose non-zero (equal to zero) elements correspond to regions occupied (not occupied) by the image. Once they are in the computer, the data can be evaluated using the methods described in the next section, where calculation of D for computer generated clusters is discussed.

The only principal question related to processing of pictures arises if two-dimensional images of objects embedded into three dimensions are considered. It has already been mentioned that the fractal dimension of the projection of an object onto a $(d-m)$-dimensional plane is the same as its original fractal dimension, if $D < d-m$.

(b) *Scattering experiments* represent a powerful method to measure the fractal dimension of structures. Depending on the characteristic length scales associated with the object to be studied, light, X-ray or neutron scattering can be used to reveal fractal properties. There are a number of possibilities to carry out a scattering experiment. One can investigate i) the structure factor of a single fractal object, ii) scattering by many clusters growing in time, iii) the scattered beam from a fractal surface, etc.

Evaluation of numerical data

Throughout this section we assume that the information about the stochastic structures is stored in the form of d-dimensional arrays which correspond to the values of a function given at the nodes (or sites) of some underlying lattice. In the case of studying geometrical scaling only, the value of the function attributed to a point with given coordinates (the point being defined through the indexes of the array) is either 1 (the point belongs to the fractal) or 0 (the site is empty). When multifractal properties are investigated the site function takes on arbitrary values. In general, such discrete sets of numbers are obtained by two main methods: i) by digitising pictures taken from objects produced in experiments, ii) by numerical procedures used for the simulation of various biological structures. For convenience, in the following we shall frequently use the terminology "particle" for a lattice site which belongs to the fractal (is filled) and cluster for the objects made of connected particles.

Below we discuss how to measure D for a single object. To make the estimates more accurate one usually calculates the fractal dimension for *many clusters* and averages over the results.

Perhaps the most practical method is to determine the number of particles $N(R) = R^D$ within a region of linear size R and obtain the fractal dimension D from the slopes of the plots $\ln N(R)$ versus $\ln R$. If the centers of the regions of radius R are the particles of the cluster, than $N(R)$ is equivalent to the integral of the density correlation function. In practice one chooses a subset of randomly selected particles of the fractal

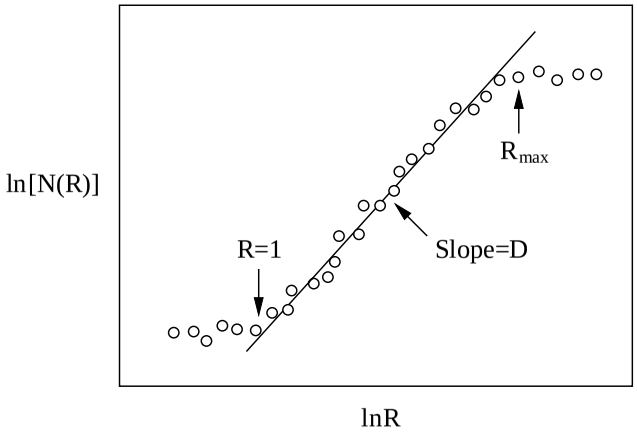


Figure 2.4: Schematic log-log plot of the numerically determined number of particles $N(R)$ belonging to a fractal and being within a sphere of radius R. If R is smaller than the particle size or larger than the linear size of the structure a trivial behaviour is observed. The fractal dimension is obtained by fitting a straight line to the data in the scaling region [3].

(as many as needed for a reasonable statistics) and determines $\langle N(R)\rangle$ for a sequence of growing R (or counts the number of particles in boxes of linear size L). In order to avoid undesirable effects caused by anomalous contributions appearing at the edge of the cluster one should not choose particles as centres close to the boundary region. The situation is shown in Fig. 2.4. Typically there is a deviation from scaling for small and large scales.

The roughness exponent H corresponding to *self-affine* fractals is usually determined from the definition 2.15. An alternative method is to investigate the scaling of the standard deviation $\sigma(l) = [\langle F^2(x)\rangle_x - \langle F(x)\rangle_x^2]^{1/2}$ of the self-affine function F

$$\langle \sigma(l)\rangle \sim l^H, \tag{2.17}$$

where the left hand side is the average of the standard deviation of the function F calculated for regions of linear size l. The roughness exponent H can be calculated by determining $\sigma(l)$ for parts of the interfaces for various l. An averaging should be made over the segments of the same length and the results plotted on a double

logarithmic plot as a function of l.

2.2 Stochastic processes

2.2.1 The physics of microscopic objects

Every biological process eventually takes place at the molecular level. The physics of this microscopic realm is fundamentally different from the physics of our macroscopic world, and requires a completely different description. First of all, as the length-scale and velocity-scale go down to molecular scales, the Reynolds number goes down too, and we approach the overdamped regime in which inertia plays no role anymore [9] and where the velocity (and not the acceleration) of the objects is proportional to the forces acting on them. Secondly, there is Brownian motion. Microscopic objects are being randomly kicked around by molecules of the surrounding medium, and the processes have an inherently stochastic nature.

The time scale of macroscopic processes is set by the velocity and acceleration of massive objects, the thermal fluctuations are negligible, and when we design a macroscopic device we try to suppress any stochastic element as much as possible. On the other hand, for microscopic objects every degree of freedom has inevitably a significant $\frac{1}{2}k_{\mathrm{B}}T$ thermal energy on average (where T denotes the absolute temperature and k_{B} is the Boltzmann coefficient), the timing of the processes is set by thermally assisted events (such as diffusion or activated transitions over energy barriers), and for the design of microscopic devices thermal fluctuations should be exploited rather than suppressed.

In general, the motion of any object in a thermal environment can be described by the *Langevin equation* [10]:

$$m\ddot{x}(t) = -\gamma\dot{x}(t) + \gamma\sqrt{2D}\xi(t) + F(x,t) \;, \tag{2.18}$$

where x, m, γ, and D denote the position, mass, viscous friction coefficient, and diffusion coefficient of the object, respectively. The three force terms on the right hand side of the equation are the viscous friction by the medium, $\gamma\dot{x}(t)$; the thermal noise coming from the molecules of the medium, $\gamma\sqrt{2D}\xi(t)$; and all the other forces unrelated to the medium, $F(x,t)$. Since the thermal noise term is a stochastic function the Langevin equation is referred to as a stochastic differential equation. The noise factor, $\xi(t)$, is usually modelled by a Gaussian white noise with zero time average,

$$\langle\xi(t)\rangle = 0 \;, \tag{2.19}$$

and autocorrelation function

$$\langle\xi(t)\xi(t')\rangle = \delta(t-t') \;. \tag{2.20}$$

Both the viscous friction and the thermal noise are exerted by the medium, and are not independent. Their magnitudes are connected by the fluctuation-dissipation theorem (or Einstein relation):

$$D = \frac{k_{\mathrm{B}}T}{\gamma} \;. \tag{2.21}$$

For biomolecules in water solution the Langevin equation can be simplified. The ratio $\tau_{\mathrm{relax}} = m/\gamma$ is a characteristic time scale of the Langevin equation (2.18), and tells us how long it takes for a particle to loose its initial velocity via viscous friction if the thermal noise and F are turned off. Multiplying this by the particle's characteristic velocity v, we get the characteristic distance $\lambda_{\mathrm{relax}} = vm/\gamma$ on which the particle comes to a halt. Comparing this distance to the characteristic size of the particle a, we get some information about the strength of the viscous damping: if $\lambda_{\mathrm{relax}}/a \ll 1$ the damping is strong, because the particle stops on a much shorter distance than its size; and if $\lambda_{\mathrm{relax}}/a \gg 1$ the damping is weak. Supposing that m is proportional to $a^3\rho$ and γ is proportional to $a\eta$ (cf. Stokes law), where ρ and η are the density and dynamic viscosity of the medium respectively, $\lambda_{\mathrm{relax}}/a$ becomes proportional to

$$R = \frac{va}{\eta/\rho} = \frac{va}{\nu} \;, \tag{2.22}$$

which is called the *Reynolds number* ($\nu = \eta/\rho$ is the kinematic viscosity). Thus, it is the Reynolds number that characterises the strength of the damping. Low Reynolds number means strong damping.

Let us now estimate the Reynolds number for biological molecules. The typical size of a protein is in the order of nanometers ($a \approx 1$ nm), the density and dynamic viscosity of water are $\rho \approx 10^3\,\mathrm{kg/m^3}$ and $\eta \approx 10^{-3}\,\mathrm{kg/s/m}$. The maximal forces acting on a protein are in the order of piconewtons (a few $k_{\mathrm{B}}T$ over a few nanometers), thus the characteristic velocity of a protein cannot be much larger than $v \approx 1\ \mathrm{pN}/(a\eta) \approx 1$ m/s. This shows that the Reynolds number for biomolecules is in the order of 10^{-3} or even smaller, i.e., we are in the strongly damped or *overdamped* regime. $R = 10^{-3}$ is a somewhat shocking result. It means that the viscous friction can stop a protein on a distance ($\sim 10^{-3}$ nm) much shorter than the size of the atoms.

In this overdamped regime when the forces change, the velocity of a particle relaxes so quickly (during τ_{relax}) and on such a small distance ($\lambda_{\mathrm{relax}} \ll a$) that the acceleration term (the derivative of the velocity with respect to time) on the left hand

side of the Langevin equation (2.18) can be neglected:

$$\dot{x}(t) = F(x,t)/\gamma + \sqrt{2D}\xi(t) \; . \tag{2.23}$$

This kind of reduction is called adiabatic elimination of the fast variables [11, 10]. Since the motion of biomolecules can be well described by the overdamped Langevin equation, from now on we will use only this version of the equation, and also the term Langevin equation will always refer to its overdamped version (2.23).

From this stochastic ordinary differential equation one can derive a deterministic partial differential equation, the *Fokker-Planck equation* (or Smoluchowski equation) [10], which describes the time evolution of the probability density $P(x,t)$ of the position of the particle:

$$\partial_t P(x,t) = -\partial_x J(x,t) \; , \tag{2.24}$$

where

$$J(x,t) = \frac{F(x,t)}{\gamma} P(x,t) - \frac{k_\mathrm{B}T}{\gamma}\partial_x P(x,t) \tag{2.25}$$

is the probability current of the particle. If the force field $F(x,t)$ is the negative gradient of a potential: $F(x,t) = -\partial_x V(x,t)$, the probability current can be written in the form

$$J(x,t) = -\frac{k_\mathrm{B}T}{\gamma}\mathrm{e}^{-V(x,t)/k_\mathrm{B}T}\partial_x\left[\mathrm{e}^{V(x,t)/k_\mathrm{B}T}P(x,t)\right] \; . \tag{2.26}$$

2.2.2 Kramers formula and Arrhenius law

In many systems Brownian particles are wiggling in deep potential wells (compared to $k_\mathrm{B}T$) for long periods of time, rarely interrupted by quick jumps into one of the neighbouring wells. If the potential is static or changes in a much longer time-scale than the duration of these jumps (which is usually the case), a *kinetic* approach can be used to describe the motion of the particles, with transition rate constants between discrete states.

The discipline of rate theory (for review see Ref. [12]) was created when Arrhenius [13] extensively discussed various reaction-rate data and showed that they vary on a logarithmic scale linearly to the inverse temperature T^{-1}. In other words, the escape (or jumping) rate constants, k, follow the Arrhenius law

$$k = \nu\mathrm{e}^{-\Delta E/k_\mathrm{B}T}, \tag{2.27}$$

where ΔE denotes the threshold energy for activation and ν is a frequency prefactor.

Using Kramers' method [14, 10], the Arrhenius law can be easily derived from the Fokker-Planck equation for an overdamped Brownian particle moving in a one-dimensional potential $V(x)$ (depicted in Fig. 2.5), if the potential well from which

the particle is trying to escape is much deeper than $k_\mathrm{B}T$. In this case the probability density near the bottom can be well approximated by its equilibrium value

$$P_\mathrm{eq}(x) = P_0\mathrm{e}^{-V(x)/k_\mathrm{B}T} \; , \tag{2.28}$$

which can be derived from Eq. (2.26) by setting its right-hand side to zero. The normalisation factor is approximately

$$P_0 = \frac{1}{\int_b^c \mathrm{e}^{-V(x)/k_\mathrm{B}T}\mathrm{d}x} \; , \tag{2.29}$$

because the vast majority of the probability falls into the interval $[b,c]$ where the potential difference from the bottom of the potential is not larger than a few ($\sim$5) $k_\mathrm{B}T$. Kramers' approach is based on the assumption that the probability current over the potential barrier between B and C is a constant J. Indeed, this condition holds, because the interval $[B,C]$ contains only a very small fraction of the total probability. Another assumption is that the probability distribution for $x \geq C$ is zero, because the particle has basically no chance to get back to the well and the escape can be considered to be completed. Thus, after rearranging and integrating Eq. (2.26) from B to C we get

$$J\frac{\gamma}{k_\mathrm{B}T}\int_B^C \mathrm{e}^{V(x)/k_\mathrm{B}T}\mathrm{d}x = -\left[\mathrm{e}^{V(x)/k_\mathrm{B}T}P_\mathrm{eq}(x)\right]_{x=B}^{C} \; , \tag{2.30}$$

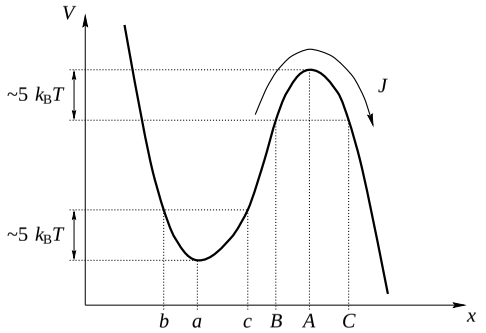


Figure 2.5: Potential $V(x)$ with a deep well and a barrier over which an overdamped Brownian particle tries to escape.

where the expression between the square brackets is P_0 for $x = B$ and zero for $x = C$. From this the current J over the barrier (which is equivalent to the escape rate constant k) can be expressed as

$$J \equiv k = \frac{D}{\int_b^c \mathrm{e}^{-V(x)/k_\mathrm{B}T}\mathrm{d}x \int_B^C \mathrm{e}^{V(x)/k_\mathrm{B}T}\mathrm{d}x} = \frac{D\,\mathrm{e}^{-[V(A)-V(a)]/k_\mathrm{B}T}}{\int_b^c \mathrm{e}^{-[V(x)-V(a)]/k_\mathrm{B}T}\mathrm{d}x \int_B^C \mathrm{e}^{-[V(A)-V(x)]/k_\mathrm{B}T}\mathrm{d}x}\,.$$

This expression has indeed the same form as that of the Arrhenius law (2.27). Here the activation energy ΔE is the height of the barrier $V(A) - V(a)$, and the frequency prefactor ν depends only on the shape of the potential near the bottom of the well and the top of the barrier.

This derivation holds even if the potential $V(x)$ changes in time but much slower then the intrawell relaxation time of the particle [15]. In this case the escape rate constant becomes also time dependent.

Most chemical reactions can also be described in terms of kinetic rate constants, so if they are present, they represent another source of stochasticity in molecular processes in addition to the thermal noise. The waiting time for any escape process or chemical reaction characterised by a rate constant k has an exponential distribution with mean value $1/k$.

2.3 Continuous phase transitions

In the following chapter we discuss the relevance of the so called self-organised critically (SOC) for biology. However, before describing this more advanced, non-equilibrium concept we give a short introduction to the closely related precursor, the second order (or continuous) phase transition occurring in equilibrium.

Phase transitions can be easily understood on a simple thermodynamic level. Let us consider a substance which can exist in two different phases like water (liquid and ice) or iron (paramagnetic and ferromagnetic). Usually one of the phases is *disordered* while the other is *ordered*. The distinction is based on the *symmetry* of the state: the symmetric (or isotropic) state is the disordered one. To characterise the strength of ordering at given values of thermodynamic parameters (like the temperature) we introduce an *order parameter*. It measures how well ordered the substance is. By convention if the order parameter is zero we speak of a completely disordered state

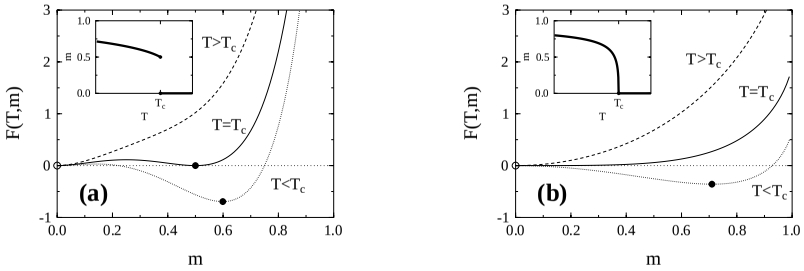


Figure 2.6: Two routes to a phase transition: (a) first order and (b) second order. The insets show the order parameter as the function of temperature.

(e.g., liquid water) and if it is non-zero then the substance is in an ordered state (e.g., ice crystal). As it is known from thermodynamics, at constant temperature the phase with the lowest free energy is stable. Since the free energy generally can also be a function of the order parameter, this energy minimum requirement will select which phase can be observed at a given temperature T. In this case the temperature plays the role of *control parameter* which allows for tuning the system to the phase transition.

In Fig. 2.6 we sketch two possible ways leading to a phase transition. When the temperature is above the *critical temperature* T_c (where the transition occurs) the only globally stable solution to the energy minimum criterion is the phase with zero order parameter ($m = 0$). This means that for $T > T_\mathrm{c}$ the system is in its disordered state. Fig. 2.6(a) shows a case when lowering the temperature, at $T = T_\mathrm{c}$ first a non-trivial ($m > 0$) minimum appears and then it shifts to larger values as the temperature is lowered (see inset in Fig. 2.6(a)). This is the scenario of a *first order* phase transition. The characteristic feature of this type of transition is a jump discontinuity in the order parameter (and certain other quantities). In contrary, Fig. 2.6(b) demonstrates another type of transition where the order parameter changes *continuously* as the temperature is lowered below T_c. This type of transition is referred as a *second order* phase transition and it has received much larger attention in the past decades. A motivation for this interest lies in the special properties of such transitions.

Without loss of generality let us consider a more specific example for a second order transition: a magnetic material at temperature T and magnetic field H. If this system shows a paramagnetic–ferromagnetic second order transition at T_c (for

$H = 0$) then it is convenient to use the *reduced temperature*

$$t = \frac{T - T_c}{T_c} \tag{2.31}$$

as control parameter instead of T.

Measuring the physical properties of the sample reveals divergences of various physical observables as the critical temperature is approached. The order parameter in this case is the zero-field ($H = 0$) magnetisation M_0 since it is zero if $T > T_c$, and non-zero below T_c. In the vicinity of T_c it behaves as

$$M_0(t) \sim |t|^{\beta}, \tag{2.32}$$

where β is the *critical exponent* of the magnetisation. Similarly, for the specific heat (which gives the change of energy for a small change of temperature)

$$C_{H=0}(t) \sim |t|^{-\alpha}, \tag{2.33}$$

and the susceptibility (which is the sensitivity of the magnetisation with respect to the external field H) scales as

$$\chi(T) \sim |t|^{-\gamma}. \tag{2.34}$$

Both of these quantities describe a response of the system to some external perturbation. Close to T_c they diverge showing that there the system is extremely sensitive: it is in a critical state.

The above defined three critical exponents (α, β, γ) are not independent of each other. A more detailed analysis shows that the *exponent relation*

$$\beta = \frac{2 - \alpha - \gamma}{2} \tag{2.35}$$

holds.

Near the transition point the spontaneous fluctuations in the system become large due the high susceptibilities. For the case of a fluid these strong fluctuations are observable as the decrease of light transmittance (*critical opalescence*). Since the length scale ξ associated with these fluctuations, i.e., the typical size of fluid droplets, also has power law divergence

$$\xi \sim t^{-\nu}, \tag{2.36}$$

at the critical point there will be no typical length scale except the trivial lower (atomic size) and upper (system size) scales. This fact is manifested via the fractal [3] structure of the fluctuations in the system and it is closely connected to other power

law divergences present at T_c. The fractality from the experimenter's point of view means that the fluctuations are statistically invariant under the transformation

$$(x, y, \ldots) \longmapsto (\lambda x, \lambda y, \ldots), \tag{2.37}$$

i.e., no typical length scale can be identified.

2.3.1 The Potts model

Using the thermodynamic approach it is possible to derive the critical exponents only heuristically, based on symmetry arguments, supposing some form of the (coarse grained) free energy near the critical point. This method is used by the Landau theory of critical phenomena. To overcome the limitations of this approach one has to introduce models which include more details about the interactions leading to the phase transition. A number of different models can be constructed depending on the level of abstraction at which interactions are handled. Here we discuss a rather general lattice model introduced by Potts in 1952 [16].

The q-state Potts model consists of a set of "spins" (or particles) $\{s_i\}$ each of which may have integer values $s_i = 0, 1, \ldots (q-1)$. These spins sit on a lattice and the Hamiltonian (the energy function) is defined as

$$H[\{s_i\}] = -\sum_{<ij>} J\, \delta_{\mathrm{Kr}}(s_i, s_j), \tag{2.38}$$

where δ_{Kr} is the Kronecker delta function and the summation goes over nearest neighbours only (short range interaction). The meaning of the energy function Eq. (2.38) is that only particles with $s_i = s_j$ "like" each other, only such combinations lower the energy of the system as illustrated in Fig. 2.7.

If the temperature is high then all the q states will be equally populated, so any quantity can serve as an order parameter which measures the difference of the distribution of spin states from uniform. At low temperatures the system organises itself into a configuration where most spins are in a randomly selected state while the other states are weakly populated.

The Potts model is related to many other lattice models in statistical physics [17]. For $q = 2$ (two states: 'up' and 'down') it is equivalent to the well known Ising model. The $q = 1$ limit reproduces the percolation problem, while $q = 0$ can be mapped to the resistor network problem.

2.3.2 Mean-field approximation

Although some of the special cases mentioned above can be solved exactly, no general solution exists to the Potts model itself. Here we present the simplest approach to

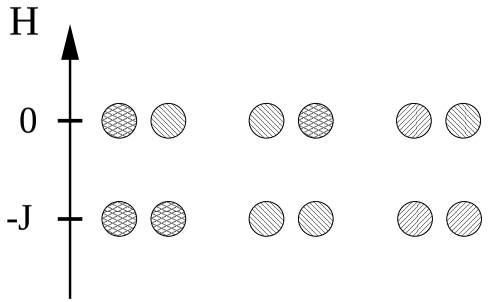


Figure 2.7: Interaction energy of particles in the Potts model.

determine the exponents [17]. Consider the following slightly modified Hamiltonian

$$H[\{s_i\}] = -\frac{zJ}{N}\sum_{i<j}\delta_{\mathrm{Kr}}(s_i, s_j), \tag{2.39}$$

where z is coordination number of the lattice (number of neighbours, i.e., $z = 4$ for a planar square lattice) and N is the total number of spins. In contrast to Eq. (2.38) this Hamiltonian allows long range interactions since interaction of every spin-pair contributes to the total energy. In other words, a spin is affected not only by its neighbours, but rather by the mean state of the whole system. For this reason such an approximation is called a *mean-field* description. This approach has several deficiencies: by introducing long range interaction it neglects the fluctuations which are essential parts of phase transitions, and therefore does not give correct results for the critical temperature and the exponents. Nevertheless, since gives a qualitatively correct picture it is worth examining.

Instead of accounting for the state of every spin we are going to consider the distribution of spins by states. Let x_i denote the fraction of spins that are in spin state $i = 0,1,\ldots(q-1)$. The probability of the Kronecker delta in Eq. (2.39) being 1 is

$$\Pr[\delta_{\mathrm{Kr}}(s_i, s_j) = 1] = \sum_{i=0}^{q-1} x_i^2. \tag{2.40}$$

Then the energy per spin is

$$e = \frac{E}{N} = -\frac{zJ}{N^2}\frac{N^2}{2}\sum_{i=0}^{q-1} x_i^2 = -\frac{1}{2}zJ\sum_{i=0}^{q-1} x_i^2 \tag{2.41}$$

and the entropy per spin is given by

$$s = \frac{S}{N} = -k_{\mathrm{B}}\sum_{i=0}^{q-1} x_i \ln x_i. \tag{2.42}$$

The free energy is then

$$f = e - Ts = -\frac{1}{2}zJ\sum_i x_i^2 + k_{\mathrm{B}}T\sum_i x_i \ln x_i. \tag{2.43}$$

Supposing that in the ordered phase the $i = 0$ state will be populated we look for a solution in the form of

$$x_0 = \frac{1}{q} + \frac{q-1}{q}m, \tag{2.44}$$

and

$$x_i = \frac{1}{q} - \frac{1}{q}m, \tag{2.45}$$

where the value of the order parameter $0 \le m \le 1$ is to be chosen as to minimize the free energy Eq. (2.43). In the disordered phase $m = 0$ since all sites are equally populated, while in the ordered state $m > 0$. The choice Eqs. (2.44) and (2.45) trivially satisfies the normalisation condition $\sum_i x_i = 1$. Using Eqs. (2.43), (2.44), and (2.45) one finds that

$$\frac{f}{k_{\mathrm{B}}T} = \frac{1+(q-1)m}{q}\ln[1+(q-1)m] + \frac{q-1}{q}(1-m)\ln(1-m) - K\frac{q-1}{2q}m^2, \tag{2.46}$$

with $K = zJ/(k_{\mathrm{B}}T)$. The equilibrium order parameter m_0 as a function of temperature T has to determined from $f'(m_0) = 0$. It is easily seen that $m_0 = 0$ is always a solution, but at sufficiently low temperatures other solutions with $m_0 > 0$ emerge which may lead a lower free energy. The critical temperature T_{c} is then defined as the temperature at which this shift of the absolute minimum of the free energy occurs.

It is instructive to expand Eq. (2.46) in powers of m around $m = 0$. Up to third order one gets

$$\frac{f}{k_{\mathrm{B}}T} = \frac{q-1}{2q}(q-K)m^2 - \frac{1}{6}(q-1)(q-2)m^3 + \ldots. \tag{2.47}$$

For $q > 2$ the coefficient in the cubic term becomes negative which indicates a possibly first order transition. If $q < 2$ then the transition in this mean-field limit is second

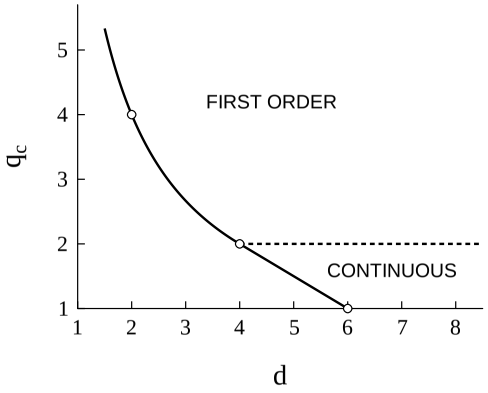


Figure 2.8: Schematic plot of $q_c(d)$, the critical value of q beyond which the transition is mean-field-like (first order for $q > 2$ and continuous for $q \leq 2$). The known points $q_c(2) = 4$, $q_c(4) = 2$ and $q_c(6) = 1$ are denoted by open circles (after [17]).

order (cf. Fig. 2.6(b)). The critical exponent for the order parameter obtained by this method is

$$\beta = \frac{1}{2}.$$

As we have already mentioned above, generally the results of such a mean-field calculation should not be regarded as the solution of the model. In fact, the nature of the transition depends also on the dimension d of the underlying lattice. The dimensionality together with the lattice topology (i.e., square, triangular, etc.) determines the number of neighbouring sites z. With increasing lattice dimension a site gets more neighbours and thus the system approaches the mean-field limit where the lattice is fully connected. Therefore, it is expected that above a certain *critical dimension* d_c (or conversely, above some critical value of q) the mean-field behaviour will be exact (Fig. 2.8).

Bibliography

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature*, Freeman, San Francisco, 1982

[2] T. A. Witten and L. M. Sander, Diffusion-limited aggregation, a kinetic critical phenomenon, *Phys. Rev.* **B27**, 5686, 1983

[3] T. Vicsek, *Fractal Growth Phenomena*, World Scientific, Singapore, 1992

[4] F. Family and T. Vicsek, editors, *Dynamics of Fractal Surfaces*, World Scientific, Singapore, 1991

[5] B. B. Mandelbrot, Self-affine fractals sets, in *Fractals in Physics*, edited by L. Pietronero and E. Tosatti, Elsevier, Amsterdam, 3–28, 1986.

[6] J. Kertész and T. Vicsek, Self-affine interfaces, in *Fractals in Sciences*, A. Bunde and S. Havlin, editors, Springer, Berlin, 89–116, 1994

[7] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procacia and B. I. Shraiman, Fractal measures and their singularities:The characterisation of strange sets, *Phys. Rev.* A33:1141–1151, 1986

[8] B. B. Mandelbrot, An introduction to multifractal distribution functions, in *Random Fluctuations and Pattern Growth* Stanley, H. E. and Ostrowsky, N., editors, Kluwer, Dordrecht, 279–291, 1988

[9] E. M. Purcell, Life at low Reynolds number, *Am. J. Phys.*, 45:3–11, 1977.

[10] H. Risken, *The Fokker-Planck Equation.* Berlin, New York: Springer-Verlag, 1989.

[11] H. Haken, *Synergetics, An Introduction.* Berlin, New York: Springer-Verlag, 1983.

[12] P. Hänggi, P. Talkner, and M. Borkovec, Reaction-rate theory: fifty years after Kramers, *Rev. Mod. Phys.*, 62:251–341, 1990.

[13] S. Arrhenius, Über die Reaktionsgeschwindigkeit bei der Inversion von Rhrzucker durch Säuren, *Z. Phys. Chem.*, 4:226–248, 1889.

[14] H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions, *Physica*, 7:284–304, 1940.

[15] I. Derényi and R. D. Astumian, Intrawell relaxation time: the limit of the adiabatic approximation, *Phys. Rev. Lett.*, 82:2623–2627, 1999.

[16] R. B. Potts. Some generalized order-disorder transformations, *Proc. Camb. Phil. Soc.*, 48:106, 1952.

[17] F. Y. Wu. The Potts model. *Rev. Mod. Phys.*, 54:235, 1982.

Chapter 3

Self-organised criticality (SOC)

Perhaps the most conspicuous feature of life is its complexity. This complexity is due to processes capable of generating elaborated structures from such relatively simple building blocks as atoms. On the other hand, we know that the laws of physics (or chemistry) are rather simple and do not contain *a priori* information about highly organised patterns. Therefore, there must be some simple mechanisms which lead to complex spatial and temporal behaviour. Since the resulting structures appear in nature without any external enforcement, purely due to the interaction of the units in a system, we call the process of generating complexity *self-organisation*.

In thermodynamics we define two distinct classes of systems: closed systems left alone for a very long time approach the so called thermal equilibrium state in which the temperature is constant everywhere and no macroscopic changes take place in time. On the contrary, in non-equilibrium systems there is motion even in the so called stationary states when the overall characteristics of the system do not change in time. It is in non-equilibrium state when structures can emerge spontaneously from an originally homogeneous medium.

A specific kind of complexity can be observed in systems slowly driven to a stationary state. For example, slow driving means the gradual addition of energy to units of a system which may also loose this energy due to their interaction. In many cases this stationary state may represent a delicate balance among the units of the system. If the interaction between two units is such that a change exceeding a critical value in one unit results in a similar change in the neighbouring unit, than large, avalanche-like series of changes may take place in the system when it is close to a (critical, balanced) state. The system becomes very sensitive to perturbations, or in other words, critical.

As will be discussed later several biological systems can be considered both self-organised and critical. In the present chapter we describe the major concepts related

to *self-organised criticality* (SOC) and present a few examples from biology which may be best interpreted in terms of SOC.

3.1 SOC model

So far we have considered systems in which the critical state was reached by tuning some parameter, typically the temperature. There are, however, systems in Nature being in critical state *without* apparent tunable parameters. Such systems include granular materials, earthquakes, fraction, river networks, complex biological (e.g., brain, evolution), and economical systems, etc. Systems which can reach their critical state without parameter tuning are termed *self-organising critical* (SOC). The concept of SOC has been widely used in the studies of complex systems.

The standard model for SOC was introduced by Bak, Tang and Wiesenfeld in 1987 [1]. To provide an intuitive insight into the phenomenon, they used a sandpile as an example. Consider the typical childhood experience when one starts building a sandpile by depositing sand on a surface. At first, sand is accumulated making a sandpile. As one adds more and more sand to the pile eventually at some place the slope of the sandpile will reach its critical value. Adding one more grain to that site will trigger an avalanche. Sand will then be redistributed and thus may create avalanches at various places of the pile. Avalanches may actually span the whole sandpile.

This phenomenon is modelled by a discrete variable $h(\mathbf{r})$ defined on a d dimensional lattice. For the sake of simplicity we are going to consider here only the $d = 2$ square lattice case. The rules for the evolution of the system are as follows

Step 1 Choose a site (x, y) at random;

Step 2 Add one "grain" to that site: $h(x, y) \leftarrow h(x, y) + 1$;

Step 3 If $h(x, y) \leq 4$ then continue on Step 1.

Step 4 The pile is too "steep" locally so sand is redistributed among the nearest neighbours:

$$h(x, y) \leftarrow h(x, y) - 4$$

and

$$h(x', y') \leftarrow h(x', y') + 1,$$

where (x', y') denotes the nearest neighbours of site (x, y).

Step 5 The previous step is repeated until no site with $h > 4$ is found and then Step 1 is performed again.

Step 1 and Step 2 represent a continuous uniform adiabatic (infinitely slow) driving which makes the system far-from-equilibrium. Step 4 and Step 5 are called the process of relaxation. The size of an avalanche is defined by the number of times Step 4 was executed after adding a single grain. Large scale computer simulations demonstrated that avalanches are power law distributed

$$P(\geq s) \sim s^{-\tau}, \tag{3.1}$$

where $P(\geq s)$ is number of avalanches larger that s and the exponent τ is found to be around 0.25.

One should note that the role of slope in this model is replaced by the height, this is why the sandpile picture is only a way to gain an intuitive gasp of the concept. Nevertheless, it is possible to construct models with critical slope instead of critical height, but this does not solve the problem of sand piles since they are *not* critical in reality [2]. It turns out that the behaviour of sandpile is more reminiscent of a first order transition due to the existence not a single but two distinct critical slopes: the maximum angle of stability (θ_m) and the angle of repose ($\theta_r < \theta_m$). When the slope of sand exceeds θ_m an avalanche starts and it does not stop until the slope goes back to θ_r. The overall behaviour then will be a rather periodic appearance of avalanches and no power law size distribution can be observed.

3.2 Applications in biology

The existence of large events that spread the whole system can be generally considered as the foresign of self-organised criticality. Such large scale phenomena have been observed when tracing the extinction events during the evolution. It turned out (see Fig. 3.1) that the extinction rate was not constant as one could conjecture from a "smooth" evolutionary process, but it shows peaks corresponding to major extinction events in our history. Although one can not rule out sudden environmental changes as the cause of mass extinction, it is also reasonable to consider the evolution as a self-organised critical process.

Below we give some theoretical support of the idea of self-organised critical evolution and at the end of this section we describe another SOC system in biology: the lungs.

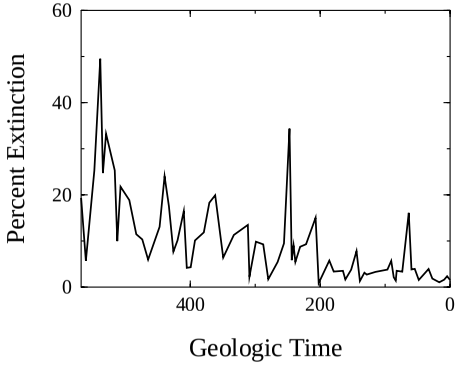


Figure 3.1: Variation of the extinction rate during the biological evolution (from [3]).

3.2.1 SOC model of evolution

In theoretical models of evolution the concept of *fitness* plays an essential role. Fitness of a species represents its ability to survive environmental changes as a function of its genetic code. There is no exact way to deduce a numerical value of fitness from a real DNA sequence so this quantity is purely heuristic. Nevertheless, it allows to construct models which grasp the main ingredients of evolution: mutation and selection. Within the framework of a model one can change the genetic code and calculate the corresponding change in fitness (*mutation*) and then if the new value of fitness is higher then accept the mutation or otherwise reject it (*selection*). Such a simple model would drive the system to a local maximum of fitness, therefore it is necessary to accept non-beneficial moves with some small probability. Nevertheless, the model described above still would lead to an equilibrium state which is not the case in real evolution, as we have discussed above.

In order to construct a SOC model for evolution [4, 5] let us consider an ecosystem of N species where the fitness of the ith species ($i = 1, 2, \ldots, N$) is a real number

$$0 \leq B_i \leq 1. \tag{3.2}$$

Species with low fitness has larger selection pressure therefore they change their fitness more rapidly then the ones with higher fitness (for this reason fitness is denoted by B as *barrier*). For the sake of simplicity we implement a coarse-grained version of this

dynamics:

1. We choose the species with the smallest fitness

$$B_j = \min_{i=1,\ldots,N} B_i. \tag{3.3}$$

2. Then this species makes an adaptive change (or mutation) which is represented by taking a random number for its fitness

$$B_j = \texttt{RND}, \tag{3.4}$$

where RND is equally distributed in [0,1].

3. Finally, the fitnesses of its two neighbours are also changed due to interaction

$$B_{j+1} = \texttt{RND}, \quad \text{and} \quad B_{j-1} = \texttt{RND}. \tag{3.5}$$

Taking only the first two steps would lead to a state where all species have the highest fitness ($B_i \equiv 1$). There would be no further dynamics possible in this state so it would correspond more to a "dead" state then to a living and evolving ecology. The interaction step (step 3) is therefore essential, it corresponds for example to the fact that neighbouring species are all consecutive parts of the food chain.

This model can be easily implemented in a computer. Simulations are started from random initial configuration of fitness. After some thousands of iterations the distribution of fitnesses tends to a non-trivial stable distribution (Fig. 3.2*a*). The distribution is flat below and above a threshold $B_c \sim 0.67$. It turns out that mutations take place through fitnesses that are smaller than B_c. This critical fitness is selected by the dynamics, not by tuning some external parameter, therefore we can say that the system has reached a self-organised state.

The activity in the systems can be displayed on a space-time plot of mutations (Fig. 3.2*b*). One can observe the intermittent nature of the evolution: after a period of localised mutations the position (i.e., the species affected by mutation) changes abruptly. In Fig. 3.2*c* we plotted the distribution of distances of consecutive mutations which characterises the correlations in the system. The straight line on the log-log plot indicates a power law distribution, i.e., there is no typical correlation length and hence the system is critical.

It is instructive to plot the local activity of a site (in this case defined as the number of mutations within 1000 iterations) as a function of time (Fig. 3.2*d*). One can

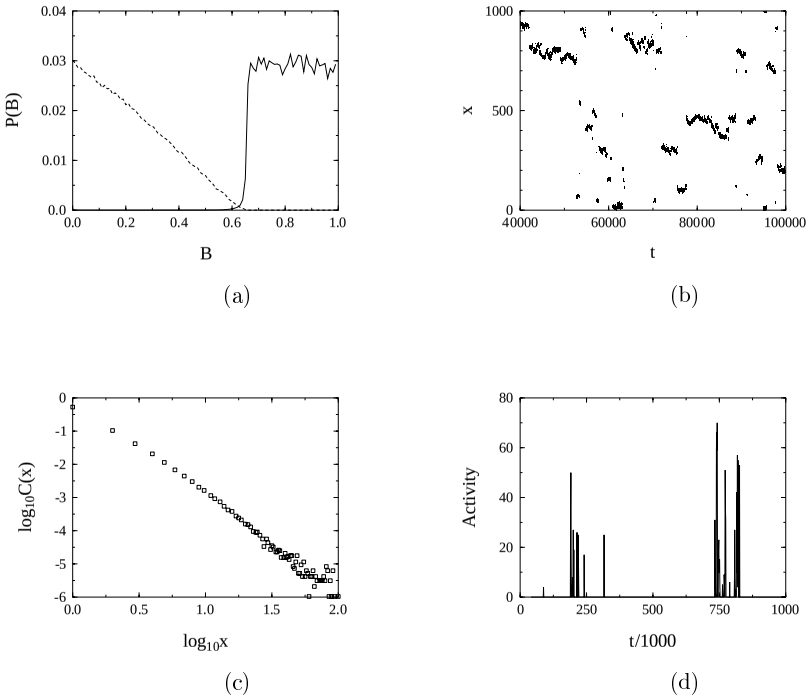


Figure 3.2: Punctuated equilibrium behaviour in the SOC model for evolution. (a) Distribution of barriers (solid curve) and of active barriers (dashed curve) in the critical state. (b) Typical mutation activity of the system displayed in the space-time domain. (c) Distribution of distances between subsequent mutations. (d) Activity of a site averaged over thousand consecutive iterations.

observe long periods of passivity interrupted by sudden bursts of activity. The temporal separation between burst is even more pronounced if one takes into account that the real (physical) time needed for a change of barrier B is scales as $\sim \exp(-B/\tau)$, where τ is the characteristic time scale for the mutations.

The model exhibits punctuated equilibrium: although the macroscopic features are stationary on a microscopic scale large sequences of mutations (avalanches) can be observed. A straightforward way to define an avalanche in this system is to consider the number of subsequent mutations below a threshold B. With this definition there is a hierarchy of avalanches, each defined by their respective threshold. During an avalanche sites may mutate more than once. For a threshold below the critical ($B < B_c$) the is a typical scale of avalanches. This size diverges approaching B_c and one observes a power law distribution of avalanche size ($P(\geq s) \sim s^{-\tau}$, where $\tau \simeq 1.07$ [6]), indicating that avalanches of all sizes are present, including catastrophic ones which span the whole system.

In real evolution, large extinction events may be thought of as large avalanches caused by the intrinsic dynamics of biology. Of course, one can not exclude the effect of large cataclysmic events like volcanic eruptions or meteorites. Nevertheless, the model indicates that there may be no need for such events to explain an intermittent evolution: the ecology is able to self-organise itself into a critical state with avalanches of all sizes.

In the model there was nearest neighbour interaction between the species. In order to formulate a solvable mean-field approximation [7] we consider *random neighbour* interaction, i.e., after having changed the barrier of a species $K-1$ other species are also assigned a random fitness. The main difference is that in this case the interacting sites are chosen randomly at every update rather than being fixed like in the original model. Therefore the $K = 3$ case is not equivalent to the original model since we are neglecting the correlations between neighbouring sites by reassigning interaction partners (Fig. 3.3). This redistribution of interactions validates the use of distribution functions for the barriers.

Let the distribution of barriers (x) in the ecology be $p(x,t)$. Then the smallest barrier is distributed as

$$p_1(x,t) = Np(x)Q^{N-1}(x,t) = -NQ^{N-1}\frac{dQ}{dx} = -\frac{dQ^N}{dx}, \tag{3.6}$$

where

$$Q(x,t) = \int_x^1 dx' p(x',t). \tag{3.7}$$

It is easy to show that

$$Q(0,t) = 1, \quad \text{and} \quad Q(1,t) = 0. \tag{3.8}$$

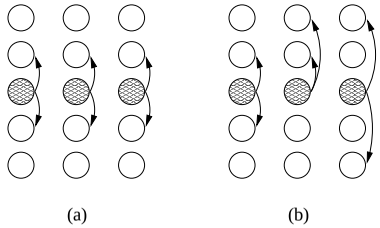


Figure 3.3: Three consequent mutations of the same site (shaded): (a) in the original model the interaction is always nearest neighbour, (b) in the mean-field version the interaction partners are chosen randomly in each iteration.

The evolution equation for $p(x,t)$ is the following

$$p(x,t+1) - p(x,t) = -\frac{1}{N}p_1(x,t) - \frac{K-1}{N-1}\left(p(x,t) - \frac{1}{N}p_1(x,t)\right) + \frac{K}{N}. \tag{3.9}$$

The first term on right-hand side of Eq. (3.9) represents the removal of the smallest barrier value. The second term takes into account the removal of $K-1$ of the $N-1$ barrier values remaining after the smallest has been removed from the set of N values. The last term corresponds to the addition of K equally distributed random barriers replacing the removed ones.

We are looking for stationary solution $p(x)$ such as that the left-hand side of Eq. (3.9) vanishes. Substituting Eq. (3.6) into Eq. (3.9) we have

$$0 = -\frac{dQ^N}{dx}\left(\frac{1}{N} + \frac{K-1}{N(N-1)}\right) + \frac{K-1}{N-1}\frac{dQ}{dx} + \frac{K}{N}. \tag{3.10}$$

Integrating from x to 1 results in

$$0 = Q^N\left(\frac{1}{N} + \frac{K-1}{N(N-1)}\right) - \frac{K-1}{N-1}Q + \frac{K}{N}(1-x), \tag{3.11}$$

or

$$0 = Q^N(N-K) + N(K-1)Q - K(N-1)(1-x). \tag{3.12}$$

This is an equation for $Q(x)$ which can not be generally solved exactly. However, it is possible to find approximate solutions in the limit $N \gg K$.

Let us now consider the two extremities when $Q \ll 1$ and $1-Q \ll 1$, i.e., when $1-x \ll 1$ and $x \ll 1$, respectively. The scale of Q (whether is can be considered

small or large) is set by the relative significance the first and the second terms in Eq. (3.12).

If Q is small then the first term is irrelevant, that is

$$Q^N(N-K) \ll N(K-1)Q, \tag{3.13}$$

giving

$$Q^{N-1} \ll \frac{N(K-1)}{N-K} \approx K-1 = O(1). \tag{3.14}$$

Thus from Eq. (3.12) we have

$$Q_a = \frac{K(N-1)}{N(K-1)}(1-x) \approx \frac{K}{K-1}(1-x). \tag{3.15}$$

The criterion of Eq. (3.14) and the solution Q_a are consistent only if

$$x - 1/K \gg O(1/N). \tag{3.16}$$

Conversely, if Q is large (close to 1) then the second term becomes negligible compared to the other terms. The solution in this case is given by

$$Q_b^N = \frac{-N(K-1)}{(N-K)} + \frac{K(N-1)}{N-K}(1-x) \approx 1 - Kx. \tag{3.17}$$

The consistency criterium now reads

$$1/K - x \gg O(1/N). \tag{3.18}$$

Using Eq. (3.7) we have

$$p(x) = \begin{cases} K/N, & \text{if} \quad x - 1/K \gg O(1/N); \\ K/(K-1), & \text{if} \quad 1/K - x \gg O(1/N). \end{cases} \tag{3.19}$$

In the limit $N \to \infty$ the distribution function of the barriers develops a discontinuity at $x = B_c = 1/K$. It vanishes below the threshold and constant above it, in accord with the results of numerical simulations (cf. solid line in Fig 3.2*a*).

The above calculation does not give correct value for B_c. From the simulations we know that it is around 0.67 but the calculations give for $K = 3$ the value of 1/3. This discrepancy is due to the fact that in the calculation we have neglected the correlations. In fact, a more realistic approximation can be constructed by assuming that the two smallest barriers and a randomly selected third value are replaced in each time step.

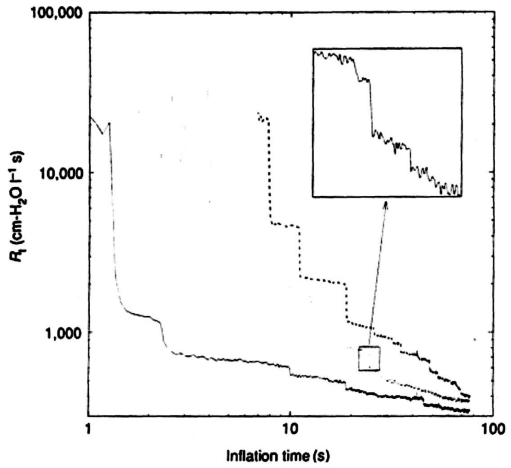


Figure 3.4: Terminal airway resistance as a function of inflation time (from [9]).

3.2.2 SOC in lung inflation

The morphology of the mammalian lung has fascinated physiologists for over a century. Recent investigations [8] showed that bronchial airways have a fractal tree–like structure which can be approximated by a branching process, i.e., the number of branches in a "generation" of the tree is about twice as large as in the previous generation (see Fig. 3.5a).

It is even more surprising that during inhalation the rate of air intake is not smoothly varying over time. In the actual experiment [9] an emptied dog lung was inflated by a constant flow and the terminal airway resistance (R_t) was measured. It was found that R_t decreases in discrete jumps (Fig. 3.4), which corresponds to a non-constant rate of inflation of the lung. Both the sizes of the jumps and the time intervals between them vary significantly within a single experiment. Magnifying a small portion of the curve one finds structures which are statistically similar to the original curve.

Therefore, the appropriate data analysis in this case is studying the distribution function of the sizes of the jumps and the time delays. To perform this analysis one needs a relatively large number of independent experiments (> 10) in order to obtain

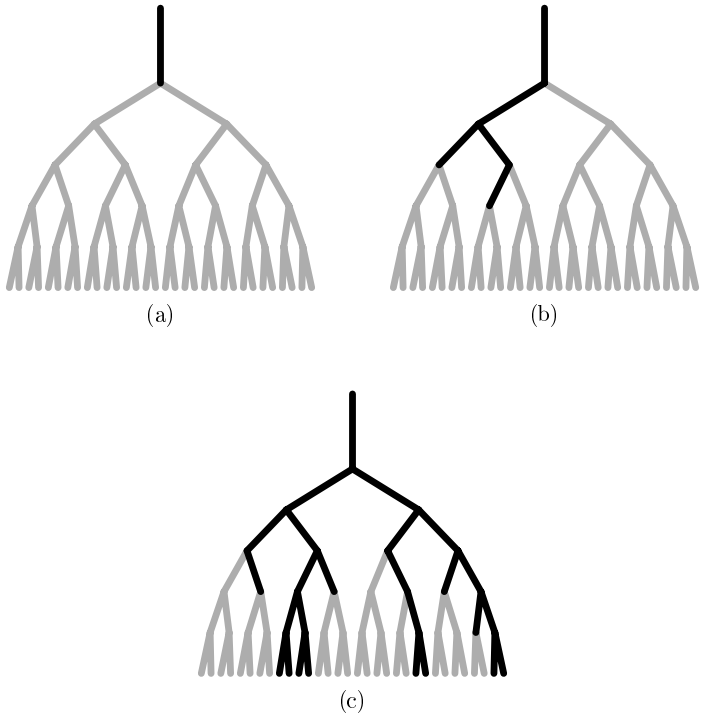


Figure 3.5: Typical development of avalanches in the lung model. (a) After the main alveola has opened, a small part of the lung also opens (b) due to lower critical pressure in that region. (c) The situation after a large avalanche: about the half of alveoli are open (after [9]).

good statistics of the data. The results showed that the distribution functions are power-law over a range of more than two decades. We have seen previously that the presence of power-law distribution indicates the lack of a characteristic scale: if there would be a typical scale then the distribution was exponential.

This result suggests the interpretation that the airways do not open independently, but rather in bursts. That is, the opening of one airway may initiate the opening of several more peripheral airways. When the air pressure implied on the lung (P_B) exceeds the opening threshold of an airway, all the daughter airways that had a smaller threshold would be opened immediately (one can neglect the time of opening an airway in this situation). The number of airways, and hence the size of the recruited alveolar volumes involved in such an opening sequence, depends on the size of subtrees, and can vary substantially. The sequential activation of alveoli triggered by overcoming a local threshold can be considered as an avalanche. The existence of avalanches often leads to power-law dependences (i.e., criticality).

To test the above interpretation of the lung opening phenomena a simple SOC model, similar to one considered for the evolution, can be constructed. Consider the branched lung structure as shown in Fig. 3.5a. Each airway is assigned an opening threshold P_i which takes a random value from the interval $[0,1]$. Then we set $P_B = 1$ and lower it until it reaches the opening threshold ($P_B = P_0$) of the main alveola. At this point the main alveola opens (Fig. 3.5*a*). Next, the two daughter airways are checked: they are opened if their threshold is lower that the actual bronchial pressure (P_B). This process is repeated until there is no more airway to open (Fig. 3.5*b*), i.e., all airways which are in contact with the opened part of the lung have $P_i > P_B$. At this point we consider this opening event (the avalanche) finished. Then P_B is incremented and the process is iterated so that larger and larger part of the lung opens. The newly opened alveolar volume in an avalanche can be significant with respect to the total lung volume, as it is demonstrated in Fig. 3.5*c*. The statistical analysis of the above model yields results which are consistent with the ones obtained experimentally. One can conclude that the process of lung inflation can be well treated as self-organised critical.

The examples reviewed in this section show that the general framework of self-organised critical phenomena proves useful for describing a range of dynamical biological processes. In many situations simple SOC models containing some sort of threshold mechanisms can be easily constructed to give a theoretical interpretation of the observed phenomena.

Bibliography

[1] P. Bak, C. Tang, and K. Wiesenfeld. Self-Organized Criticality: An Explanation of $1/f$ Noise *Phys. Rev. Lett.*, 59:381, 1987.

[2] S. R. Nagel. Instabilities in a sandpile. *Rev. Mod. Phys.*, 64:321–325, 1992.

[3] D. Raup. *Extinction: Bad Genes or Bad Luck?* W. W. Norton & Company, 1992.

[4] P. Bak and K. Sneppen. Punctuated equilibrium and criticality in a simple model of evolution. *Phys. Rev. Lett.*, 71:4083–4086, 1993.

[5] K. Sneppen, P. Bak, and M. J. Jensen. Evolution of a self-organized critical phenomenon. *Proc. of the Nat. Acad. of Sci.*, 92:5209, 1995.

[6] M. Marsili, P. De Los Rios, and S. Maslov. Expansion around the mean field solution of the bak-sneppen model. *Phys. Rev. Lett.*, 80:1457, 1998.

[7] H. Flyvbjerg, K. Sneppen, and P. Bak. Mean field theory for a simple model of evolution. *Phys. Rev. Lett.*, 71:4087–4090, 1993.

[8] M. F. Shlesinger and B. J. West. Complex fractal dimension of the bronchial tree. *Phys. Rev. Lett.*, 67:2106–2108, 1991.

[9] B. Suki, A.-L. Barabási, Z. Hantos, F. Peták, and H. E. Stanley. Avalanches and power-law behaviour in lung inflation. *Nature*, 368:615–618, 1994.

Chapter 4

Patterns and correlations

4.1 Bacterial colonies

4.1.1 Introduction

Bacteria are often said to be typical unicellular organisms and thought to live independently from each other. However, in their natural living places like on soil or rock surfaces bacteria form colonies and, as we will see, these colonies alter their form depending on the environmental conditions. This response cannot be understood by assuming that individual cells are simply staying close to each other as the consequence of being difficult to move. In fact, some of the complex biochemical processes performed by bacteria could not be possible without forming organised groups [?, ?, ?]. In some cases bacteria differentiate into specialised cell types (Fig. 4.1) within the colonies resulting in increased resistance to antibacterial agents or ability of invading eukaryotic hosts.

In addition to these interesting microbiological aspects, the study of bacterial colony formation can give certain insight into the formation of self-organised biological structures, in general. When trying to understand the function and operation of complex interacting networks like the nervous system or the intracellular signal transduction pathways, we often lack the appropriate experimental methods and abstract formalism. In the case of colonies, however, we can hope that the interactions are still simple enough to be captured by quantitative models and the collective behaviour can be explored with available computational power.

A basic problem of such studies is related to the very different length scales involved in the problem: The self-organised structures are macroscopic, consisting up to 10^{10} bacteria with a typical length scale of a few cm. However, these patterns are determined by the bacterial behaviour and interactions at the microscopic scale

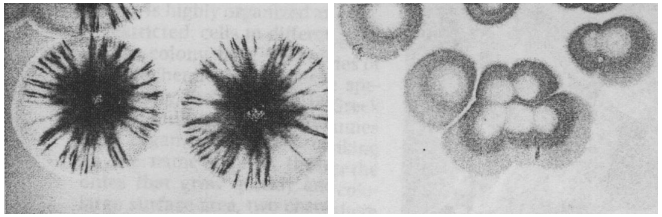


Figure 4.1: Examples for inhomogeneous protein expression in bacterial colonies grown on agar surface. (a) Some cells of *Chromobacterium violaceum* produces the pigment violacein. (b) These *Escherichia coli* colonies carry genetically engineered DNA sequences encoding the enzyme β-galactosidase; the colony turns blue where the enzyme is expressed. (After [?].)

of the individual cells. Given this large difference in the length scales, it is usually not straightforward to predict the effect of a certain microbiological feature on the macroscopic behaviour of the colony. In fact, as we will show, many of the microscopic details of the underlying system can turn out to be irrelevant to the dynamics of the large scale structures. This *universality* of the pattern formation process explains the fact that the same patterns may develop in completely different systems.

As a case study, within this chapter we would like to describe and understand the colony formation of a few bacterial strains displaying somewhat similar macroscopic behaviour. First we overview some of the microbiological background (after [?]), then discuss the relevant experimental observations. In the remaining parts we investigate one by one the development of the various morphologies and the corresponding models.

4.1.2 Bacteria in colonies

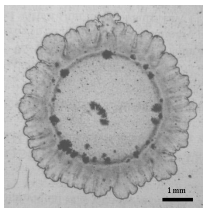


Figure 4.2: A growing bacterial colony (*Bacillus subtilis 168*) on the surface of a dry agar gel. Cells were inoculated in the central circular region of the picture. (After [?]).

In the simplest experimental studies of colony formation, bacteria are grown on the surface of agar gels as Fig. 4.2 shows. If the surface is dry and the cells cannot move then the time needed to spread over the substrate can take several weeks. Since this time period is long compared to the duplication time of the cells, the *proliferation* of the bacteria is the key factor determining the colony morphology. In the other extreme case, i.e., if the gel is soft or the bacteria produce surfactants, the colony spreads over the substrate in a few hours and the bacterial *motion* and its *chemotactic* modulation are important. In a general case, both of these effects must be considered to understand the response of the colonies to changes in their environments.

Microbiological background

Proliferation. As revealed by experiments in fermentors where microorganisms are grown in a steady, homogeneous liquid environment, the increase of the number and total mass of bacteria strongly depends on the nutrient concentration. In fact, the *rate* of growth (number of cell divisions within a population of unit size during a unit time interval) increases with the nutrient concentration in a hyperbolic manner reminiscent to Michaelis - Menten kinetics (Fig. 4.3). This finding probably reflects that the nutrient uptake is mediated by transport proteins. There is, however, a minimal concentration of nutrients required for bacterial proliferation, since not *all* the uptaken free energy is converted to growth as there is a finite amount required to maintain the intracellular biochemical processes [?, ?].

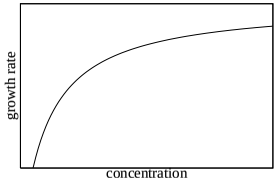


Figure 4.3: Schematic representation of the bacterial growth rate vs the nutrient concentration in the environment.

Motility. Procaryotes move in aquatic environments by rotating their flagella, which are rather rigid membrane-bound helical protein polymers. Bacteria can have either only one flagellum (monotrichous), a pair of flagella at the opposite cell poles (amphitrichous), clusters of flagella at the poles (lophotrichous), or the flagella can be uniformly distributed over the cell membrane (peritrichous).

Detailed microscopic investigations revealed that the direction of flagellar rotation determines the nature of bacterial movement. A single, polar flagellum rotates counterclockwise (when viewed from outside the cell) during normal forward movement

while the cell body rotates in the opposite direction. In general, multiflagellated cells operate in a similar fashion: each flagellum rotates counterclockwise, and they form a bundle that propels them forward. This forward motion is interrupted by short intervals of "*tumbling*" when most flagella rotate in the opposite direction disrupting the bundles which results in a random change in the orientation of the cell.

A very different type of motility, the so called *gliding* is employed by many strains when moving on surfaces. Surprisingly little is known about this type of motion as there are no visible cellular structures associated with it. In fact, gliding motility varies greatly in the nature of the motion which may indicate that more than one mechanism for gliding motility exists. E.g., bacteria such as *Myxococcus* and *Flexibacter* glide along in a direction parallel to the long axis of the cells. Others display a screwlike motion (e.g., *Saprospira*) or even move in a direction perpendicular to the long axis (*Simonsiella*). Gliding bacteria secrete viscous extracellular glycoconjugates (slime) but it does not appear to propel them directly; rather, it probably attaches them to the substrate and lubricates the surface for more efficient movement. There is some evidence that differences in surface tension can propel *Myxococcus xanthus*, which may secrete a surfactant at its end opposite to the direction of the movement.

Chemotaxis. Bacteria rarely move aimlessly since they are attracted by nutrients like sugars and amino acids and are repelled by many harmful substances and metabolic waste products. They can also respond to other environmental factors like temperature, light or oxygen concentration. This movement toward chemical attractants or away from repellents is known as *chemotaxis*.

The chemotactic response is a stochastic process, since bacteria exposed to chemical gradients modulate their tumbling frequency. Tumbling is repressed when moving towards higher concentration of chemoattractants, i.e., they continue to move forward for a more extended period of time. In fact, the swimming behaviour is influenced by changes of the chemical concentration *in time*: The molecular machinery in a bacterium compares its current environment with that experienced a few moments previously. The underlying biochemical processes have been identified in detail for *Escherichia coli*.

Morphology diagram

The number of environmental parameters influencing colony development (temperature, humidity, chemical composition of the substrate, etc.) is rather large, and various sets of parameters can give rise to different morphologies even for the very same bacterial strain. In systematic experiments aimed at exploring the effect of environmental factors on colony behaviour, the concentration of the *agar* and *nutri-*

ents were changed systematically in the substrate [?, ?, ?]. The former component is related to the consistency of the gel and thus, to the motility of the bacteria and the diffusibility of the nutrients. As we have seen, the nutrient concentration determines the proliferation rate of the cells (Fig. 4.3). As a result of these experiments, *morphology diagrams* were constructed in which a characteristic colony shape was assigned for each pair of the parameters.

The first such experiments have been performed with a wild type *Bacillus subtilis* strain [?, ?] (later denoted as OG-01 [?], Fig. 4.4a), a non motile mutant of the same strain (OG-01b [?], Fig. 4.4b) and a strain[1] [?] denoted after its characteristic morphology as *Paenibacillus dendritiformis* (Fig. 4.5). The morphology diagrams of these three strains show certain similarities. The growth is compact at high concentrations of the nutrients and becomes branching under nutrient deficient conditions. However, the diagrams do differ in many important aspects.

P. dendritiformis colonies exhibit dense branching morphology (*DBM*) characterised by thin, radial branches below 1g/l peptone concentration. When grown on hard agar (above 20g/l), the branches are *curved*. In contrast, *B. subtilis* OG-01 cells are non-motile on hard agar. On softer agar, the colony can display periodic growth patterns and *DBM*, but not at low nutrient concentrations. These differences demonstrate that certain microbiological features do have a strong influence on the colony morphology, hence, it is not determined entirely by the environmental conditions. At the same time, the pure existence of such diagrams and the reproducibility of the patterns suggests that the behaviour of the colonies is still determined by quantitative laws.

Statistical measures of morphology

To distinguish the various stochastic structures and classify their morphology appropriate statistical measures are needed. One widely applied measure is the *density-density correlation* function defined as the conditional probability

$$C_\varrho(r) \equiv \frac{\langle \varrho(r\vec{e}+\vec{a})\varrho(\vec{a})\rangle_{\vec{e},\vec{a}}}{\langle \varrho(\vec{a})\rangle_{\vec{a}}} \tag{4.1}$$

where $\vec{e}$ is an arbitrary unit vector and $\varrho(\vec{a})$ denotes the "density" of the object at the point $\vec{a}$: ϱ is 1 if $\vec{a}$ belongs to the object and zero otherwise. We use the notation $\langle f(\vec{a})\rangle_{\vec{a}}$ for the average of any quantity f over all possible values of $\vec{a}$. Then, by definition, $C_\varrho(0) = 1$ and if the object has a finite size L, then $C_\varrho(r > L) = 0$.

[1]This strain was first thought to be a variant ("$\mathcal{T}$ morphotype") of *B. subtilis* 168 [?], but recently it has been genetically identified as a new strain [?] coexisting in the original stock with *B. subtilis* [?].

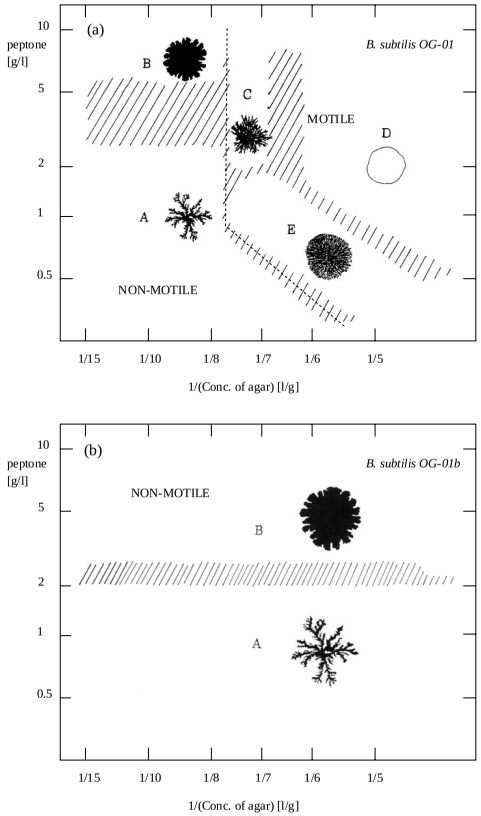


Figure 4.4: Morphology diagram of the strain *Bacillus subtilis* OG-01 (a) and a non motile mutant OG-01b (b) as a function of peptone and agar concentration. The dashed line in (a) indicates the boundary of the active movement of bacterial cells inside the colonies. The morphologies are classified as follows: fractal (A), compact with rough boundary (B), branching with periodic growth phases (C), compact with diffuse boundary (D) and dense branching (E). In the case of the non motile strain the regions A and B seen in (a) expand laterally, while regions C, D and E disappear. (After [?].)

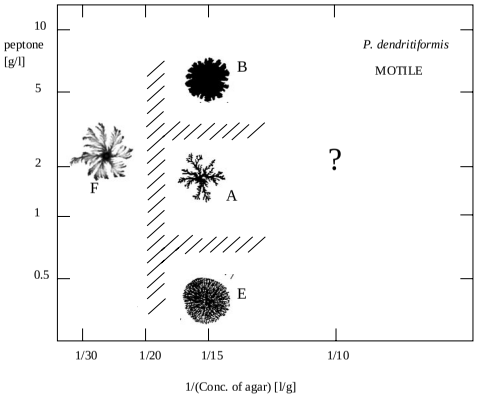


Figure 4.5: Morphology diagram of *Paenibacillus dendritiformis* ($\mathcal{T}$ morphotype) colonies as a function of peptone and agar concentration, drawn using data published in [?]. On hard substrate a new, twisted morphology appears (F). On soft agar even the *phenotype* of the bacteria changes as discussed in Sec.4.1.5.

In many cases C_ϱ decays as a power-law

$$C_\varrho(r) \sim r^{-\alpha} \tag{4.2}$$

for $1 \ll r \ll L$. Within these limits such objects are *scale-invariant* (fractal), since the morphology, i.e., the correlation function (4.2), is invariant under the $r' = br$ rescaling as

$$C'_\varrho(r') = C_\varrho(br) \sim b^{-\alpha} r^{-\alpha} \sim C_\varrho(r). \tag{4.3}$$

where C'_ϱ denotes the correlation function of the rescaled object. As described in Sec 2.1, the fractal dimension D is related to the exponent α as $D = 2 - \alpha$ [?]. The calculated fractal dimensions of *P. dendritiformis* colonies are shown in Fig. 4.6. Since both the compact and *DBM* morphologies are characterised by $D \approx 2$, in some cases further measures complementing C_ϱ are needed.

To characterise the irregular surface of compact objects (Fig 4.7), the *height-height correlation* function can be used, defined as

$$C_h^2(x) \equiv \left\langle [h\,(x+a) - h(a)]^2 \right\rangle_a, \tag{4.4}$$

where the surface is denoted by $h(x)$, and it is measured as a distance from a certain line or point of reference. Again, the quantity $C_h(x)$ has a clear meaning: it is the

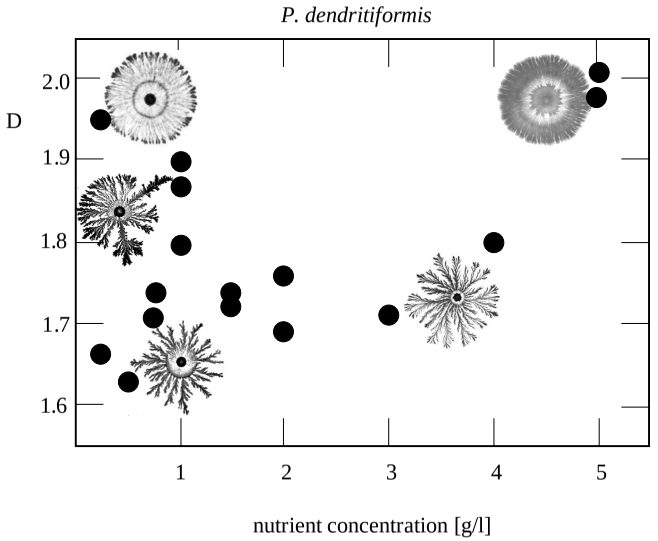


Figure 4.6: Quantitative characterisation of the morphology of *P. dendritiformis* colonies grown on substrates containing 20 g/l agar and various amount of nutrients. The filled circles show the fractal dimension (D) of the colonies vs the initial nutrient concentration. Representative colony images are placed in the vicinity of the position corresponding to their fractal dimension and growth condition. On nutrient rich substrates the colonies are compact ($D \approx 2$), while lowering the nutrient concentration fractal branches develop ($D \approx 1.7$). Below 1 g/l peptone concentration the fractal patterns coexist with the dense branching morphology characterised by thin radial branches and $D \approx 2$. This coexistence of the morphologies can sometimes be observed within the same colony as the leftmost inset demonstrates. The formation of these remarkable patterns is discussed in Sec. 4.1.4.

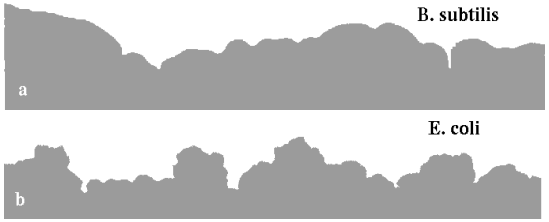


Figure 4.7: Typical rough colony boundaries of various bacterial strains (a: *Bacillus subtilis* 168, after [?] and b: *Escherichia coli* , after [?]) grown on a solid substrate rich in nutrients. a:

expected change in h over a distance of x. Another possibility is to calculate the *average standard deviation* w of the surface in a window of linear size x, i.e.,

$$w^2(x) \equiv \left\langle \left\langle h^2(a) \right\rangle_{u \le a \le u+x} - \left\langle h(a) \right\rangle^2_{u \le a \le u+x} \right\rangle_u . \tag{4.5}$$

Fig. 4.8 shows some typical examples of the correlation functions calculated for the boundary of bacterial colonies. In many cases, irrespective of the bacterial strain,

$$w(x) \sim C_h(x) \sim x^H \tag{4.6}$$

was found with

$$H = 0.8 \pm 0.05 \tag{4.7}$$

within quite a broad interval of x, i.e., the interface is a *non-trivial self-affine* curve with long range correlations decaying as a power-law. The insensitivity of H to the underlying biological system gives the major support for the following models predicting "universal" behaviour. As it was mentioned in the introduction, in our context *universality* means that the large scale patterns are self-organised, and not influenced by many details of the underlying systems.

4.1.3 Compact morphology

As was shown seen in the previous section, if the nutrients are abundant in the substrate then *compact* colonies with either smooth or irregular perimeter are formed. In this section we investigate this simplest case of colony morphogenesis in detail. The results will serve as a starting point for the studies of more sophisticated colonies where such detailed mathematical analysis cannot be carried out.

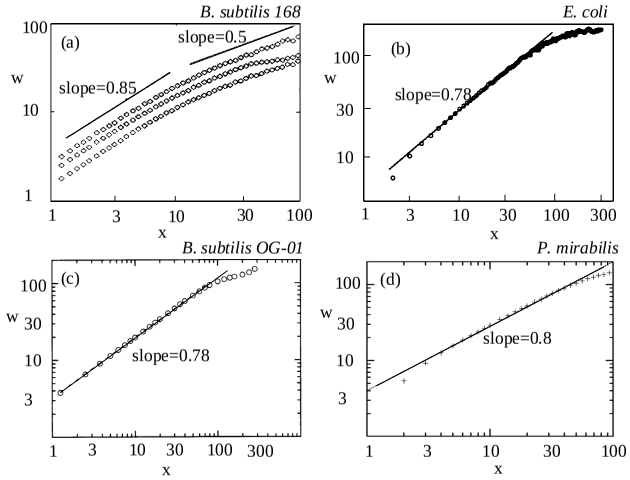


Figure 4.8: The average standard deviation w of surface segments of length x calculated for colonies of *B. subtilis* 168 (a, after [?]), *Escherichia coli* (b, after [?]), *Bacillus subtilis* OG-01 (c, after [?]) and *Proteus mirabilis* (d, calculated by the author from images published in [?]). The linear regime on the double logarithmic plots indicates the scaling behaviour (4.6). The plotted values are given in arbitrary units (pixels).

Spreading of bacteria: The Fisher-Kolmogorov equation

In the simplest example of bacterial spreading, nutrients are abundant and the substrate is wet enough to allow bacterial motion. Under such circumstances the cells proliferate and migrate across the dish within a few hours (see, e.g., region D in Fig. 4.4a). Videomicroscopic tracking of the bacteria revealed that their trajectories are *random walks* [?]. Therefore, the interaction between the cells must be negligible and the time dependence of the bacterial density ϱ (number of cells in a unit surface are) can be described by the Fisher-Kolmogorov equation [?]

$$\partial_t \varrho = D_\varrho \nabla^2 \varrho + f(\varrho). \tag{4.8}$$

The first term at the rhs of (4.8) describes the random *translocation* of bacteria. The diffusion coefficient D_ϱ is related to the microscopically measurable mean squared displacement $d^2(t)$ of the individual cells during a time period t as

$$\overline{d^2(t)} = 2D_\varrho t, \tag{4.9}$$

where the overline represents ensemble average among cells.

The bacterial *multiplication* is incorporated in f. When the bacterial density is small, the cells proliferate with a fixed rate resulting in an exponential growth. Thus, $f(\varrho) = r(c)\varrho$ for small ϱ, where r depends on the local nutrient concentration c. As we discussed in Sec. 4.1.2, $r \sim c$ holds for small values of c. In practice, even with unlimited nutrient supply, the cell density cannot increase above a certain threshold ϱ_*, e.g., because of the accumulation of toxic metabolites. Throughout this section we measure the cell density in units such that $\varrho_* = 1$. Thus, the growth rate must decrease as $\varrho \to 1$, hence $f(1) = 0$. The particular form of f will be unimportant, but where it is needed, we use the logistic function $f(\varrho) = r\varrho(1-\varrho)$ as an example. Although equation (4.8) is very simple, we discuss it here because it serves as a prototype of such problems, and it is still far from being trivial even in one dimension.

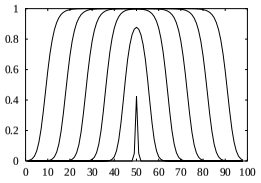


Figure 4.9: Typical result of the numerical integration of the Fisher equation (4.8) starting from a localised perturbation ($D_\varrho = 1$, $f(x) = x - x^2$, i.e., $r = 1$). The $\varrho(x,t)$ curves are plotted for $t = 0, 5, 10, 15, 20$ and 25. The domain grows with a stationary speed of $v = 2$.

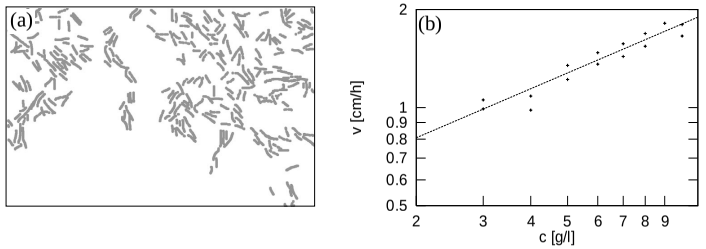


Figure 4.10: a: Microscopic image of the fuzzy interface of a *B. subtilis* colony. The scale of the figure is 0.2 mm in width. b:The growth speed v of the interface vs the nutrient concentration. The dashed line shows the expected square root-like behaviour (After [?]).

Numerical integration of (4.8) starting from a localised (e.g., Gaussian) perturbation results in a growing domain expanding with a constant speed $v \approx v_*$ where

$$v_* = 2\sqrt{rD_\varrho} \tag{4.10}$$

(see Fig. 4.9). When, in order to calculate v, the expanding domain of $\varrho(x,t) \sim 1$ is written in a moving frame of reference as $\tilde{\varrho}(u,t) = \varrho(x,t)$ where $u = x - vt$ and without loss of generality $v > 0$, $\varrho(-\infty) = 1$ and $\varrho(\infty) = 0$ is assumed one obtains

$$\partial_t \tilde{\varrho} = D_\varrho \tilde{\varrho}'' + v\tilde{\varrho}' + f(\tilde{\varrho}), \tag{4.11}$$

where the prime denotes differentiation with respect to u.

Surprisingly, Eq. (4.11) has a *stationary* solution for *any* value of $v \geq v_*$, i.e. v can not be derived by the stationarity condition imposed on $\tilde{\varrho}$. This feature constitutes the "velocity selection problem" associated with similar reaction-diffusion equations. In the case of Fisher equation the dynamics selecting the front propagating with $c = c_*$ from the set of possible stationary solutions of (4.11) was published in [?] and discussed in detail in [?].

Experimental results. The velocity selection in real bacterial spreading has been experimentally investigated with *B. subtilis* grown in region E of the morphology diagram (Fig.4.4a). The spreading speed of the colony was measured for various nutrient concentrations [?], which determines the growth rate as $r \sim c$. The bacterial

diffusion parameter D_ϱ was calculated using (4.9) and was found to be fairly constant while changing c within an order of magnitude. Thus, (4.10) predicts

$$v \sim \sqrt{c}. \tag{4.12}$$

As described in [?], colonies started growing about eight hours after the inoculation. As the colony expanded, its contour became circular, its growth speed increased and asymptotically approached a finite value. This asymptotic growth speed was measured when the diameter of the colony reached 50 mm. The obtained experimental data are shown in Fig. 4.10, and within the limited, experimentally accessible range are in agreement with (4.12).

Cell-cell interactions. In most interesting cases, however, individual bacteria are not independent during the spreading of the colony. E.g., when the cells are not motile, then there is an abrupt change in the cell density at the border of the colony as Fig. 4.11 shows. In this case the propagation of the boundary is driven mainly by the expansion of the cell volumes *inside the colony*. If the bacteria cannot expand to their preferred size then they exert mechanical pressure on their surrounding as discussed in [?]. Thus, for large cell densities this pressure p is proportional to $\varrho - \varrho_0$, where ϱ_0 is the threshold density determined by the close-packed configuration of the cells on the substrate. Since the attachment of the bacteria to the substrate cannot withstand arbitrarily large stresses, thus, at least for large p, the displacement of the cells will be proportional to ∇p as described by

$$v = D_0 \nabla(\varrho - \varrho_0). \tag{4.13}$$

The D_0 prefactor plays a similar role in the following as D_ϱ in (4.8). Hence, in this case the time development of the bacterial density is described by a modified form of the Fisher-Kolmogorov equation

$$\partial_t \varrho = \begin{cases} D_0 \nabla^2 \varrho + f(\varrho) & \text{for } \varrho > \varrho_0 \\ f(\varrho) & \text{otherwise.} \end{cases} \tag{4.14}$$

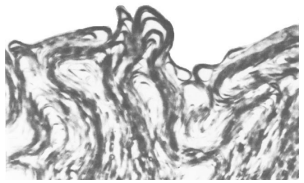


Figure 4.11: Long bundles of chains of individual cells consist the colony of the *B. subtilis* OG-01 strain grown on hard agar. Note the abrupt change in cell density at the boundary. (After [?]).

Again, in one dimension (4.14) gives travelling fronts with a constant velocity v: In the moving frame of reference $u = x - vt$ for the stationary solution of (4.14) we obtain

$$D_0 \tilde{\varrho}'' + v\tilde{\varrho}' + f(\tilde{\varrho}) = 0 \tag{4.15}$$

inside the "colony", i.e. for $u < 0$. At $u = 0$ the boundary conditions are

$$\tilde{\varrho}(0) = \tilde{\varrho}_0 \tag{4.16}$$

$$v\tilde{\varrho}_0 = D_0 \tilde{\varrho}'(0), \tag{4.17}$$

which ensures the conservation of $\tilde{\varrho}$. In such cases, however, as Figs. 4.7 and 4.11 demonstrate the colony boundary does not remain smooth as the growth is strongly influenced by the various inhomogeneities of the environment.

Self-affine interface roughening

The experimental finding that the boundary of bacterial colonies shows self-affine properties was not completely unexpected, since one of the earliest methods to generate self-affine objects had related biological motivation. In the variants of the Eden model – published in 1961 [?] – some assembly of "cells" is grown on a lattice with the following rules (see Fig. 4.12). In each step of the process one of the lattice sites next to the "populated" areas is chosen randomly and occupied.[2] Although this stochastic dynamics is very simple, the resulting objects are quite irregular as Fig. 4.13 shows. Large scale numerical studies revealed that for the surface of the two dimensional Eden "colony" (4.6) holds with $H = 1/2$ [?].

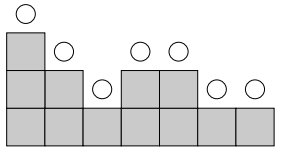


Figure 4.12: Generation of the Eden cluster. In each step of the process one of the lattice sites (marked by circles) next to the "colony" (gray) is selected with uniform probability and filled.

Figure 4.13: A typical colony in the Eden model grown on a strip of 256 lattice units. (After [?])

[2] In fact, the model has slightly different versions assigning different probabilities for occupying the perimeter sites depending on the local configuration of the colony. However, the statistical features of the developing clusters are the same for all variants of the model in the asymptotic limit [?].

The Kardar-Parisi-Zhang (KPZ) equation. To understand this frequent occurrence of self-affine curves in surface roughening phenomena, an approach based on stochastic differential equations can be applied. It turns out to be fruitful to simplify the problem defined by (4.14) and focus our attention upon the interface itself. The motion of interfaces in the presence of external fluctuations or in an inhomogeneous environment can be described [?] by

$$\partial_t h = \nu \partial_x^2 h + \frac{\lambda}{2}(\partial_x h)^2 + v + \eta, \tag{4.18}$$

where the terms at the rhs of (4.18) have the following physical meaning:

The surface tension term, $\nu\partial_x^2 h$, tends to smoothen the interface and does not permit discontinuities (large jumps) in h.

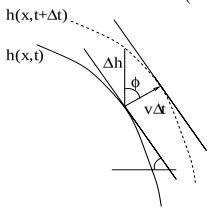


Figure 4.14.

The nonlinear $\lambda(\partial_x h)^2 + v$ terms reflects the *isotropy* of the growth. If the displacement is *perpendicular* to the interface, then using the notations of Figure 4.14,

$$\Delta h \approx v\Delta t\sqrt{1+tg^2\phi} \approx v\Delta t + \frac{v\Delta t}{2}(\partial_x h)^2 \tag{4.19}$$

holds for small slopes.

Finally, $\eta(x,t)$ represents some sort of *noise*. Various type of models can be distinguished based on the assumed properties of this term: the simplest choice is a noise *uncorrelated* both spatially and temporally defined as

$$C_\eta(x,t) \equiv \langle \eta(x',t')\eta(x+x',t+t')\rangle_{x',t'} = 2D\delta(x)\delta(t), \tag{4.20}$$

where the average $\langle ...\rangle_{x',t'}$ is calculated over all possible values of x' and t', and the $\delta(x)$ measure satisfies $\int_{-\infty}^{\infty}\delta(x) = 1$ and $\delta(x) = 0$ for all $x \neq 0$. The infinite standard deviation ($C_\eta(0,0) \equiv \langle\eta^2\rangle$) is only a formal consequence of the continuum limit[3].

[3] If x and t are discrete variables, then (4.20) can be written in the more obvious form of $C_\eta(x,t) = \delta_{x0}\delta_{t0}$, where the Kronecker symbol $\delta_{ij} = 1$ if $i = j$ and zero otherwise. The continuum limit (4.20) can be derived from the requirement that the standard deviation of the stationary noise averaged over a finite interval must be finite and proportional to the length of the interval, i.e.,

$$\left\langle \left(\int_x^{x+\ell} \eta(a)da \right)^2 \right\rangle_x = \left\langle \int_0^\ell da \int_0^\ell db\eta(x+a)\eta(x+b) \right\rangle_x =$$
$$\int_0^\ell da \int_0^\ell db \langle \eta(x+a)\eta(x+b)\rangle_x = \int_0^\ell da \int_0^\ell db\, C_\eta(a-b) \sim \ell. \tag{4.21}$$

In the case of the Eden model the noise comes from the random choice of growth segments, which is an uncorrelated process, by definition.

There are quite a few analytical methods available to solve (4.18). Probably the simplest way is based on the calculation of the probability distribution of the steady-state height fluctuations [?]. If the stochastic variables $\vec{y}(t) = \{y_i(t)\}$, $i = 1,\ldots,n$ are governed by the Langevin equation

$$\frac{d\vec{y}}{dt} = \vec{A}(\vec{y}) + \vec{\eta}(t) \tag{4.22}$$

where $\vec{A}$ is any – usually non-linear – vector-vector function and $\vec{\eta}(t)$ is an uncorrelated noise

$$C_{ij}(t) \equiv \langle \eta_h(t')\eta_j(t+t')\rangle_{t'} = 2D\delta(t)\delta_{ij} \tag{4.23}$$

in analogy with (4.20), then the time evolution of the probability distribution $P(\vec{y},t)$ is given by the Fokker-Planck equation

$$\partial_t P(\vec{y},t) = -\mathtt{div}[\vec{A}(\vec{y})P(\vec{y},t) - D(\nabla P)(\vec{y},t)] \tag{4.24}$$

where $\mathtt{div}\vec{A} = \sum_{i=1}^n \partial_{y_i} A_i$ and $(\nabla P)_i = \partial_{y_i} P$.

Eq. (4.24) can be generalised to describe the time evolution of the probability density of a continuous $h(x,t)$ function instead of a finite number of variables represented by $\vec{y}(t)$. The gradient, which assigns a vector field to a scalar field can be replaced by the functional derivative $\mathcal{D}$ which assigns a function to an $I(f)$ functional as

$$I(f+\epsilon) = I(f) + \int (\mathcal{D}I)(x)\epsilon(x)dx \tag{4.25}$$

for arbitrary $\epsilon(x)$, in the $\epsilon(x) \to 0$ limit. In a similar fashion, the derivative $\mathcal{DF}$ of an operator $\mathcal{F}$ is defined through

$$\lim_{\epsilon(x)\to 0}[\mathcal{F}(f+\epsilon)](x) = [\mathcal{F}(f)](x) + \int [(\mathcal{DF})(f)](x,y)\epsilon(y)dy, \tag{4.26}$$

since for a fixed x the $[\mathcal{F}(f)](x)$ quantity is a functional of f. Finally, the generalisation of the $\mathtt{div}$ operation is

$$(\mathtt{div}\mathcal{F})(f) = \lim_{\delta\to 0}\int dx \int_{x-\delta}^{x+\delta} dy\ [(\mathcal{DF})(f)](x,y). \tag{4.27}$$

With these notations (4.24) can be written in the form of

$$\partial_t P(h,t) = -\mathtt{div}\left[\mathcal{F}(h)P(h) - D(\mathcal{D}P)(h)\right], \tag{4.28}$$

where $\mathcal{F}(h) = \partial_x^2 h + \frac{\lambda}{2}(\partial_x h)^2$.

In general, it is very hard to solve equations like (4.28). However, in our particular case one can show that

$$P(h) = \exp\left(-\frac{1}{2}\int (\partial_x h)^2\right) \tag{4.29}$$

is a stationary solution, i.e., in the stationary regime of (4.18) the probability of finding a particular $h(x)$ curve is proportional to $P(h)$ as given in (4.29). To see this let us substitute (4.29) into (4.28). In the $\epsilon \to 0$ limit

$$P(h+\epsilon) = \exp\left[-\frac{1}{2}\int (\partial_x h)^2 - \int (\partial_x h)(\partial_x \epsilon)\right] = P(h) - P(h)\int (\partial_x h)(\partial_x \epsilon) \tag{4.30}$$

holds. Performing a partial integration in the last term of (4.30) and using definition (4.25)

$$(\mathcal{D}P)(h) = P(h)\partial_x^2 h \tag{4.31}$$

is obtained. Thus the second term in the rhs of (4.28) cancels the λ-independent contribution of the first term. The remaining, λ-dependent expression is

$$-\frac{\lambda}{2}\mathtt{div}\left[(\partial_x h)^2 P(h)\right] = -\frac{\lambda}{2}\int (\partial_x h)^2 \mathcal{D}P(h) - \frac{\lambda}{2}P(h)\mathtt{div}(\partial_x h)^2. \tag{4.32}$$

By substituting (4.31) into (4.32) and performing partial integration one can show that the first term at the rhs of (4.32) is identically zero. A similar statement holds for the remaining term as well. Based on (4.26),

$$\left[\mathcal{D}(\partial_x h)^2\right](x,y) = 2(\partial_x h)(x)\delta'(x-y). \tag{4.33}$$

Therefore,

$$\begin{aligned}\mathtt{div}(\partial_x h)^2 &= 2\int dx \int dy\ (\partial_x h)(x)\delta'(x-y)\\ &= -2\int dx \int dy\ (\partial_x^2 h)(x)\delta(x-y)\\ &= -2\int dx (\partial_x^2 h)(x) = 0.\end{aligned}$$

The probability distribution (4.29) tells us that the consecutive slopes of h are independent variables, therefore in the $t \to \infty$ limit $h(x)$ is a random walk characterised by $H = 1/2$. Moreover, the solution (4.29) remains the same if the $(\partial_x h)^2$ term in (4.18) is replaced with any linear combination of nonlinear terms of the type

$(\partial_x h)^{2k}$ with a positive integer k. This robust analytical result for surface roughening driven by temporally uncorrelated fluctuations illustrates the best what we mean under universality and explains why the various versions of the Eden model with slightly different rules give the same asymptotic behaviour. Unfortunately, the $H = 1/2$ result is not compatible with the experimental finding (4.7).

Quenched noise. For bacterial colonies – and for many other physical systems as well – the assumption of an uncorrelated noise is not realistic. If the major factor modulating the interface velocity is the local inhomogeneity of the agarose gel, then this might be modelled as uncorrelated in space, but it is certainly *correlated in time*. In fact, η is then a function of the position of the interface

$$\eta(x,t) = 2D\tilde{\eta}(x, h(x,t)) \tag{4.34}$$

with

$$C_{\tilde{\eta}}(x,y) = \Delta(x)\Delta(y), \tag{4.35}$$

where $\Delta(u) \sim 1$ holds for some finite interval of u around zero.

If such a "quenched" noise is incorporated into Eq. (4.18), then the behaviour of the equation is significantly changed. By appropriate choice of the time and length units and using (4.19) all the parameters λ, ν and v can be transformed out, i.e., it is sufficient to investigate the $\lambda = \nu = v = 1$ case

$$\partial_t h = \partial_x^2 h + \sqrt{1 + (\partial_x h)^2} + \eta, \tag{4.36}$$

where the full non-linear term representing the lateral growth is included. The only remaining parameter of the problem is D in (4.34), which determines the average magnitude of the noise as $\sqrt{C_{\tilde{\eta}}(0,0)} = \sqrt{2D}$. Now, two extreme cases can be distinguished.

If $D > D_* \sim 1$ then the interface encounters points where it becomes pinned and it will experience that particular condition for a long time until it is released by the additional "pulling" force caused by the advancement of the neighbouring segments of the interface. If the density of such pinning sites is high enough then eventually the propagation of the whole surface can be blocked. The shape of the frozen interface is then determined by the distribution of the pinning sites in the substrate and becomes independent of the dynamics of the growth. In fact, the surface roughening can be mapped onto a specific percolation problem (See, e.g. [?]): Since the interface is blocked at each possible segment, it is adjacent to a cluster of pinning sites. Due to

the single-valuedness of h, the cluster is *directed*: it spans in a certain direction, turns up and down, but never back.[4]

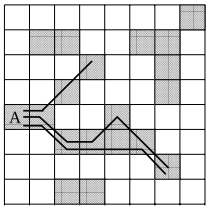


Figure 4.15: Examples of directed percolation clusters leaving site A. Such clusters consist of filled neighbour sites and extend from left to right.

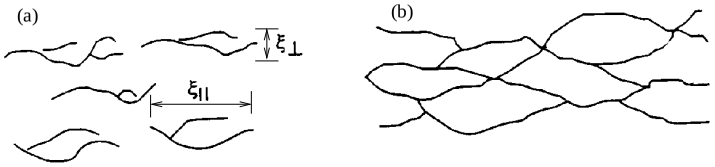


Figure 4.16: Directed percolation clusters below (a) and above (b) the critical density of the filled lattice sites at which a cluster spanning the entire system appears. Below this percolation threshold the directed percolation clusters have a typical length $\xi_{\parallel}$ (parallel to the distinguished direction) and a typical width $\xi_{\perp}$.

Such a *directed percolation cluster* is characterised by two correlation lengths, being parallel ($\xi_{\parallel}$) and perpendicular ($\xi_{\perp}$) to the interface (Figs. 4.15, 4.16). As exhaustive studies [?] revealed, these correlation lengths diverge in the vicinity of the threshold density of the pinning sites p_c as

$$\xi_{\parallel} \sim |p - p_c|^{-\nu_{\parallel}}, \qquad \xi_{\perp} \sim |p - p_c|^{-\nu_{\perp}} \tag{4.37}$$

with

$$\nu_{\parallel} = 1.733 \quad \text{and} \quad \nu_{\perp} = 1.097. \tag{4.38}$$

Complete blocking of the interface propagation appears when $\xi_{\parallel}$ becomes equal to the system size L. The width w of such a path is in the order of $\xi_{\perp}$, hence

$$L^H \sim w \approx \xi_{\perp} \sim |p - p_c|^{-\nu_{\perp}} \sim \xi_{\parallel}^{\nu_{\perp}/\nu_{\parallel}} \approx L^{\nu_{\perp}/\nu_{\parallel}}, \tag{4.39}$$

[4]The description using single-valued functions can be valid only for process resulting in rather flat surfaces, i.e., for $H < 1$. Otherwise "overhangs" can be generated which are ignored here.

i.e., for the blocked interface one obtains

$$H = \frac{\nu_{\perp}}{\nu_{\parallel}} = 0.633 \tag{4.40}$$

in accord with the numerical simulation of Eq. (4.36) with $D \gg D_*$ [?].

In the $D \ll D_* \sim 1$ case the interface is never pinned and advances with a steady velocity. Since it experiences a fluctuating noise with some finite temporal correlations, in this limit the previously described (standard) KPZ equation can be applied resulting in interfaces with $H = 1/2$.

We are, however, interested in the moving, but noise dominated regime. In this case the interface consists of both blocked and freely moving, almost straight segments, thus one can expect a *higher* roughness exponent than that of the blocked interface. By extensive numerical simulations it has been shown that in this case the resulting surfaces (Fig. 4.17) have self-affine exponents ranging from 0.71 [?, ?] to 0.75 [?], which values are reasonably close to the experimentally obtained (4.7). The agreement between the experimental findings and the behaviour of Eqs. (4.18), (4.34) and (4.36) is even better than this. As recently has been reported in [?, ?], colonies grown on soft agarose gels (region D in Fig. 4.4a) – where the pinning of the interface is less dominant – indeed show standard KPZ-like behaviour characterised by $H = 1/2$.

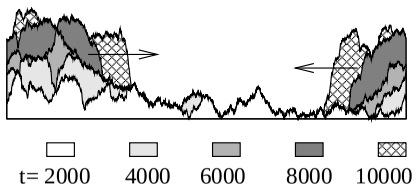
4.1.4 Branching morphology

If the experiments are performed on *nutrient-poor* agar substrates then complex, branching colonies develop as first reported by Fujikawa and Matsushita [?, ?]. In their study a few weeks' growth of a *Bacillus subtilis* colony resulted in beautiful branching structures as demonstrated in Fig. 4.18.

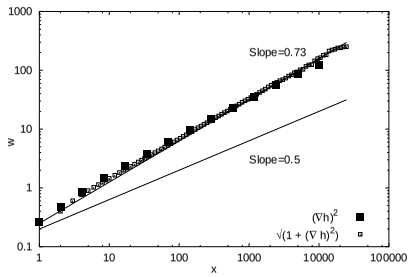
Similar patterns (Fig. 4.19) were observed in experiments with *Citrobacter*, *Escherichia*, *Klebsiella*, *Salmonella*, *Serratia* and *Proteus* strains as well [?]. According to a recent extensive study [?], many strains belonging to the *Bacilli* species also display tip splitting growth. For a – probably not complete – list of the reported strains involved in branching colony formation, see Table 4.1 which demonstrates that branching colony formation, though not exhibited by *all* the strains, is really not rare under nutrient-deficient conditions.

Many of the reported findings can be explained assuming that the growth of the colony is *diffusion-limited*, e.g., the multiplication of the bacteria is determined by the locally available amount of nutrients: At the beginning of colony development, the local nutrient supply is sufficient to maintain the growth. After further bacterial multiplication, however, nutrient deprivation will progress in and around the colony.

(a)



(b)



(c)

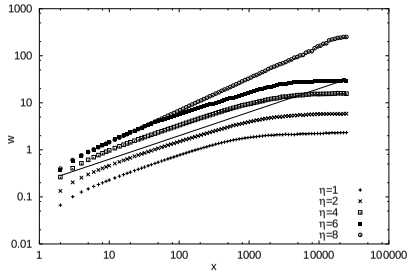


Figure 4.17: Results of the numerical integration of the quenched KPZ equation in the noise dominated regime. (a) Subsequent snapshots of the evolving surface. The interface consists of blocked segments and quite straight, moving parts marked by arrows. (b) The surface width w vs the length x over which its average was calculated. The straight part of the double logarithmic plot indicates self-affine scaling with an exponent $H = 0.73$. The universality of H is demonstrated by simulations with two distinct nonlinear terms. (c) For small noise amplitudes $H = 1/2$ is recovered (solid line). (After [?, ?, ?].)

Bacterial strain	Substrate	Reference
Bacillus alvei		
PB1875	BP	[?]
NRRL B-383		
Bacillus amyloliquefaciens		
H	BP	[?]
Bacillus circulans		
ATCC 4513	BP	[?]
NRRL B-378		
Bacillus licheniformis		
FDO1	BP	[?, ?]
Bacillus macerans		
BKM B-51	BP	[?]
Bacillus popillae		
NRRL B-2309	BP	[?]
Bacillus subtilis		
OG-01	BP	[?, ?, ?]
W23,BD170,BD79,SB25,IS58,	BP	[?]
IS56,YS1,RM125,OI1085,		
OI2836,OI1055,OI3180 *		
Citrobacter freundii		
NPC 3003	BTB	[?]
Escherichia coli		
ATCC 25922	VB	[?]
W 3110	VB	
	TS	
Klebsiella pneumoniae		
Ful	VB	[?]
Paenibacillus dendritiformis		
"T morphotype"	BP	[?, ?]
Proteus mirabilis		
NPC 3007	NAS	[?]
Salmonella anatum		
KS 200	VB	[?]
KS 700	VB	
Salmonella typhimurium		
LT2	VB	[?]
	BTB	
ATCC 14028	VB	
Serratia marcescens		
NS 25	VB	[?]
NS 38	VB	

Table 4.1: Bacterial strains developing branching colonies under nutrient-deficient conditions. BP: Bacto-Peptone agar, $\approx$ 1 g/l peptone; VB: Vogel-Bonner agar, $\approx$ 0.4% glucose; BTB: bromthymol blue agar, TS: tryptic soy agar, NAS: soft nutrient agar, $\approx$ 0.4% glucose. (*)As suggested in [?], these *B. subtilis* strains probably coexist with group I *Bacilli* which are responsible for the tip-splitting growth.

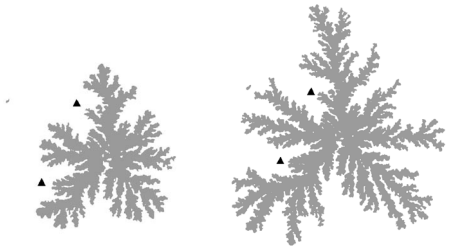


Figure 4.18: The growth of a *Bacillus subtilis* colony, under nutrient-poor conditions. The pictures were taken 8 (a) and 19 (b) days after inoculation. Many inner branches, typical examples of which are pointed by black triangles, are seen to stop growing. (After [?].)

Thus, bacterial growth will depend on the diffusive transport from distant regions of the Petri dish. This assumption was supported by experiments with non motile *B. subtilis* bacteria seeded into strongly inhomogeneous initial nutrient distribution [?]. If the nutrient was localised in a certain region of the agar gel, then growth occurred towards that region only (Fig. 4.20). Since the particular cells were non-motile, this directed growth could not be a chemotactic effect.

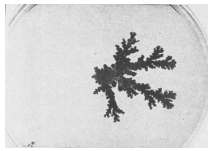


Figure 4.20: A colony of *Bacillus subtilis* showing a clear tendency to grow toward the nutrient (peptone), which was initially put at the right half of the dish. (After [?].)

Mullins-Sekerka instability in diffusion-limited systems

Diffusion-limited pattern formation has been extensively studied in non-living systems (Fig. 4.21) because of its relevance in processes like crystal growth, electrochemical deposition or viscous fingering (see, e.g., [?]). The growth of the objects in diffusion-limited systems is determined by an external field u (e.g., the concentration) as

$$v_n = \vec{n}\nabla u, \tag{4.41}$$

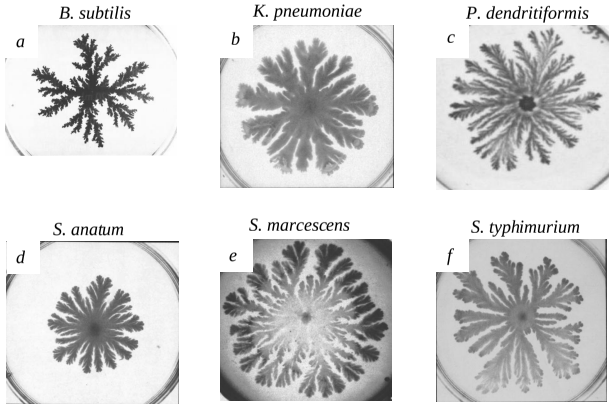


Figure 4.19: A few example of strains exhibiting tip-splitting growth. (a,e: after [?]; b,d: after [?]; c: after [?]; f: after [?])

where v_n and $\vec{n}$ denote the dimensionless velocity and the normal vector of the surface. The dynamics of the field is governed by the diffusion equation

$$\partial_t u = D\nabla^2 u \tag{4.42}$$

with boundary conditions $u(x \to \infty) = u_\infty = const$ and $u(x) = u_\Gamma(x)$ at the interface. This latter boundary condition can incorporate some features of the microscopic dynamics, e.g., the anisotropy of the crystal structure. For the most simple, isotropic case u_Γ can be written in the form of

$$u_\Gamma = d_0\kappa, \tag{4.43}$$

where d_0 is proportional to the capillary tension and κ is the curvature of the interface.[5] If the growth of the cluster is so slow that the field u can be considered stationary for a given configuration of the cluster then we arrive to the so called Laplacian growth. In such cases the field satisfies the Laplace-equation

$$0 = \nabla^2 u. \tag{4.44}$$

[5]As we will show, this dependence of the local curvature is needed to avoid the formation of arbitrary thin needles in the system. Instead of (4.43), one could also introduce a constraint limiting the maximal curvature of the interface.

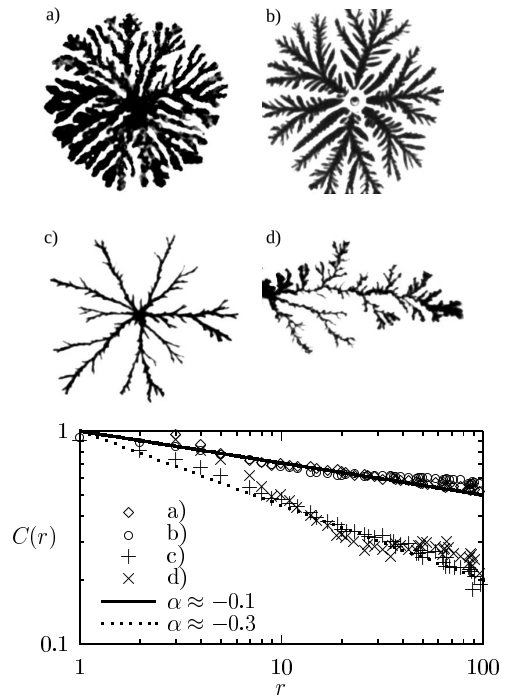


Figure 4.21: Irrespective of the underlying physical system, the morphologies resulting by diffusion-limited growth are very similar. The dense branching morphologies (a: crystallisation in amorphous $Al_{0.4}Ge_{0.6}$ thin films at 230°C, after [?]; b: viscous fingering in liquid crystal, after [?]) have a fractal dimension $D \approx 2$. In the case of Laplacian growth (c: viscous fingering in a non-newtonian liquid where the viscosity decreases with the shear [?]; and d: crystallisation in an amorphous $GeSe_2$ thin film at 220°C, after [?]) a sparse object develops with $D \approx 1.73$.

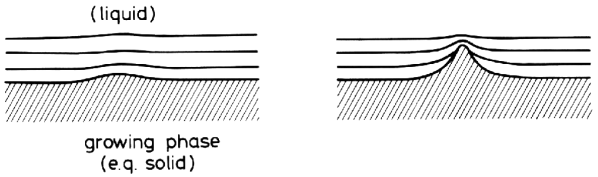


Figure 4.22: The Mullins-Sekerka instability. If a bump is created by some fluctuation then the gradient of the field is larger at the tip of the bump, so the tip's growth will be further enhanced.

Although Eqs. (4.41) and (4.42) are linear, the moving boundary condition for u at the interface makes the problem rather hard to study both analytically and numerically.

To show that the interface is *unstable* against perturbations [?] (Fig. 4.22), let us consider a growing circle of radius R in a diffusive field with a boundary condition (4.43) and $u(R_\infty) = 1$, where $R_\infty \gg R$. The field $u(r, \phi) = u_0(r)$ satisfies (4.44) which reads in polar coordinates as

$$u_0'' + \frac{u_0'}{r} = 0 \tag{4.45}$$

yielding

$$u_0 = 1 + B \ln \frac{r}{R_\infty}, \quad \text{with} \quad B = \frac{d_0/R - 1}{\ln R/R_\infty} > 0, \tag{4.46}$$

and the velocity of the interface is

$$\frac{dR}{dt} \equiv V(R) = u_0'(R). \tag{4.47}$$

After [?], the linear stability analysis of this system will be performed in the *quasistationary* approximation (for the general solution see, e.g. [?, ?]). If the growing surface is perturbed as $R(t) = R_0(t) + R_1(t, \phi)$, then u is changed as well. In general, (4.42) gives rise to memory effects: a deformation of the interface causes a perturbation of the diffusion field which, in turn, affects the motion of the interface at later times. However, if the interface moves slowly enough then it remains effectively stationary during the time needed for the relaxation of the diffusion field. Thus it is reasonable to solve the problem approximately by, first, solving the Laplace equation (4.44) with the appropriate boundary conditions, and then inserting this field into (4.43) to obtain the time development for R_1.

Now we write the perturbation in the form of

$$R_1(\phi,t) = \hat{R}_1(t)\cos m\phi, \tag{4.48}$$
$$u(r,\phi,t) = u_0(r,t) + \hat{u}_m(t)u_m(r,\phi) \tag{4.49}$$

with $\hat{R}_1 \ll R_0$, $\hat{u}_m \ll 1$ and $u_m = r^{-m}\cos m\phi$. Both u_0 and u_m is a solution of (4.44) and with this particular form the boundary condition (4.43) can be also satisfied. The curvature κ in polar coordinates is given as

$$\kappa = \frac{R^2 + 2R'^2 - RR''}{(R^2+R'^2)^{3/2}} \approx \frac{1}{R_0} - \frac{R_1''}{R_0^2}, \tag{4.50}$$

where now the prime denotes derivation in respect to ϕ. Thus in a linear approximation (4.43) reads as

$$d_0\left(\frac{1}{R_0} + \frac{\hat{R}_1 m^2\cos m\phi}{R_0^2}\right) = u(R(\phi),\phi), \tag{4.51}$$

where

$$u(R(\phi),\phi) \approx u(R_0,\phi) + (\partial_r u)(R_0,\phi)\hat{R}_1\cos m\phi \tag{4.52}$$

resulting in the

$$\left(\frac{m^2 d_0}{R_0^2} - V(R_0)\right)\hat{R}_1 = \frac{\hat{u}_m}{R_0^m} \tag{4.53}$$

relation between $\hat{R}_1$ and $\hat{u}_m$. Substituting (4.53) back into (4.41) and keeping the linear terms of $\hat{R}_1$ one finds

$$\frac{d\hat{R}_1}{dt} = \Lambda_m(R_0)\hat{R}_1 \tag{4.54}$$

with

$$\Lambda_m(R_0) = \frac{m}{R_0}V(R_0) + V'(R_0) - \frac{m^3 d_0}{R_0^3}. \tag{4.55}$$

Since Λ_m is the amplification rate of the perturbation, its sign determines the stability of the interface in respect with perturbations characterised with the $\frac{m}{R_0}$ wave number. For $\Lambda > 0$ the perturbation grows exponentially in time, while it dies out quickly if $\Lambda < 0$. It can be seen from Fig. 4.23 that there is a region where the amplification rate is positive. The upper cutoff is due to the surface tension represented by d_0. If $d_0 = 0$, the growth rate increases indefinitely for arbitrary short wavelengths, and the problem is ill-posed from the physical point of view.

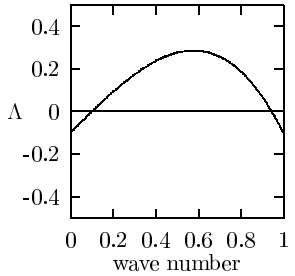


Figure 4.23: The growth rate Λ of a perturbation vs the wave number m/R_0 for $V = 1$, $V' = 0.1$ and $d_0 = 1$. Deformations with $\Lambda > 0$ grow in an unstable manner. We expect that the unstable interface is characterised by the curvature for which Λ is maximal.

Finally, we expect the dynamics to select the curvature for which Λ is maximal. Thus, to obtain tip splitting growth the particular shape of the Λ_m curve is not important until it has a positive maximum value at some finite wave number, which can explain the frequent occurrence of the tip-splitting morphology.

Diffusion-limited aggregation

The first model reproducing the branching fractal clusters characteristic for Laplacian growth is due to Witten and Sander [?]. Their model (DLA) is somewhat similar in spirit to the Eden model: Particles released one by one perform random walks until absorbed by a growing, non-moving cluster. When a particle is absorbed, far from the cluster a new particle is released. A typical example of the clusters obtained from computer simulations is shown in Fig. 4.24. As the power-law decay of the two-point correlation function indicates, the DLA clusters are scale invariant and are characterised by the fractal dimension

$$D_{DLA} = 1.715. \tag{4.56}$$

It is easy to see that the DLA model is indeed related to the diffusion-limited processes described above. Let us denote by $p(x,y)$ the probability of the event that – during a long enough time – a given particle passes the lattice site (x,y). If the cluster was not absorb the particles then $p(x,y)$ would be constant everywhere. Since the particles are absorbed at the surface, inside the cluster $p = 0$, while at large distances from the cluster $p = const$ holds. The growth probability of the various parts of the cluster is determined by p near the boundary. Since at the boundary $p = 0$, on the average the growth is proportional to ∇p at the surface. Finally, the stationary distribution of p satisfies the Laplace equation as shown by the local

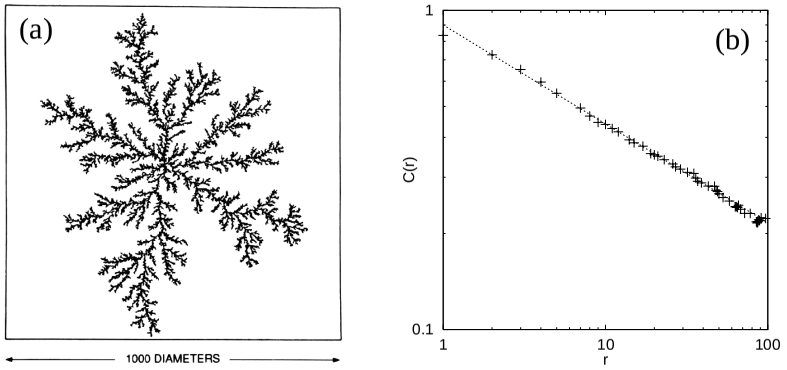


Figure 4.24: (a) Typical DLA cluster with 50,000 particles, reproduced from [?]. (b) The object is self-similar with $D \approx 1.7$ as the power-law decay (dotted line) in the density-density correlation function shows.

balance equation of the lattice version of the DLA model relating $p(x,y)$ to p at the neighbour sites as

$$p(x,y) = \frac{1}{4}\left[p(x-1,y) + p(x+1,y) + p(x,y-1) + p(x,y+1)\right], \qquad (4.57)$$

since the probability of stepping in any direction is 1/4.

Many theoretical attempts has been made to understand how does the nontrivial fractal structure of Fig. 4.24 emerge from the simple models of diffusion-limited growth. Although there is no established theory yet, the most promising approaches are based on some sort of *renormalisation transformation*, see, e.g. [?].

Many natural processes can be represented better by growing clusters from a pool of random walkers that initially occupy a finite fraction p of the lattice. In the numerical simulations the particles are selected randomly and moved to a nearest neighbour site. If the new site is on the perimeter of the growing cluster then the particle becomes the part of the cluster.

The clusters generated by this model also have DLA-like structure on short length scales $1 \gg r \gg \xi$ and are uniform for $\xi \gg r$ (Fig. 4.25). The crossover length ξ is determined by imposing the condition that the average density of the cluster and that of the pool of particles must be the same. At the same time ξ is also the width of the

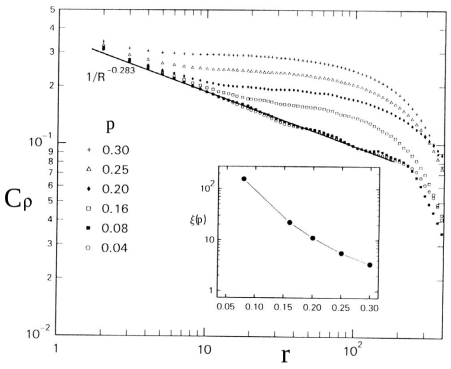


Figure 4.25: Density-density correlation function for multi-particle DLA aggregates at different concentrations p of the diffusing particles. At low values of p the standard DLA fractal behaviour is recovered (solid line). The inset shows how the correlation length ξ varies with p. (After [?].)

depleted region around the growing pattern. Thus, the density-density correlation function (4.1) has the form of

$$C_\varrho = \begin{cases} r^{-\alpha} & \text{for } r \gg \xi \\ const. & \text{for } \xi \gg r \end{cases} \tag{4.58}$$

Models resolving individual bacteria

To understand the diversity of bacterial patterns and their microbiological regulation, more elaborated models are needed than the previously presented continuum description of the diffusion-limited growth or the DLA model. We have two major classes of methods at our hand both with advantages and shortcomings: the continuum description by differential equations and the models based on the simulation of individual organisms/particles.

As we have seen, in case of working with differential equations a coarse-grained description is given which is valid for larger length scales than that of the microscopic cell-cell interactions. Hence, in this framework it is usually not straightforward to find the most suitable formulas representing the various basic microbiological processes. However, in some cases these models might be treated analytically yielding valuable information on the robustness and universality of the description.

The other way is to model the *individual* bacteria by some idealised units, defining their behaviour and interactions, and obtain the large-scale behaviour of the system by computer simulations. Models for single bacterial cells have been published for *E. coli* [?, ?] and *B. subtilis* [?]. The purpose of these studies was to integrate subcellular processes into whole-cell models, involving the use of much information on the biochemistry of these bacteria. However, in our case a much simpler description is sufficient, since bacterial growth on the cellular scale can be described by rather simple dynamics [?]. When modelling individual units, it is convenient to incorporate into the model details like the cell cycle or the differentiation of the cells, allowing the exploration of more sophisticated behaviour. At the same time such models are not really suitable for analytical treatment.

Now we focus on some variants of the individual-based models, originally introduced in [?, ?] and – as we show in the next section – later successfully applied for other colony formation processes as well. Such models are constructed as *minimal* models with the following guide lines: both the *microscopic* "rules" and the calculated/simulated behaviour of the whole system on the *macroscopic* level should be in accord with experimental observations. In fact, the models are often built in an iterative manner: the morphology diagram is approximated with the inclusion of more and more microscopic details, until satisfactory agreement has been reached at the

macroscopic level. Of course, at the microscopic level the model always gives a poor representation of the biological entities, but based on our experience with different models in the previous sections, we have a good chance that several such details are irrelevant in the pattern formation process. The resulting picture can be then verified by additional experimental or analytical studies.

Modelling non-motile bacteria. In the simplest version of such models, each particle i is characterised by its *position* x_i and *cell cycle state* E_i.

i) Changes in E_i control the *sporulation* and *division* of the cells as follows. If E_i decays below a threshold value (which can be specified as zero) then the particle becomes inactive (sporulates)[6]. Above an other threshold value, say 1, the model bacterium divides, and both of the daughter cells receive an initial value of $0 < E_* < 1$. It is known [?] that long lasting starvation initiate the sporulation process, and the average multiplication rate is influenced by the available amount of nutrients. Thus, in the model, changes in E_i are coupled to the nutrient consumption rate ω_i as

$$\frac{dE_i}{dt} = \kappa\omega_i - \epsilon \tag{4.59}$$

where κ is a conversion factor relating the maximal nutrient consumption rate with the shortest cell cycle time, and ϵ is the generic "maintenance" term incorporating all free energy expenditures not directly contributing to growth [?, ?].

ii) The *nutrient uptake* is limited by both the local concentration c and enzymatic rates. It is approximated by a Michaelis-Menten-like kinetics (see Fig. 4.3) as

$$\varrho(x_i)\omega_i = \min\left[\omega_{max}\varrho(x_i), \omega_0 c(x_i)\right]. \tag{4.60}$$

where ϱ is the local cell density, ω_{max} is the maximal uptake rate of the cells, and $\omega_0 c$ is the maximal diffusive transport from the substrate to the cells.

iii) Changes in c are given by the diffusion equation with the appropriate sink terms at the position of the active particles as

$$\frac{dc}{dt} = D_c\nabla^2 c - \sum_i \omega_i \delta(x - x_i). \tag{4.61}$$

[6] In all of the examples considered here this process is practically irreversible, since sporulation occurs deep inside the colony, where the nutrient-deficient condition remains through the colony growth.

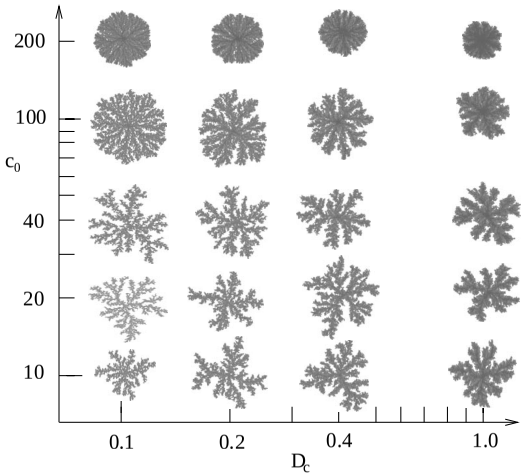


Figure 4.26: Morphology diagram generated by the model with non-motile particles as a function of the initial nutrient concentration c_0 and nutrient diffusivity D_c. The colonies were grown (in the computer) until either their size or the number of bacteria reached a threshold value.

By appropriate choice of units, $\kappa = 1$ and $\omega_{max} = 1$ can be assumed. Thus, the remaining control parameters of the model are E_*, ω_0, ϵ, D_c, and the initial nutrient concentration c_0. As the simplest choice, here we investigate the $E_* = 1/2$ and $\epsilon = 0$ case, i.e., when all the consumed nutrient is transformed into biomass. We fix $\omega_0 = 0.1$ and systematically explore the parameter space in respect of D_c and c_0. Note that in the experiments c_0 is directly controllable while D_c is determined by the hardness of the agar substrate. Since the colonies can consist of up to 10^{10} cells, even with the present computational power it is impossible to treat that many units. Thus, we might think of the simulated entities as if they represent ca. 0.1% of the total population.

Fig. 4.26. shows the morphology diagram obtained by numerical simulations. When the food concentration is high, the resulting patterns are compact, while they become branching for more hostile conditions in agreement with the experimentally determined morphology diagram of the non-motile *B. subtilis* OG-01b strain (Fig. 4.4b). The calculation of the C_ϱ density-density correlation function (4.1) reveals (Fig. 4.27) that the scaling characteristic for diffusion-limited objects ($D \approx 1.7$) is limited by both an upper cutoff determined by the finite diffusion length of the nutrients similarly to the case of multiparticle-DLA (Fig. 4.25) and a lower cutoff due to the finite width of the branches. This latter quantity is increasing with D_c as one can see from the following simple argument: Inside the colony the cell density is high and the nutrient uptake is approximately $\omega_0 c$. Therefore, c decreases as

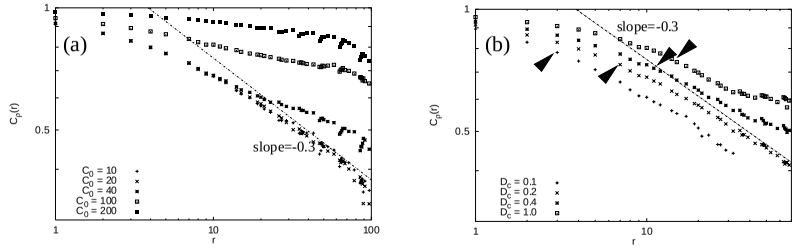


Figure 4.27: Density-density correlation functions calculated for the morphologies shown in Fig. 4.26. a: When increasing the initial nutrient concentration c_0 at $D_c = 0.2$, the DLA-like fractal behaviour ($\alpha = 0.3$, lines) changes to compact morphology (compare with Fig. 4.25). b: For a given c_0 ($c_0 = 20$) one can observe the parallel increase of the nutrient diffusivity D_c and the lower cutoff length (marked by arrowheads) corresponding to the width of the branches.

$$c \sim e^{-z\sqrt{\frac{\omega_0}{D_c}}} \tag{4.62}$$

with z being the distance measured from the colony boundary. Since the growing branches have a stationary shape, the infiltration depth of the nutrients $\sqrt{D_c/\omega_0}$ and the width of the branches are proportional.

This behaviour of $C_\varrho(r)$ also explains the lack of tip-splitting growth in certain colonies. If the penetration depth of the nutrients is too large or the diffusion length outside the colony is too small then the upper and lower cutoff values can be arbitrary close which results in compact growth within the experimentally accessible range of parameters.

Modelling motile bacteria. Now let us turn our attention to the morphology diagram exhibited by *P. dendritiformis* (Fig. 4.5). As microscopic observations revealed, in this case the bacteria are motile. However, they still cannot migrate on

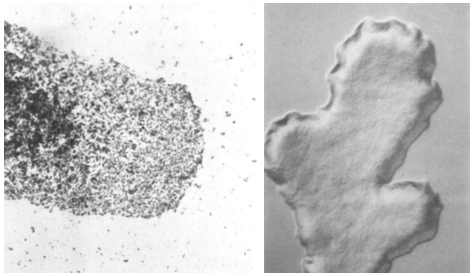


Figure 4.28: Optical microscopic observation of the branches of *P. dendritiformis* colonies. a: Staining visualises the short, rod-like bacteria in the growing tips of the colonies. The stronger-stained objects behind the tip are spores. b: Interference microscopy using Nomarsky optics reveals difference in optical depth due to secreted extracellular fluid. (After [?].)

dry agar gel surfaces, only where some surfactant was secreted (Fig 4.28). To explore this system, we keep the rules i – iii of the previous model, but also include new ones describing the motion of the bacteria.

iv) The active particles move randomly within a *boundary* described by

$$\frac{dx_i}{dt} = v_0 \vec{e}, \tag{4.63}$$

where $\vec{e}$ denotes a random vector of unit length and uniform angular distribution.

v) The propagation of the boundary is assumed to be proportional to the local density of active bacteria. In fact, collisions of the particles with the boundary was counted and when a threshold value (N_c) was reached the new neighbouring cell could be occupied. This rule mimics the surfactant or extracellular slime secretion and deposition in a crude fashion.

The resulting morphology diagram is shown in Fig. 4.29. Fig. 4.29 certainly resembles to the morphology diagram in Fig. 4.5 within a limited region of the parameters, but it fails to predict the formation of the thin, straight radial branches and the corresponding increase in the fractal dimension at very low food concentrations (Fig. 4.6). To explain these structures a repulsive chemotaxis signalling mechanism

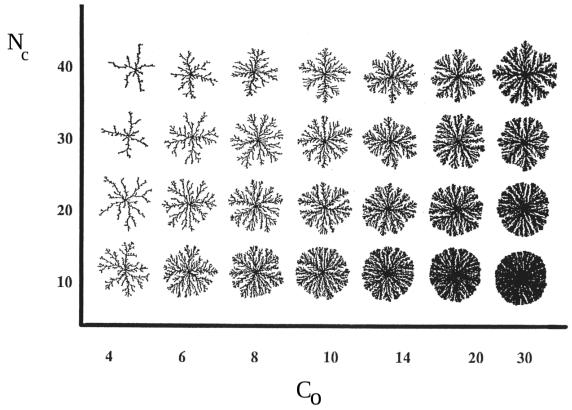


Figure 4.29: Morphology diagram generated by the model with motile bacteria as the function of the initial nutrient concentration and agar gel "hardness", i.e. the threshold value for the colony borderline displacement. (After [?].)

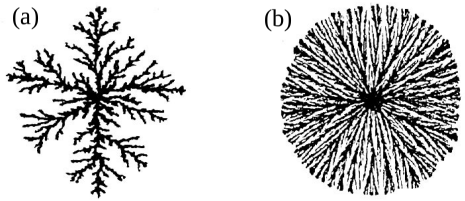


Figure 4.30: Effect of repulsive chemotaxis signalling in the model with motile bacteria. Results are shown of simulations without (a) and with (b) chemotaxis signalling. The pattern becomes denser and shows radial thin branches and a well defined circular envelope in agreement with experimental observations. (After [?].)

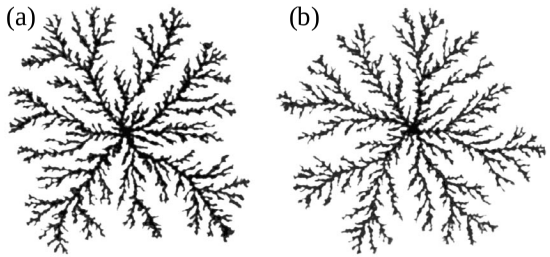


Figure 4.31: The effect of food chemotaxis in the same model. Results of simulations without (a) and with (b) chemotaxis signalling. In the latter case the growth velocity is twice as large as that of (a), yet the patterns have the same morphology. (After [?]).

was suggested in [?]. It was assumed that the bacteria which are in the sporulation phase emit certain diffusive chemicals acting as *chemorepellents*, i.e., the random swimming of the bacteria becomes biased: the cells move less toward the chemical gradient (Fig. 4.30).

Additional studies [?] with such models revealed further possible role of bacterial chemotaxis signalling in the colony development. As Fig. 4.31 demonstrates, *chemotaxis towards nutrients* does not change the morphology but significantly enhances the colony growth. Thus, based on such investigations and experimental studies of the mixed morphologies shown in Fig. 4.6, it was suggested that different chemotactic effects dominate the growth at various parts of the morphology diagram: While above a certain threshold nutrient concentration ($c_* \approx 1$g/l) the food chemotaxis is relevant, below that concentration the repulsive signalling is the main factor affecting the morphology. Furthermore, since these changes in the morphology are not inheritable it was assumed that below c_* the food chemotaxis "shuts down". Unfortunately, molecular biological studies on the chemotaxis system of the *P. dendritiformis* strain have not been performed yet. However, experiments with an other strain, *S. marcescens* support the hypothesis that chemotaxis is really involved in the generation of dense radial branching patterns (Fig. 4.32).

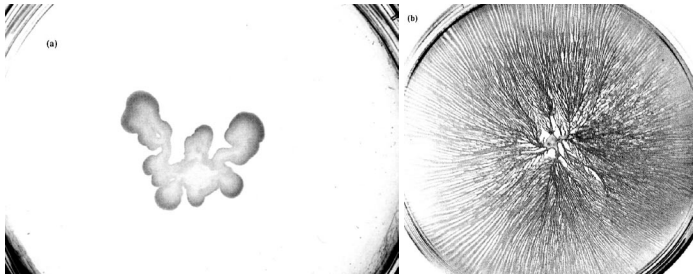


Figure 4.32: Chemotaxis deficient mutant of *S. marcescens* (a) vs the wild type (b). Note that the dense branching morphology disappears. (After [?].)

4.1.5 Chiral and rotating colonies

When bacterial motion is strongly influenced by cell-cell interactions, i.e., if the cells cannot be treated as random walkers with a bias towards some external concentration gradients, then the colony formation process can be even more complex than that shown in the previous sections. Here we give some examples for such a behaviour and demonstrate how the modelling approach can help understanding these systems.

Chiral patterns

Under certain conditions the strain *Paenibacillus dendritiformis* can *reproducibly* change its behaviour resulting in a distinct morphology diagram from that of shown in Fig. 4.5. As described in [?, ?] and independently verified in [?], after a growth of about two days on soft substrates (agar concentrations below 10g/l), a sudden burst of branches with a new, *chiral* morphology can be observed: each branch is curved and has the same handedness (see Fig. 4.33). New colonies inoculated with bacteria from these branches exhibit chiral morphology without any adaptation period. The transition of growth morphology is reversible: after some incubation on harder substrate the chiral growth reverts back to the much studied tip-splitting morphology. Therefore, at agar concentrations around 10g/l, the developed patterns are not unambiguously determined by the current environment, but also depend upon the conditions experienced by the bacteria in the past. Recent molecular biological [?] and microbiological classification studies [?] failed to find any difference between the bacteria taken from the two type of colonies, hence they are assumed to be two distinct phenotypes of the same strain.

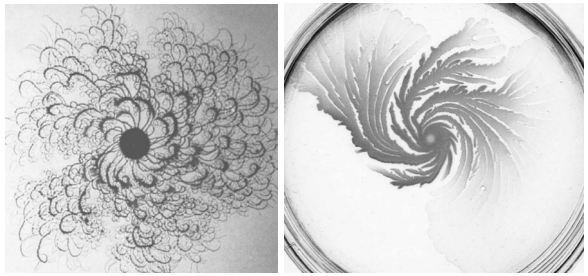


Figure 4.33: Chiral growth of bacterial colonies. (a) The chiral form of *Paenibacillus dendritiformis* on soft agar (after [?]). (b) Serrawettin-deficient mutant of *Serratia marcescens* (after [?]).

Albeit it is less frequent than tip-splitting patterns, the chiral morphology of *P. dendritiformis* is not unique: It has been reported for certain *Bacillus subtilis*, *Bacillus macerans* [?] and *Serratia marcescens* [?] strains as well. As the same handedness develops in an upside-down grown colony [?], the chirality is not an effect of some external (e.g., magnetic or Coriolis) force. When looking for possible mechanisms for the chiral symmetry breaking in the case of *P. dendritiformis*, microscopic observations revealed that the cells are very elongated and their orientation is parallel to each other (Fig. 4.34). Furthermore, as described in [?], the motion of the bacteria is strongly correlated with their orientation, they move approximately parallel to their axis. Already this highly coordinated *gliding* motion of the cells shows a certain level of chirality. In addition, individual bacteria can leave a "trail" which can guide the motion of the following cells. Less microscopic detail has been published for the other strains exhibiting chiral morphology. The *S. marcescens* strain is known to be deficient in producing serrawettin, a surfactant agent. In that case the lack of sufficient lubrication layer may enhance the coordination between the individual cells and give rise to similar microscopic behaviour than that of the chiral form of *P. dendritiformis*.

The effect of the above features on the colony morphology was explored with model calculations [?]. In the simulations the particles were characterised not only by their location x_i or cell-cycle state E_i but by their orientation ϕ_i as well. The food consumption, cell activity and colony boundary advancement rules were the same as described in the previous section. However, the motion of the particles was not a biased or unbiased random walk, instead, their displacement was parallel to their orientation ϕ_i. Finally, ϕ_i was changed by two factors: the intrinsic chirality of the

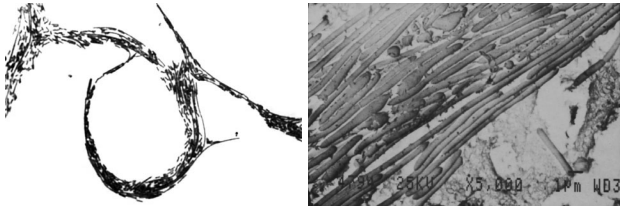


Figure 4.34: Optical (a) and electron- (b) microscopic images of the chiral form of *P. dendritiformis*. The bacteria are long and exhibit a high level of orientational correlations. (After [?].)

motion and the alignment to a local orientation field Φ representing neighbour cells and trails on the surface. In fact, the time development of ϕ_i was defined as

$$\phi_i(t+\Delta t) = P(\phi_i(t), \Phi(x_i(t))) + \varphi + \xi_i(t), \tag{4.64}$$

where φ represents the intrinsic chirality of the motion of the bacteria, ξ is a random noise with an amplitude less than $\hat{\xi}$ and the interaction with the local orientation field is described by the projector function P as $P(\alpha,\beta) = \alpha + (\beta - \alpha) \mathrm{mod} \pi$. The value of Φ was set at the occupation of the particular lattice site and remained unchanged. Since Φ represents the tracks in the agarose gel, its initial value was chosen to be the orientation of the first walker entering that field.

In Fig. 4.35 a detailed drawing of the model is presented and results of the numerical simulations with various parameter values are shown demonstrating that the microscopic chiral behaviour can be indeed amplified to macroscopic scales. In some sense it is a similar phenomenon to the appearance of the hexagonal crystal structure of ice in the snowflake patterns and it is strongly dependent on the preservation of the orientational correlations among the adjacent particles. If the noise amplitude $\hat{\xi}$ is increased then the chiral nature of the pattern disappears and tip-splitting growth takes place. These results are in complete agreement with experimental observations. Bacteria are shorter and less aligned when grown on nutrient rich substrates, where the chirality of the colonies is indeed lost [?].

Self-generated vortices

The rotation of disk shaped aggregates consisting of a few or millions of bacteria, or their peculiar motion along ring shaped trajectories has been reported long ago [?, ?, ?]. Although such types of motion have commonly been associated with *Bacillus*

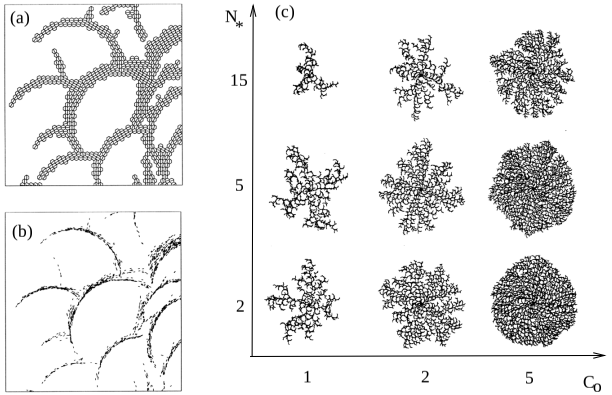


Figure 4.35: Computer simulation of chiral colony formation. (a) The occupied lattice sites with their orientation. (b) The position and orientation of the particles within the colony. (c) Colony shapes for various values of the initial nutrient concentration c_0 and agar hardness N_c. (after [?]).

circulans, similar phenomena have been seen in colonies of *Archangium violaceum, Chondromyces apiculatus, Clostridium tetani* and in colonies of the recently identified strain of *Paenibacillus vortex* [?, ?] as well.

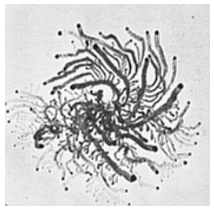


Figure 4.36: A typical colony of the strain *Paenibacillus vortex*. Each branch is formed by a rotating droplet of many bacteria moving together in a correlated manner at the branch tips. (After [?]).

In Fig. 4.36 a typical *P. vortex* colony is shown. As described in [?, ?], the bacteria spread via droplets (the darker dots in the figure) leaving a trail mostly filled with spores behind. Each droplet consists of many cells moving at the perimeter with typical velocities of 10μm/s (Fig. 4.37). Depending on the growth conditions and the location inside the colony, the number of bacteria in a single vortex can vary from a couple of hundreds to many thousands, and the vortex can consist of both single and

multiple layers of bacteria. Usually, the "pioneering" droplets are larger, while the smaller ones fill the empty areas left behind by the advancing front. Within a single colony, both clockwise and anti-clockwise rotating vortices can be observed.

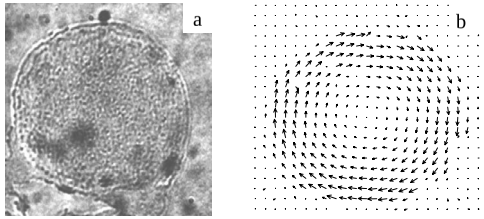


Figure 4.37: Bright field micrograph of a single rotating droplet with a magnification of 500× (a) and the corresponding velocity field obtained by digitising the videomicroscopic recordings (b). (After [?].)

In Section 6.1 we show that geometrical constrains of the boundaries can enforce a circular collective motion in a system of locally interacting self-propelled particles. This result suggests that a radial inward force acting on the migrating bacteria can explain the observed vortices.

Since collectively migrating bacteria do not "tumble", their response to chemotactic agents must differ from that of swimming cells. As Fig. 4.38 demonstrates, such chemotactic effect certainly exist although its molecular bases are not yet known. Thus, in the following part of this section we will use the words "chemotaxis" and "chemoattractant" in a broad sense, the particular response we consider can easily result from passive physical forces like surface tension or changing efficiency of translocation with the amount and quality of the extracellular slime deposited. In [?] such a special kind of chemotactic response was proposed: If the propulsion force of the individual cells is modulated depending on the local chemical concentrations, then it results a torque acting on the whole group.

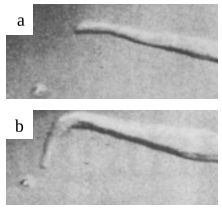


Figure 4.38: "Purposeful movement" of *Myxococcus xanthus* cells is shown in consecutive frames with 7 minutes time difference. The cells collectively migrate toward a latex bead at the bottom of the images. (After [?].)

To see this, let us consider a group of bacteria moving as a rigid body in a concentration field c_A. In a linear approximation the propulsion forces F are distributed as

$$\vec{F}(\vec{r}) = \vec{F}_0(1 + a\vec{r}\nabla c_A) \tag{4.65}$$

where $\vec{F}_0$ is the average propulsion force, $\vec{r}$ is measured from the center of the group and a is a coefficient. Since at the microscopic length scales the inertial effects are negligible, for the velocity $\vec{v}$ of the whole group one can assume

$$\vec{v} \sim \int \vec{F}(\vec{r}) d^2r \sim \vec{F}_0 \ , \tag{4.66}$$

while for its rotation

$$\vec{\omega} \sim \int [\vec{r} \times \vec{F}(\vec{r})] d^2r \sim \vec{F}_0 \times \nabla c_A \tag{4.67}$$

holds. Since the velocity and the propulsion forces are strongly correlated with the orientation of the cells (see the previous section)

$$\frac{d\vec{F}_0}{dt} \sim \frac{d\vec{v}}{dt} = \vec{\omega} \times \vec{v} \ , \tag{4.68}$$

for the equation of motion of a single coherent bacterial group we obtain

$$\frac{d\vec{v}}{dt} \sim -\vec{v} \times (\vec{F}_0 \times \nabla c_A) \tag{4.69}$$

which expression is independent of the size of the group [?].

To study the interactions of the bacteria and understand the spectacular self-organised vortices, we expand (4.69) to include the orientational interaction between the neighbouring cells and also add a term proportional to the cell density gradient preventing the accumulation of walkers into arbitrary dense spots as

$$\frac{\mathrm{d}\vec{v}_i}{\mathrm{d}t} = \mu\Big(\langle\vec{v}\rangle_{i,\epsilon} - \vec{v}_i\Big) - \nu\vec{v}_i + F\frac{\vec{v}_i}{v_i} - \kappa\nabla\varrho - \frac{\chi_A F}{v_i}\vec{v}_i \times (\vec{v}_i \times \nabla c_A) + \vec{\zeta}. \tag{4.70}$$

The first term at the rhs of (4.70) incorporates both the orientational interaction and the momentum transfer between the neighbour cells (see Sec. 6.1). The $\langle\vec{v}\rangle_{i,\epsilon}$ average is calculated over the particles within a distance ϵ from the walker i. The next two terms express the self-propulsion and the overdamped nature of the motion. The last term in (4.70) is an uncorrelated noise.

The time evolution of the diffusing chemoattractant is described by

$$\frac{\partial c_A}{\partial t} = D_A\nabla^2 c_A + \Gamma_A\varrho - \lambda_A c_A. \tag{4.71}$$

The first and the third term at the rhs represents the diffusion and a constant rate decay respectively, while the second (source) term assumes that bacteria produce the chemoattractant with a constant rate. This is a useful assumption if we investigate the formation of a single vortex, but must be refined when we also intend to describe the morphogenesis of the whole colony.

The simulation of the above model was carried out through the numerical integration of Eqs. (4.70) and (4.71) as described in [?]. The *positive feedback* of the attractive chemotaxis breaks the originally homogeneous spatial distribution of the particles and creates dense aggregates: If a fluctuation increases the density locally, then the emission of chemoattractants (being proportional to ϱ) is also increased. The increased concentration gradient attracts even more particles to the aggregate. As a consequence of the local alignment interaction, rotation develops in a spontaneously selected direction. In such an aggregate the difference of the $\nabla\varrho$ term and the attractive force of chemotaxis provides the appropriate centripetal acceleration. Typical vortices formed in the system are shown in Fig. 4.39, being in a good agreement with experimental observations (Fig. 4.37). With slightly different parameter values the model yields rotating rings reported also in experiments (Fig. 4.40). The qualitative results discussed here turned out to be insensitive to the exact form of Eq. (4.70) until a cohesive force with co-alignment interaction is represented.

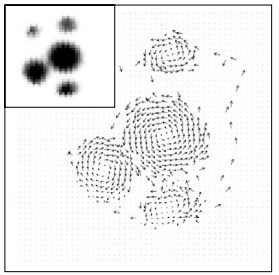


Figure 4.39: Typical result of model simulations for vortex formation. The positive chemotactic feedback breaks the originally homogeneous density resulting in aggregation of the cells. The flow field is represented by arrows of a magnitude proportional with the local velocity. The inset shows the concentration distribution of the chemoattractant. (After [?].)

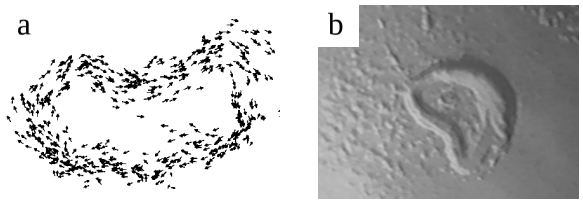


Figure 4.40: a: For slightly different value of the parameter μ (providing stronger velocity-velocity interaction) than that of used to generate Fig.4.39, rotating rings develop. (After [?].) b: This phenomenon was also documented in a *B. circulans* colony ([?]).

Thus, the model defined through Eqs (4.70) and (4.71) explains many experimentally observed features in the colonies of *P. vortex*, *B. circulans* and related strains such as collective migration of groups or formation of rotating dense aggregates in which the direction of the spinning was selected by spontaneous symmetry breaking. However, the model is not complete since it does not describe colony formation on the macroscopic level, which would be certainly needed to understand the possible (evolutionary) benefits of this strange behaviour. To expand the model to describe colony morphogenesis, further details were considered in [?]: extracellular slime influencing the motion of the organisms, nutrient diffusion and consumption determining the growth rate of the colony and additional chemical regulation of the movement of the individual vortices (Fig. 4.41).

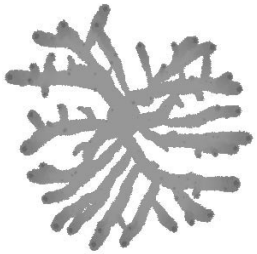


Figure 4.41: Typical colony obtained from numerical simulations of the model incorporating nutrient diffusion, reproduction and sporulation of bacteria, chemotactic regulation, velocity-velocity interaction and extracellular fluid deposition. The simulation is performed in a 600×600 system with ca. 10000 particles. In excellent agreement with the experimental observations, the colony grows via rotating droplets at the tips of the branches. Smaller vortices also emerge inside the colony, sometimes giving rise to a new side branch. In the figure the gray scale represents the density of the active bacteria. (After [?].)

The model calculations [?] also revealed an interesting benefit of vortex formation. As beyond the vortex (inside the colony) the nutrient concentration is rather small, vortices are exposed to a nutrient field with a strong gradient. In time average, cells in the rotating vortices experience less inhomogeneous concentration field, thus *both* their reproduction and sporulation rate is reduced. Therefore, by reducing their expenditure on reproductive and sporulation processes, they can spend on the average more energy to enhance the speed of their propagation. Under diffusion-limited conditions the faster expansion increases the inflow of nutrients.

The strategy reducing the rate of multiplication and using the energy for colonising surfaces is not a unique feature of vortex forming bacteria. In the case of swarmer strains similar behaviour was reported [?]: the differentiated swarmer cells are specialised to expand the colony, and many of their metabolic pathways (including the multiplication processes) are repressed.

The question of the *robustness* of such complex models arises in a natural way, and – according to our opinion – cannot be answered completely yet. However, one can certainly find supporting arguments. First, the experimental data for various external conditions and organisation levels (ranging from the possible microscopic interactions through the flow fields in the intermediate length scales to features of the macroscopic colony formation) yields, in fact, strict constrains for the models. Second, some of the generic features such as nutrient consumption, multiplication or sporulation of the bacteria and their mathematical representation can be found and verified in various other systems as well. Third, according to the experience with numerical models, usually a broad range of parameters yield qualitatively the same behaviour, i.e., in many cases no "fine tuning" of the parameters was required.

Chapter 5

Microscopic mechanisms of biological motion

The origin of biological motion can be traced back to the function of tiny *motor proteins* operating at the *molecular scale* (see in Refs. [1, 2]). During the last few decades numerous different families of these molecular motors have been discovered, and the newest experimental techniques developed in recent years have allowed to study *in vitro* the operation of *individual* motor proteins (for review see Ref. [3]). Such studies have revealed that the motors move in discrete, unidirectional steps with step sizes in the range of a few nanometers, and exert piconewton forces. The motors use the *chemical energy* stored in adenosine triphosphate (ATP) molecules or in proton gradient as fuel and convert it into *mechanical work*.

5.1 Characterisation of motor proteins

5.1.1 Cytoskeleton

Eukaryotic cells contain a dynamic protein network, the *cytoskeleton* (Fig. 5.1), which maintains the shape of the cell, anchors organelles, moves the cell, and serves as highways for intracellular transport. It consists of three kinds of filamentous structures: the *microtubules*, the *actin filaments* (also known as microfilaments) and the *intermediate filaments*. Along these polymeric filaments molecular motor proteins can move, consuming the chemical energy stored in ATP molecules. Three different motor protein families have been identified by now: *kinesins*, *dyneins* which move along the microtubules, and *myosins* which move along the actin filaments. There is no (known) motor proteins associated with the intermediate filaments.

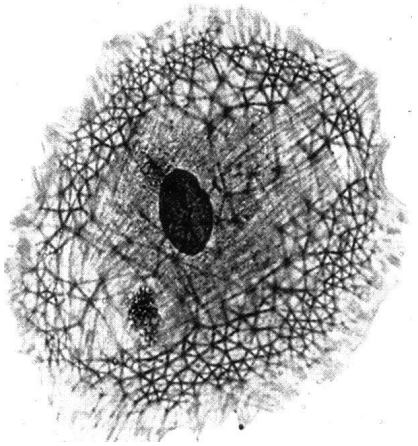


Figure 5.1: The cytoskeleton. A cell has been fixed and stained with a general stain for proteins. A variety of filamentous structures that extend through the cell can be seen. (After [2])

Microtubules are straight, hollow tubes (24 nm in diameter), the wall of which is made up of tubulin heterodimers arranged in 13 longitudinal rows called protofilaments (Fig. 5.2). A tubulin heterodimer is 8 nm long and consists of two globular proteins about 4 nm in diameter: α- and β-tubulin. The dimers bind head-to-tail giving the polarity to the protofilaments. The microtubule has a helical surface lattice with two possible configurations: the adjacent protofilaments can be shifted in the longitudinal direction by about either 5 nm (A-type lattice) or 1 nm (B-type lattice). Electron micrograph measurements of microtubules decorated with kinesin head fragments [4, 5, 6, 7] indicate the predominance of the B-type lattice, in addition, show that kinesin heads can bind only to the β-tubulin (meanwhile weakly interacting with the α-tubulin as well).

In most cells, the "minus" (slow growing) ends of the microtubules are stabilised and embedded in a structure near the nucleus, called centrosome. In addition to the structural role of the microtubules, they serve as tracks for the microtubule associated motor proteins that carry organelles (such as mitochondria, see Fig. 5.3 from Ref. [8]), vesicles (e.g., from the Golgi apparatus to their destination) or large molecules (chromosomes during cell division) inside the cytoplasm. Kinesins usually walk towards the

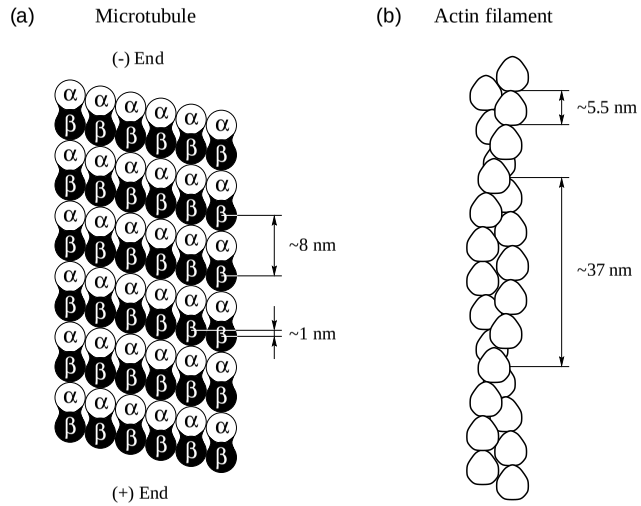


Figure 5.2: (a) The structure of the B-type helical surface lattice of the microtubule. Kinesin moves along the protofilaments (vertical rows) and its heads bind mainly to the β monomers. (After [69]) (b) The double-helical structure of the actin filament.

"plus" (fast growing) end of the microtubules, while dyneins walk towards the minus end. Nonclaret disjunctional (ncd) proteins, which are kinesin-related microtubule motor proteins, are also minus end motors.

Microtubules play the major role in the beating of eukaryotic cilia and flagella. The basis of these external cellular structures is formed by a ring of nine doublet microtubules, with two singlet microtubules at the center. A doublet microtubule consists of 23 protofilaments in a figure-of-eight arrangement. The peripheral doublets are connected by dyneins. The motion of these dyneins results in a relative sliding of adjacent doublets, inducing the cilia and flagella to bend.

Actin filaments are assembled from the globular protein actin and are the thinnest (7 nm in diameter) of the cytoskeleton fibres. Individual actin monomers polymerise into a chain and two chains twist together to form a helix with 13.5 molecules (74 nm) per turn (Fig. 5.2). Many actin filaments are attached to proteins within the plasma membrane. By this attachment, the filaments can help maintain the shape of cells. Where the actin interacts with its motor protein, the myosin, the generated force

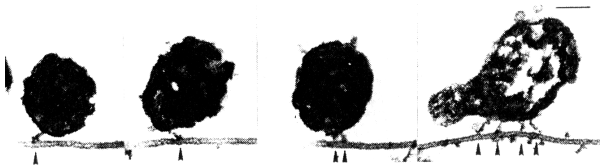


Figure 5.3: The picture taken by electron microscopy shows some mitochondria (large objects) as they are carried by several motor proteins (indicated by arrow heads) along a microtubule. (After [8])

can change the shape of the cell, allowing it to put out pseudopodia (for amoeboid movement), to invaginate particles (during phagocytosis), to cleave the cell (during cell division), or to cause *muscle contraction* (see Section 5.1.2 for details). This system also plays some role in intracellular transport processes.

Intermediate filaments are intermediate in size (8-12 nm in diameter) between microtubules and actin filaments. They are made of different kinds of proteins in different types of cells. While microtubules and actin filaments can be polymerised and depolymerised rapidly, intermediate filaments tend to be more stable and help maintain the shape of the cell over long periods of time. The nucleus is held in place in the cytoplasm by a particular form of intermediate filaments.

Kinesin

Kinesin was discovered in the mid-1980's [9]. Further studies have led to the identification of a large number of proteins that are related in structure to kinesin and constitute the kinesin superfamily of motor proteins. Native kinesin is a dimeric molecule with a tail and two globular ($\sim$9×3×3 nm) head domains [10, 11] (Fig. 5.4a). The heads are highly conserved, and each contains an ATP and a tubulin binding site. The more or less extended tail interacts with the cargo.

Recent experimental studies of individual kinesin molecules in *in vitro* motility assays (see Section 5.1.4) have revealed the following properties of kinesin movement:

- kinesin moves unidirectionally parallel to the protofilaments towards the plus end of the microtubule [9, 12];
- under an increasing load the speed of the kinesin decreases almost linearly [13, 14] (see Fig. 5.15);

- under its stall force (about 5 pN) kinesin still consumes ATP at an elevated rate [14];
- in the absence of ATP (in rigor state) kinesin binds to the microtubule very strongly: it supports forces in excess of 10 pN [14];
- the step size ($\sim$ 8 nm) is identical with the periodicity of the protofilaments [15, 16];
- kinesin hydrolyses one ATP per each 8-nm step [17];
- occasionally kinesin takes backward steps [14, 16];
- the displacement variance at saturating ATP and at low load increases linearly with time, but at (only somewhat more than) half of the rate of a single Poisson stepping process with 8 nm step size, implying that one step consists of two sequential subprocesses with comparable limiting rates [18, 19].

Ncd, which is also a member of the kinesin superfamily, moves in the opposite direction, towards the minus end of the microtubule (Fig. 5.4b). This is a surprising phenomenon, because the motor domains of kinesin and ncd are structurally very similar and bind on the microtubule with similar orientations [11]. This mystery has been deepened by a recent elegant experiment [20], in which a chimera was formed by attaching the motor domain of ncd to the neck region of kinesin. Surprisingly, the resulting motor catalysed plus-end-directed motion characteristic of kinesin from which the neck (and not the motor) region was taken. Attaching the motor domain of kinesin to the neck region of ncd [21] led to similar results: the directionality of the chimera was determined by the neck region. Moreover, changing a few amino acids in the neck can reverse the directionality of the motor [21, 22]. Structural studies also show that it is the neck region, where kinesin and ncd differ the most [22].

5.1.2 Muscle contraction

Muscle is an organ specialising in force generation and movement at the macroscopic scale. There are many types of muscles. Vertebrate muscles fall into three categories: skeletal muscle (or striated muscle, see Fig. 5.5), responsible for locomotion, flight, etc.; cardiac muscle, which has a vital role and is able to function for a century or more without ever taking a break; and smooth muscle (or involuntary muscle) which lines the walls of the arteries to control blood pressure, or controls the digestion of food by causing movement of the intestine. Some insect flight muscles are adapted for speed and can contract 1000 times a second, while molluscan adductor muscles are

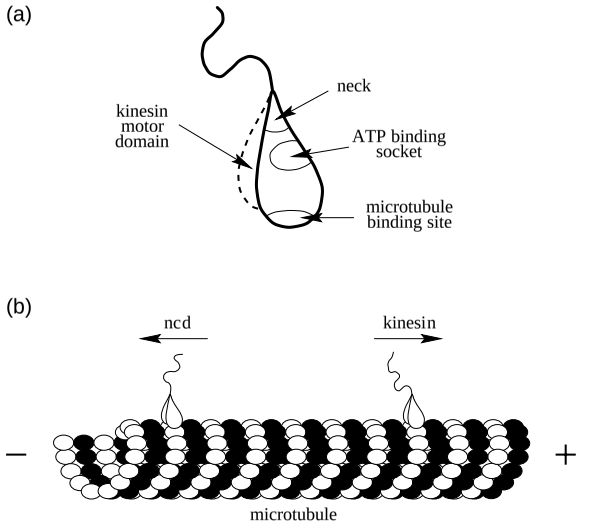


Figure 5.4: (a) Schematic drawing of the two-headed kinesin molecule. (b) Kinesin and ncd are structurally very similar proteins and bind on the microtubule with similar orientations, yet, they move in opposite directions.

built for strength and can lift a 10 kg weight per square centimetre cross-section. In spite of this large diversity the basic mechanism of force generation and contraction is common in all types of muscles: *myosin cross-bridges* extending from the thick myosin filaments attach to the binding sites of the thin actin filaments and exert force on them, inducing a relative sliding of the actin and myosin filaments. This process is fuelled by the hydrolysis of ATP molecules.

Myosin cross-bridge

Myosins, similarly to kinesins, are also dimeric molecules with a long tail and two globular heads (cross-bridges), but probably only one of the two heads can interact with the actin filament at a time. Each head contains an ATP and an actin binding site. The tails of many myosins attach to each other, forming the thick myosin filament. The mechanical force and motion is generated, when a head which is bound

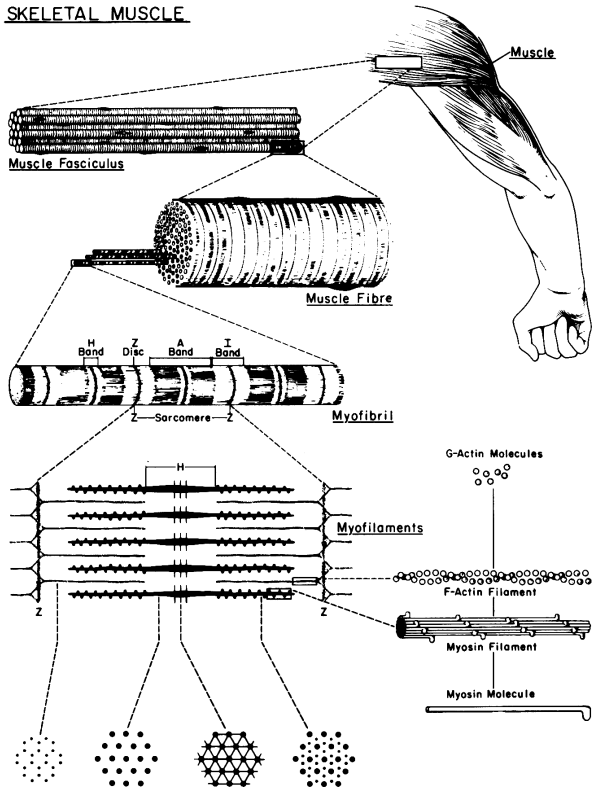
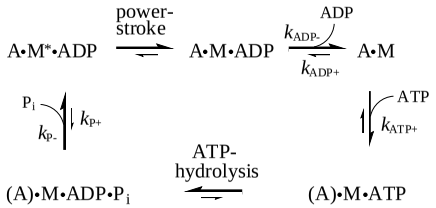


Figure 5.5: The structure of the vertebrate skeletal muscle. (After [23])

to the actin undergoes a conformational change. Though individual heads usually spend only a small fraction of their time strongly bound to actin, the large number of cross-bridges between the actin and myosin filaments prevents them from diffusing away.

The widely accepted cross-bridge scheme for the states of the actomyosin ATPase, which was originally proposed by Lymn and Taylor [24], is



(For review see Refs. [25, 26].) Initially, the myosin cross-bridge is in a weakly bound, non-force-producing $(A) \cdot M \cdot ADP \cdot P_i$ state. A and M denote actin and myosin respectively, and (A) indicates that myosin is weakly bound to actin. Releasing the phosphate P_i the cross-bridge gets into a strongly bound, force-producing $A \cdot M^* \cdot$ ADP state. Then the cross-bridge executes a *power-stroke* [see Fig. **??**(a)] producing mechanical force on actin, a conformational change in myosin, and a relative sliding between adjacent actin and myosin filaments. This new $A \cdot M \cdot ADP$ state cannot readily bind P_i any more, in contrast to the preceding one. At the end of the power-stroke, when continued attachment would resist useful work, ADP is released with transition to the $A \cdot M$ state, which is still strongly bound. ADP may then rebind, resulting in continued counter-productive attachment, or ATP may bind, resulting in rapid cross-bridge detachment. The hydrolysis step occurs in this weakly bound $(A) \cdot M \cdot ATP$ state resulting the $(A) \cdot M \cdot ADP \cdot P_i$ cross-bridge state, which is free to participate in a new working cycle.

Recent *in vitro* experimental studies of individual myosin molecules have revealed that a cross-bridge can generate a force up to 3-4 pN, and the power-stroke working distance is about 11 nm [27]. This distance is also consistent with the structural studies of myosin [28, 29, 30, 31]. The conformations of the cross-bridge states was determined by EPR spectroscopy [32]. These data together with fluorescence polarisation measurements [33] indicate that the 11 nm-long power-stroke is the result of the tilt and twist of the elongated neck (light-chain) domain of the myosin's head.

5.1.3 Rotary motors

The motor proteins discussed above (kinesin, ncd, dynein, myosin) can be called *linear motors*, since they move along straight, one-dimensional structures. In contrast, *rotary motors* do not advance large distances, but they rotate their shaft (or rotate around their shaft).

One example is the *bacterial flagellar motor* which is anchored in the cell membrane and rotates its flagellar filament that propels the cell. The motor can rotate in both direction and it is fuelled by the electrochemical gradient between the two side of the membrane. This gradient is typically a proton (H^+) gradient, but it can also be accomplished using sodium ions (Na^+), particularly by bacteria living in highly alkaline environments where there are few protons, or in marine environments where there is an abundance of sodium ions.

Other motors (typically DNA motors) move unidirectionally along their polymeric track, while they are also rotating around it. Thus, their overall motion is rather helical than linear.

ATP synthase

Perhaps the most peculiar rotary motor system is the enzyme *ATP synthase* which is present in the membranes of bacteria, chloroplasts and mitochondria, and has been remarkably conserved through the evolution. The function of the ATP synthase is to couple the synthesis of ATP with a proton flux across the membrane down the proton gradient. The flux goes from the protochemically positive P-side (high proton electrochemical potential) to the protochemically negative N-side. In bacteria the P-side is the outside, the N-side is the cytoplasm; in mitochondria the P-side is the intermembrane space, and the N-side is the mitochondrial matrix; in chloroplasts the P-side is the lumen and the N-side is the stroma. The operation of the ATP synthase is fully reversible, it can generate proton flux up the proton gradient by ATP hydrolysis. In some bacteria this is the main function of the enzyme.

The ATP synthase is a mushroom-shaped enzyme (see Fig. 5.6), consisting of two parts: the F_0 portion (stalk) is the proton channel that spans the membrane, and the F_1 portion (button) is a cytoplasmic domain containing the active catalytic sites. Both parts are *rotary molecular motors* working against each other via their common shaft. The F_0 part is driven by the proton gradient, while the F_1 part by the ATP hydrolysis, and the "stronger" motor forces the "weaker" one to rotate in the reverse direction. This is the way how ATP synthase is able to synthesise ATP molecules from ADP and P_i consuming the energy stored in the proton gradient or vice versa [34, 35, 36, 37].

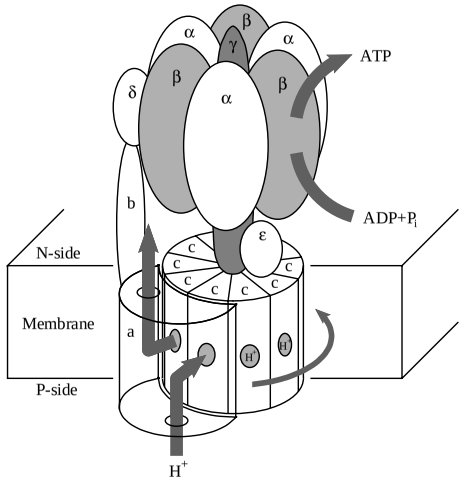


Figure 5.6: Schematic picture of the structure and function of the ATP synthase. The proton gradient across the membrane induces the rotation of the F_0 component (lower part) of the enzyme, which then forces the F_1 component to synthesise ATP from ADP and P_i.

The F_0 portion consists of 3 subunits in a stoichiometry of 1a:2b:9-12c. The twin α-helix-shaped c-subunits form a cylinder, which is believed to be the "rotor", while the a-subunit consists of 5 transmembrane helices, and constitutes the "stator". Each of the c-subunits has a conserved acidic residue half-way across the membrane. The idea for the operation of the F_0 motor suggested by Wolfgang Junge [37] is that the a-subunit provides a port for entry of the protons from the P-side, and a port for exit to the N-side. When a proton enters through the P-size port, it neutralises the acidic residue of the c-subunit. Only in this neutral form can the c-subunit complex rotate away from association with the a-subunit. Rotation brings a neutral c-subunit to the exit port, allowing it to lose the proton, and associate with the a-subunit. Successive protonations allow the c-subunit complex to rotate by $1/N_c \times 360°$ for each proton, where $N_c = 9 - 12$ is the stoichiometry of the c-subunit per ATP synthase.

The F_1 portion (which is also known as F_1 ATP-ase) contains 5 subunits, in a stoichiometry of 3α:3β:1γ:1δ:1ε. The α- and β-subunits in alternating order form a hexagonal ring (stator), surrounding the γ-subunit (rotor or shaft). The 3 active cat-

alytic ATP binding sites are in the β-subunits, and the additional nucleotide binding sites in the α-subunits are regulatory. The F_1 portion alone catalyses ATP hydrolysis, but not ATP-synthesis. It operates through a mechanism in which the three active sites undergo a change in binding affinity for the reactants of the ATP-ase reaction, ATP, ADP and P_i. The change in affinity accompanies a change in the position of the γ-subunit relative to the α-β-ring, which involves a rotation of the one relative to the other [38, 39], and a 120° phase-shift in the catalytic cycle of the 3 active sites relative to one another. This so-called *binding change mechanism* (Fig. **??**), which was first proposed by Paul Boyer [34], is confirmed by two direct evidences: (i) John Walker's group solved the crystal structure of the F_1 ATP-ase [40], and found that the 3 active sites are indeed in 3 different states of the catalytic cycle; and (ii) the rotational motion was captured in dramatic videos from the laboratory of Masasuke Yoshida [39] by tethering the α-β-ring to a glass surface and detecting the rotation of a fluorescently labelled actin filament that was attached to the γ-subunit.

The exact role of the b-, δ-, and ε-subunits is unknown, they are probably structural elements: the b- and δ-subunits may link the stators of the F_0 and F_1 portions, preventing the α-β-ring from moving, while the ε-subunit may help connecting the rotors.

5.1.4 Motility assay

In vitro motility assays (see, e.g., Refs. [41, 42]) have truly revolutionised the study of the operation of linear motor proteins. Purified motors can be examined as they move along a single complementary filament (actin or microtubule) in the presence

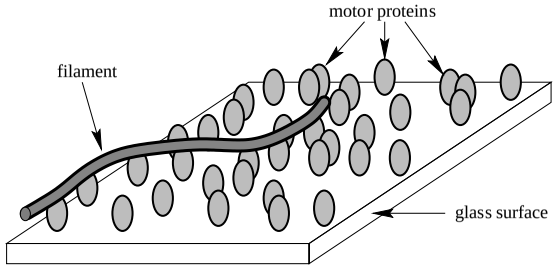


Figure 5.7: An example of the *in vitro* motility assays where a lawn of immobilised motor proteins moves a complementary filament.

of controlled amount of substances (such as ATP) and in the complete absence of other components of the cell. In such systems both the biochemical (such as ATP consumption) and mechanical (force, velocity, step size) properties of motor proteins can be measured quantitatively. In addition, kinesin and cytoplasmic dynein were identified also in motility assays. On the other hand, the cellular role of numerous motors is yet undetermined, for the most part.

In a conventional motility assay a quartz or glass surface is coated with motor proteins, and then a solution of the complementary filaments is added. The motors grab the filaments and move them unidirectionally across the surface (Fig. 5.7). If the concentration of the motors is low enough the motion of single molecules can be discerned. The motion of fluorescently labelled filaments can be observed directly by using fluorescent microscopy. Two possible way of controlling the load force against which the motors work are (i) attaching a microneedle to the filament; or (ii) applying a viscous drag force to the moving filament by about 100-fold increasing the viscosity of the medium for the filament.

A great advance in the experiments was the use of *optical traps* (sometimes called *optical tweezers*) to facilitate the examination of *individual* motor proteins.

Optical trap

Optical trap uses a highly focused laser light to grab and manipulate microscopic dielectric objects. Its unique trapping capability come from the three-dimensional gradient in light intensity found near the focus. When a moderately powerful laser is focused to a diffraction-limited spot in the specimen plane of a microscope, a steep gradient in light intensity is produced in the focal region. Small dielectric objects, such as latex or silica micro-spheres, or biological material, including cells, chromosomes, and organelles, experience a form of radiation pressure, called the gradient force, that tends to draw them towards the center of that region, where their dielectric energy is minimal. The dielectric energy density in the object is

$$\rho(\boldsymbol{x}) = -\frac{1}{2}\Delta\alpha E(\boldsymbol{x})^2 \,, \tag{5.1}$$

where $E(\boldsymbol{x})$ is the electric field of the light, and $\Delta\alpha$ is the polarisability of the object relative to the suspending solution. Consequently, this energy density is minimal in the center of the focal region, where the light has the maximum intensity.

For a spherical dielectric object (bead) the gradient force can also be understand easily from the law of conservation of momentum. Fig. 5.8 illustrates that if the bead is displaced out of its equilibrium position, the total momentum of the light changes during the refraction, and this momentum transfer between the bead and the light results in an effective force that draws the bead back to its equilibrium position.

Another form of radiation pressure, usually called the scattering force, arises from reflection or absorption of light, and it tends to push objects down along the beam, in the direction of the propagation of the light. Stable, three-dimensional trapping takes place when the effect of gradient force is sufficiently large to overcome the scattering force. In practice, such a condition can be achieved by using microscope objectives with the highest possible numerical aperture, which generate the steepest gradients.

The laser beam can serve not only to confine small dielectric objects, but also to measure their position. In addition, because the optical trap behaves like a spring, i.e., the trap force grows in proportion to the displacement of the object from the center, the force acting on the object can also be determined.

Svoboda *et al.* [15] used first this techniques in motility assay. They added kinesin to 0.6-micron silica beads at a concentration so low that, on average, less than one molecule of kinesin clung to each bead. Then they coated a substratum with microtubules and placed a kinesin-carrying bead on a microtubule, using the optical tweezers, and measured the displacement of the bead that was pulled by the kinesin molecule (Fig. 5.9). The results of this experiment have been discussed in

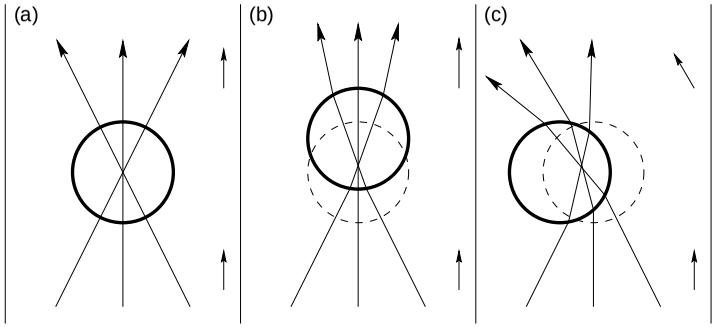


Figure 5.8: Illustration of the trapping capability of an optical trap by tracing three typical rays of a highly focused laser light. (a) The equilibrium position of the bead. (b) If the bead is displaced out of its equilibrium position in the longitudinal direction, the bead transfers momentum to the light because of the refraction, resulting in a force that draws the bead back to its equilibrium position. The total momentum of the light is indicated by small arrows on the right-hand side of the figures. (c) If the bead is displaced out of its equilibrium position in the transversal direction, the force arising from the momentum transfer also draws back the bead.

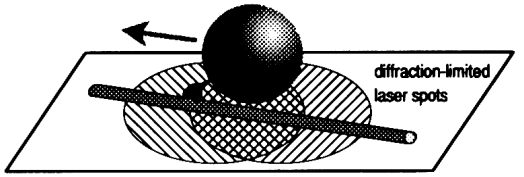


Figure 5.9: A kinesin molecule walks along an immobilised microtubule and carries a silica bead which is grabbed by an optical trap. The displacement of the bead is sensed optically. (After [15])

detail in Section 5.1.1.

To measure the mechanical properties of single myosin molecules, which remains anchored to actin for only a small fraction of the stepping cycle, Finer *et al.* [27] used optical traps to capture the filaments rather than the motor molecules (Fig. 5.10). Their main results have been summarised in Section 5.1.2.

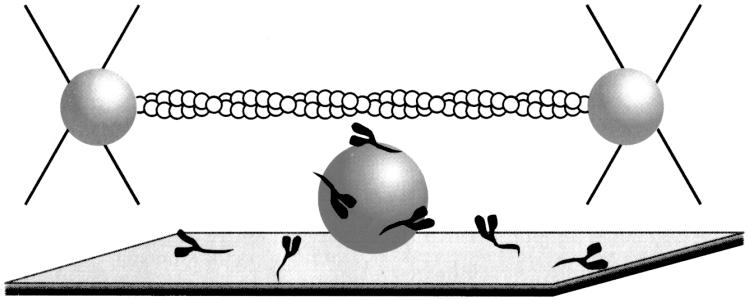


Figure 5.10: An immobilised myosin molecule tries to pull the actin filament, the ends of which are attached to polystyrene beads and held in place by optical traps. (After [27])

5.2 Fluctuation driven transport

The wide range of new experimental results on motor proteins and the desire to understand the basic mechanism of their operation have led to the development of

a new statistical physical principle, called *fluctuation driven transport*. It is based on the discovery that nonequilibrium fluctuations, whether generated externally or by a chemical reaction far from equilibrium, can drive directed motion along a periodic and spatially asymmetric structure (ratchet), without thermal gradients or net macroscopic forces, simply by biasing Brownian motion. Systems operating on this principle are often referred to as *thermal ratchets* or *Brownian ratchets* (for review see Refs. [43, 44, 45, 46, 47]). Since living cells maintain their ATP concentration far from equilibrium, the ATP hydrolysis can be considered as the source of nonequilibrium fluctuations for most motor proteins, and the polymeric filaments along which these motors move are the required periodic and asymmetric (polar) structures.

On symmetry grounds breaking the temporal and spatial symmetry simultaneously is enough for the existence of a velocity vector. In a Brownian ratchet system temporal symmetry is automatically broken because of the dissipative (viscous) environment, and the spatial symmetry is broken by the asymmetry (polarity) of the structure. A simple choice for this structure – inspired by Feynman's ratchet – is a sawtooth shaped periodic potential [Fig. 5.11(a)]. In his *Lectures* [48], Feynman used the "ratchet and pawl" engine to illustrate some implications of the second law of thermodynamics, in particular, that useful work cannot be extracted from equilibrium fluctuations (thermal noise). This justifies that indeed, additional nonequilibrium fluctuations are necessary for a Brownian motion to be rectified, resulting in net motion (and work if some external load is attached to the particle). The two

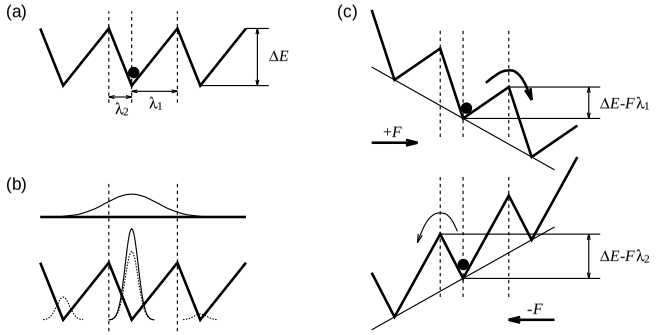


Figure 5.11: The motion of a Brownian particle (a) in a sawtooth shaped potential due to the effect of (b) a fluctuating potential and (c) a fluctuating external force. (After [47])

basic types of nonequilibrium fluctuations are the fluctuating potential (or "flashing ratchet") [49, 50, 51, 52], where the potential is time dependent and fluctuates between two or more different states; and the fluctuating force (or "rocking ratchet") [53, 50, 52], where the potential is static but a fluctuating external force with zero time average is applied.

The Langevin equation (2.23) for the motion of a particle in such systems can be written as

$$\gamma\dot{x} = -\partial_x V(x,t) + F(t) + F_{\rm load} + \gamma\sqrt{2D}\xi(t) \ , \tag{5.2}$$

where $V(x,t)$ denotes the periodic ratchet potential (which can fluctuate in time), $F(t)$ is the fluctuating external force (if there is any), and $F_{\rm load}$ is the load force if the particle has to move against some external load.

5.2.1 Basic ratchet models

Fig. 5.11(b) is an example for the flashing ratchets and illustrates the time evolution of the probability density of the position of a particle when a sawtooth potential is switched on and off repeatedly with an appropriate switching rate. When the sawtooth potential is switched on and the temperature is small enough compared to the height of the barriers, the particle stays near the bottom of a well. After switching off the sawtooth potential the particle starts to diffuse on the flat potential. Switching back to the sawtooth potential, the particle has a larger probability to be captured by the neighbouring well on the left than by the neighbouring well on the right because of the asymmetry of the potential. Thus, a net motion to the left can be observed.

Fig. 5.11(c) is an example for the rocking ratchets, where an external force alternates between $+F$ and $-F$ with an appropriate characteristic frequency with zero time average. When the force points to the right ($F(t) = +F$), the energy barrier for jumping to the right becomes shorter by F-times the longer distance λ_1. When the same force points to the left ($F(t) = -F$), the energy barrier for jumping to the left becomes shorter but only by F-times the shorter distance λ_2. Thus, a net motion to the right is expected.

The frequency response for the two types of fluctuations is very different [50] (Fig. 5.12). The fluctuating force causes a net flow at low frequencies. With increasing frequency the flow decreases slowly, but exceeding the inverse of the intrawell relaxation time the flow approaches zero rapidly. On the other hand, for fluctuating potential a maximum flow occurs in an intermediate frequency range and the flow vanishes at both high and low frequencies.

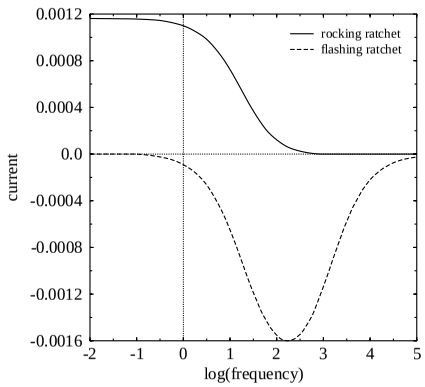


Figure 5.12: An example for the net current in a rocking and a flashing ratchet as a function of the frequency of the fluctuations.

5.2.2 Brief overview of the models

The transport induced by fluctuating forces has recently been studied for various statistics of the fluctuation [54, 55, 56, 57, 58]. In several cases a surprising phenomenon, the reversal of the direction of the motion has been observed for changing some parameters of the statistics. The spatial symmetry of the system can be broken not only by the asymmetry of the potential, but an asymmetrically fluctuating force (with zero mean) can also result in net flow even if the potential is spatially symmetric [59, 60, 61].

In the case of the fluctuating potential the transition rate constants between the different potentials (states) may depend on the position of the particle, and the spatial asymmetry can be embedded not only in the shape of the potential, but also in the transition rates. The most efficient transport with fluctuating potentials can be achieved when the extrema of the two (or more) potentials are shifted relative to each other, and a continuous downslope motion of the particle can be attained [62] (Fig. 5.13). If the potential fluctuates between two states that are the inverses of one another the direction and velocity of the motion are very sensitive to the shape of the potential, the time characteristics of the fluctuations, and some externally applied small forces [63, 64].

Nonuniform periodic temperature distribution alone can also induce net flow

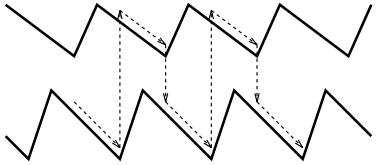


Figure 5.13: An example how a continuous downslope motion can be attained if the transition between the two potentials is allowed only in the vicinity of the minima, and the extrema of the potentials are shifted relative to one another.

("blow torch effect" of Landauer [65, 66, 67]), because the system is out of equilibrium.

5.2.3 Illustration of the second law of thermodynamics

At the beginning of this chapter it has been argued that thermal noise alone can never produce rectified motion in a ratchet potential because of the second law of thermodynamics. This can also be proven easily for an overdamped Brownian particle moving in an arbitrary one-dimensional periodic potential $V(x)$ with period λ. In this case the motion of the particle is described by the Fokker-Planck equation (2.24) with probability current (2.26). Since the potential $V(x)$ is periodic, it is enough to consider the motion in one period ($0 \le x < \lambda$) with periodic boundary conditions. In the steady state, when the probability density $P(x,t)$ and the current $J(x,t)$ have already reached their stationary values $P_{\rm st}(x)$ and $J_{\rm st}(x)$, the Fokker-Planck equation leads to $\partial_x J_{\rm st}(x) = 0$, i.e., the stationary current is constant in space: $J_{\rm st}(x) = J_{\rm st}$. After rearranging and integrating Eq. (2.26) from 0 to λ in the steady state we get

$$J_{\rm st}\frac{\gamma}{k_{\rm B}T}\int_0^\lambda \mathrm{e}^{V(x)/k_{\rm B}T}\mathrm{d}x = -\left[\mathrm{e}^{V(x)/k_{\rm B}T}P_{\rm st}(x)\right]_{x=0}^{\lambda} . \qquad (5.3)$$

Since the period of the system is λ, the right-hand side of this equation is zero. Because the integral on the left-hand side is always positive, independently of the potential $V(x)$, the stationary current must also be zero:

$$J_{\rm st} = 0 , \qquad (5.4)$$

which means that indeed, no net motion can be gained from thermal fluctuations. Note that the stationary (or equilibrium) solution of the probability density is the

Boltzmann distribution:

$$P_{\rm st}(x) = \frac{{\rm e}^{-V(x)/k_{\rm B}T}}{\int_0^\lambda {\rm e}^{-V(x)/k_{\rm B}T}{\rm d}x}\,, \tag{5.5}$$

which follows from Eq. (2.26) with zero current.

5.3 Realistic models

Although the fluctuation driven transport sheds light on the basic mechanism of motion generation and energy conversion by molecular motors, specific models are needed to describe real biological systems.

5.3.1 Kinesin

There are two basic mechanisms for the stepping of the two-headed kinesin molecule [68]. The first one is the so-called "Two-Step" model (see Fig. 5.14) in which both heads of the kinesin advance 8 nm in rapid succession for the hydrolysis of only one ATP by one of the heads. The second mechanism is the "Long-Stride" model (see Fig. 5.17) in which the back head detaches, advances 16 nm passing by the bound front head, and reattaches to the nearest binding site in front of the bound head. Then the heads change their roles and a new step may take place. With both mechanisms the center of the kinesin advances 8 nm for each step.

Two-Step model

Several versions of the Two-Step model can be distinguished with respect to the relative positions of the heads of the kinesin. If both heads track the same protofilament they are not able to pass by each other, one of the heads always trails the other one, and their distance alternates between 8 and 16 nm during each step. If the heads track adjacent protofilaments their relative position (looking from the side) alternates most probably between ≈ 0 and ≈ 8 nm, and the front head in a step may become the back head in another step.

The ratchet model [69] we present here describes the whole family of the Two-Step models. We suppose that each of the two Brownian heads moves along its own one-dimensional periodic potential with period $L = 8$ nm in an overdamped environment. These potentials represent the interaction between the heads and the protofilaments. Each period contains a deep potential well corresponding to the binding site of the β-tubulin and the potentials everywhere else are flat. Each well has an asymmetric

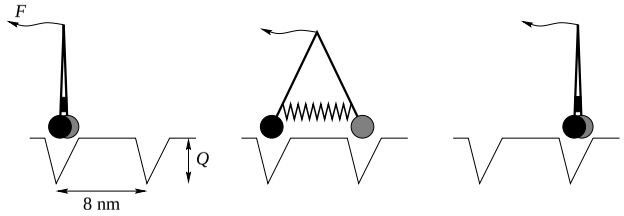


Figure 5.14: Schematic picture of the ratchet potential and the two substeps of the kinesin molecule in the Two-Step model (time passes from left to right). (After [69])

"V" shape (see Fig. 5.14) reflecting the polarity of the protofilaments: the slope in the backward direction (towards the minus end of the microtubule) is steeper and $0.5 - 1$ nm long, while the other slope (towards the plus end) is $1.5 - 2$ nm long. These ranges are in the order of the Debye length. The potentials for the two heads can be shifted relatively to each other by an arbitrary distance, depending on version of the Two-Step.

In this specific model the heads are connected at a hinge, and a spring acts between them [70] (Fig. 5.14). At the beginning of the mechanochemical cycle both heads are sitting in their wells waiting for an ATP molecule, and the spring is unstrained. Any configuration of the model is mathematically equivalent to the one in which the shift between the two potentials is zero, the heads are at the same position and, therefore, the rest length of the spring is zero. (Further on we will consider only this version of the model.) After one of the heads binds an ATP, its hydrolysis causes a conformational change in this head, more precisely, induces the head to take a 8 nm forward step. In the language of the model this means that the rest length of the spring changes from zero to 8 nm right after the hydrolysis. Then, as the first rate-limiting subprocess, the strained spring is trying to stretch pushing one of the heads to the neighbouring well. Reaching its new 8 nm rest length another conformational change occurs in the head as a consequence of the ADP release: the rest length of the spring changes back to zero. Then, as the second rate-limiting subprocess, the spring is trying to contract pulling one of the heads to the position where the other one is located. Completing the contraction a new cycle can begin. Due to the asymmetry of the potentials an expansion and contraction in most instances result in an 8 nm forward step. Hirose *et al.* [7] provided evidence that the kinesin·ADP complex has indeed a different conformation near the junction of the heads. The model is also consistent with the scheme of Gilbert *et al.* [71] for the pathway of the kinesin ATPase, and naturally explains the low displacement variance of the kinesin's

motion [18, 19]. An important cooperative feature of this model is that only one ATP hydrolysis occurs during each cycle.

We can take the load force into account in a natural way. Since in the experiments of Block's group [15, 14] a large ($\sim$0.5 μm in diameter) and therefore slow (compared to the kinesin heads) silica bead was linked to the hinge of the kinesin molecule by a relatively weak elastic tether, we can apply a constant force F to the hinge. We arrive at the same conclusion by considering the experiments of Hunt *et al.* [13], where a viscous load acted on the moving microtubule while the long tail of the kinesin was fixed.

The motion of the heads can be described by two coupled Langevin equations

$$\gamma\dot{x}_1 = -\partial_x V(x_1) - F_1^{\rm load} - K\left[x_1 - x_2 - l(t)\right] + \gamma\sqrt{2D}\xi_1(t) \ ,$$
$$\gamma\dot{x}_2 = -\partial_x V(x_2) - F_2^{\rm load} + K\left[x_1 - x_2 - l(t)\right] + \gamma\sqrt{2D}\xi_2(t) \ ,$$

where x_1 and x_2 denote the positions of the front and back heads respectively, $V(x)$ is the periodic ratchet potential, K is the stiffness of the spring, $F_1^{\rm load} = F_2^{\rm load} = F/2$ are the load forces acting on the heads, and the two thermal noise terms $\xi_1(t)$, $\xi_2(t)$ are uncorrelated. The load force F can be distributed between $F_1^{\rm load}$ and $F_2^{\rm load}$ in different ways, but due to the robustness of the model they lead practically to the same outcome. $l(t)$ represents the rest length of the spring, which is not an explicit function of time, but alternates stochastically between 0 and 8 nm as ATP is hydrolysed or ADP is released.

In order to obtain results we can compare with the experiments first the values of the input parameters have to be specified. The friction coefficient γ for a single head can be calculated from the Stokes law yielding $\gamma \approx 6\times10^{-11}$ kg/s. The free energy that can be gained from the ATP hydrolysis is about $25k_{\rm B}T$ or 100×10^{-21} J. During a stepping cycle the two conformational changes consume the free energy of the ATP, as the rest length of the spring changes by 8 nm in both cases. This means that $25k_{\rm B}T/2 \approx K\,(8\text{ nm})^2/2$, from which we get $K \approx 1.5$ pN/nm.

A lower limit for the depth Q of the potential wells can be determined from the fact that the kinesin in rigor state supports forces in excess of 10 pN. If we try to pull out the two-headed kinesin molecule (with drag coefficient 2γ) from a $2Q$ deep potential well with a force F, a lower limit for the escape rate is $(F^2/k_{\rm B}T2\gamma)\exp(-2Q/k_{\rm B}T)$ as follows from Eq. (2.31). Assuming that the escape rate is smaller than 10^{-2} 1/s we get $13k_{\rm B}T$ as a lower limit of Q. In reality the two heads cannot be handled as one larger head, because they are not fixed to each other rigidly, therefore, this is a very low limit. $Q \approx 20k_{\rm B}T$ is expected to be a better estimate.

We can assume that at the beginning of a stepping cycle both heads are sitting in the same well of the potential. After ATP hydrolysis the strained spring is trying

to stretch, pushing one of the heads to a neighbouring well. When the load force is small it has a large probability $p_{0L}^+ = k_{0L}^+/(k_{0L}^+ + k_{0L}^-)$ that the front head jumps to the forward direction, because of the asymmetry of the potential wells, and it has only a small probability $p_{0L}^- = k_{0L}^-/(k_{0L}^+ + k_{0L}^-)$ that the back head jumps backward. k_{0L}^+ and k_{0L}^- denote the corresponding jumping rate constants. Increasing the load the probability of the forward jump decreases while that of the backward jump increases. The average lifetime $t_{0L} = 1/(k_{0L}^+ + k_{0L}^-)$ of the stressed state also slightly increases. Completing this subprocess the second conformational change occurs: the spring is trying to contract. Now for low load force the probability $p_{L0}^+ = k_{L0}^+/(k_{L0}^+ + k_{L0}^-)$ that the backward head jumps forward to the next well, where the other head is sitting, is close to 1, while the probability $p_{L0}^- = k_{L0}^-/(k_{L0}^+ + k_{L0}^-)$ that the forward head jumps backward is very small. Increasing load has similar effect on these jumping rate constants as in case of stretching. Thus, under low load force the kinesin molecule steps 8 nm forward during almost each mechanochemical cycle. But increasing the load the probability that the molecule remains at the same place or even takes a backward 8 nm step increases. Reaching the stall load the average displacement of the kinesin becomes zero.

Since the potential wells are deep compared to $k_{\rm B}T$, the jumping rate constants can be calculated from the Kramers formula (2.31) with a potential that is the sum of the ratchet potential, the quadratic spring potential, and the linear potential of the load force. From the jumping rate constants one can calculate the average displacement $d = (p_{0L}^+ p_{L0}^+ - p_{0L}^- p_{L0}^-)L = (k_{0L}^+ k_{L0}^+ - k_{0L}^- k_{L0}^-)[1/(k_{0L}^+ + k_{0L}^-)][1/(k_{L0}^+ + k_{L0}^-)]L$ and duration $t = t_{0L} + t_{L0} = 1/(k_{0L}^+ + k_{0L}^-) + 1/(k_{L0}^+ + k_{L0}^-)$ of a step. Increasing the load the average displacement decreases due to the increasing probabilities of remaining and slipping back, and the average time slightly increases as a manifestation of the Fenn effect [72]. (At stall load it is about three times larger than without load.)

At saturating ATP [Fig. 5.15(a) and (c)]concentration the only rate-limiting factor is the stepping process, therefore dividing the average displacement by the average time gives the average velocity $v = d/t$ of the kinesin. However, at low ATP concentration [Fig. 5.15(b)] the rate-limiting factor is the diffusion of the ATP to the kinesin, thus, the average velocity is proportional to the average displacement during one cycle multiplied by the ATP consumption rate

$$\nu(c_{\rm ATP}) = \frac{\nu_{\rm sat}\, c_{\rm ATP}}{K_{\rm m} + c_{\rm ATP}} \approx \frac{\nu_{\rm sat}}{K_{\rm m}}\, c_{\rm ATP} \approx \text{const}\ c_{\rm ATP} \ , \qquad (5.6)$$

where $c_{\rm ATP}$ denotes here the ATP concentration, $\nu_{\rm sat} = 1/t$ is the inverse of the cycle time, and $K_{\rm m}$ is the mechanochemical Michaelis-Menten constant. The difference between the shapes of the force-velocity curves at saturating and low ATP concentrations shows the Fenn effect unambiguously.

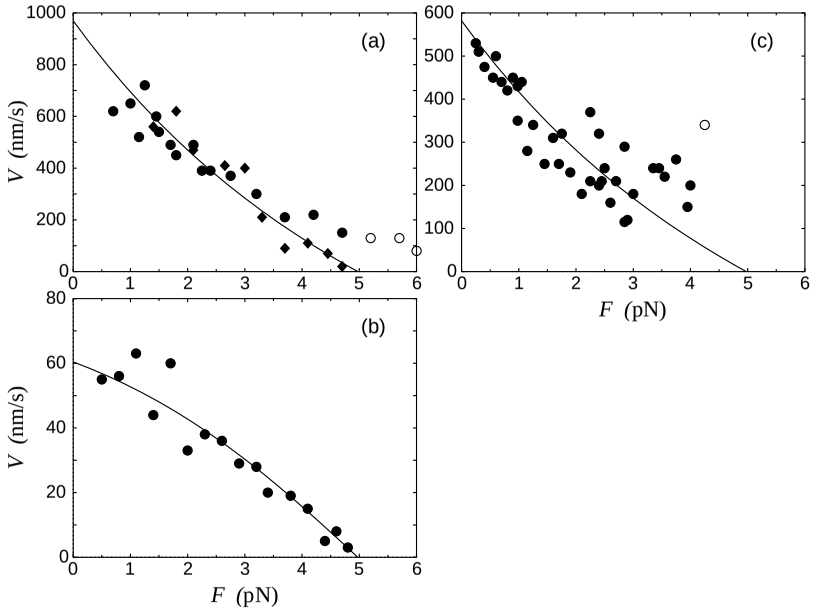


Figure 5.15: The force-velocity curves for an individual kinesin molecule at saturating ($\gg 90$ μM) ATP in (a) and (c), and at low ($\sim$10 μM) ATP concentration in (b). The experimental data (scattered symbols) are from Svoboda and Block [14] in (a) and (b) obtained by using optical tweezers, and from Hunt *et al.* [13] in (c) applying viscous load. The open circles correspond to the simultaneous effect of multiple kinesin motors in the authors' interpretation [13, 14]. The circles and diamonds in (a) mean two different set of the measured data. To get the best fit (solid lines) we chose 0.7 nm for the backward length of the potential wells and 1.75 nm for the forward length, the depth of the wells was $Q = 20k_BT$. In plot (b) we multiplied the average displacement by the ATP consumption rate $\nu(c_{ATP}) = 10$ 1/s for the best fit. In plot (c) we used the same curve as in (a) but multiplied by a factor of 0.6, which accounts for the increased viscosity. (After [69])

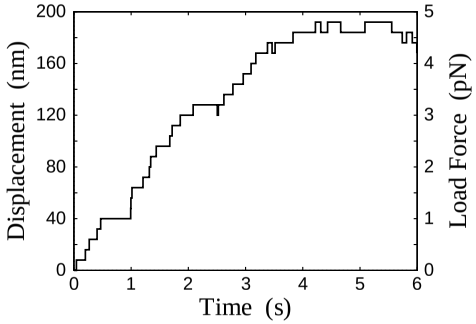


Figure 5.16: A typical simulated trajectory of kinesin in an optical trap with strength 0.04 pN/nm.

Note that during the entire calculation the concentration of the ATPase products (ADP and P_i) has been neglected. Though this assumption is supported by the *in vitro* experiments, thermodynamically complete description of the model cannot be given without taking into account the reverse processes of the ATP hydrolysis.

A typical trajectory of the model with a load force that increases linearly with the displacement of the kinesin (as in optical trapping), can be seen in Fig. 5.16. This curve is very similar to the experimental data [15, 14]. For large forces (i.e., for large displacements) the backward steps can be observed.

Long-Stride model

A Long-Stride model (Fig. 5.17)for the motion of kinesin was introduced by Peskin and Oster [73]. In this model the heads are also connected at a hinge and move along a single protofilament. The heads are initially bound to two nearest β-tubulin sites of the protofilament, separated by 8 nm. It is supposed that ATP hydrolysis by either head weakens its affinity to the microtubule, but after releasing P_i (or ADP) the head recovers its high affinity. The key constituents of the model that drive the motor forward are that (i) the preferred configuration of a bound head is the one in which the head leans forward (the rest length of the spring in Fig. 5.17 is the short one), and (ii) and the hydrolysis rate is greater for the back head than for the front head (e.g. because the strain induced in the front head adversely affects its hydrolysis site, slowing or arresting its catalytic affinity).

Thus, as the ATP hydrolysis in the back head weakens its affinity to the mi-

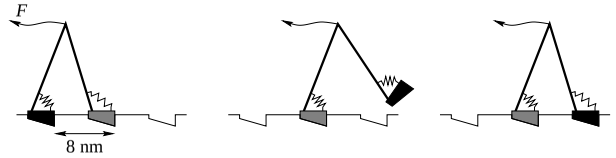


Figure 5.17: Illustration of the stepping of kinesin in the Long-Stride model.

crotubule, the back head detaches allowing the front head to lean forward into its preferred configuration. The diffusive motion of the detached head is restricted to the interval [-8 nm, 8 nm] with respect to the position of the bound head. After releasing the P_i (or ADP) the free head recovers its high affinity and binds to the first empty β-tubulin site that it encounters. The two sites that are within reach are the ones 8 nm in front and 8 nm in back of the bound head. Binding is biased towards the back site by the applied load force, but towards the front site by the forward-leaning tendency of the bound head.

Choosing reasonable parameters, much of the experimental data can be fitted with this model, too.

Duke and Leibler [74] introduced another version of the Long-Stride model in which the heads are capable of undergoing conformational changes and an elastic element connects each head to the rest of the molecule. They investigated cases where the chemical cycles of the heads were either correlated or uncorrelated and found good agreement with the experimental results.

A general drawback of the Long-Stride models is that during the long 16 nm diffusion of the free head it encounters several *beta*-tubulin sites on the same or neighbouring protofilaments, where the head may bind. This would lead to non-uniform step sizes and sidling motion. The Two-Step models give more stability to the protofilament tracking and yields uniform step sizes.

Chemically reversible model

In the above models the direction of the kinesin's motion is determined either by the asymmetry of the interaction between the head and the microtubule or by the conformation of the heads in different chemical states. Reversing the direction of motion would require a change in these characteristics of the models. As shown by recent experiments [21, 22] changing a few amino acids in the neck region of the heads is enough to reverse the directionality of both kinesin and ncd, and it is hard to imagine that these tiny changes would affect the interaction potentials or the

conformations so much.

Based on the fact that the neck region is close to and interacts with the ATP binding socket [22] we present here another ratchet model (which belongs to the Two-Step models), where the direction of motion is controlled by subtle differences in the chemical mechanism of ATP hydrolysis, rather than by differences in the structure of the motor or in its interaction with the microtubule [75, 76]. Fig. 5.18 illustrates the interaction potential between the head and the protofilament, which depends on the chemical state of the head. In the E^{ADP+P_i} and E^{ADP} states of the molecule there are two potential wells within each period: one with a relatively high energy minimum (H) and the other one with a relatively low energy minimum (L). When the motor gets into either of these chemical states there is first a fast local equilibration in the H well, followed by a slower relaxation from H to L. When the motor is in the E^{ADP+P_i} state, a transition from H to L to the right is allowed, while the transition to the left is blocked by a high energy barrier. When the motor is in the E^{ADP} state, a transition from H to L to the left is allowed, but the transition to the right is blocked.

In this model the direction of motion changes when the relative rates for releasing P_i and ADP change. The solid arrows trace the case for slow P_i and fast ADP release. Because P_i release is slow, the motor in the E^{ADP+P_i} state relaxes with high probability to L to the right before P_i dissociates. Following the release of P_i, the motor rapidly equilibrates in the H well of the E^{ADP} state. But because ADP release is fast, ADP most probably dissociates before relaxation to L can occur. Thus, completing one chemical cycle takes the motor one period to the right.

On the other hand, if P_i is fast and ADP release is slow, the motor moves one period to the left for every cycle (dashed arrows). So simply by changing the chemical transition rate constants relative to the mechanical ones, one can reverse the direction of the motion.

But there is a second, even simpler way of reversing the directionality, by changing the chemical specificity of the motor in the H and L positions. It is easy to see that if L is specific for P_i release in the E^{ADP+P_i} state and H is specific for ADP release in the E^{ADP} state, the motor moves to the right (solid arrows). Or, if H is specific for P_i release and L is specific for ADP release, the motor moves to the left.

Introducing some cooperativity between the two heads of the molecule (e.g., ATP hydrolysis at one head induces ATP binding to the other, but ATP hydrolysis at this second head cannot proceed until ADP dissociates from the first head) and choosing reasonable parameters, the results of the model are in very good agreement with the experiments. Note that the low energy barriers in the E^{ATP} state are necessary for a Two-Step motion, because it allows a passive 8 nm step of one head while the second head takes an active 8 nm step (by completing its chemical cycle) and tries to drag the first head along.

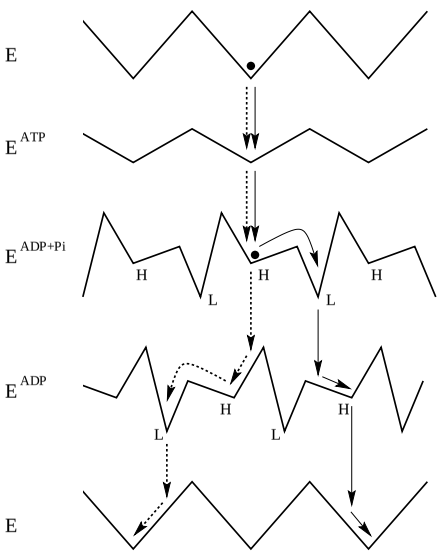


Figure 5.18: Ratchet mechanism for chemically reversible motion. The motor concomitantly cycles through its chemical states while catalysing ATP hydrolysis (in the vertical direction) and translocates through space along a microtubule (horizontally). (After [76])

While in the preceding models the presence of the two heads of kinesin (or ncd) was crucial to the motion, this chemically reversible model predicts that a single headed molecule should display some processivity. Very recent experiments have indeed confirmed that a single-headed kinesin construct can move processively [77].

Bibliography

[1] J. Darnell, H. Lodish, and D. Baltimore, *Molecular Cell Biology*. New York: Scientific American Books, 1990.

[2] B. Alberts, D. Bray, M. Raff, K. Roberts, and J. Watson, *Molecular Biology of the Cell*. New York: Garland, 1994.

[3] B. G. Levi. Measured steps advance the understanding of molecular motors. *Phys. Today* 48, no. 4:17–19, 1995.

[4] Y.-H. Song and E. Mandelkow. Recombinant kinesin motor domain binds to beta-tubulin and decorates microtubules with a B surface lattice. *Proc. Natl. Acad. Sci. USA* 90:1671–1675, 1993.

[5] A. Hoenger, E. P. Sablin, R. D. Vale, R. J. Fletterick, and R. A. Milligan. Three-dimensional structure of a tubulin–motor-protein complex. *Nature* 376:271–274, 1995.

[6] M. Kikkawa, T. Ishikawa, T. Wakabayashi, and N. Hirokawa. Three-dimensional structure of the kinesin head–microtubule complex. *Nature* 376:274–277, 1995.

[7] K. Hirose, A. Lockhart, R. A. Cross, and L. A. Amos. Nucleotide-dependent angular change in kinesin motor domain bound to tubulin. *Nature* 376:277–279, 1995.

[8] A. Ashkin, K. Schütze, J. M. Dziedzic, U. Euteneuer, and M. Schliwa. Force generation of organelle transport measured in vivo by an infrared laser trap. *Nature* 348:346–348, 1990.

[9] R. D. Vale, T. S. Reese, and M. P. Sheetz. Identification of a novel force-generating protein, kinesin, involved in microtubule-based motility. *Cell* 42:39–50, 1985.

[10] F. J. Kull, E. P. Sablin, R. Lau, R. J. Fletterick, and R. D. Vale. Crystal structure of the kinesin motor domain reveals a structural similarity to myosin. *Nature* 380:550–555, 1996.

[11] E. P. Sablin, F. J. Kull, R. Cooke, R. D. Vale, and R. J. Fletterick. Crystal structure of the motor domain of the kinesin-related motor NCD. *Nature* 380:555–559, 1996.

[12] S. Ray, E. Meyhöfer, R. A. Milligan, and J. Howard. Kinesin follows the microtubule's protofilament axis. *J. Cell. Biol.* 121:1083–1093, 1993.

[13] A. J. Hunt, F. Gittes, and J. Howard. The force exerted by a single kinesin molecule against a viscous load. *Biophys. J.* 67:766–781, 1994.

[14] K. Svoboda and S. M. Block. Force and velocity measured for single kinesin molecules. *Cell* 77:773–784, 1994.

[15] K. Svoboda, C. F. Schmidt, B. J. Schnapp, and S. M. Block. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* 365:721–727, 1993.

[16] C. M. Coppin, D. W. Pierce, L. Hsu, and R. D. Vale. The load dependence of kinesin's mechanical cycle. *Proc. Natl. Acad. Sci. USA* 94:8529–8544, 1997.

[17] M. J. Schnitzer and S. M. Block. Kinesin hydrolyses one ATP per 8-nm step. *Nature* 388:386–390, 1997.

[18] K. Svoboda, P. P. Mitra, and S. M. Block. Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl. Acad. Sci. USA* 91:11782–11786, 1994.

[19] K. Visscher, M. J. Schnitzer, and S. M. Block. Single kinesin molecules studied with a molecular force clamp. *Nature* 400:184–189, 1999.

[20] U. Henningsen and M. Schliwa. Reversal in the direction of movement of a molecular motor. *Nature* 389:93–96, 1997.

[21] S. A. Endow and K. W. Waligora. Determinants of kinesin motor polarity. *Science* 281:1200–1202, 1998.

[22] E. P. Sablin, R. B. Case, S. C. Dai, C. L. Hart, A. Ruby, R. D. Vale, and R. J. Fletterick. Direction determination in the minus-end-directed kinesin motor ncd. *Nature* 395:813–816, 1998.

[23] W. Bloom and D. W. Fawcett, *A Textbook of Histology, 2nd edition*. Philadelphia: W.B. Saunders, 1975.

[24] R. W. Lymn and E. W. Taylor. Mechanism of adenosine triphosphate hydrolysis by actomyosin. *Biochemistry* 10:4617–4624, 1971.

[25] Y. E. Goldman. Kinetics of actomyosin ATPase in muscle fibers. *Ann. Rev. Physiol.* 49:629–654, 1987.

[26] C. R. Bagshaw, *Muscle Contraction*. London: Chapman & Hall, 1993.

[27] J. T. Finer, R. M. Simmons, and J. A. Spudich. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature* 368:113–119, 1994.

[28] I. Rayment, W. R. Rypniewski, K. Schmidtbase, R. Smith, D. R. Tomchick, M. M. Benning, and D. A. Winkelmann. 3-dimensional structure of myosin subfragment-1 – a molecular motor. *Science* 261:50–58, 1993.

[29] I. Rayment, H. M. Holden, M. Whittaker, C. B. Yohn, M. Lorenz, K. C. Holmes, and R. A. Milligan. Structure of the actin-myosin complex and its implication for muscle contraction. *Science* 261:58–65, 1993.

[30] A. J. Fisher, C. A. Smith, J. Thoden, R. Smith, K. Sutoh, H. M. Holden, and I. Rayment. Structural studies of myosin: nucleotide complexes: a revised model for the molecular basis of muscle contraction. *Biophys. J.* 68:19s–28s, 1995.

[31] R. Dominguez, Y. Freyzon, K. M. Trybus, and C. Cohen. Crystal structure of a vertebrate smooth muscle myosin motor domain and its complex with the essential light chain: Visualization of the pre-power stroke state. *Cell* 94:559–571, 1998.

[32] E. M. Ostap and D. D. Thomas. Transient detection of spin-labeled myosin's internal dynamics and global orientation during ATP hydrolysis. *Biophys. J.* 68, p. 335s, 1995.

[33] J. E. T. Corrie, B. D. Brandmeier, R. E. Ferguson, D. R. Trentham, J. Kendrick-Jones, S. C. Hopkins, U. A. van der Heide, Y. E. Goldman, C. Sabido-David, R. E. Dale, S. Criddle, and M. Irving. Dynamic measurement of myosin light-chain-domain tilt and twist in muscle contraction. *Nature* 400:425–430, 1999.

[34] P. D. Boyer. The binding change mechanism for ATP synthase – some probabilities and possibilities. *Biochim. Biophys. Acta* 1140:215–250, 1993.

[35] R. H. Fillingame. Coupling H^+ transport and ATP synthesis in F_1F_0-ATP synthases: glimpses of interacting parts in a dynamic molecular machine. *J. Exp. Biol.* 200:217–224, 1997.

[36] R. Cross and T. Duncan. Subunit rotation in F_1F_0-ATP synthases as a means of coupling proton transport through F_0 to the binding changes in F_1. *J. Bioenerg. Biomem.* 28:403–408, 1996.

[37] W. Junge, H. Lill, and S. Engelbrecht. ATP synthase: An electro-chemical transducer with rotatory mechanics. *Trends Biochem. Sci.* 22:420–423, 1997.

[38] D. Sabbert, S. Engelbrecht, and W. Junge. Intersubunit rotation in active F-ATPase. *Nature* 381:623–625, 1996.

[39] H. Noji, R. Yasuda, M. Yoshida, and K. Kinosita. Direct observation of the rotation of F_1-ATPase. *Nature* 386:299–302, 1997.

[40] J. P. Abrahams, A. G. W. Leslie, R. Lutter, and J. Walker. Structure at 2.8 åresolution of F_1-ATPase from bovine heart mitochondria. *Nature* 370:621–628, 1994.

[41] S. J. Kron and J. A. Spudich. Fluorescent actin filaments move on myosin fixed to a glass surface. *Proc. Natl. Acad. Sci. USA* 83:6272–6276, 1986.

[42] L. Bourdieu, M. O. Magnasco, D. A. Winkelmann, and A. Libchaber. Actin filaments on myosin beds: the velocity distribution. *Phys. Rev. E* 52:6573–6579, 1995.

[43] R. D. Astumian and M. Bier. Mechanochemical coupling of the motion of molecular motors to ATP hydrolysis. *Biophys. J.* 70:637–653, 1996.

[44] P. Hänggi and R. Bartussek. Brownian rectifiers: how to convert Brownian motion into directed transport. in *Nonlinear Physics of Complex Systems, Lecture Notes in Physics* (J. Parisi, S. C. Müller, and W. Zimmermann, eds.) 467:294–308, Berlin: Springer, 1996.

[45] R. D. Astumian. Thermodynamics and kinetics of a Brownian motor. *Science* 276:917–922, 1997.

[46] F. Jülicher, A. Ajdari, and J. Prost. Modelling molecular motors. *Rev. Mod. Phys.* 69:1269–1281, 1997.

[47] I. Derényi and T. Vicsek. Realistic models of biological motion. *Physica A* 249:397–406, 1998.

[48] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman Lectures on Physics*. Reading, MA: Addison-Wesley, 1966.

[49] A. Ajdari and J. Prost. Drift induced by a spatially periodic potential of low symmetry: pulsed dielectrophoresis. *C. R. Acad. Sci. Paris* 315:1635–1639, 1992.

[50] R. D. Astumian and M. Bier. Fluctuation driven ratchets: molecular motors. *Phys. Rev. Lett.* 72:1766–1769, 1994.

[51] J. Prost, J.-F. Chauwin, L. Peliti, and A. Ajdari. Asymmetric pumping of particles. *Phys. Rev. Lett.* 72:2652–2655, 1994.

[52] C. S. Peskin, G. B. Ermentrout, and G. F. Oster. The correlation ratchet: a novel mechanism for generating directed motion by ATP hydrolysis. in *Cell Mechanics and Cellular Engineering* (V. C. Mow *et al.*, eds.), New York: Springer-Verlag, 1994.

[53] M. O. Magnasco. Forced thermal ratchets. *Phys. Rev. Lett.* 71:1477–1480, 1993.

[54] M. M. Millonas and D. I. Dykman. Transport and current reversal in stochastically driven ratchets. *Phys. Lett. A* 185:65–69, 1994.

[55] C. R. Doering, W. Horsthemke, and J. Riordan. Nonequilibrium fluctuation-induced transport. *Phys. Rev. Lett.* 72:2984–2987, 1994.

[56] R. Bartussek, P. Hänggi, and J. G. Kissner. Periodically rocked thermal ratchets. *Europhys. Lett.* 28:459–464, 1994.

[57] M. M. Millonas. Self-consistent microscopic theory of fluctuation-induced transport. *Phys. Rev. Lett.* 74:10–13, 1995.

[58] R. Bartussek, P. Reimann, and P. Hänggi. Precise numerics versus theory for correlation ratchets. *Phys. Rev. Lett.* 76:1166–1169, 1996.

[59] A. Ajdari, D. Mukamel, L. Peliti, and J. Prost. Rectified motion induced by ac forces in periodic structures. *J. Phys. I France* 4:1551–1561, 1994.

[60] D. R. Chialvo and M. M. Millonas. Asymmetric unbiased fluctuations are sufficient for the operation of a correlation ratchet. *Phys. Lett. A* 209:26–30, 1995.

[61] J. Łuczka, R. Bartussek, and P. Hänggi. White noise induced transport in spatially periodic structures. *Europhys. Lett.* 31:431–436, 1995.

[62] J.-F. Chauwin, A. Ajdari, and J. Prost. Force-free motion in asymmetric structures: a mechanism without diffusive steps. *Europhys. Lett.* 27:421–426, 1994.

[63] J.-F. Chauwin, A. Ajdari, and J. Prost. Current reversal in asymmetric pumping. *Europhys. Lett.* 32:373–378, 1995.

[64] M. B. Tarlie and R. D. Astumian. Optimal modulation of a Brownian ratchet and enhanced sensitivity to a weak external force. *Proc. Natl. Acad. Sci. USA* 95:2039–2043, 1998.

[65] R. Landauer. Motion out of noisy states. *J. Stat. Phys.* 53:233–248, 1988.

[66] M. Büttiker. Transport as a consequence of state-dependent diffusion. *Z. Phys. B* 68:161–167, 1987.

[67] I. Derényi and R. D. Astumian. Efficiency of Brownian heat engines. *Phys. Rev. E* 59:R6219–R6222, 1999.

[68] S. M. Block and K. Svoboda. Analysis of high resolution recordings of motor movement. *Biophys. J.* 68:230s–241s, 1995.

[69] I. Derényi and T. Vicsek. The kinesin walk: a dynamic model with elastically coupled heads. *Proc. Natl. Acad. Sci. USA* 93:6775–6779, 1996.

[70] A. Ajdari. Force-free motion in an asymmetric environment: a simple model for structured objects. *J. Phys. I France* 4:1577–1582, 1994.

[71] S. P. Gilbert, M. R. Webb, M. Brune, and K. A. Johnson. Kinesin crossbridge detachment occurs after ATP hydrolysis. *Nature* 373:671–676, 1995.

[72] W. O. Fenn. The relation between the work performed and the energy liberated in muscular contraction. *J. Physiol.* 184:373–395, 1924.

[73] C. S. Peskin and G. Oster. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophys. J.* 68:202s–211s, 1995.

[74] T. Duke and S. Leibler. Motor protein mechanics: A stochastic model with minimal mechanochemical coupling. *Biophys. J.* 71:1235–1247, 1996.

[75] R. D. Astumian and I. Derényi. Fluctuation driven transport and models of molecular motors and pumps. *Eur. Biophys. J.* 27:474–489, 1998.

[76] R. D. Astumian and I. Derényi. A chemically reversible brownian motor: application to kinesin and ncd. *Biophys. J.* 77:993–1002, 1999.

[77] Y. Okada and N. Hirokawa. A processive single-headed motor: kinesin superfamily protein KIF1A. *Science* 283:1152–1157, 1999.

[78] A. F. Huxley. Muscle structure and theories of contraction. *Prog. Biophys.* 7:255–318, 1957.

[79] T. L. Hill. Theoretical formalism for the sliding filament model of contraction of striated muscle, part I. *Prog. Biophys. Mol. Biol.* 28:267–340, 1974.

[80] E. Eisenberg, T. L. Hill, and Y. Chen. Cross-bridge model of muscle contraction. *Biophys. J.* 29:195–227, 1980.

[81] E. Pate and R. Cooke. A model of crossbridge action: the effect of ATP, ADP and P_i. *J. Muscle Res. Cell Motil.* 10:181–196, 1989.

[82] E. Pate, H. White, and R. Cooke. Determination of the myosin step size from mechanical and kinetic data. *Proc. Natl. Acad. Sci. USA* 90:2451–2455, 1993.

[83] N. J. Córdova, B. Ermentrout, and G. F. Oster. Dynamics of single-motor molecules: the thermal ratchet model. *Proc. Natl. Acad. Sci. USA* 89:339–343, 1992.

[84] S. Leibler and D. A. Huse. Porters versus rowers: a unified stochastic model of motor proteins. *J. Cell. Biol.* 121:1357–1368, 1993.

[85] D. A. Smith and M. A. Geeves. Strain-dependent cross-bridge cycle for muscle. *Biophys. J.* 69:524–537, 1995.

[86] D. A. Smith and M. A. Geeves. Strain-dependent cross-bridge cycle for muscle II. Steady-state behavior. *Biophys. J.* 69:538–552, 1995.

[87] T. A. J. Duke. Molecular model of muscle contraction. *Proc. Natl. Acad. Sci. USA* 96:2770–2775, 1999.

[88] A. V. Hill. The heat of shortening and the dynamic constants of muscle. *Proc. R. Soc. Lond.* B126:136–195, 1938.

[89] K. A. P. Edman and J. C. Hwang. The force-velocity relationship in vertebrate muscle fibres at varied tonicity of the extracellular medium. *J. Physiol.* 269:255–272, 1977.

[90] R. Cooke and W. Bialek. Contraction of glycerinated muscle fibers as a function of the MgATP concentration. *Biophys. J.* 28:241–258, 1979.

[91] R. Cooke and E. Pate. The effect of ADP and phosphate on the contraction of muscle fibers. *Biophys. J.* 48:789–798, 1985.

[92] Y. Harada, A. Noguchi, A. Kishino, and T. Yanagida. Sliding movement of single actin filaments on one-headed myosin filaments. *Nature* 326:805–808, 1987.

[93] E. Pate, M. Lin, K. Franks-Skiba, and R. Cooke. Contraction of glycerinated rabbit slow-twitch muscle fibers as a function of MgATP contraction. *Am. J. Physiol.* 262:C1039–C1046, 1992.

[94] B. Katz. The relation between force and speed in muscle contraction. *J. Physiol.* 96:45–64, 1939.

[95] J. Sleep and H. Glyn. Inhibition of myofibrillar and actomyosin subfragment 1 adenosinetriphosphatase by adenosine 5'-diphosphate and adenyl-5'-yl imidodiphosphate. *Biochemistry* 25:1149–1154, 1986.

[96] A. F. Huxley and R. M. Simmons. Proposed mechanism of force generation in striated muscle. *Nature* 233:533–538, 1971.

[97] A. Ishijima, H. Kojima, T. Funatsu, M. Tokunaga, H. Higuchi, H. Tanaka, and T. Yanagida. Simultaneous observation of individual ATPase and mechanical events by a single myosin molecule during interaction with actin. *Cell* 92:161–171, 1998.

[98] C. Veigel, L. M. Coluccio, J. D. Jontes, J. C. Sparrow, R. A. Milligan, and J. E. Molloy. The motor protein myosin-I produces its working stroke in two steps. *Nature* 398:530–533, 1999.

[99] K. Kitamura, M. Tokunaga, A. H. Iwane, and T. Yanagida. A single myosin head moves along an actin filament with regular steps of 5.3 nanometres. *Nature* 397:129–134, 1999.

[100] I. Derényi and T. Vicsek. Unified quantitative theory for the ATP synthase. *Europhys. Lett.*, 1999. (submitted).

[101] T. Elston, H. Wang, and G. Oster. Energy transduction in ATP synthase. *Nature* 391:510–513, 1998.

[102] H. Wang and G. Oster. Energy transduction in the F1 motor of ATP synthase. *Nature* 396:279–282, 1999.

[103] P. Dimroth, H. Wang, M. Grabe, and G. Oster. Energy transduction in the sodium F-ATPase of propionigenium modestum. *Proc. Natl. Acad. Sci. USA* 96:4924–4929, 1999.

[104] R. Yasuda, H. Noji, K. Kinosita, and M. Yoshida. F_1-ATPase is a highly efficient molecular motor that rotates with discrete 120° steps. *Cell* 93:1117–1124, 1998.

[105] I. Derényi, P. Tegzes, and T. Vicsek. Collective transport in locally asymmetric periodic structures. *Chaos* 8:657–664, 1998.

[106] I. Derényi and T. Vicsek. Cooperative transport of Brownian particles. *Phys. Rev. Lett.* 75:374–377, 1995.

[107] I. Derényi and A. Ajdari. Collective transport of particles in a "flashing" periodic potential. *Phys. Rev. E* 54:R5–R8, 1996.

[108] F. Jülicher and J. Prost. Cooperative molecular motors. *Phys. Rev. Lett.* 75:2618–2621, 1995.

[109] Z. Csahók, F. Family, and T. Vicsek. Transport of elastically coupled particles in an asymmetric periodic potential. *Phys. Rev. E* 55:5179–5183, 1997.

[110] M. R. Evans, D. P. Foster, C. Godrèche, and D. Mukamel. Spontaneous symmetry breaking in a one dimensional driven diffusive system. *Phys. Rev. Lett.* 74:208–211, 1995.

[111] F. Jülicher and J. Prost. Spontaneous oscillations of collective molecular motors. *Phys. Rev. Lett.* 78:4510–4513, 1997.

Chapter 6

Collective motion

6.1 Flocking: collective motion of self-propelled particles

Introduction

The collective motion of various organisms like the flocking of birds (Fig. 6.1) [1], swimming of schools of fish [2], motion of herds of quadrupeds, migrating bacteria (Sec. 4.1), ants [3] or pedestrians (Sec. 6.2) is a fascinating phenomenon capturing our eyes when we observe our natural environment. In addition to the aesthetic aspects, studies on this collective behaviour can have interesting applications as well, e.g., modelling the motion of a crowd of people can help urban designers. Here we address the question whether there are some global, perhaps universal features of this type of behaviour when many organisms are involved and parameters like the level of *perturbations* or the mean *distance* between the individuals are changed.

These studies are also motivated by recent developments in areas related to statistical physics. Concepts originated from the physics of phase transitions in equilibrium systems [4, 5] such as scale invariance and renormalisation have also been shown to be useful in the understanding of various non-equilibrium systems, typical in our natural and social environment. Motion and related transport phenomena represent characteristic aspects of many non-equilibrium processes and they are essential features of most living systems.

To understand the collective motion of large groups of organisms in Ref. [6] the concept of *self propelled particle* (SPP) models was introduced. As the motion of flocking organisms is usually controlled by interactions with their neighbours, the SPP models consist of locally interacting particles with an intrinsic driving force, hence

Figure 6.1: Flocking of birds as a typical manifestation of collective motion in biology

with a finite steady velocity. Because of their simplicity, such models represent a statistical approach complementing other studies which take into account more details of the actual behaviour [1, 7, 8], but treat only a moderate number of organisms and concentrate less on the large scale behaviour.

In spite of the analogies with *ferromagnetic* models, the general behaviour of SPP systems can be quite different from those observed in equilibrium magnets. In particular, equilibrium ferromagnets possessing continuous rotational symmetry do not have ordered phase at finite temperatures in two dimensions [9]. However, in 2d SPP models an ordered phase can exist at finite noise levels (temperatures) as it was first demonstrated by simulations [6, 10] and explained by a theory of flocking developed by Toner and Tu [11]. Further studies revealed that modelling collective motion leads to interesting specific results in all of the relevant dimensions (from 1 to 3).

6.1.1 Models and simulations

A generic model for two dimensional SPP systems

The simplest model, introduced in [6], consists of particles moving on a plane with periodic boundary condition. The particles are characterised by their (off-lattice) location $\vec{x}_i$ and velocity $\vec{v}_i$ pointing in direction ϑ_i. The self-propelled nature of the particles is manifested by keeping the magnitude of the velocity fixed to v_0. Particles

interact through the following local rule: at each time step a given particle assumes the *average direction of motion* of the particles in its local neighbourhood $S(i)$ (e.g., in a circle of some given radius centered at the position of the ith particle) with some uncertainty, as described by

$$\vartheta_i(t+\Delta t) = \langle \vartheta(t) \rangle_{S(i)} + \xi, \tag{6.1}$$

where the noise ξ is a random variable with a uniform distribution in the interval $[-\eta/2, \eta/2]$. The locations of the particles are updated as

$$\vec{x}_i(t+\Delta t) = \vec{x}_i(t) + \vec{v}_i(t)\Delta t \tag{6.2}$$

with $|\vec{v}_i| = v_0 = const$, see Fig. 6.2.

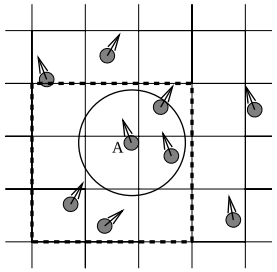


Figure 6.2: Schematic illustration of the model. Particles move off-lattice on a plane and interact with other particles located in the local surrounding, which can be either a circle or 9 neighbouring cells in an underlying lattice. We plot these interaction areas for particle A with a solid and dashed line, respectively.

The model defined by Eqs. (6.1) and (6.2) is a transport related, non-equilibrium analog of the ferromagnetic models [12]. The analogy is as follows: the Hamiltonian tending to align the spins in the same direction in the case of equilibrium ferromagnets is replaced by the rule of aligning the direction of motion of particles, and the amplitude of the random perturbations can be considered proportional to the temperature for $\eta \ll 1$. From a hydrodynamical point of view, in SPP systems the momentum of the particles is *not* conserved. Thus, the flow field emerging in these models can considerably differ from the usual behaviour of fluids.

6.1.2 Scaling properties

The model defined through Eqs. (6.1) and (6.2) was studied by performing large-scale Monte-Carlo simulations in [10]. Due to the simplicity of the model, only two control parameters should be distinguished: the (average) density of particles ϱ and the amplitude of the noise η. Depending on the value of these parameters the model can exhibit various type of behaviours as Fig. 6.3 demonstrates.

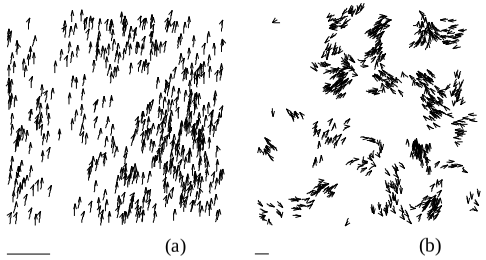


Figure 6.3: Typical configurations of SPPs displayed for various values of density and noise. The actual velocity of a particle is indicated by a small arrow, while its trajectory for the last 20 time step is shown by a short continuous curve. For comparison, the radius of the interaction is displayed as a bar. (a) At high densities and small noise ($N = 300$, $L = 5$ and $\eta = 0.1$) the motion becomes ordered. (b) For small densities and noise ($N = 300$, $L = 25$ and $\eta = 0.1$) the particles tend to form groups moving coherently in random directions. (After [6].)

For the statistical characterisation of the system a well-suited order parameter is the magnitude of the *average momentum* of the particles:

$$\phi \equiv \frac{1}{N} \left| \sum_j \vec{v}_j \right|. \tag{6.3}$$

This measure of the net flow is non-zero in the ordered phase, and vanishes (for an infinite system) in the disordered phase. Since the simulations were started from a random, disordered configuration, $\phi(t=0) \approx 0$. After some relaxation time a steady state emerges indicated by the convergence of the cumulative average $(1/\tau)\int_0^\tau \phi(t)dt$. The stationary values of ϕ are plotted in Fig. 6.4a vs η for $\varrho = 2$ and various system sizes L. For weak noise the model displays an ordered motion, i.e. $\phi \approx 1$, which disappears in a continuous manner by increasing η.

As $L \to \infty$, the numerical results show the presence of a phase transition described by

$$\phi(\eta) \sim \begin{cases} \left(\frac{\eta_c(\varrho)-\eta}{\eta_c(\varrho)}\right)^\beta & \text{for } \eta < \eta_c(\varrho) \\ 0 & \text{for } \eta > \eta_c(\varrho) \end{cases}, \tag{6.4}$$

where $\eta_c(\varrho)$ is the critical noise amplitude that separates the ordered and disordered phases and

$$\beta_{2d} = 0.42 \pm 0.03, \tag{6.5}$$

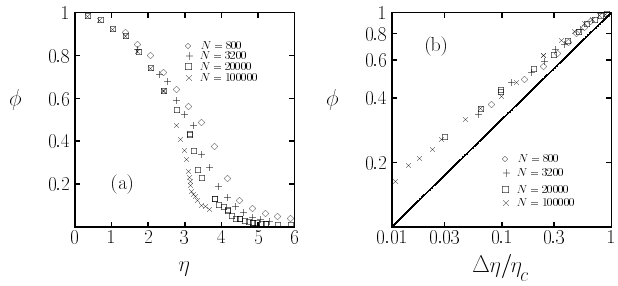


Figure 6.4: (a) The average momentum of the 2d SPP model in the steady state vs the noise amplitude η for $\varrho = 2$ and four different system sizes [($\diamond$) $N = 800$, $L = 20$; ($+$) $N = 3200$, $L = 40$; ($\circ$) $N = 20000$, $L = 100$ and ($\times$) $N = 10^5$, $L = 223$]. (b) The order present at small η disappears in a continuous manner reminiscent of second order phase transitions: $\phi \sim [(\eta_c(L) - \eta)/\eta_c(L)]^\beta \equiv (\Delta\eta/\eta_c)^\beta$, with $\beta = 0.42$, different from the mean-field value 1/2 (solid line). (After [10].)

found to be *different* from the the mean-field value 1/2 (Fig 6.4b) and *insensitive* to the specific choice of the interaction range $S(i)$.

Next we discuss the role of density. In Fig. 6.5a, $\phi(\eta)$ is plotted for $L = 100$ and various values of ϱ. One can observe that the long-range ordered phase is present for any ϱ, but for a fixed value of η, ϕ vanishes with decreasing ϱ. All of the $\phi(\eta, \varrho)$ functions can be collapsed to the same function $\tilde{\phi}(x)$ by rescaling η with $\eta_c(\varrho)$:

$$\phi(\eta, \varrho) = \tilde{\phi}(\eta/\eta_c(\varrho)), \tag{6.6}$$

where $\tilde{\phi}(x) \sim (1 - x)^\beta$ for $x < 1$, and $\tilde{\phi}(x) \approx 0$ for $x > 1$, see Fig. 6.5b. The critical line $\eta_c(\varrho)$ in the $\eta - \varrho$ parameter space was found to follow

$$\eta_c(\varrho) \sim \varrho^\kappa, \tag{6.7}$$

with $\kappa = 0.45 \pm 0.05$ (see Fig. 6.5c). This critical line is qualitatively different from that of the diluted ferromagnets, since here the critical density at $\eta \to 0$ (corresponding to the percolation threshold for diluted ferromagnets, see, e.g., [12]) is vanishing.

Note, that Eq.(6.6) also implies that the exponent β', defined as $\phi \sim (\varrho - \varrho_c)^{\beta'}$ for $\varrho > \varrho_c$ [6], must be equal to β, since

$$\phi(\eta, \varrho_c(\eta) + \epsilon) = \tilde{\phi}\left(\frac{\eta}{\eta_c[\varrho_c(\eta) + \epsilon]}\right) \approx \tilde{\phi}\left(1 - \frac{1}{\eta}\frac{\partial \eta_c}{\partial \varrho}\Big|_{\varrho_c} \epsilon\right) \sim \epsilon^\beta, \tag{6.8}$$

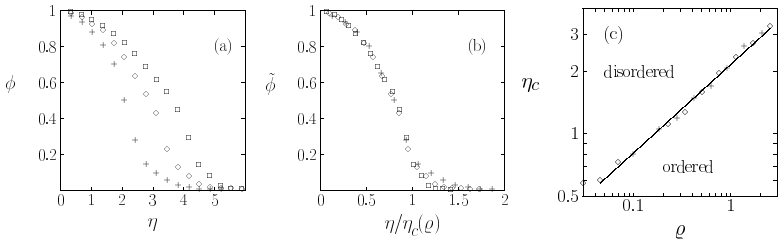


Figure 6.5: (a) The average momentum of the system in the steady state vs η for $L = 100$ and three different densities [($\circ$) $\varrho = 4$, ($\diamond$) $\varrho = 2$ and ($+$) $\varrho = 0.5$]. (b) The $\phi(\eta)$ functions parameterised by various ϱ can be collapsed onto a single curve $\tilde{\phi}[\eta/\eta_c(\varrho)] \equiv \phi(\eta, \varrho)$. (c) The critical line in the $\eta - \varrho$ phase space is a power-law in the examined regime: $\eta_c(\varrho) \sim \varrho^\kappa$ with $\kappa \approx 0.45$ (solid line) for a system of size $L = 100$. (After [10].)

where $\eta_c(\varrho)$ denotes the inverse function of $\varrho_c(\eta)$ as $\eta \equiv \eta_c[\varrho_c(\eta)]$.

As a further analogy with equilibrium phase transitions, the fluctuations of the order parameter also increase on approaching the critical line. Fig. 6.6a shows the standard deviation σ of the total momentum vs the rescaled noise amplitude $x \equiv \eta/\eta_c(\varrho)$ for various densities and $L = 100$. The tails of the curves are symmetric, and decay as power-laws (see Fig. 6.6b),

$$\sigma(x) \sim |1 - x|^{-\gamma}, \tag{6.9}$$

with an exponent γ close to 2, which value is, again, different from the mean-field result.

These findings indicate that SPP systems can be quite well characterised using the framework of classical critical phenomena, but also show surprising features when compared to the analogous equilibrium systems. The velocity v_0 provides a control parameter which switches between the SPP behaviour ($v_0 > 0$) and an XY ferromagnet ($v_0 = 0$). Indeed, for $v_0 = 0$ Kosterlitz-Thouless vortices [13] can be observed in the system, which are *unstable* (Fig. 6.7.) for any nonzero v_0 investigated in [10].

3d SPP system

In two dimensions an effective long range interaction can build up because the migrating particles have a considerably higher chance to get close to each other and interact than in three dimensions (where, as is well known, random trajectories do

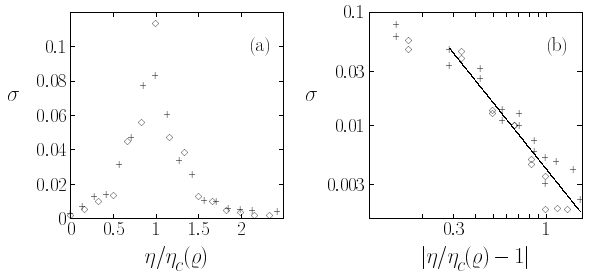


Figure 6.6: (a) The rms deviation of the order parameter σ in the steady state, for $L = 100$, and two independent data sets ($\diamond, +$), each averaged over runs with $\varrho = 0.3, 0.6, 0.9, 1.2, 1.5$ and 2 . (b) The divergence is symmetric and its tail decays as $\sigma(x) \sim |1-x|^{-2}$ (solid line), where x denotes the rescaled noise amplitude $\eta/\eta_c(\varrho)$. (After [10].)

not overlap). The less interaction acts against ordering. On the other hand, in three dimensions even regular ferromagnets order. Thus, it is interesting to see how these two competing features change the behaviour of 3d SPP systems. The convenient generalisation of (6.1) for the 3d case can be the following [14]:

$$\vec{v}_i(t+\Delta t) = v_0\ \mathbf{N}(\ \mathbf{N}(\langle \vec{v}(t)\rangle_{S(i)}) + \vec{\xi}), \tag{6.10}$$

where $\mathbf{N}(\vec{u}) = \vec{u}/|\vec{u}|$ and the noise $\vec{\xi}$ is uniformly distributed in a sphere of radius η.

Generally, the behaviour of the system was found [14] to be similar to that of described in the previous section. The long-range ordered phase was present for any ϱ, but for a fixed value of η, ϕ vanished with decreasing ϱ. To compare this behaviour to the corresponding diluted ferromagnet, $\phi(\eta, \varrho)$ was determined for $v_0 = 0$, when the model reduces to an equilibrium system of randomly distributed "spins" with a ferromagnetic-like interaction. Again, a major difference was found between the SPP and the equilibrium models (Fig. 6.9): in the static case the system *does not order* for densities below a critical value close to 1 which corresponds to the percolation threshold of randomly distributed spheres in 3d.

1d SPP system

Since in 1d the particles cannot get around each other, some of the important features of the dynamics present in higher dimensions are lost. On the other hand, motion in

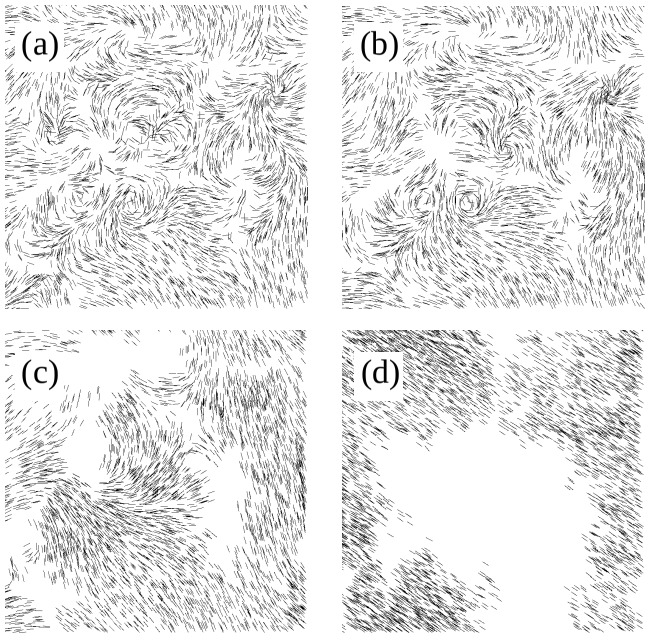


Figure 6.7: Snapshots of the time development of a system with $N = 4000$, $L = 40$ and $v_0 = 0.01$ at 50 (a), 100 (b), 400 (c) and 3000 (d) Monte-Carlo steps. First the behaviour is reminiscent of the equilibrium XY model, where the long range order is missing since vortices are present in the system. Here the vortices are unstable, and finally a self-organised long-range order develops. (After [10].)

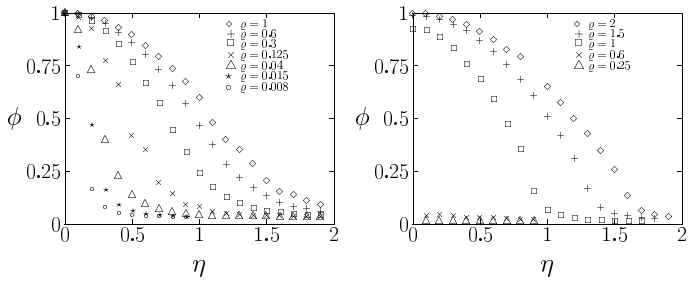


Figure 6.8: (a) The order parameter ϕ vs the noise amplitude η (N=1000) for the 3d SPP system. (b) As a comparison, when $v_0 = 0$ the behaviour of the model is similar to diluted ferromagnets: ϕ vanishes below the percolation threshold ($\rho^* \simeq 1$). (after [14].)

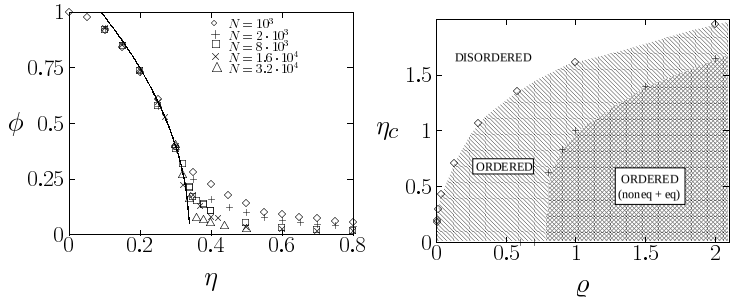


Figure 6.9: (a) The order parameter ϕ vs the noise amplitude η for the 3d SPP model for various system sizes. In these simulations the density was fixed and the system size (number of particles N) was increased to demonstrate that for any system size the ordered phase disappears in a continuous manner beyond a size dependent critical noise. (b) Phase diagram of the 3d SPP and the corresponding ferromagnetic system. The diamonds show our estimates for the critical noise for a given density for the SPP model and the crosses show the same for the static case. The SPP system becomes ordered in the whole region below the curved line connecting the diamonds, while in the static case the ordered region extends only down to the percolation threshold $\rho \simeq 1$. (After [14].)

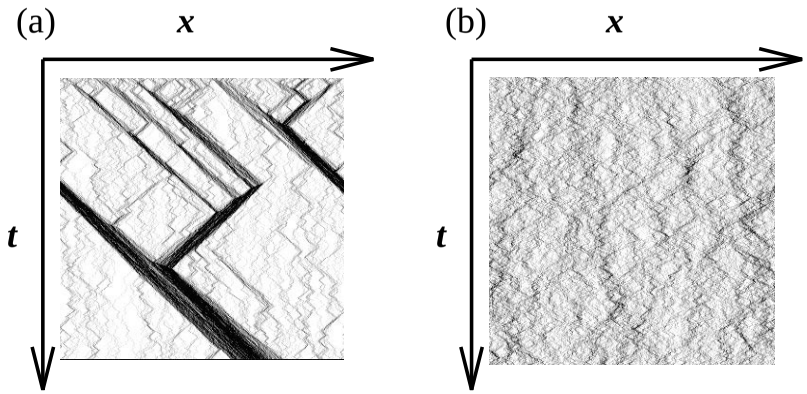


Figure 6.10: The first 3000 time steps of the 1d SPP model [$L = 300$, $N = 600$, $\eta = 2.0$ (a) and $\eta = 4.0$ (b)]. The darker gray scale represents higher particle density. (After [15].)

1d implies new interesting aspects (groups of the particles have to be able to change their direction for the opposite in an organised manner) and the algorithms used for higher dimensions should be modified to take into account the specific crowding effects typical for 1d (the particles can slow down before changing direction and dense regions may be built up of momentarily oppositely moving particles).

In a way the system studied below can be considered as a model of people, moving in a narrow channel. Imagine that a fire alarm goes on, the tunnel is dark, smoky, everyone is extremely excited. People are both trying to follow the others (to escape together) and behave in an erratic manner (due to smoke and excitement).

Thus, in [15] N off-lattice particles along a line of length L have been considered. The particles are characterised by their coordinate x_i and dimensionless velocity u_i updated as

$$x_i(t + \Delta t) = x_i(t) + v_0 u_i(t) \Delta t, \tag{6.11}$$

$$u_i(t + \Delta t) = G\Big(\langle u(t) \rangle_{S(i)}\Big) + \xi_i. \tag{6.12}$$

The local average velocity $\langle u \rangle_{S(i)}$ for the ith particle is calculated over the particles located in the interval $[x_i - \Delta, x_i + \Delta]$, where $\Delta = 1/2$. The function G incorporates both the "propulsion" and "friction" forces which set the velocity to a prescribed

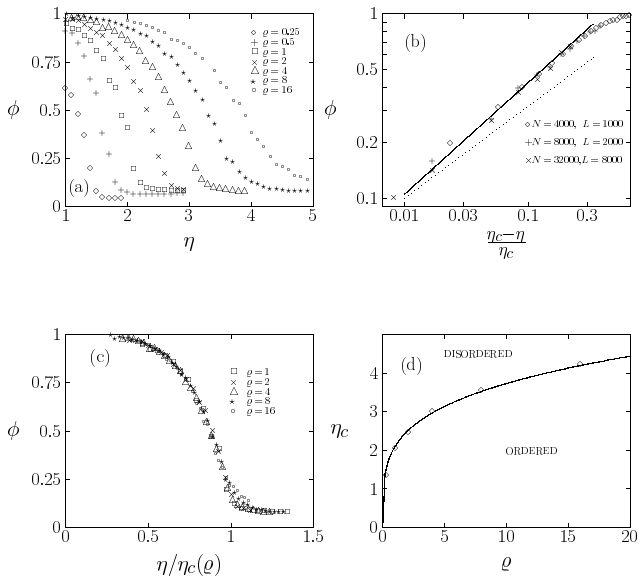


Figure 6.11: (a) The order parameter ϕ vs η for the 1d SPP model ($L = 1000$). (b) ϕ vanishes as a power-law in the vicinity of $\eta_c(\rho)$. Note the increasing scaling regime with increasing L. The solid line is a power-law fit with an exponent $\beta = 0.6$, while the dotted line shows the mean-field slope $\beta = 1/2$ as a comparison. (c) As in the 2d case, the $\phi(\eta, \varrho)$) functions can be collapsed onto a single curve. (d) Phase diagram in the $\rho - \eta$ plane, the critical line follows $\eta_c(\rho) \sim \rho^\kappa$. The solid curve represents a fit with $\kappa = 1/4$. (After [15].)

value v_0 on the average: $G(u) > u$ for $u < 1$ and $G(u) < u$ for $u > 1$. In the numerical simulations [15] one of the simplest choices for G was implemented as

$$G(u) = \begin{cases} (u+1)/2 & \text{for } u > 0 \\ (u-1)/2 & \text{for } u < 0, \end{cases} \tag{6.13}$$

and random initial and periodic boundary conditions were applied.

In Fig. 6.10 we show the time evolution of the model for $\eta = 2.0$. In a short time the system reaches an ordered state, characterised by a spontaneous broken symmetry and clustering of the particles. In contrast, for $\eta = 4.0$ the system remains in a disordered state.

The order parameter for various values of ϱ is shown in Fig. 6.11. As in the two and three dimensional cases, the emergence of the ordered phase was observed through a second order phase transition (Fig. 6.11b) with

$$\beta_{1d} = 0.60 \pm 0.05, \tag{6.14}$$

which is different from both the the mean-field value 1/2 and $\beta_{2d} \approx 0.4$ found in 2d. Fig. 6.11c shows that the various $\phi(\eta, \varrho)$ curves can be collapsed onto a single function $\phi_0(x)$, where $x = \eta/\eta_c(\varrho)$, just like in 2d. The $\rho - \eta$ phase diagram is shown in Fig. 6.11d. The critical line, $\eta_c(\varrho)$, follows (6.7) with

$$\kappa_{1d} \approx 1/4. \tag{6.15}$$

6.1.3 Further variants of SPP models

Effect of boundary conditions. In various simulations [17, 18, 16] the boundary conditions were changed to *reflective circular walls* to gain insight into the importance of the boundary conditions in the SPP models. To avoid singular behaviour at the boundaries, i.e., the aggregation of particles in a narrow zone, a *short range "hard-core" repulsion* was also incorporated into the models: E.g., in the simulations described in [16], if the distance between the particles is smaller than a certain ϵ^* value they repel each other and the direction of their motion will be given by the following expression instead of Eq. 6.1:

$$\vartheta_i(t + \Delta t) = \Phi\Big(- \sum_{\substack{j \neq i \\ |\vec{x}_j - \vec{x}_i| < \epsilon^*}} \mathbf{N}(\vec{x}_j(t) - \vec{x}_i(t)) \Big), \tag{6.16}$$

as demonstrated in Fig. 6.12.

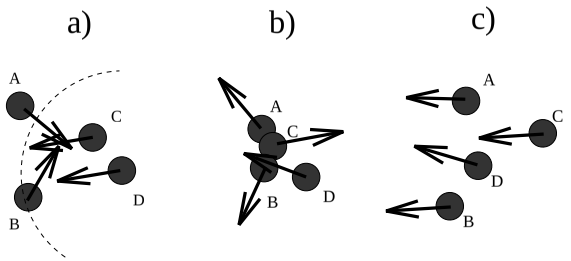


Figure 6.12: Schematic illustration of the model with both hard-core interaction and ferromagnetic like coupling of the velocities. The figure shows the position and velocity (arrows) of the particles at time t (a), $t+\Delta t$ (b) and $t+2\Delta t$ (c). The range of the hard-core repulsion (ϵ^*) is represented as a filled circle. The dashed circle on Fig. (a) indicates the range of velocity interaction for particle D. In the collision at time $t+\Delta t$ particles A,B and C repel each other, while the new velocity of particle D was determined by the average direction of motion of B,C and D. At time $t+2\Delta t$ no collision occurs, each particle is moving approximately with the local average velocity. (After [16].)

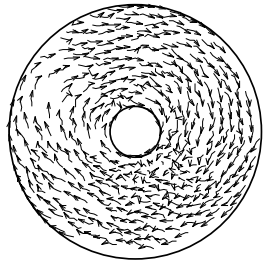


Figure 6.13: A possible stationary state of the model represented in Fig. 6.12 with reflective boundary conditions. (After [16])

In such simulations rotation of the particles develop (see Fig. 6.13) in the high density, low noise regime. The direction of the rotation is selected by spontaneous symmetry breaking, thus both clockwise and anti-clockwise spinning "vortices" can emerge. This rotating state should be distinguished from the Kosterlitz-Thouless vortices [13], since in our case a *single* vortex develops irrespective to the system size. In Sec. 4.1 we give examples of such vortices developing in nature.

Lattice gas approach. As discrete cellular automata models often have the advantages that they allow both an approximate analytic and large-scale numerical investigations, in [19] the following discrete model for SPP systems was introduced and studied: Each site of a triangular lattice can be either empty or occupied by one or more particles. If more than one particle is present at a site then only one of them, called the "lowest", will interact with other particles. The particles are characterised by their position $\vec{r}_i$ and velocity $\vec{v}_i$ ($i = 1 \dots N$) which is of unit length ($|\vec{v}_i| = 1$) and can point in any of the six lattice directions ($\vec{u}_\alpha$). At one time step the positions and velocities of all particles are updated simultaneously according to the following rules:

1. (a) for non-interacting particles a random direction is assigned;
   (b) otherwise a new velocity $\vec{u}_\alpha$ is chosen with a Boltzmann distribution as

$$P(\vec{u}_\alpha) = \frac{1}{\mathcal{Z}} \exp(-\beta \vec{u}_\alpha \sum_{j \in \mathrm{lnn}} \vec{v}_j), \tag{6.17}$$

   where $\mathcal{Z}$ is a normalising factor so that $\sum_{\alpha=1}^{6} P(\vec{u}_\alpha) = 1$, and β is $1/T$ ($k_B = 1$). Thus, in this case the there is a direct correspondence between the noise strength of the former models and some temperature T. The summation goes over the nearest neighbours which are in lowest position (lnn). This rule is equivalent to having an equilibrium spin system with the Hamiltonian

$$H = -\frac{1}{2} \sum_{i,j \in \mathrm{lnn}} \vec{v}_i \vec{v}_j. \tag{6.18}$$

2. every particle is moved one lattice unit in direction of its velocity:

$$\vec{r}_i \leftarrow \vec{r}_i + \vec{v}_i.$$

This system in closely related to the 6-state Potts model [20] since we have $q = 6$ possible velocity states for a particle, but unlike the Potts model these states are not orthogonal, and the system is not in equilibrium.

As initial configuration random distribution for the position and velocity of the particles was chosen in [19]. Typical configurations of the system for various temperatures T and particle densities ϱ are shown in Fig. 6.14. In a similar fashion to the behaviour of the corresponding models with rotational symmetry, it can be seen that at low temperatures the particles tend to form ordered clusters.

Fig. 6.15a. demonstrates the order parameter as a function of the temperature for $\varrho = 0.9$. Note the discontinuity characteristic for first order transitions at T_c.

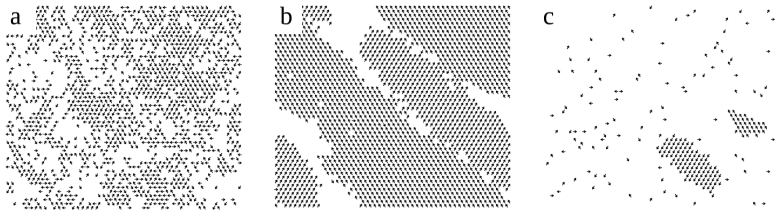


Figure 6.14: Typical snapshots of the system at (a) high temperature, (b) low temperature at $\varrho = 0.9$ and (c) intermediate temperature at a lower density. Note the appearance of ordered clusters. (Only the particles in the lowest position are drawn.) (After [19].)

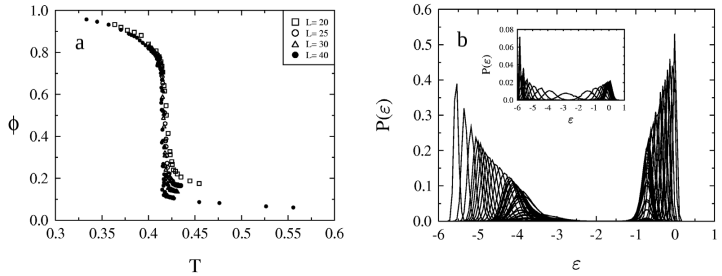


Figure 6.15: Statistical characterisation of the transition between ordered and disordered systems. (a) The order parameter (ϕ) as a function of the temperature (T) for density $\varrho = 0.9$. (b) The energy distribution ($P(\varepsilon)$) for various temperatures below and above the transition ($\varrho = 0.9$). Note the energy gap between $\varepsilon \approx -1$ and $\varepsilon \approx -3.5$. Inset shows the distribution for Ising spins instead of mobile particles where the transition is continuous. (After [19].)

A strong evidence supporting the presence of a first order transition is shown in Fig. 6.15b, where the distribution $P(\varepsilon)$ of the energy of the particles is plotted for a number of different temperatures below and above T_c. One can clearly see a *gap* in the distributions at intermediate energies which is a unique feature of first order phase transitions [21]. The inset in the figure shows the distribution for Ising type interaction of non-moving particles in the same system where the transition is known to be second order. This transition was found to be present even for very small particle densities although the position of the critical temperature decreases with the density just as in the case of the off-lattice models.

SPP model in SOC state. In [22] another variant of the 2d model was investigated where the noise amplitude was dependent on the local particle density ρ_ℓ as

$$\eta = \pi \rho_\ell^{-\alpha}, \tag{6.19}$$

instead of being an external control parameter. In addition to the Eqs (6.1) and (6.2) the number of particles were determined by local population dynamics rules.

The angular correlation plot is shown in Fig. 6.16: for $\alpha \to 0$ there are many flocks in the system moving independently, while for large α values all the particles are organised into a single flock. The transition was found to be continuous between the two regimes.

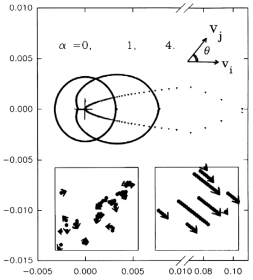


Figure 6.16: Polar plot of the angular correlation function of the velocities of the migrating individuals for three values of α. The inserts below show configurations with $\alpha = 0$ (left) with random migration and $\alpha = 4$ (right) where ordered motion develops.

Deterministic model for the motion of organisms. Many attempts have been made to describe the actual behaviour of various organisms using effective interaction potentials or forces. In [7] a deterministic model was introduced describing the dynamics of the individuals in terms of their velocity, position, and a *heading unit*

vector $\vec{n}$ as

$$m\dot{\vec{v}_i} = -\gamma\vec{v_i} + a\vec{n_i} + \sum_{i\neq j}\alpha_{ij}\vec{f}_{ij} + \vec{g_i} \tag{6.20}$$

$$\tau\dot{\vec{n}_i} = \vec{n_i}\times\mathbf{N}(\vec{v_i})\times\vec{n_i} \tag{6.21}$$

where Eq.(6.20) is Newton's equation of motion for particles of mass m; γ is the damping coefficient based on Stokes's law. The locomotive force of magnitude a acts in the heading direction $\vec{n}$. The $\vec{f}_{ij}$ force between particles j and i incorporates effects providing a mechanical equilibrium at a given distance r_0, and in general, $\vec{f}_{ij} \neq \vec{f}_{ji}$. The term $\vec{g}_i$ represent a force toward the center of the group. $\vec{v}$ and $\vec{n}$ are assumed to relax parallel with relaxation time τ according to Eq. (6.21).

The model is fairly complex and exhibits several distinct types of collective behaviour (Fig. 6.17). For example, a continuous transition was found as a function of the corresponding dimensionless parameters between a disordered (chaotic) and an ordered (non chaotic) phase, resembling the transition found in the stochastic 2d SPP model.

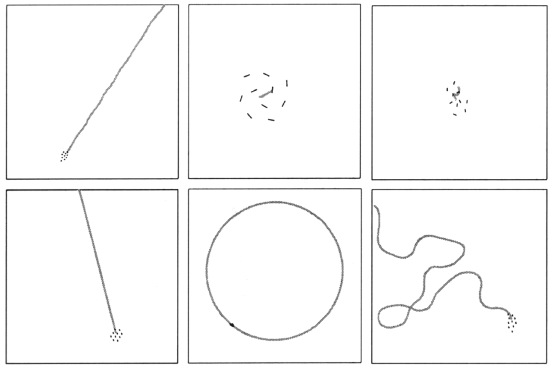


Figure 6.17: Patterns of clusters and trajectories of their centers for various values of the control parameters. Solid bars represent the motive elements while the gray line is the trajectory of the center of the cluster.

Theoretical results

To understand the properties of the ordered phase or the phase transitions observed in the models described in the previous section, efforts have been made to set up a consistent continuum theory in terms of $\mathbf{v}$ and ρ, representing the coarse-grained velocity and density fields, respectively. The first such approach was made by J. Toner and Y. Tu [11]. Here, however, we overview the efforts not in chronological order, but in a progressive fashion starting from the most simple mean-field lattice theory through the one dimensional system and arriving to the most difficult higher dimensional cases.

As we discussed in the previous section, in Ref. [19] a cellular automata SPP model was introduced and studied. The major advantage of such cellular automata models is the possibility of an approximate analytic investigation which, as we demonstrate here following [19], often gives surprisingly good numerical results and provides a general understanding of the system.

As the first step of building a mean-field theory, a *mean-field Hamiltonian* can be introduced [23] instead of Eq. (6.18) as

$$H_{\mathrm{MF}} = -\frac{1}{2}\sum_{i,j\in\mathrm{l}}\mathbf{v}_i\mathbf{v}_j, \tag{6.22}$$

where the summation goes over *all lowest* particles, not only the nearest neighbours. The mean field energy function can be written as

$$E_{\mathrm{MF}} = -\frac{1}{2}N\varrho_{\mathrm{eff}}\sum_{\alpha,\gamma=1}^{6}x_\alpha U_{\alpha\gamma}x_\gamma, \tag{6.23}$$

where $\varrho_{\mathrm{eff}} = 1-\exp(-\varrho)$ is the effective density, i.e., a site has on average $6\varrho_{\mathrm{eff}}$ occupied neighbouring sites, x_α is the fraction of particles travelling in the lattice direction α ($\sum_{\alpha=1}^{6}x_\alpha = 1$) and $U_{\alpha\gamma} = \mathbf{u}_\alpha\mathbf{u}_\gamma$ which for the case of the Potts model would be simply $U_{\alpha\gamma} = \delta_{\alpha\gamma}$. Then, the average energy per particle is

$$\varepsilon_{\mathrm{MF}} = \frac{E_{\mathrm{MF}}}{N} = -\frac{1}{2}\varrho_{\mathrm{eff}}\sum_{\alpha,\gamma=1}^{6}x_\alpha U_{\alpha\gamma}x_\gamma. \tag{6.24}$$

while the entropy per particle is

$$s_{\mathrm{MF}} = -\sum_{\alpha=1}^{6}x_\alpha\ln x_\alpha, \tag{6.25}$$

so for the free energy per particle we obtain

$$\beta f_{\rm MF} = \beta \frac{F_{\rm MF}}{N} = \beta \varepsilon_{\rm MF} - s_{\rm MF} = \sum_{\alpha=1}^{6} \left(x_\alpha \ln x_\alpha - \frac{1}{2} \varrho_{\rm eff} \beta x_\alpha \sum_{\gamma=1}^{6} U_{\alpha\gamma} x_\gamma \right). \qquad (6.26)$$

We intend to find the configuration x_α which minimises the free energy $f_{\rm MF}$. Since all the lattice directions are equivalent we can look for a solution in the form of

$$x_1 = \frac{1}{6} + \frac{5}{6} \phi_{\rm MF} \qquad (6.27)$$

and

$$x_{\alpha>1} = \frac{1}{6} - \frac{\phi_{\rm MF}}{6}, \qquad (6.28)$$

where $\phi_{\rm MF}$ is the mean field order parameter which satisfies the relation

$$\phi_{\rm MF} = \left| \sum_{\alpha=1}^{6} x_\alpha \mathbf{u}_\alpha \right|.$$

Substituting Eq. (6.27) and Eq. (6.28) to Eq. (6.26) after a bit of algebra one gets for the mean field free energy

$$\begin{aligned} \beta f_{\rm MF} = & -\frac{1}{2} \varrho_{\rm eff} \beta \ \phi_{\rm MF}^2 - \log \frac{1}{6} + \frac{5(1-\phi_{\rm MF})}{6} \log \frac{1-\phi_{\rm MF}}{6} \\ & + \frac{1+5\phi_{\rm MF}}{6} \log \frac{1+5\phi_{\rm MF}}{6}. \end{aligned} \qquad (6.29)$$

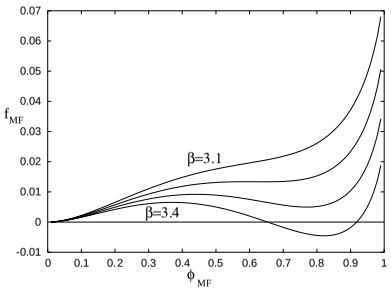


Figure: 6.18: The mean-field free energy $f_{\rm MF}$ as given by Eq. (6.29) vs the order parameter $\phi_{\rm MF}$ for $\varrho_{\rm eff} = 1$ and for various β values in the vicinity of the phase transition.

For high temperatures ($T > T_c$) this function has its minimum at $\phi_{\rm MF} = 0$ which corresponds to the disordered state of the system. At the critical temperature, which

can be derived from $f_{\rm MF}$ and in our case it is

$$T_c \approx \frac{\varrho_{\rm eff}}{3.353}, \qquad (6.30)$$

a non-trivial minimum appears as demonstrated in Fig. 6.18. The phase transition, like in the 6-state Potts model [24], is first order. The jump in the order parameter in this approximation is given exactly by

$$\Delta \phi_{\rm MF} = 0.8,$$

which is in reasonable agreement with the numerical findings presented in 6.15a.

It is also possible to carry out a mean-field analysis of the problem without assuming thermal equilibrium as was first described in [25]. The specific model we consider here is slightly different from the one studied in the previous section: Each node $\vec{r}$ can contain up to four particles in different "velocity channels" corresponding to nearest neighbour sites. The state of the entire lattice at time t is specified by the occupation numbers $s_i(\vec{r}, t) = 0, 1$ denoting the absence or presence of a particle in the channel $\vec{c}_i$ at node $\vec{r}$.

Similarly to the previous lattice model, the evolution from time t to $t+1$ proceeds in two stages: first an interaction step is performed during which the preinteraction states $\{s_i(\vec{r}, t)\}$ are replaced by a postinteraction state $\{\sigma_i(\vec{r}, t)\}$ according to stochastic rules that are applied to each node $\vec{r}$ independently. The interaction step is then followed by a propagation step during which particles move to nearest neighbour sites in the direction of their velocity as

$$s_i(\vec{r} + \vec{c}_i, t+1) = \sigma_i(\vec{r}, t). \qquad (6.31)$$

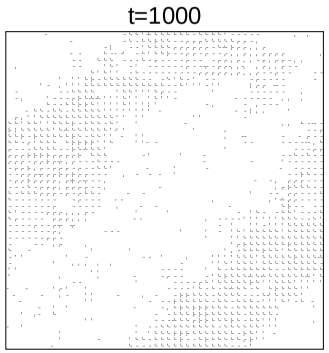


Figure: 6.19: Swarming behaviour in a cellular automaton model. Shown are the occupied velocity channels of the cellular automaton organised into a symmetry-broken state after 1000 time steps. (After [25])

The local alignment interaction is implemented through calculating the average flux of particles at the nearest neighbours

$$\vec{D}(\vec{r},t) = \sum_{p=1}^{4}\sum_{i=1}^{4} \vec{c}_i s_i(\vec{r}+\vec{c}_p,t). \tag{6.32}$$

The number of particles at each node, $\rho(\vec{r},t) = \rho[s(\vec{r},t)] \equiv \sum_{i=1}^{4} s_i(\vec{r},t)$, is conserved during interaction. Thus, if $\vec{J}(\sigma) = \sum_{i=1}^{4} \vec{c}_i\sigma_i$ denotes the particle flux immediately after interaction then the transition probability from $s(\vec{r},t)$ to $\sigma(\vec{r},t)$ in the presence of $\vec{D}(\vec{r},t)$ is given by

$$A[s\to\sigma|\vec{D}] = \frac{1}{\mathcal{Z}}\delta[\rho(\sigma),\rho(s)]\exp\left[\beta\vec{D}\cdot\vec{J}(\sigma)\right], \tag{6.33}$$

where the normalisation factor $\mathcal{Z}$ is chosen such that $\sum_\sigma A[s\to\sigma|\vec{D}] = 1$ for all s. Note that this interaction rule minimises the angle between the director field $\vec{D}$ and the postinteraction flux $\vec{J}(\sigma)$ in a similar manner to (6.17). Figure 6.19 shows the time evolution of an initially random distribution for $\beta > \beta_c$. The formation of locally aligned patches can clearly be observed.

The behaviour of such models can be analysed by considering the time evolution of a statistical ensemble of systems as discussed in [26]. For the mean-field description the ensemble-averaged occupation numbers $f_i(\vec{r},t) \equiv \overline{s_i(\vec{r},t)}$ are introduced. It is assumed that at each time step just before interaction the probability distribution is completely factorised, i.e., the probability $\mathcal{P}(\{s_i(\vec{r})\})$ to find a microstate $\{s_i(\vec{r})\}$ at time t is given by

$$\mathcal{P}(\{s_i(\vec{r})\}) = \prod_{\vec{r}}\prod_{i=1}^{4}[f_i(\vec{r},t)]^{s_i(\vec{r})}[1-f_i(\vec{r},t)]^{1-s_i(\vec{r})}. \tag{6.34}$$

We denote averaging using the above, factorised statistical weights by $\langle\cdots\rangle_{\rm MF}$. Averaging (6.31) and replacing the real ensemble average by $\langle\cdots\rangle_{\rm MF}$, i.e., neglecting all correlations between occupation numbers, we obtain a closed evolution equation for $f_i(\vec{r},t)$:

$$f_i(\vec{r}+\vec{c}_i,t+1) = f_i(\vec{r},t) + I_i(\vec{r},t). \tag{6.35}$$

Here the term $I_i(\vec{r},t) \equiv \langle\sigma_i(\vec{r},t) - s_i(\vec{r},t)\rangle_{\rm MF}$, taking values between -1 and 1, equals the average change in the occupation number of channel $(\vec{r},\vec{c}_i)$ during interaction.

As a consequence of the conservation of particle number, $\sum_i I_i = 0$, combined with the invariance of the interaction rules under discrete rotations and translations, a possible solution to Eq. (6.35) is

$$f_i(\vec{r},t) = f^* = \rho^*/4. \tag{6.36}$$

To assess the stability of this spatially homogeneous and stationary solution with respect to fluctuations

$$\delta f_i(\vec{r},t) = f_i(\vec{r},t) - f^*, \tag{6.37}$$

Eq. (6.35) can be linearised resulting in

$$\delta f_i(\vec{r}+\vec{c}_i,t+1) = \delta f_i(\vec{r},t) + \sum_{j,\vec{r}'}\left.\frac{\partial I_i}{\partial f_j(\vec{r}')}\right|^* \delta f_j(\vec{r}',t). \tag{6.38}$$

Since the interaction is local, (6.38) can be simplified to

$$\delta f_i(\vec{r}+\vec{c}_i,t+1) = \delta f_i(\vec{r},t) + \sum_{j,p}\Omega_{ij}^p\delta f_j(\vec{r}+\vec{c}_p,t) \tag{6.39}$$

with $p\in\{0,...,4\}$, $\vec{c}_0 = 0$ and

$$\Omega_{ij}^p = \left.\frac{\partial I_i}{\partial f_j(\vec{r}+\vec{c}_p)}\right|^*. \tag{6.40}$$

Fourier transforming (6.39) as $\hat{\delta f}_i(\vec{k},t) = \sum_{\vec{r}} e^{-i\vec{k}\cdot\vec{r}}\hat{\delta f}_i(\vec{r},t)$ we obtain

$$\hat{\delta f}_i(\vec{k},t+1) \simeq \sum_{j=1}^{4}\Gamma_{ij}(\vec{k})\hat{\delta f}_j(\vec{k},t) \tag{6.41}$$

with

$$\Gamma_{ij}(\vec{k}) = e^{-i\vec{k}\cdot\vec{c}_i}\left[\delta_{ij} + \sum_{p=0}^{4} e^{i\vec{k}\cdot\vec{c}_p}\,\Omega_{ij}^p\right]. \tag{6.42}$$

It can be shown [26, 25] that $\delta_{ij}+\Omega_{ij}^0 = 1/4$ for all i,j; this is a consequence of the fact that the outcome $\sigma(\vec{r})$ of an interaction step only depends on $s(\vec{r})$ through $\rho(\vec{r})$. For $1\le p\le 4$ the elements $\Omega_{ij}^p \equiv \omega_{ij}$ do not depend on p, and for symmetry reasons $(\omega)_{ij}$ is a cyclic matrix whose first row has the structure $(a+b,-b,-a+b,-b)$. To determine $a(\beta,\rho^*)$ and $b(\beta,\rho^*)$ for given values of the sensitivity β and the average density ρ^* the expression

$$\begin{aligned}\omega_{ij} = &\sum_{\{s(\vec{r}+\vec{c}_p)\}}\sum_{\sigma(\vec{r})}(\sigma_i(\vec{r}) - s_i(\vec{r}))\frac{s_j(\vec{r}+\vec{c}_1)-f^*}{f^*(1-f^*)}\\ &\times A[s\to\sigma|\vec{D}(\{s(\vec{r}+\vec{c}_p)\})]\prod_{p'=0}^{4}F(s(\vec{r}+\vec{c}_{p'})),\end{aligned} \tag{6.43}$$

can be evaluated (this is done numerically because of the highly nonlinear dependence on f_i and $\beta\vec{D}$, combined with the large number of terms) where $F^*(s) = \prod_{i=1}^{4} f^{*s_i}(1-f^*)^{1-s_i}$ is the factorised distribution of the homogeneous equilibrium system.

To investigate the stability of the spatially uniform state, i.e. $\vec{k}=0$, note that $\Gamma_{ij}(\vec{k}=0)$ has an eigenvalue $\lambda_1 = 1$ with corresponding eigenvector $e_1 = (1,1,1,1)$, reflecting the fact that the total density is conserved. Furthermore there is a twofold degenerate eigenvalue $\lambda_{x,y} = 8a$ with an eigenspace spanned by $e_x = (1,0,-1,0)$ and $e_y = (0,1,0,-1)$, corresponding to the x- and y components of the total particle flux. The remaining eigenvector $e_{x^2-y^2} = (1,-1,1,-1)$ has eigenvalue $\lambda_{x^2-y^2} = 16b$, corresponding to the difference between the number of horizontally and vertically moving particles. Numerically b is found [25] to be about two orders of magnitude smaller than a, so that the onset of instability of the homogeneous state is determined by the condition $\lambda_{x,y} = 1$. The location of the critical line in the $(\beta,\bar{\rho})$ parameter plane is shown in Fig. 6.20, which was obtained by numerically solving the equation $a(\beta,\bar{\rho}) = 1/8$. The resulting phase diagram is in qualitative agreement for $\rho^* \sim 1$ with the one shown in Fig. 6.5c.

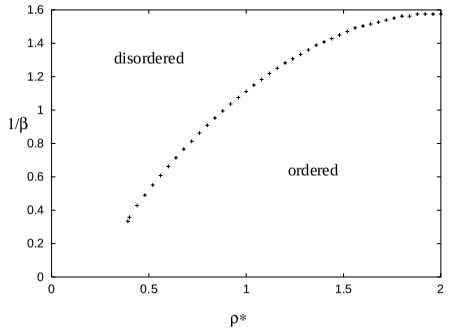


Figure: 6.20: Phase diagram for swarming model. Shown are the regions where the disordered state is stable and unstable against small fluctuations, as a function of "noise" $1/\beta$ and average density ρ^*. (After [25].)

6.1.4 Continuum equations for the 1d system

Now let us focus on the continuum approaches. As the simplest case, let us first investigate following [15] the 1d SPP system described in the previous section. We denote by $n(u,x,t)dudx$ the number of particles moving with a velocity in the range of $[v_0 u, v_0(u+du)]$ at time t in the $[x,x+dx]$ interval. The particle density $\rho(x,t)$ is then given as

$$\rho = \int n du, \tag{6.44}$$

while the local dimensionless average velocity $U(x,t)$ can be calculated as

$$\rho U = \int n u du. \tag{6.45}$$

According to the microscopic rule of the dynamics (6.12), in a given time interval $[t,t+\tau]$ all particles choose a certain velocity $v = v_0[G(\langle u\rangle)+\xi]$ and travel a distance $v\tau$.

It can be shown that in the $v_0\tau \ll 1$, and $\rho \gg 1$ limit

$$\partial_t \rho = -v_0 \partial_x(\rho U) + D\partial_x^2 \rho \tag{6.46}$$

with

$$D = v_0^2 \tau \sigma^2/2. \tag{6.47}$$

This equation corresponds to the conservation of the particles, and we have established the connection between its "macroscopical" parameters v_0, D and the parameters of the underlying microscopical dynamics. Notice, that the appearance of the diffusion term is a consequence of the non-vanishing correlation time τ.

Similarly, one can derive that

$$\partial_t U = f(U) + \mu^2 \partial_x^2 U + \alpha \frac{(\partial_x U)(\partial_x \rho)}{\rho} + \zeta, \tag{6.48}$$

where $f(U) = (G(U)-U)/\tau$, $\mu^2 = (dG/dU)/(6\tau)$, $\alpha = 2\mu^2$ and

$$\zeta = \frac{1}{\rho\tau}\int \nu u du. \tag{6.49}$$

Note, that $f(U)$ is an antisymmetric function with $f(U) > 0$ for $0 < U < 1$ and $f(U) < 0$ for $U > 1$. The noise ζ is *not* negligible in this case, and satisfies

$$\overline{\zeta} = 0 \tag{6.50}$$

$$\overline{\zeta^2} = \sigma^2/\rho\tau^2. \tag{6.51}$$

At this point we consider Eqs. (6.46) and (6.48) with the coefficients μ, α, σ, v_0 and D as the continuum theory describing a large class of SPP models. The the nonlinear coupling term $(\partial_x U)(\partial_x\rho)/\rho$ is *specific* for such SPP systems: it is responsible for the slowing down (and eventually the "turning back") of the particles under the influence of a larger number of particles moving oppositely. When two groups of particles move in the opposite direction, the density locally increases and the velocity decreases at the point they meet. Let us consider a particular case, when

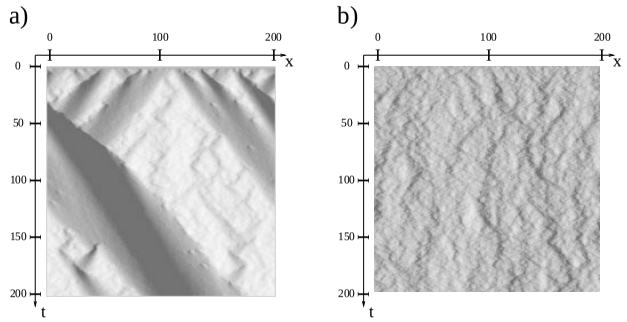


Figure 6.21: Results of the numerical integration of Eqs. (6.46) and (6.48). In this space-time (horizontal-vertical) plot the regions of higher densities are indicated with darker shades. For small enough noise (a) the system is organised into one high-density group travelling in a spontaneously selected direction. In contrast, for higher noise the system remains homogeneous at large length scales (b). (After [27].)

particles move from left to right and the velocity is locally decreasing while the density is increasing as x increases (particles are moving towards a "wall" formed between two oppositely moving groups). The term $(\partial_x u)$ is less, the term $(\partial_x \rho)$ is larger than zero in this case. Together they have a negative sign resulting in the slowing down of the local velocity. This is a consequence of the fact that there are more slower particles (in a given neighbourhood) in the forward direction than faster particles coming from behind, so the average action experienced by a particle in the point x slows it down.

For $\alpha = 0$ the dynamics of the velocity field U is independent of ρ, and Eq.(6.48) is equivalent to the time dependent Ginsburg-Landau Φ^4 model of spin chains, where domains of opposite magnetisation develop at finite temperatures [4]. As numerical simulations demonstrate (Fig. 6.21), for large enough α and low noise the initial domain structure breaks and the system is organised into a single major group travelling in a spontaneously selected direction.

6.1.5 Hydrodynamic formulation for 2D

To get an insight into the possible analytical treatment of SPP systems in higher dimensions, now we apply the Navier-Stokes equations for a "fluid" of SPPs, as described in [28]. The two basic equations governing the dynamics of the *noiseless*

"self-propelled fluid" are the continuity equation

$$\partial_t \rho + \nabla(\rho \vec{v}) = 0 \tag{6.52}$$

and the equation of motion

$$\partial_t \vec{v} + (\vec{v}\cdot\nabla)\vec{v} = \vec{F}(\vec{v}) - \frac{1}{\tau_1}\vec{v} + \frac{1}{\tau_2}\left(\langle\vec{v}\rangle - v\right) - \frac{1}{\rho^*}\nabla p + \nu\nabla^2\vec{v}, \tag{6.53}$$

where ρ^* is the effective density of the particles, p is the pressure, ν is the kinematic viscosity, $\vec{F}(\vec{v})$ is the *intrinsic* driving force of biological origin and τ_1, τ_2 are time scales associated with velocity relaxation resulting from interaction with the environment and the surrounding SPPs, respectively. Let us now go through the terms of Eq. 6.53.

The self-propulsion can be taken into account as a constant magnitude force acting in the direction of the velocity of the particles as

$$\vec{F} = \frac{v_0}{\tau_1}\frac{\vec{v}}{|\vec{v}|}, \tag{6.54}$$

where v_0 is the speed determined by the balance of the propulsion and friction forces, i.e., v_0 would be the speed of a homogeneous fluid.

Taking Taylor expansion of the expression $\frac{1}{\tau_2}\left(\langle\vec{v}\rangle - v\right)$ describing local velocity alignment yields

$$\begin{aligned}\langle\vec{v}\rangle_\epsilon - \vec{v} &= \frac{\int_{|\vec{\xi}|<\epsilon} \mathrm{d}\vec{\xi}\Big(\vec{v}\rho + (\vec{\xi}\nabla)\vec{v}\rho + \frac{1}{2}(\vec{\xi}\nabla)^2\vec{v}\rho + \ldots\Big)}{\int_{|\vec{\xi}|<\epsilon} \mathrm{d}\vec{\xi}\Big(\rho + (\vec{\xi}\nabla)\rho + \frac{1}{2}(\vec{\xi}\nabla)^2\rho + \ldots\Big)} - \vec{v} \\ &= \frac{\epsilon^2}{6}\Big(\frac{\nabla^2(\vec{v}\rho)}{\rho} - \vec{v}\frac{\nabla^2\rho}{\rho}\Big) + \ldots \\ &= \frac{\epsilon^2}{6}\Big(\nabla^2\vec{v} + 2(\nabla\vec{v})\frac{\nabla\rho}{\rho}\Big) + \ldots \end{aligned} \tag{6.55}$$

similarly to the 1d case (??). In the following we will consider cases where the density fluctuations are forced to be small. Hence the velocity-density coupling term in (6.55) is negligible, which is a major difference compared to the previously studied 1d system. Thus, if the relative density changes are small, the velocity alignment can be incorporated with an effective viscosity coefficient ν^* into the viscous term of (6.53),

In the simplest cases the pressure can be composed of an effective "hydrostatic" pressure, and an externally applied pressure as

$$p = g^*\rho + p_{\mathrm{ext}} \tag{6.56}$$

where g^* is a parameter related to the compressibility of the fluid.

Combining (6.52), (6.53), (6.54) and (6.56) one gets the following final form for the equations of the SPP flow:

$$\partial_t \rho + \nabla(\rho \vec{v}) = 0, \tag{6.57}$$

and

$$\partial_t \vec{v} + (\vec{v}\cdot\nabla)\vec{v} = \frac{v_0}{\tau_1}\frac{\vec{v}}{|\vec{v}|} - \frac{1}{\tau_1}\vec{v} + \nu^*\nabla^2\vec{v} - g^*\nabla\rho + -\frac{1}{\rho^*}\nabla p_{\rm ext}. \tag{6.58}$$

It will be useful in the following to introduce the characteristic length

$$\lambda = \sqrt{\nu^*\tau_1} \tag{6.59}$$

and characteristic time

$$T = \frac{\lambda}{v_0} \tag{6.60}$$

and write (6.57) and (6.58) in a dimensionless form of

$$\partial_{t'}\rho + \nabla'(\rho\vec{v}') = 0, \tag{6.61}$$

$$\frac{\tau_1}{T}\left(\partial_{t'}\vec{v}' + (\vec{v}'\cdot\nabla')\vec{v}'\right) = \frac{\vec{v}'}{|\vec{v}'|} - \vec{v}' + \nabla'^2\vec{v}' - g_1'\nabla'\rho + g_2'\nabla' p_{\rm ext}. \tag{6.62}$$

with $t' = t/T$, $x' = x/\lambda$, $v' = v/v_0$ and the ∇' operator derivates with respect to $\vec{r}' = \vec{r}/\lambda$. We also introduced the notations $g_1' = T\tau_1 g^*/\lambda^2$ and $g_2' = g_1'/g^*$. In the following we will drop the prime for simplicity.

For certain simple geometries it is possible to obtain analytical stationary solutions of the above equations supposing incompressibility ($\rho = const.$). Taking $\partial_t \equiv 0$ in Eqs. 6.61 and 6.62, the following dimensionless equations are obtained:

$$\nabla\vec{v} = 0 \tag{6.63}$$

and

$$\frac{\tau_1}{T}(\vec{v}\nabla)\vec{v} = \frac{\vec{v}}{|\vec{v}|} - \vec{v} + \nabla^2\vec{v} - g_2\nabla p_{\rm ext}. \tag{6.64}$$

A simple, but relevant boundary condition is realized when the system is confined to a circular area of radius R. Since this is a finite geometry, in the stationary state no net flux is possible, i.e., the flow field must include vortices. In [28] it has been demonstrated that a single vortex is indeed a possible stationary configuration of the system: The velocity field can be expressed in polar coordinates (r, ϕ) as

$$\vec{v} = v(r)\ \vec{e}_\phi, \tag{6.65}$$

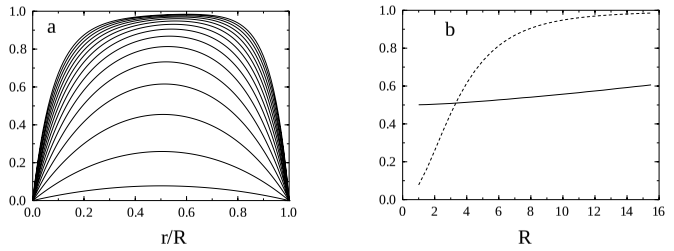


Figure 6.22: (a) Velocity profiles in a vortex for various values of the dimensionless system size, R (bottom curve: $R = 1$, top curve: $R = 16$). (b) Position (solid line) and value (dashed line) of the maximal velocity for the profiles shown in (a) as a function of R. (After [28].)

where $\vec{e}_\phi$ is the tangential unit vector. Since $\nabla = \partial_r + \frac{1}{r}\partial_\phi$, one can easily see that (6.63) is satisfied for any velocity profile $v(r)$. Substituting (6.65) into (6.64) yields two ordinary differential equations:

$$0 = r^2\frac{\mathrm{d}^2 v}{\mathrm{d}r^2} + r\frac{\mathrm{d}v}{\mathrm{d}r} - v(1+r^2) + r^2, \tag{6.66}$$

and

$$-\frac{1}{r^2}v^2 = -\frac{1}{\rho^*\lambda}\frac{\partial p_{\rm ext}}{\partial r}. \tag{6.67}$$

The first equation gives the velocity profile while the second determines the pressure to be applied for maintaining the constant ρ condition.

The boundary conditions for the velocity profile are either $v(0) = 0$ and $v(R) = 0$ (closed boundary at $r = R$) or $v(0) = 0$ and $\frac{\mathrm{d}v}{\mathrm{d}r}|_R = 0$ (free boundary at $r = R$). The solution of (6.66) with the above boundary conditions is given by $I_1(r)$, the modified Bessel function of order one and $L_1(r)$, the modified Struve function of order one [29]. Thus the velocity profile of a stationary single vortex is

$$v(r) = \alpha I_1(r) - \frac{\pi}{2}L_1(r). \tag{6.68}$$

The parameter α should be chosen to satisfy the boundary condition at $r = R$.

In Fig. 6.22a we show velocity profiles for different dimensionless system size R. The maximal velocity decreases for decreasing R (Fig. 6.22b). From Fig. 6.22 it is also

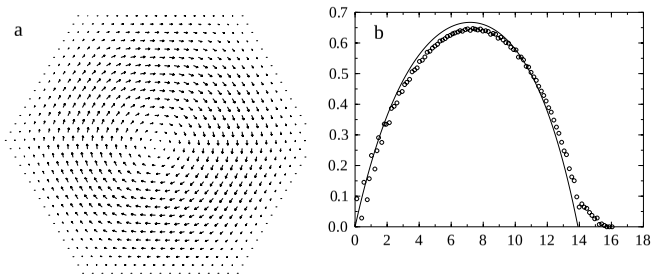


Figure 6.23: (a) A numerically generated vortex for dimensionless system size $R = 4.4$ and $g_1 = 750$. The length of the arrows is proportional to the local velocity, while their thickness is proportional to the density. (b) The measured (circles) and the theoretical (solid line) velocity profile for the vortex shown in (a). (After [28].)

clear that the minimal size of a vortex is in the order of λ, i.e., one in dimensionless units. Therefore, if $R \gg 1$ then initially many Kosterlitz-Thouless-like vortices are likely to be present in the system.

For further investigations numerical solutions of Eq.6.57 and Eq. 6.58 was also calculated in [28] with finite compressibility and $p_{\text{ext}} = 0$. The only remaining dimensionless quantity characterising the system is τ_1/T which relates the various relaxation times and v_0. The system can be considered overdamped if $\tau_1 \ll T$ holds. Figure 6.23 shows the stationary state for the overdamped equations in the high compressibility ($g_1 = 750$) limit. The length and direction of the arrows show the velocity, while the thickness is proportional to the local density of the fluid. In Fig. 6.23b the radial velocity distribution is presented for the vortex shown in Fig. 6.23a together with the velocity profile given by (6.68). Rather good agreement is seen; the differences are due to the fact that the numerical system is not perfectly circular. In simulations with $R \gg 1$ multi-vortex states develops (Fig. 6.24), which transforms into a single vortex after a long enough time.

Thus, we demonstrated that – in contrast to the 1d case – the ordering in 2d is not due to the density-velocity coupling term in the expansion of $\langle v \rangle$: the viscosity and the internal driving force terms are sufficient to destabilise the originally present vortices and organise the system into a globally ordered phase. The stability of that ordered phase against a finite amount of fluctuations has been shown by Tu and Toner as we discuss in the next section.

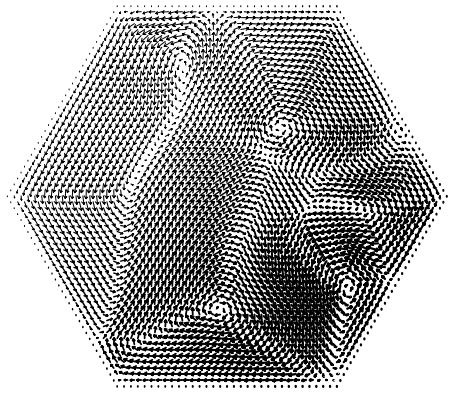


Figure 6.24: Transient multiple vortex state at $R = 25.3, g_1 = 128$. (After [28].)

6.1.6 The existence of long-range order

The first theory describing the full nonlinear higher dimensional dynamics was presented in [11, 30]. Tu and Toner followed the historical precedent [31] of the Navier-Stokes equation by deriving the continuum, long wavelength description *not* by explicitly coarse graining the microscopic dynamics as we did it for the one dimensional system, but, rather, by writing down the most general continuum equations of motion for $\vec{v}$ and ρ consistent with the symmetries and conservation laws of the problem. This approach allows to introduce a few phenomenological parameters, (like the viscosity in the Navier-Stokes equation) whose numerical values will depend on the detailed microscopic behaviour of the particles. The terms in the equations describing the large-scale behaviour, however, should depend only on symmetries and conservation laws, and *not* on the microscopic rules.

The only symmetry of the system is rotation invariance: since the particles lack a compass, all direction of space are equivalent to other directions. Thus, the "hydrodynamic" equation of motion cannot have built into it any special direction picked "a priori"; all directions must be spontaneously selected. Note that the model does *not* have Galilean invariance: changing the velocities of all the particles by some constant boost $\vec{v}_b$ does *not* leave the model invariant.

To reduce the complexity of the equations of motion still further, a spatial-temporal gradient expansion can be performed keeping only the lowest order terms in gradients and time derivatives of $\vec{v}$ and ρ. This is motivated and justified by the aim to consider *only* the long distance, long time properties of the system. The resulting set of equations is

$$\begin{aligned} \partial_t \mathbf{v} + \lambda_1(\mathbf{v}\nabla)\mathbf{v} + \lambda_2(\nabla\mathbf{v})\mathbf{v} + \lambda_3\nabla(|\mathbf{v}|^2) &= \alpha\mathbf{v} - \beta|\mathbf{v}|^2\mathbf{v} - \nabla P + \\ & D_L\nabla(\nabla\mathbf{v}) + D_1\nabla^2\mathbf{v} + D_2(\mathbf{v}\nabla)^2\mathbf{v} + \vec{\xi} \\ \partial_t\rho + \nabla(\rho\mathbf{v}) &= 0, \end{aligned} \tag{6.69}$$

where the $\alpha, \beta > 0$ terms make $\mathbf{v}$ have a nonzero magnitude, $D_{L,1,2}$ are diffusion constants and $\vec{\xi}$ is an uncorrelated Gaussian random noise. The λ terms on the left hand side of Eq. (6.69) are the analogs of the usual convective derivative of the coarse-grained velocity field $\vec{v}$ in the Navier-Stokes equation. Here the absence of Galilean invariance allows all *three* combinations of one spatial gradient and two velocities that transform like vectors; if Galilean invariance *did* hold, it would force $\lambda_2 = \lambda_3 = 0$ and $\lambda_1 = 1$. However, Galilean invariance does *not* hold, and so all three coefficients can be non-zero phenomenological parameters whose values are determined by the microscopic rules. The pressure P depends on the local density only, as given by the expansion

$$P(\rho) = \sum_n \sigma_n(\rho - \rho_0)^n. \tag{6.70}$$

It is possible to treat the whole problem analytically using dynamical renormalisation group and show the existence of an ordered phase in 2D, and extract exponents characterising the velocity-velocity and density-density correlation functions [30]. The most dramatic result is that an intrinsically non-equilibrium and nonlinear feature, namely, convection, suppresses fluctuations of the velocity $\vec{v}$ at long wavelengths, making them much smaller than the analogous fluctuations found in ferromagnets, for all spatial dimensions $d < 4$. Specifically, the measure of the fluctuations, $C_C(\vec{R})$, defined as:

$$C_C(\vec{R}) = C\left(\vec{R}\right) - \lim_{|\vec{R}'|\to\infty} C\left(\vec{R}'\right) \quad , \tag{6.71}$$

with

$$C(\vec{R}) \equiv \left\langle \vec{v}(\vec{R}+\vec{r},t)\cdot\vec{v}(\vec{r},t)\right\rangle \tag{6.72}$$

and

$$C(\vec{R}\to\infty) = \phi^2, \tag{6.73}$$

decays to zero much more rapidly, as $|\vec{R}| \to \infty$, than the analogous correlation function in magnets. Quantitatively, for points whose separation $\vec{R} \equiv \vec{R}_\perp$ lies perpendicular to the mean direction of motion

$$C_C(\vec{R}) \propto R_\perp^{2\chi} \tag{6.74}$$

holds, where the universal exponent is

$$\chi = -\frac{1}{5} \tag{6.75}$$

exactly, in $d = 2$, and is smaller than its equilibrium value, $1 - \frac{d}{2}$, in magnetic systems for all $d < 4$.

The physical mechanism for this suppression of fluctuations is easy to understand: increased fluctuations in the direction of motion of different parts of the system actually *enhance* the exchange of information between those different parts. This exchange, in turn, suppresses those very fluctuations, since the interactions between particles tend to make them all move in the same direction.

These non-equilibrium effects also lead to a spatial anisotropy of scaling between the direction along ($\|$) and those orthogonal to ($\perp$) the mean velocity $\langle\vec{v}\rangle$. The physical origin of the anisotropy is also simple: if particles make small errors $\delta\theta$ in their direction of motion, their random motion *perpendicular* to the mean direction of motion $\langle\vec{v}\rangle$ is much larger than that *along* $\langle\vec{v}\rangle$; the former being $\propto \delta\theta$, while the later is proportional to $1 - \cos\delta\theta \sim \delta\theta^2$. As a result, *any* equal-time correlation function in the system of *any* combination of fields crosses over from dependence purely on $|\vec{R}_\perp|$ to dependence purely on $R_\|$ when

$$\frac{R_\|}{\ell_0} \approx \left(\frac{|\vec{R}_\perp|}{\ell_0}\right)^\zeta \quad , \tag{6.76}$$

where ℓ_0 is the interaction range. The universal anisotropy exponent is

$$\zeta = \frac{3}{5} \quad , \tag{6.77}$$

exactly in $d = 2$, and is < 1 for all $d < 4$.

Bibliography

[1] C. W. Reynolds. Flocks, herds, and schools: a distributed behavioral model. *Comp. Graph.*, 21:25–34, 1987.

[2] J. K. Perrish and L. Edelstein-Keshet. Complexity, pattern and evolutionary trade-offs in animal aggregation. *Science*, 284:99–101, 1999.

[3] E. M. Rauch, M. M. Millonas, and D. R. Chialvo. Pattern formation and functionality in swarm models. *Phys. Lett. A*, 207:185, 1995.

[4] H. E. Stanley. *Introduction to phase transitions and critical phenomena*. Oxford University Press, Oxford, 1971.

[5] S.-K. Ma. *Statistical mechanics*. World Scientific, Singapore, 1985.

[6] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen, and O. Shochet. Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75:1226–1229, 1995.

[7] N. Shimoyama, K. Sugawara, T. Mizuguchi, Y. Hayakawa, and M. Sano. Collective motion in a system of motile elements. *Phys. Rev. Lett.*, 76:3870–3873, 1996.

[8] D. Helbing, J. Keltsch, and P. Molnar. Modelling the evolution of human trail systems. *Nature*, 388:47–50, 1997.

[9] N. D. Mermin and H Wagner. Absence of ferromagnetism or antiferromagnetism in one- or twodimensional isotropic Heisenberg models. *Phys. Rev. Lett.*, 17:1133, 1966.

[10] A. Czirók, H. E. Stanley, and T. Vicsek. Spontaneously ordered motion of self-propelled particles. *J. Phys. A*, 30:1375–1385, 1997.

[11] J. Toner and Y. Tu. Long-range order in a two-dimensional dynamical xy model: How birds fly together. *Phys. Rev. Lett.*, 75:4326–4329, 1995.

[12] R. B. Stinchcombe. *Phase Transitions and Critical Phenomena*, volume 7. Academic Press, New York, 1983.

[13] J. M. Kosterlitz and D. J. Thouless. Ordering, metastability and phase transitions in two-dimensional systems. *J. Phys. C*, 6:1181, 1973.

[14] A. Czirók, M. Vicsek, and T. Vicsek. Collective motion of organisms in three dimensions. *Physica A*, 264:299–304, 1999.

[15] A. Czirók, A.-L. Barabási, and T. Vicsek. Collective motion of self-propelled particles: kinetic phase transition in one dimension. *Phys. Rev. Lett.*, 82:209–212, 1999.

[16] A. Czirók, E. Ben-Jacob, I. Cohen, and T. Vicsek. Formation of complex bacterial patterns via self-generated vortices. *Phys. Rev. E*, 54:1791–1801, 1996.

[17] Y. L. Duparcmeur, H. J. Herrmann, and J. P. Troadec. Spontaneous formation of vortex in a system of self motorized particles. *J. Phys. (France) I*, 5:1119–1128, 1995.

[18] J. Hemmingsson. Modellization of self-propelling particles with a coupled map lattice model. *J. Phys. A*, 28:4245–4250, 1995.

[19] Z. Csahók and T. Vicsek. Lattice-gas model for collective biological motion. *Phys. Rev. E*, 52:5297, 1995.

[20] R. B. Potts. *Proc. Camb. Phil. Soc.*, 48:106, 1952.

[21] K. Binder, K. Vollmayr, H.-P. Deutsch, J. D. Reger, and M. Sheucher. *Int. J. Mod. Phys. C*, 5:1025, 1992.

[22] E. V. Albano. Self-organized collective displacements of self-driven individuals. *Phys. Rev. Lett.*, 77:2129 – 2132, 1996.

[23] F. Y. Wu. The Potts model. *Rev. Mod. Phys.*, 54:235, 1992.

[24] K. Binder. Static and dynamic critical phenomena of the two-dimensional q-State potts model. *J. Stat. Phys.*, 24:69, 1981.

[25] H. J. Bussemaker, A. Deutsch, and E. Geigant. Mean-field analysis of a dynamical phase transition in a cellular automaton model for collective motion. *Phys. Rev. Lett.*, 78:5018–5021, 1997.

[26] H. J. Bussemaker. Analysis of a pattern-forming lattice-gas automaton: Mean-field theory and beyond. *Phys. Rev. E*, 53:1644, 1996.

[27] A. Czirók. *Models of collective behaviour in biology*. (PhD thesis, Eötvös University, Budapest, 2000).

[28] Z. Csahók and A. Czirók. Hydrodynamics of bacterial motion. *Physica A*, 243:304, 1997.

[29] M. Abramowitz and I. A. Stegun, editors. *Handbook of Mathematical Functions*. Dover Publication, New York, 1970.

[30] J. Toner and Y. Tu. Flocks, herds, and schools: A quantitative theory of flocking. *Phys. Rev. E*, 58:4828, 1998.

[31] D. Forster, D. R. Nelson and M. J. Stephen. Large-distance and long-time properties of a randomly stirred fluid. *Phys. Rev. A*, 16:732, 1977.

[32] K. L. Schick, A. A. Verveen. $1/f$ noise with a low frequency white noise limit. Nature 251:599, 1974;

J. A. C. Gallas, H. J. Herrmann, S. Sokołowski. Convection cells in vibrating granular media. *Phys. Rev. Lett.* 69: 1371-1374, 1992.

H. A. Makse, S. Havlin, P. R. King, H. E. Stanley. Spontaneous stratification in granular mixtures. *Nature* 386:379-382, 1997.

[33] H. J. Herrmann. In: P.L. Garrido, J. Marro (eds.) *3rd Granada Lectures in Computational Physics* (Springer, Heidelberg, 1995);

[34] E. Ben-Jacob, O. Shochet, A. Tenenbaum, I. Cohen, A. Czirók and T. Vicsek. Generic modelling of cooperative growth patterns in bacterial colonies. *Nature* 368:46-49, 1994.

D.A. Kessler, H. Levine. Pattern formation in Dictyostelium via the dynamics of cooperative biological entities. *Phys. Rev. E* 48:4801-4804, 1993.

[35] D. Helbing, F. Schweitzer, J. Keltsch, P. Molnár. Active walker model for the formation of human and animal trail systems. *Phys. Rev. E* 56:2527-2539, 1997.

[36] W. Weidlich and G. Haag. *Concepts and Models of a Quantitative Sociology* (Springer, Berlin, 1983);

W. Weidlich. Physics and social science - the approach of synergetics. Phys. Rep. 204:1, 1991;

D. Helbing. *Quantitative Sociodynamics. Stochastic Methods and Models of Social Interaction Processes*. (Kluwer Academic, Dordrecht, 1995).

[37] *Highway Capacity Manual,* Chap. 13 (Transportation Research Board, Special Report 209, Washington, D.C., 1985).

[38] U. Weidmann. *Transporttechnik der Fußgänger*. (Schriftenreihe des Instituts für Verkehrsplanung, Transporttechnik, Straßen- und Eisenbahnbau Nr. 90, ETH Zürich, 1993).

[39] D. Oeding. *Verkehrsbelastung und Dimensionierung von Gehwegen und anderen Anlagen des Fußgängerverkehrs* (Straßenbau und Straßenverkehrstechnik, Heft 22, Bonn, 1963);

S. J. Older. Movement of Pedestrians on Footways in Shopping Streets. Traffic Eng. Control 10:160, 1968;

P. D. Navin, and R. J. Wheeler. Pedestrian-flow characteristics. Traffic Eng. 39:31, 1969.

[40] A. J. Mayne. Some further results in the theory of pedestrians and road traffic. Biometrica 41:375, 1954.

N. Ashford, M. O'Leary, P. D. McGinity. Stochastic modelling of passanger and baggage flows through an airport terminal. Traffic Eng. Control 17: 207, 1976.

S. J. Yuhaski Jr. and J. M. Smith. Modelling circulation systems in buildings using state dependent queueing models. Queueing Systems 4:319, 1989;

G. G. Løvås. In A. Pave (ed.) *Modelling and Simulation 1993* (Society for Computer Simulation International, Ghent, Belgium, 1993).

[41] A. Borgers, H. J. P. Timmermans. City centre entry points, store location patterns and pedestrian route choice behaviour: A microlevel simulation model. Socio-Economic Planning Science 20:25, 1986;

H. Timmermans, X. van der Hagen, and A. Borgers. Transportation systems, retail enviroments and pedestrian trip chaining behaviour: Modelling issues and applications *Transportation Res. B* 26:45-59, 1992.

[42] L. F. Henderson. On the fluid mechanics of human crowd motion. Transportation Res. 8:509, 1974.

[43] D. Helbing. A Fluid-Dynamic Model for the Movement of Pedestrians. Complex Systems 6:391, 1992;

D. Helbing. *Stochastische Methoden, nichtlineare Dynamik und quantitative Modelle sozialer Prozesse*. (Shaker, Aachen, Germany, 1993).

[44] D. Helbing. *Verkehrsdynamik. Neue physikalische Modellierungskonzepte* (Springer, Berlin, 1997).

[45] D. Helbing. A mathematical model for the behavior of pedestrians. Behavioral Science 36:298, 1991.

[46] D. Helbing, P. Molnár. Social force model for pedestrian dynamics. Phys. Rev. E 51:4282, 1995.

[47] P. Molnár. *Modellierung und Simulation der Dynamik von Fußgängerströmen* (Shaker, Aachen, Germany, 1996).

[48] D. Helbing. In D. E. Wolf et al. (eds.) *Traffic and Granular Flow* (World Scientific, Singapore, 1996).

[49] D. Helbing, P. Molnár. In F. Schweitzer (ed.) *Self-Organization of Complex Structures: From Individual to Collective Dynamics* (Gordon and Breach, London, 1997).

[50] P. G. Gipps and B. Marksjö. A micro-simulation model for pedestrian flows. Mathematics and Computers in Simulation 27:95, 1985.

[51] Video films by T. Arns, Wannenstr. 22, 70199 Stuttgart, Germany.

[52] L. F. Henderson. The Statistics of Crowd Fluids. Nature 229:381, 1971;

L. F. Henderson, D. J. Lyons. Sexual Differences in Human Crowd Motion. Nature 240:353, 1972;

L. F. Henderson, D. M. Jenkins. Response of pedestrians to traffic challenge. Transportation Res. 8:71, 1973.

[53] J. S. Coleman, and J. James. The Equilibrium Size Distribution of Freely-Forming Groups. *Sociometry* 24:36, 1961.

J. S. Coleman. *Introduction to Mathematical Sociology.* (The Free Press of Glencoe, New York, 1964).

[54] S. B. Santra, S. Schwarzer, and H. Herrmann. Fluid-induced particle-size segregation in sheared granular assemblies. Phys. Rev. E 54:5066, 1996;

[55] K. Yoshikawa, N. Oyama, M. Shoji, and S. Nakata. Use of a saline oscillator as a simple nonlinear dynamical system: Rhythms, bifurcation, and entrainment. Am. J. Phys. 59:137, 1991;

X.-L. Wu, K. J. Måløy, A. Hansen, M. Ammi, and D. Bideau. Why hour glasses tick. Phys. Rev. Lett. 71:1363, 1993;

T. L. Pennec, K. J. Måløy, A. Hansen, M. Ammi, D. Bideau, and X.-L. Wu. Ticking hour glasses: Experimental analysis of intermittent flow. Phys. Rev. E 53:2257, 1996.

[56] K. Lewin. *Field Theory in Social Science* (Harper and Brothers, New York, 1951).

[57] D. Helbing. Boltzmann-like and Boltzmann-Fokker-Planck equations as a foundation of behavioral models. Physica A 196:546, 1993; and J. Math. Sociology 19:189, 1994.

[58] D. Helbing, and T. Vicsek. Optimal Self-Organization. *New J. Physics* 1:13.1-13.17, 1999.

[59] D. Helbing, I. Farkas, and T. Vicsek. Freezing by Heating in a Driven Mesoscopic System. *Phys. Rev. Lett.* 84:1240-1243, 2000.

[60] D. Helbing. In *Natural Structures,* Part II (Sonderforschungsbereich 230, Stuttgart, Germany, 1992) 93.

[61] I. Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution* (Frommann-Holzboog, Stuttgart, 1973).

[62] K. Bolay. *Nichtlineare Phänomene in einem fluid-dynamischen Verkehrsmodell* (Master's thesis, University of Stuttgart, Germany, 1999).

[63] J. Klockgether, H. P. Schwefel. In *Proc. 11. Symp. on Eng. Aspects of MHD* (Cal. Inst. Techn., 1970) 141.

[64] K. Humpert et al. In K. Teichmann, and J. Wilke (eds.) *Prozeß und Form "Natürlicher Konstruktionen". Der Sonderforschungsbereich 230.* (Ernst & Sohn, Berlin, 1996) 172.

[65] E. Schaur. *Ungeplante Siedlungen / Non-Planned Settlements* (Krämer, Stuttgart, Germany, 1991).

[66] J. Keltsch. *Selbstorganisation von Wegen durch "Active Walkers"* (Master's thesis, University of Stuttgart, Germany, 1996).

[67] D. Helbing. In M. Schreckenberg, and D. Wolf (eds.) *Traffic and Granular Flow '97* (Springer, Singapore, 1998).

[68] B. Davis. Reinforced Random Walk. *Probab. Th. Rel. Fields* 84:203, 1990;

R. D. Freimuth and L. Lam, In L. Lam and V. Naroditsky (eds.) *Modeling Complex Phenomena* (Springer, New York, 1992);

D. R. Kayser, L. K. Aberle, R. D. Pochy and L. Lam. Active walker models: tracks and landscapes. *Physica A* 191:17, 1992;

R. D. Pochy, D. R. Kayser, L. K. Aberle and L. Lam. "Boltzmann active walkers and rough surfaces. *Physica D* 66:166, 1993;

L. Lam, and R. Pochy. Active-walker models: Growth and form in nonequilibrium systems. *Computers in Physics* 7:534, 1993;

L. Lam. Active walker models for complex systems. *Chaos, Solitons & Fractals* 6:267, 1995;

F. Schweitzer. In L. Schimansky-Geier, and T. Pöschel (eds.) *Lectures on Stochastic Dynamics* (Springer, Berlin, 1997).

[69] F. Schweitzer and L. Schimansky-Geier. Clustering of "active" walkers in a two-component system. *Physica A* 206:359, 1994.

[70] L. Schimansky-Geier, M. Mieth, H. Rosé, and H. Malchow. Surface formation by active Brownian particles. *Phys. Lett. A* 207:140, 1995;

L. Schimansky-Geier et al. In F. Schweitzer, (ed.) *Self-Organization of Complex Structures: From Individual to Collective Dynamics* (Gordon and Breach, London, 1997) 101.

[71] A. Stevens. *Mathematical Modeling and Simulations of the Aggregation of Myxobacteria* (PhD thesis, Ruprecht-Karls-University Heidelberg, 1992);

A. Stevens. and F. Schweitzer. In W. Alt et al. (eds.) *Dynamics of Cell and Tissue Motion* (Birkhäuser, Basel, 1997) 183.

[72] F. Schweitzer and J. Steinbrink. In F. Schweitzer (ed.) *Self-Organization of Complex Structures: From Individual to Collective Dynamics* (Gordon and Breach, London, 1997) 501.

[73] F. Otto. *Die natürliche Konstruktion gewachsener Siedlungen* (Sonderforschungsbereich 230, Stuttgart, Germany, 1991) Heft 37.

[74] B. Hölldobler and E. O. Wilson. *The Ants* (Belknap, Cambridge, MA, 1990).