Computer Modeling in Physics

Oleg M. Braun

Institute of Physics
National Academy of Sciences of Ukraine

Contents

Preface

This book is based on the one-term lecture course (about twenty lectures) which has been given by one of the authors in 1988–1992 at Kiev University for forth–year students (that corresponds to last–year undergraduate or first–year graduate students) of the Radiophysical Department. The aim of the course was to explain *in what a way* a computer can be used for solution of physical problems. It is natural, therefore, that special attention in the course is devoted to the main methods of computer simulation, namely the Molecular Dynamics (MD) method and the Monte Carlo (MC) method, and also to the Stochastic Equations (SE) method which may be considered as an intermediate one between the MD and MC methods. Note that these methods were discussed also in other books, for example, in the book of Heerman [1] or in the book of Gould and Tobochnik [2] (see also [191]).

However, besides the answer on the question *in what a way,* the principal aim of this book is an attempt to answer two additional important questions: *when* and *why* we should use computer modeling. Evidently that modeling in a general sense forms a base for development of any field of science from mathematics to philosophy. Indeed, the global aim of science is to find and explain evolution laws of the Universe. But because any phenomenon in the world is extremely complicated, the only way of solution of the problem is the following: we must maximally simplify the given phenomenon (object, process, *etc.*), i.e., to throw out all inessential details and to leave the most significant ones only. That is just what is called *to invent a model.* The invitation of new models is an art. The appearance of a new model is always the outstanding event in science, and the model is called usually by the name of the scientist who invented it. Clearly, it is impossible to teach how to invent models. But it is possible, and namely this is done in the present book, to tell about the most known models used in physics. Then, combining the known models in an appropriate way, one will be able to solve successfully almost any of standard physical problems.

The modeling in a general sense is too wide topic even for a series of many books. In the present book we set up a more modest task: to tell on those models only, for investigation of whose a computer is necessary or very desirable at least. Namely, such a situation corresponds to models from two extremely interesting divisions of physics, the stochastic theory and the solitonic theory, and also from the old (but very interesting as well) division, the statistical physics. We cannot say categorically that most problems from these divisions of physics could not be solved without computers. But it is evident that these solutions would be found much more latter (perhaps, for decades).

List of questions considered in this book can be found in Contents. We would like to mention only the questions which are not included in this book while they could be included in the book with such a title. Excepting few words in Introduction, it is absent a discussion of computers and formalize languages. We do not discuss numerical methods and do not give texts of computer programs. All these questions may be found in other textbooks, the references are given in the corresponding sections. The presentation is accompanied by a large number of examples taken, as a rule, from original papers published last time. An essential number of examples is taken from surface physics, where one of the authors of this book worked for a while. But all the examples are of a general character and may be successfully used in other divisions of physics as well.

We would like to thank our colleagues Igor Aranson, Yuri Kivshar, Alexander Kovalev, Igor Lyuksyutov, and Michel Peyrard for discussions of different topics of the present book. We would like to acknowledge also Irina Zelenskaya for her assistance in preparing the figures.

Chapter 1

Introduction

1.1 Stages of modeling

The global aim of modeling is to explain or predict some phenomenon or process, i.e., to develop a corresponding theory. In order to achieve this aim we have, first of all, to invent a model or to select a suitable one, then to work out a method for its solution, then to carry out an investigation of the solution and, finally, to reveal laws or regular trends of the phenomenon under investigation. Let us consider these stages in more detail.

A. Setting of the problem. The first stage of modeling is to invent a corresponding model. The very limited number of scientists is a success to contrive a new model. As a rule, in such case the model takes the name of the scientist who contrived it. Recall, for example, the widely known Ising model, Baxter model, Anderson model, Frenkel-Kontorova model which became the classical ones. Clearly that to teach somebody to invent new models is impracticable task. However, we will try to acquaint a reader of this book with the main models used most often, so she/he could use them after some modifications in own investigations.

Let us set out basic requirements that must be imposed upon a physical model.

(a) *The model must be maximally simple.* Beginner investigators always try to set the problem in the most general form taking into account as many as possible factors from the very beginning. This is completely mistaken way, and it *never* leads to a success. The model must be maximally simplified, all factors of secondary importance should be omitted, and only the relevant factors, i.e. the ones which are representative of the main qualitative points of the phenomenon under investigation, have to be kept. And only at the last stage of modeling, additional factors which help to describe the phenomenon in more detail, may be included one by one.

(b) *The model has to have a solution,* or, in other words, the problem should be set *correctly.* Moreover, from the very beginning we should intuitively know or, at least, should be clear in mind the final as well as all intermediate results, whereas the purpose of modeling is to prove that the intuition was right. Of course, sometimes results of modeling may occur to be far from those being expected. But in this case we will stay to win and to make a discovery.

(c) *The model should be elegant.* Being trivial, this requirement is the necessary one for the modeling to be a success.

(d) *Variables of the model should be dimensionless,* i.e., the used system of units has to be natural for the given problem. As a rule, we can introduce three dimensionless variables. These variables may be chosen more or less arbitrary, but usually the units of time, distance, and energy (or charge, or mass) are made dimensionless. Often beginners try to use seconds for time, centimeters for distance, ergs or joules for energy, *etc.* In a result, the calculations are "overloaded" by redundant numbers. To avoid this inconvenience, the units should be related to values which are inherent for the given system. For example, the period of intrinsic oscillations of the system has to be taken as the unit of time, the Bohr radius or the mean interparticle distance, as the unit of distance, the electron charge, as the unit of charge, *etc.*

B. Developing of a computer program for the solution of the settled problem. In this book we will consider only the models where the computer plays a central role in the investigation. Moreover, the computer is used not as a great calculator for an approximate solution of complicated equations, but as an "experimental device", i.e., the computer with the program appears as the instrument and, at the same time, as the system under investigation. For computer modeling in physics two methods are mainly used, the Molecular Dynamics (MD) method and the Monte Carlo (MC) method. It is interesting that both methods have been sampled for the first time in 1950-th when the first, very unsophisticated computers appeared. The MD and MC methods are considered in Chapters 3 and 5 correspondingly. Below we also give some practical advices on the choice of numerical methods and the computer as well as on the organization of the computer program and its de-bugging.

C. Investigation of the solution, or the computer simulation itself. This stage is the analog of the carrying out of a laboratory physical experiment. Namely, we vary the input program parameters acting on the state of the physical system, and study the resulting system behavior, i.e., its response to these variations. When a sufficient number of dependences have been obtained, we can proceed to the final (and the main) stage of modeling.

D. Reveal of the theory. This stage has no difference from the corresponding stage in experimental or theoretical physics. Note only that namely at this stage we may and must complicate and generalize the model in order to clear up a role of the factors which were assumed to be of secondary importance at the first stage.

1.2 Numerical methods

A computer can carry out only a finite number of the before-handed deterministic operations with a finite (discrete) set of rational numbers. Infinities, continuum variables, real or random numbers, *etc.* do not exist in numerical methods. These points must be clearly understood at the first stage of modeling when the model is chosen. For example, in the theoretical investigation of problems of the solid state physics, the continuum approximation is often used, where a discrete variable $l = 0, \pm 1, \ldots$ which numerates the atoms, is replaced by the continuum variable x: $l \to x_l = la \to x$, where a is the lattice constant. The resulting model is described by differential equations. However, at numerical investigation of these equations we must discretize them, i.e., we have to replace the continuum variable x by the discrete variable j: $x \to x_j = j\Delta x \to j$, where Δx is the integration step. Of course, for this problem the investigation of the discrete model from the very beginning will be more natural as well as more correct.

In modeling of physical problems we usually need the following numerical methods:

(a) Practically always a matrix algebra, e.g., addition and multiplication of matrices is used. More rarely we need to calculate determinants or to inverse a matrix. Often physical problems lead to eigenvalue matrix equations. All these methods are developed quite well, and some of them are even included into syntax of the standard FORTRAN or other formalize languages; besides, there are several well-developed libraries.

(b) The solution of an algebraic equation $f(x) = 0$ is carried out by well-developed methods such as the method of cutting the interval as well as various iteration methods, for example, the method of sections, the Newton-Raphson method, *etc.*

(c) The numerical integration is well developed too (recall the method of rectangulars, the method of trapezoidals, the Simpson method, the Gauss method, *etc.*).

(d) The numerical differentiation is a more complicated operation which should be avoided if possible. Sometimes it would be necessary to smooth the function before its differentiation.

(e) For the solution of differential equations, usually the Verlet algorithm or different variants of the Runge-Kutta (RK) method are used. The most popular ones are the forth-order RK method and the fifth-order predictor-corrector RK method with the automatic choice of the integration step. Note that for some problems (for example, for solution of the Langevin equation, see Sec. 4.2) the step of integration must be kept fixed.

(f) Partial differential equations appear in theoretical physics quite often, but their solution requires special methods, typically rather complicated ones [192, 120]. An example is presented in Sec. 4.5.2, although this topic is too wide to be discussed in the present course.

(g) A need to solve an integral equation occurs not too often. Usually the integral equation is reduced to the matrix one, the latter being solved by the method of consecutive iterations.

(h) In physical modeling we often meet with the necessity to find an extremum (usually minimum) of a many-variable function. As a classical example we recall that the thermodynamically equilibrium state must correspond to the minimum of the system free energy. The standard method of solution of this problem, the so-called steepest-descent method, consists in a search (usually by the MC method) of the direction along which the function decreases most steeply.

(i) Often one arises the necessity to use a generator of pseudo-random numbers. This problem is briefly discussed in Chapter 5.

(j) Finally, let us say a few words about computer arithmetics. The numbers in computer may be integer ("fixed point") and real ("floating point"). Arithmetics between integers is exact, provided the answer is not outside of the available range (which typically is $\sim 2^{15} \approx 3.28 \times 10^4$ or $\sim 2^{31} \approx 2.15 \times 10^9$ if the integer is declared as "long"); also recall that division produces an integer too (e.g., 5/2 gives 2). Arithmetics among real umbers is not exact, it is determined by the "machine accuracy" ϵ (defined as the smallest number which, when added to 1.0, produces a result different from 1.0; for the 32-bit computer $\epsilon \sim 3 \times 10^{-8}$). A nonzero machine accuracy leads to the so-called "roundoff errors" which accumulate during calculation — if an operation is performed n times, the total error will be $\sim \epsilon\sqrt{n}$ is the best case and $\sim \epsilon n$ in the most unfavorable case.

A detailed discussion of numerical methods may be found in the handbook of Korn and Korn [3] as well as in a number of textbooks and monographs [4, 5, 6, 7]. In choosing the numerical method it is to be taken into account that usually the more simple is the method, the more reliable is it; at the same time a more simple method requires the more processor time in order to achieve the same accuracy. Besides, the priority should be given to those methods which are realized by the reliable standard subroutines (e.g., see [8]); the best set of numerical methods is collected, of course, in the famous "Numerical Recipes" [9].

1.3 Choice of a computer

Naturally that the choice of the computer is an important stage of modeling because the computer should be suitable for the given problem. In this choice we have to take into account such computer parameters as the speed of calculation, the word length, the active storage, the availability of a convenient and reliable software (standard libraries, translators, editors) and external devices (first of all for computer graphics).

Modern computers, even personal laptops, already have the active storage (i.e., the memory which could be directly accessed from the central processor) of a few gigabites (1 Gb = 10^9 bites), which could be enough for physical simulation. Much more worse is the situation with the speed of calculation, which now is of the order of 2 gigahertz (1 GHz = 10^9 operations per second) and already approaches to the maximally achievable limit. Parallel (vector, array) computers, or supercomputers, may be one-two orders of magnitude faster, but the acceleration may be achieved only for programs that can be split into parallel independent parts, and they need special methods of programming.

The calculation speed is closely connected with another important parameter of the computer, the word length. The 16-bit cell used ten years ago, is sufficient for bookkeeping calculations, but typically is insufficient for mathematical modeling, because in the latter case we have to use the double precision (i.e., two cells for storage of one number), and in a result the calculation speed may be decreased in up to ten times. For modeling problems one should prefer a computer with the 32-bit cell or better with the 64-bit cell. Note that the two parameters, the calculation speed and the available memory, are coupled, and that the chosen method of calculation should depend on these characteristics of the computer. For example, each program has a set of values (sometimes, arrays) that are used many times. If the computer storage is large

but computation speed is small, these values can be calculated one time and then they have to be stored; otherwise, for small storage and high computation speed, it will be better to calculate these values each time when they are needed.

Thus, the computer must correspond to the setting problem. Sometimes successful results may be obtained with the help of a small computer, up to a simple pocket calculator (as, for example, it occurs in developing of the Feigenbaun theory, see below Sec. 2.4). As a rule, however, we meet with the opposite situation, when the computer should be as powerful as possible. Let us present some typical examples. For sufficiently simple models such as one-dimensional systems studied by the MD method or the percolation two-dimensional lattices studied by the MC method, it is enough to have a modern laptop. For MD modeling of two-dimensional systems, it would be better to use a cluster. For MD investigation of three-dimensional systems as well as for MC studying of thermodynamical properties, a supercomputer is to be preferable.

Unfortunately, it is not so simple to find the necessary computer for the given problem, because, in particular, there is a number of physical problems which are in need of so powerful computer that simply does not exist at all. Therefore we have to choose the problem under consideration appropriately for the accessible computer, or to refuse the modeling from the very beginning. Indeed, let us assume that we start the investigation of the thermodynamics of a system by the MC-Metropolis method with a personal computer. At the first stage we, may be, will get some realistic results. But as soon as we will try to carry out the calculations with a higher accuracy, or with a more complicated program, we will immediately meet with the fact that a single calculation step will take, e.g., one hour, the calculation of a single point will need one day, and a single dependence, one month, *etc.* Clearly that in this case the chosen computer is inappropriate for the setting problem, and it will be much better to throw out such a problem as early as possible.

1.4 Some practical advices

In writing and de-bugging of computer programs, we may recommend to hold on the following rules.

(1) The program has to be written as simple as possible. Complex constructions, especially with control passings, and nonstandard peculiarities of a given formalize language which can be used for the given translator only, must be avoided.

(2) The text of the program has to be followed by detailed comments, and also all used variables are to be declared explicitly (even when the implicit declaration of variable types is available).

(3) Because the program will usually run many times during simulation, the sufficient attention should be spared to input and output of information, preferably by writing this blocks as separated subroutines. Accessing to the program should be simple, the input parameters should include only "external", i.e. physical variables (such as, e.g., temperature, concentration, *etc.*) and, may be, a "key" which determines the regime of program operation (for example, the body of information which is displayed or printed). The output information must be followed by comments (the title and date of the simulation, the name of the printed variable, units, *etc.*). Besides, the results have to be easy-to-interpreted, so the graphical presentation is preferable.

(4) The program should be split into more or less independent units (blocks) and written in the form of subroutines. Often it is very convenient to show the interplay between the units by a block-scheme.

(5) De-bugging of the program has to be started from de-bugging of each subroutine (unit, block); for this purpose special control examples should be written for each subroutine. When a standard subroutine or the subroutine taken from a library is used, it must be checked with the help of a control example written specially for this case.

(6) When we are sure that all subroutines proceed without errors, we can link them together into the whole program. Usual errors at this stage are that the types of variables, or array dimensions, or sizes of common blocks (global variables) do not match one another. Recall that in the main FORTRAN program the common blocks and arrays must be declared with the maximal lengthes and dimensions.

(7) Next, the "training" of the program should be done. The program operation should be checked for an essentially wider range of intrinsic and external parameters comparing with those that will be used. In particular, simulation results must remain to be practically unchanged when the intrinsic parameters (such as, e.g., the integration step) are increased as well as decreased in two-three times.

(8) The last very important advise is the following: files have a "property" to be damaged, rubbed out, lost, *etc.* Therefore, after each stage of de-bugging, the text of the program together with the example of its proceeding has to be saved on external disks, better in two-three copies.

Experience shows that de-bugging usually takes more than 50% of the whole simulation time.

Chapter 2

Stochastic Theory

In classical mechanics the two-body problem always has the exact analytical solution in quadratures (e.g., see [10]). One naturally arises a question about existence of an analogous solution for the three-body problem. The final answer on this question was found in a recent time only, and, we emphasize, it was found namely due to using of the computer modeling. It turns out to be that, excepting exotic cases (which, however, play the extremely important role in physics, see below Chapter 6), the exact solution does not exist, and the system motion is stochastic. The history of developing of the stochastic theory is a classical example how, extremely simplifying the models and then studying them with the help of a computer, it becomes possible to solve the problem. In the present Chapter we describe these stochastic models, concentrating attention on the methods of their computer investigation. For a more detailed acquaintance with the stochastic theory we may recommend to refer to the monographs of Arnold and Avez [11], Sinai [12], Chirikov [13], Lichtenberg and Liberman [14], Rabinovich and Trubetskov [15], Zaslavsky [16], Schuster [17], and Arnold [18].

2.1 Hénon-Heiles model

Three-dimensional motion of a system consisting of three point particles interacting via central forces is the problem with nine degrees of freedom. However, because we always can associate a plane with three points, the problem reduces to one with six degrees of freedom. Also we can exclude two degrees of freedom, if we will use the coordinate system coupled with the center of mass of the system. However, a more radical simplification can be obtained, if we artificially "pin" two particles, and allow to move to the third particle only (see Fig. 2.1). Such a procedure may be physically justified, if we assume that one of three particles is essentially lighter, than the two other particles, for example, that it corresponds to a planet moving in the field of a double star. In such a way we come to the so-called *restricted three-body problem*, i.e., to the problem of the two-dimensional motion of a particle subjected to an external potential $V(x, y)$.

To proceed further, we have to take the concrete form for the function $V(x, y)$. Let us expand the potential $V(x, y)$ in Taylor series in vicinity of one of its minima, and take the coordinates of this minimum as the beginning of coordinates. The expansion must include the cubic terms as the minimal approximation, because the square expansion leads to the problem of two harmonic oscillators which has the trivial solution. Thus, the expansion has the form

$$\begin{aligned} V(x,y) = V_0 + V'_x x + V'_y y + \frac{1}{2}V''_{xx}x^2 + \frac{1}{2}V''_{yy}y^2 + V''_{xy}xy \\ + \frac{1}{6}V'''_{xxx}x^3 + \frac{1}{2}V'''_{xxy}x^2 y + \frac{1}{2}V'''_{xyy}xy^2 + \frac{1}{6}V'''_{yyy}y^3 + \cdots \end{aligned} \tag{2.1}$$

Let us take the origin of the energy scale so that $V_0 = 0$. Because the expansion of $V(x, y)$ has been done at the minimum, we have $V'_x = V'_y = 0$. Choosing appropriate directions for the x and y axes, we always can make $V''_{xy} = 0$ for the crossing term in Eq. (2.1) (such coordinates are known as the *normal coordinates*). Last, let us suppose that the "pinned" particles are identical each other, so that the potential $V(x, y)$ is even

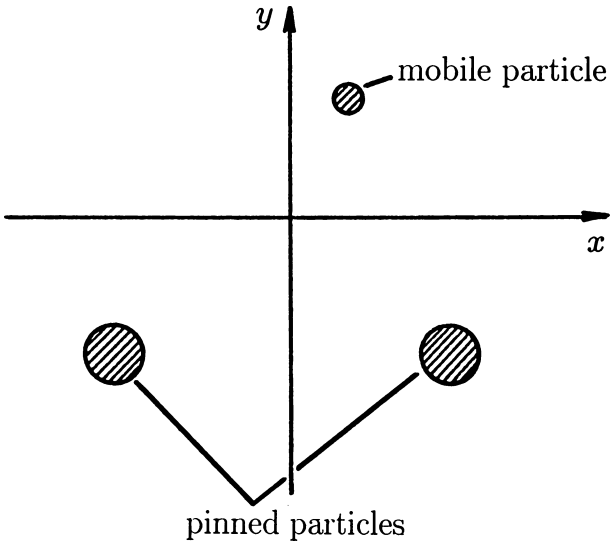

Figure 2.1: The restricted three-body problem.

in x (see Fig. 2.1), that gives $V'''_{xxx} = 0$ and $V'''_{xyy} = 0$. In the result we obtain the most simple Hamiltonian which saves all main features of the original three-body problem.

The next step is to introduce the dimensionless units. It is natural to take the mass of the mobile particle as the unit of mass. It is convenient also to choose the time unit in a way that the frequency of normal vibrations along the x axis is equal 1, and the unit of length in such a way that the coefficient at the x^2y term (which is the main term in the given problem) being equal 1 too. Finally we get the Hamiltonian (the dot denotes the derivative with respect to time)

$$H(x, \dot{x}, y, \dot{y}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + x^2 + \omega_y^2 y^2) + x^2 y + \mu y^3, \tag{2.2}$$

which depends on two parameters $\omega_y^2 = V''_{yy}(0,0)$ and $\mu = \frac{1}{6}V'''_{yyy}(0,0)$ only. For numerical investigation of the model we have to take numerical values for these parameters, and then to test how the results will be changed with variation of these parameters. Hénon and Heiles [19] who first proposed and investigated the model (2.2), have put $\omega_y = 1$ and $\mu = -\frac{1}{3}$.

Using Hamiltonian (2.2), we can write the motion equations:

$$\left\{ \begin{array}{rcl} \ddot{x} & = & -\partial V/\partial x\,, \\ \ddot{y} & = & -\partial V/\partial y\,. \end{array} \right. \tag{2.3}$$

Instead of the system of two differential second-order equations (2.3), usually it is more convenient to solve the system of four differential first-order equations

$$\left\{ \begin{array}{rcl} \dot{x} & = & v_x\,, \\ \dot{v}_x & = & -\partial V/\partial x\,, \\ \dot{y} & = & v_y\,, \\ \dot{v}_y & = & -\partial V/\partial y\,. \end{array} \right. \tag{2.4}$$

Then, choosing some initial conditions, we have to solve these equations for a time $t_{\max}$. However, it is too problematic to find directly from the shape of the trajectory of the system motion, wether the given problem has the solution in an integral form, or has not. Therefore, to finish the setting of the model, we have also to invent a method how we can find the answer on this question from the trajectory of the system motion.

In classical mechanics there exists the theorem which states: In order for a Hamiltonian system with N degrees of freedom to have an analytical solution, it must have N integrals of motion (the proof of

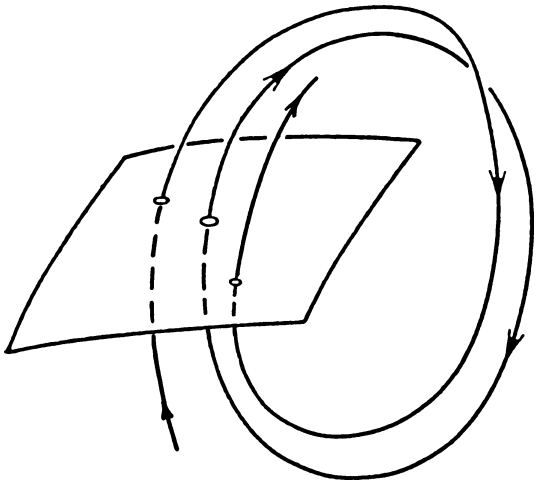

Figure 2.2: Construction of the Poincaré map.

this theorem may be found, e.g., in [11]). For example, if $V(x,y) = V^{(x)}(x) + V^{(y)}(y)$, the problem has two integrals of motion (the energies of motion along the x and y axes correspondingly), and it has the analytical solution, because the problem is split into two independent problems connected with the separated one-dimensional motions along the x and y axes. For Hamiltonian system we always have one integral of motion, that is the total energy of the system,

$$H(x,\dot{x},y,\dot{y}) = E = \mathrm{Const}_E. \tag{2.5}$$

Thus, in order for the system (2.4) to have the analytical solution, it must exist a function $J(x,\dot{x},y,\dot{y})$ (and $J \not\equiv H$) such that

$$J(x,\dot{x},y,\dot{y}) = \mathrm{Const}_J \tag{2.6}$$

provided $x(t)$, $\dot{x}(t)$, $y(t)$, and $\dot{y}(t)$ are solutions of the system (2.4).

Evolution of the system with two degrees of freedom may be described as a motion of a point in the four-dimensional phase space. Due to existence of the energy conservation law, this point must move on some three-dimensional hypersurface defined by Eq. (2.5). Let us assume now that it exists the second integral of motion J. Then, Eq. (2.6) will define the second three-dimensional hypersurface, and the point associated with the system coordinates in the four-dimensional phase space, must be confined to both hypersurfaces (2.5) and (2.6) simultaneously, i.e., the point must move on the two-dimensional surface which is the intersection of two hypersurfaces (2.5) and (2.6). Then, let us take a plane (called the *section plane*) which is intersecting with this two-dimensional surface (for example, Hénon and Heiles [19] used the $Y0\dot{Y}$ plane). The intersection will occur along a curve. It is easily to see that the points of intersection of the trajectory of the system motion in the phase space with the section plane must be confined to this curve.

Thus, the problem of computer modeling reduces to the computation of coordinates of the points where the trajectory of system motion intersects with the section plane, and the latter may be chosen more or less arbitrary (see Fig. 2.2). The construction described above is known as the *Poincaré map*. If this map consists of isolated points, or if all the intersection points form a curve, the problem, may be, has the second integral of motion and, therefore, it could have an analytical solution (which, however, still has to be found). But if we can not associate a curve with the intersection points, then, most probably, the analytical solution does not exist. Notice that usually the Poincaré map has to be calculated with $\geq 10^3$ points.

The obtained by Hénon and Heiles [19] results of computer modeling are shown in Fig. 2.3. As seen, as far as the system energy E is lower than some "critical" value E_{crit} ($E_{\mathrm{crit}} = 0.115$ for the given choice of the parameters), one can plot the curves described above (different curves in Fig. 2.3a correspond to different

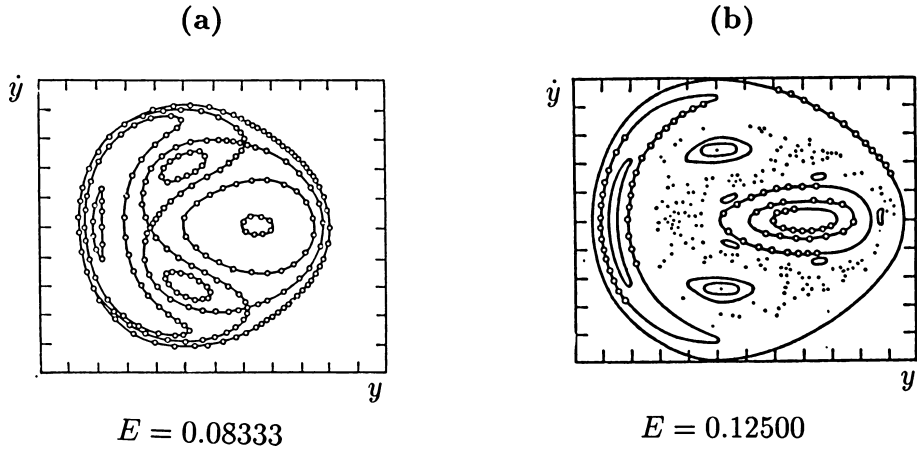

Figure 2.3: Poincaré map for the Hénon-Heiles model (2.2) at (a) $E = 0.08333$ and (b) $E = 0.12500$ (after [19]).

initial conditions). However, when $E > E_{\rm crit}$, for some initial conditions we can plot a curve (this is the so called *islands of stability*), but for other initial conditions the sequence of points of the Poincaré map covers some region (the so-called *stochastic region*), and it is impossible to associate a smooth curve which connects these points (see Fig. 2.3b). Moreover, when we take as the initial condition different points outside the islands of stability, we always obtain the picture which seems to be practically the same. According to the established terminology, the motion within the islands of stability is called the *regular motion*, while the motion outside the islands is named by the *stochastic motion*. Below we will show that the stochastic motion really corresponds to "true chaos".

Thus, the computer modeling of the Hénon-Heiles model showed that at $E < E_{\rm crit}$ the motion seems to be regular and an analytical solution seems to exist, while at $E > E_{\rm crit}$ the second constant of motion is "destroyed", the stochastic region appears, and the area of the stochastic region increases (while the total area of the islands of stability decreases) with increasing of E (notice that in fact the behavior of Hamiltonian systems is more complicated, see below Sec. 2.6). However, it remains unclear *why* and *how* does the stochastic motion emerge, and *what does it look like*. To make clear these questions, the model with two degrees of freedom occurs to be too complicated, and we have to invent a more simple model which at the same time should save the main features of the phenomenon under investigation. On the other hand, the system with one degree of freedom exhibits no chaos, because it always can be exactly integrated. Thus, we have to invent the model with $1\frac{1}{2}$ degrees of freedom, i.e. a system which evolutes according to three first-order differential equations. Such a system, however, will be not a conservative system but the dissipative one.

2.2 Driven pendulum

In the previous Section we have considered in fact the system of two nonlinearly coupled oscillators. Now let us assume that one of these oscillators is much lighter than the another. In this case we may assume that in the first approximation the light oscillator practically does not disturb the motion of the heavy oscillator, and that the heavy oscillator vibrates according to the law $y(t) = r \sin \omega_0 t$ with the amplitude and frequency being constant. Owing to existence of the coupling between the oscillators, the heavy oscillator will act on the light one as a periodic external force, and we come to the problem of motion of an oscillator influenced by an external periodic force. However, we have to take into account two important points. First, in order to state the problem correctly, we have to introduce some damping, because otherwise the amplitude of vibrations may increase to infinity. Second, the oscillator must be anharmonic, because for the harmonic

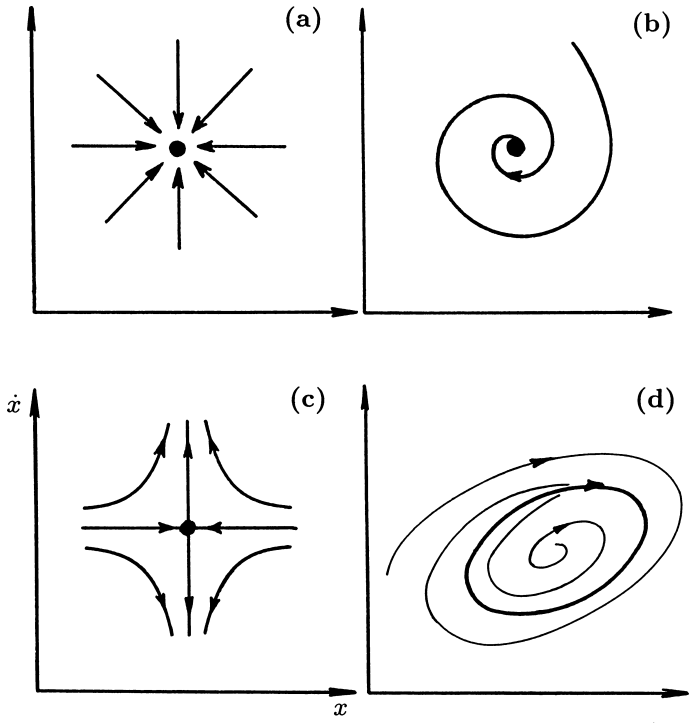

Figure 2.4: Critical points and curves on a plane: (a) the nodal point, (b) the focal point, (c) the saddle point, and (d) the limit cycle.

oscillator the problem is integrable (e.g., see [10]). In the result we come to the *driven pendulum model* which is described by the motion equation

$$\ddot{x} + \eta\dot{x} + \sin x = r\sin\omega_0 t\,, \tag{2.7}$$

where r is the amplitude of the external force and η is the viscous friction coefficient. Introducing the variables $y = \dot{x}$ and $z = \omega_0 t$, Eq. (2.7) can be rewritten as the system of three first-order differential equations

$$\left\{\begin{array}{rcl} \dot{x} &=& y\,, \\ \dot{y} &=& -\sin x - \eta y + r\sin z\,, \\ \dot{z} &=& \omega_0\,, \end{array}\right. \tag{2.8}$$

so that the system evolution is described by the trajectory in three-dimensional phase space.

In numerical investigation of the model (2.8) we have to put ω_0 and η to be fixed (e.g., $\omega_0 \sim \eta \sim 1$), and to study the trajectories of system motion for different values of the parameter r which is an analog of the energy parameter of the Hénon-Heiles model.

Recall that in qualitative investigation of dynamics of a dissipative system it is sufficient to find its peculiarities (i.e., to find critical points, curves, surfaces, *etc*). For example, when the system evolution is described by an autonomous second-order differential equation or by the set of two first-order differential equations, the system phase space is a plane, and on this plane only the following peculiarities may exist:

(a) the stable (or unstable) *nodal point* is such a critical point that all phase trajectories approach it (or go away from it) as shown in Fig. 2.4a;

(b) the stable (unstable) *focal point* distinguishes from the nodal point by that in the neighborhood of the focal point all trajectories swirls also (see Fig. 2.4b);

(c) *the saddle point* is such a singular point that two trajectories approach it, while two others go away from it (Fig. 2.4c);

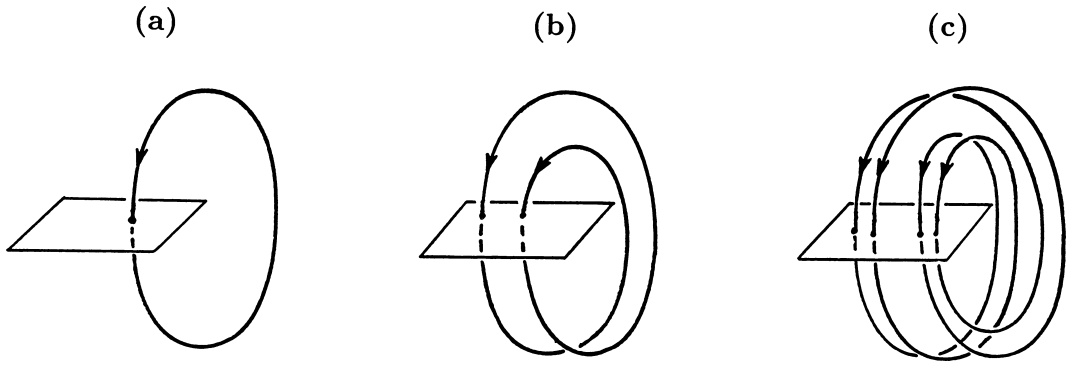

Figure 2.5: Qualitative phase trajectories for the driven pendulum: (a) the single cycle, (b) the cycle of period two, and (c) the cycle of period four.

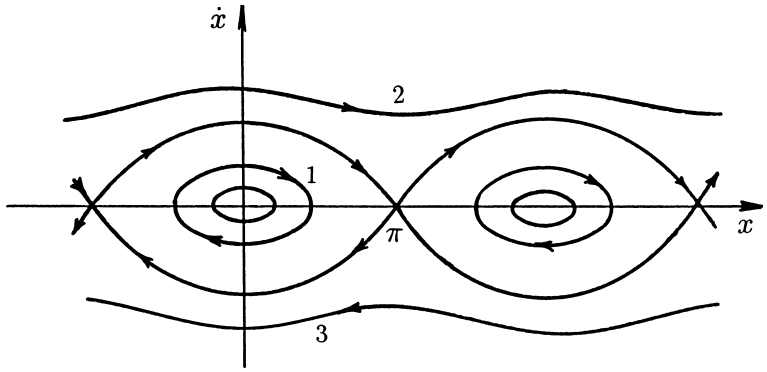

Figure 2.6: Phase portrait of the physical pendulum.

(*d*) the stable (or unstable) *limit cycle* is the closed trajectory in the phase space such that all trajectories approach it (or go away from it) as is shown in Fig. 2.4d.

When the dimension of the phase space is more than two, there may exist additionally such peculiarities as different kinds of saddle points, limit torus, *etc.* If all trajectories from some neighborhood of the phase space approach to a stable manifold, this manifold is called the *attractor*.

To find the attractor in computer simulation, it is enough to calculate the trajectory of the system motion for a time which is sufficient for the system to reach the attractor. It is convenient to present the results of modeling with the help of the stroboscopic map (an analog of the Poincaré map used for dissipative systems), i.e., to fix the positions of the phase trajectory in discrete time moments $t_n = nT$, $n = 0, 1, 2, \ldots$, which are multiplies of the period of the driven force, $T = 2\pi/\omega_0$. The results of investigation of the driven pendulum (2.7) (see [20]) show that at small amplitude of the external force ($r \ll 1$) the stroboscopic map consists of a single point, i.e., the system attractor is the orbitally stable limit cycle with the period T (see Fig. 2.5a). When the parameter r raises, i.e. when the average energy of the oscillator increases above a certain value, the single point of the stroboscopic map splits into two points (see Fig. 2.5b), i.e. the so-called *period-doubling bifurcation* takes place. With further increasing of r these two points split into four points (see Fig. 2.5c), then into eight points, and so on up to infinity. Finally, at $r > r_{\rm crit}$, where $r_{\rm crit}$ is some critical value, the stroboscopic map corresponds to some set on the $X0\dot{X}$ plane which looks like the stochastic region on the Poincaré map of the Hénon-Heiles model. This set was called the *strange attractor* (Ruelle and Takens [21]).

To make clear the reason why does the chaos emerge, let us consider an isolated physical pendulum, i.e. let us put $r = \eta = 0$ in Eq. (2.7). The phase space for the physical pendulum is shown in Fig. 2.6. The system

has two types of stationary states: the vortex points at $\dot{x} = 0$, $x = 2n\pi$, $n = 0, \pm 1, \pm 2, \ldots$, which correspond to the stable states, and the saddle points at $\dot{x} = 0$, $x = (2n + 1)\pi$, which correspond to the unstable state of the pendulum at the top position. The ellipses around the vortex points (see curve 1 in Fig. 2.6) describe the pendulum vibrations, while the periodic curves (see curves 2 and 3 in Fig. 2.6), its clockwise or anticlockwise rotations. The regions of the phase space corresponded to these qualitatively different types of system motion, are separated by the peculiar curve called the *separatrix*. The separatrix is the trajectory which connects saddle points. The motion along the separatrix has two essential features. First, this motion is *unstable*, because even the very small external perturbation will move the pendulum either to vibrational regime or to rotational regime depending on the direction of the external force acting on the pendulum when the latter is at the top (unstable) position. Second, the motion along the separatrix takes place for an infinite long time (physically this means that the pendulum "hangs up" at the top position). Thus, when the trajectory of the system motion is too close to the separatrix, it becomes impossible to predict exactly, either even or odd will be the number of half-periods of the external force during one vibration or rotation cycle for the pendulum motion from one top position to the next one. Therefore, we can not predict the direction of the external force at the top position, and, consequently, can not predict the type of the pendulum motion (either rotation or vibration) for the next cycle. Namely this instability of motion in the neighborhood of the separatrix is the reason why the chaotic motion does exist. Therefore, with increasing of the amplitude of the external force r, the amplitude of the pendulum vibrations increases too, and the phase trajectory comes to the vicinity of the separatrix, where the system motion becomes chaotic. For a more detailed investigation of the driven pendulum model which in fact is much more rich and complicated than it has been described above, we can refer to papers [22, 23, 24, 25].

Thus, the driven pendulum model helped us to understand the reason for the appearing of chaotic motion (note that the instability of the system motion in the stochastic regime was shown by Hénon and Heiles in their original paper [19]). But the question on the mechanism of creation of the chaos still remains unanswered, as well as the question of the nature of the chaotic motion itself, i.e. on a structure of the strange attractor. Therefore, it is very desirable to make a further simplification of the model under investigation. For example, let us suppose that the external force does not act continuously but consists of the discrete δ-pulses, i.e., let us put $\sin \omega_0 t \to T^{-1} \sum_{n=-\infty}^{\infty} \delta(t - nT)$ in Eq. (2.7). For this model known as the *periodically kicked rotator model*, the trajectory of system motion between the sequent points of the stroboscopic map can be calculated analytically. However, it will be better to act more radically: let us simply invent a new, more simple model. Namely, let us invent a simple rule, $x_{n+1} = F(x_n)$, with the help of which we can find the $(n + 1)$-th point of the Poincaré map from the n-th point. The simplest function $F(x)$ which leads to the chaotic motion, is the square function. In this way we come to the so-called *logistic map*.

Note that the reasoning with the help of which we were going from the three-body problem to the restricted three-body problem, then to the Hénon-Heiles model, then to the driven pendulum model, and finally to the logistic map, should not be considered as the deduction of one model from the another one. New models have to be invented but not deduced, while the reasons given above may only help to find a possible way of thinking.

2.3 Logistic map

Let us consider the following simple one-parameter map of unit interval into itself known as the *logistic map*:

$$x_{n+1} = F(x_n)\,, \quad F(x) = rx(1 - x)\,. \tag{2.9}$$

Here the index n plays the role of the discrete time, and the parameter r plays the same role as the amplitude of the external force in the driven pendulum model. To have the problem correctly stated, we also have to define the regions of x and r variation. It is natural to take $x \in (0, 1)$; then the variation of the parameter r should be restricted to the interval $1 < r \leq 4$, because otherwise the values of x may go out the limits of the interval $(0, 1)$.

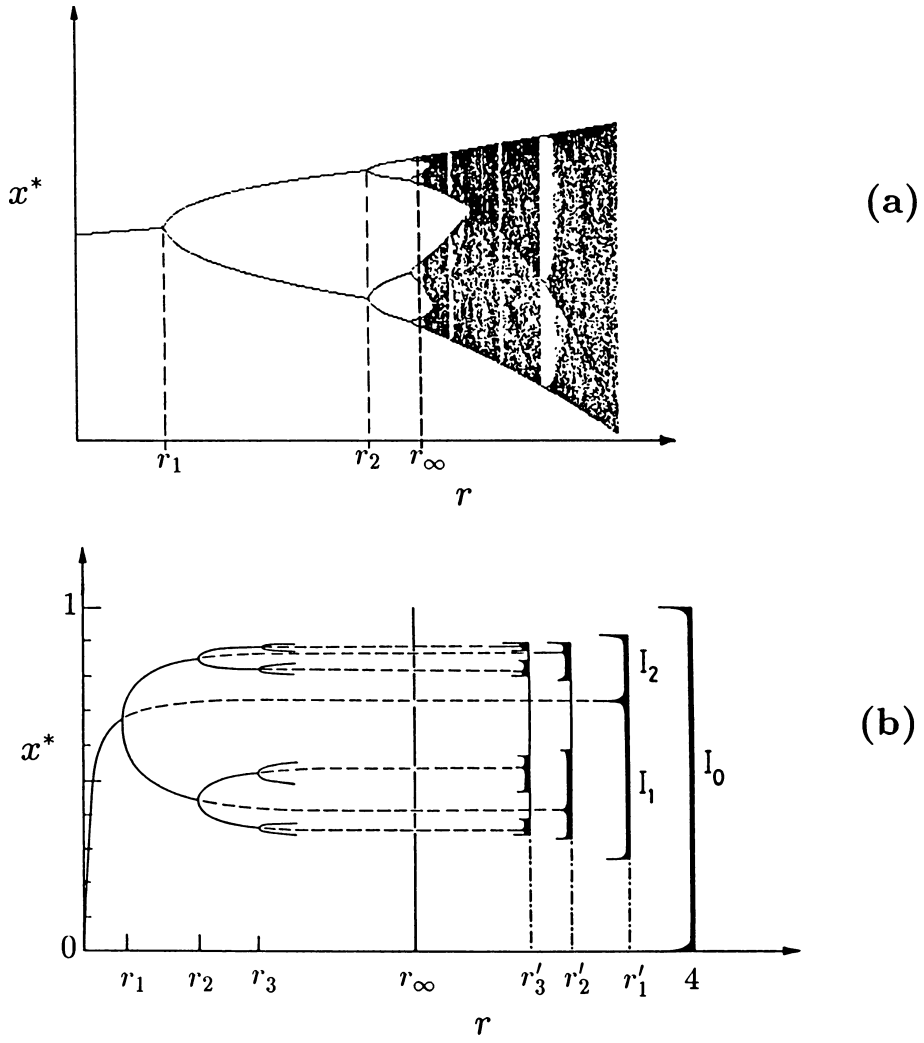

Figure 2.7: Iterates of the logistic map (2.9) (from [14]).

Investigation of the map (2.9) reduces mainly to looking for its attractor x^* for a fixed value of the parameter r. Namely, starting from an arbitrary point x_1, we have to calculate sequently the points x_2, x_3, ..., and then to find the point (or the set of points) x^* to which the points x_n are approaching at $n \to \infty$. Investigations of the logistic map when the parameter r varies within the interval $1 < r \leq 4$, have led to a rather complicated picture [26, 27, 28]. Results of numerical modeling of the map (2.9) are shown in Fig. 2.7. We see that it exists some critical value r_∞ ($r_\infty = 3.5699456\ldots$ for the model (2.9)) such that for $r < r_\infty$ the attractor consists of a finite number of points (one, two, four, *etc*), while for $r > r_\infty$ it contains an infinite number of points, and at $r \to 4$ the attractor occupies the whole interval (0,1), i.e., the whole region where the map (2.9) is defined.

Fortunately, at $r = 4$ the logistic map can be investigated analytically. Namely, using the substitution

$$x_n = \frac{1}{2}(1 - \cos 2\pi\theta_n)\,, \tag{2.10}$$

Eq. (2.9) at $r = 4$ can be transformed to the equation $\theta_{n+1} = 2\theta_n$. Here, however, we must take into account that the values θ and $\theta + i$, where i is an arbitrary integer number, determine the same value of x. Therefore, in order for the transformation (2.10) to be single valued, we must restrict the variation of θ by a unit interval, for example, $0 \leq \theta < 1$. Thus, the logistic map for $r = 4$ reduces to the so called *Bernoulli*

shift:

$$\theta_{n+1} = \{K\theta_n\}, \quad K = 2\,. \tag{2.11}$$

The map (2.11) is nonlinear due to the operation of taking of the fractional part,

$$\{x\} = x - \mathrm{int}(x) = x \,(\mathrm{mod}\ 1)\,,$$

and it is unstable owing to $K > 1$.

To study the map (2.11), let us start from an arbitrary number θ_1, for example, $\theta_1 = 0.703125\ldots$, and write it in binary (because $K = 2$) representation:

$$\theta_1 = \frac{1}{2} + \frac{1}{8} + \frac{1}{16} + \frac{1}{64} + \cdots = 0.101101\ldots \tag{2.12}$$

The consequent points generated by the map (2.11), are the following:

$$\begin{array}{rcl} \theta_2 &=& \left\{1 + \frac{1}{4} + \frac{1}{8} + \frac{1}{32} + \cdots\right\} = \{1.01101\ldots\} = 0.01101\ldots, \\ \theta_3 &=& 0.1101\ldots, \\ \theta_4 &=& 0.101\ldots, \end{array} \tag{2.13}$$

etc. Thus, the operation (2.11) reduces to the shift of the whole number on one bit to the left and to discarding of the integer part when it arises.

Now, if we take as the initial conditions the two close points θ_1 and θ_1' which have, for example, the identical first 100 binary digits but the consequent digits are different, then during the first 100-th steps the trajectories θ_n and θ_n' will be close to one another but, starting from the 101-st step, the points θ_n and θ_n' will separate, and further they will move independently from each other. In other words, even if we set the initial point with the very high precision, there always will occur a moment after which the system evolution will be completely unpredictable. Clearly that in this case the system dynamics is irreversible in time, because in order to have the reversibility we must specify the initial conditions with the infinite accuracy, but this is impossible in principle.

Now let us show why it is natural to name the sequence of point θ_n of the map (2.11) as the chaotic one. Indeed, let us take a chaotic sequence of zeros and ones which can be obtained, for example, with the help of the throwing out a coin, and let us use this sequence as the initial point θ_1. Clearly that the statistical characteristics of the resulting trajectory θ_n will be undistinguishable in principle from the characteristics of the random sequence of the throwing out the coin.

Describing the chaotic dynamics, it is natural to use not a single trajectory $x(t)$ but the probability distribution $\rho(x,t)$. According to the definition, the product $\rho(x,t)\,\Delta x$, where $\Delta x \to 0$, is equal the probability that the system is within the interval $(x, x + \Delta x)$ at time t. Using this definition, we can define chaos in mathematically rigorous way.

(a) The chaos exists in the system, if it exists the *limit probability distribution* (or the *invariant measure*) in the phase space, $\rho_\infty(x)$, such that almost any initial distribution approaches to $\rho_\infty(x)$ at $t \to \infty$.

Sometimes, for very simple systems, the function $\rho_\infty(x)$ can be found analytically. For example, for the logistic map at $r = 4$ we have $\rho_\infty(\theta) = 1$, or $\rho_\infty(x) = 1/\pi\sqrt{x(1-x)}$ (the proof can be found, e.g., in the book [17]). Usually, however, to prove the existence of the limit probability distribution and, all the more, to find it, is the extremely difficult problem, thus we have to use less rigorous statements.

Let $\tilde{f}(x)$ be a function defined in the phase space Γ. Then, if $x(t)$ is a trajectory of the system motion, the function $f(t; x_0) \equiv \tilde{f}[x(t)]$ will depend on time as well as on the initial point $x_0 \equiv x(0)$. Let us define the time averaging as

$$\langle f(t; x_0)\rangle_t = \lim_{T\to\infty} \frac{1}{T} \int_0^T dt\, f(t; x_0)\,. \tag{2.14}$$

Now we can make the following statements:

(b) If the chaos exists in the system, then

(b1) the time average (2.14) exists, and

(b2) it does not depend on the initial condition x_0.

Clearly that in this case the time mean has to coincide with the average over the invariant distribution,

$$\langle f(t;x_0)\rangle_t = \langle \tilde{f}(x)\rangle_\Gamma \equiv \int_\Gamma dx\, \tilde{f}(x)\rho_\infty(x)\,. \tag{2.15}$$

The statement **(b1)** is known as the *ergodic hypothesis*, and the statement **(b2)**, as the *Boltzmann hypothesis*, or the *hypothesis of molecular chaos*.

The simplest way to determine a character of system motion in the computer experiment is to calculate the correlation or spectral functions. The time (self-) correlation function is defined by the relationship

$$Q_f(\tau) = \lim_{T\to\infty} \frac{1}{T}\int_0^T dt\, f(t+\tau)f(t)\,, \tag{2.16}$$

and the spectral function, as its Fourier transformation

$$F_f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\tau\, e^{-i\omega\tau} Q_f(\tau)\,. \tag{2.17}$$

[Notice two important computational details. First, for the calculation of the function $Q_f(\tau)$ within the interval $0 \le \tau \le \tau_{\max}$ we have to find in advance the function $f(t)$ within the more broad interval $0 \le t \le T + \tau_{\max}$. Second, in calculation the integral (2.17) it is usually convenient to employ the fast Fourier transform. To do this, the time step $\Delta\tau$ should be chosen in a way that the total number of points being equal to 2^m with an integer m.] The following statement is valid:

(c) If the chaos exists in the system, its time correlation function has to decay according to exponential law. More rigorously, there have to exist two constants C and Λ which do not depend on the function f, such that

$$\mid Q_f(\tau) \mid \le C\exp(-\Lambda\tau) \quad \text{for} \quad \tau\to\infty. \tag{2.18}$$

The statement **(c)** is equivalent to the following statement:

(d) If the chaos exists in the system, its power spectrum is "dense", or, more rigorously, the spectral function is absolutely continuous.

Emphasize that the statements **(a)** and **(b)** are the necessary and sufficient conditions, while the fulfilment of the conditions **(c)** and **(d)** is necessary but not sufficient for existence of chaotic dynamics.

From the described above investigation of the logistic map it is clear that in order for the chaos to exist, two conditions must be fulfilled: first, the phase space must be continuous (in the discrete phase space the dynamical chaos cannot exist) and, second, the motion equations must have an instability. Owing to the first condition the initial conditions will always have some microscopic uncertainty, because it is impossible to specify the position of the initial point with the infinite accuracy (the infinite accuracy needs the infinite body of information). Due to the second condition the deterministic motion equations "unwrap" this microscopic initial uncertainty into the macroscopic uncertainty of the trajectory of the system motion.

The attractor of the map (2.11) fills the whole region where the map is defined. Within the interval (0,1) the chaotic trajectories form the manifold of the uncountable cardinality, namely, the class of all irrational numbers. Thus, if we take as the initial condition a point from the interval (0,1) in a random way, we will obtain with the probability one an irrational number which will lead to the chaotic trajectory. At the same time, this attractor contains the infinite (but countable) set of rational numbers, and each rational number will lead to the regular periodic trajectory.

Thus, the investigation of the logistic map allows us to achieve the main goal of modeling, namely to simplify the model to such a degree that it becomes possible to explain the phenomenon under investigation "by fingers". Additionally it allows us to understand not only the nature of the dynamical chaos, but, as will be shown in the next Sec. 2.4, also the mechanism of the transition from the regular to chaotic motion.

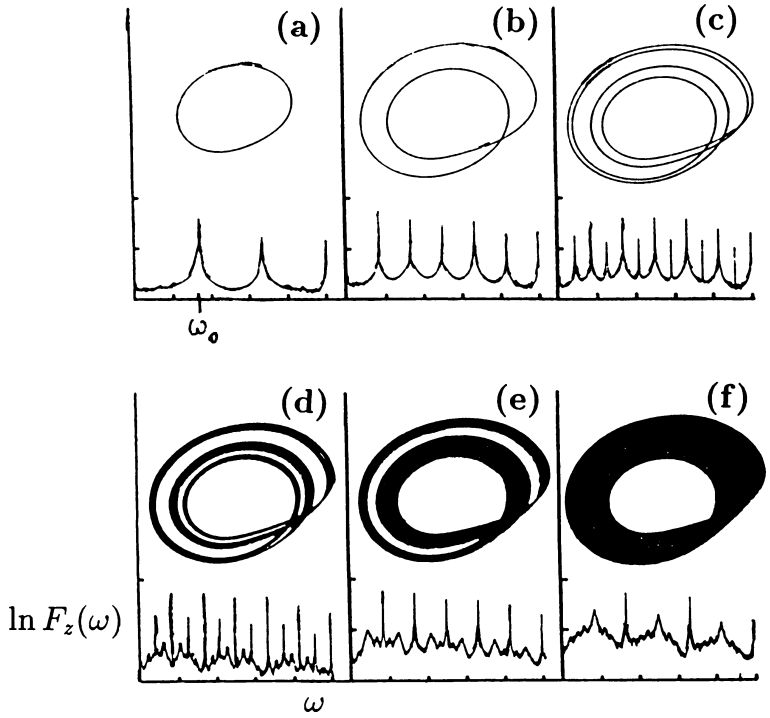

Figure 2.8: Period-doubling bifurcations for the Rössler attractor (2.22) as projected onto the $X0Y$ plane at (a) $r = 2.6$, (b) $r = 3.5$, (c) $r = 4.1$, (d) $r = 4.23$, (e) $r = 4.30$, and (f) $r = 4.60$ as calculated in [29]. The corresponding power spectral density of $z(t)$ is shown in the down parts of the figures.

2.4 Feigenbaum theory

Scenario of the transition from the regular to chaotic motion with the increase of the parameter r is clear from Fig. 2.7. At small r, $r < r_1$ ($r_1 = 3$ for the logistic map), the attractor consists of the single point x^* which is the stable solution of the equation

$$x = F(x)\,. \tag{2.19}$$

[The solution x^* is known as the *fixed point* of the map (2.9). The fixed point x^* is stable if $|F'(x^*)| < 1$.] In a real physical system this situation corresponds to existence of the orbitally stable limit cycle with the period $T = T_0$, and the Fourier spectrum of the system consists of the single harmonic with the frequency $\omega = \omega_0$ (see Fig. 2.8a). Recall that the parameter r of the map (2.9) plays the similar role as the amplitude of the external force in the driven pendulum model or the total energy in the Hénon-Heiles model. With r increasing above the point $r = r_1$, the so-called *period-doubling*, or *pitchfork bifurcation* takes place. Namely, at $r > r_1$ the solution x^* of Eq. (2.19) becomes unstable, i.e. for two sequential points of the map, $x_n = x^* + \Delta x_n$ and $x_{n+1} = x^* + \Delta x_{n+1}$, we now have $|\Delta x_{n+1}| > |\Delta x_n|$. However, instead of the single critical point x^*, there emerge two new points x_1^* and x_2^* which are solutions of the equation

$$x = \tilde{F}_2(x) \equiv F(F(x))\,. \tag{2.20}$$

This solution is stable within the interval $r_1 < r < r_2$, where $r_2 = 3.4495\ldots$ for the map (2.9). In a real physical system the period-doubling bifurcation corresponds to the "splitting" of the limit cycle (see Fig. 2.8b) so that in a result the period of the new limit cycle becomes equal $T = 2T_0$, and the new harmonic with the frequency $\omega = \frac{1}{2}\omega_0$ appears in the frequency spectrum. With further increasing of r at $r = r_2$ the next period-doubling bifurcation takes place (see Fig. 2.8c), then, at $r = r_3$, the next bifurcation, and so on up to the value $r = r_\infty - \delta$ (where $\delta \to 0$) when the attractor becomes to be consisting of the infinite (countable) number of points, and the frequency spectrum, of the same number of harmonics.

With r increasing in the region $r > r_\infty$, the isolated points of the attractor "broadens" to small intervals which then sequentially merge with each other (see Fig. 2.7). In the previous Sec. 2.3 we have shown that at $r = 4$ the attractor of the logistic map takes the whole interval from 0 to 1 (see Fig. 2.8f). When r decreases starting from the $r = 4$ value, the interval occupied by the attractor shrinks, and at $r = r'_1$ the so-called *inverse period-doubling bifurcation* takes place, so that the "blob" of trajectories splits into two more narrow "blobs", or "branches". Within the region $r'_2 < r < r'_1$ the attractor consists of two "branches" as is shown in Fig. 2.8e, the each second point of the stroboscopic map belongs to the same subregion of the strange attractor (i.e., to one of the "branches") and fills it randomly, while the sequential points of the map change the "branches" in the regular way. At the same time, in the frequency spectrum which was dense for $r > r'_1$, at $r < r'_1$ the discrete harmonic with the frequency $\omega = \frac{1}{2}\omega_0$ appears, and the amplitude of the continuous contribution decreases (see Fig. 2.8e). At $r = r'_2$ the next inverse period-doubling bifurcation takes place, the "branches" of trajectories split again so that the attractor becomes consisting of four "branches", and the new harmonics with the frequencies $\omega = \frac{1}{4}\omega_0$ and $\omega = \frac{3}{4}\omega_0$ appear in the frequency spectrum (see Fig. 2.8d). With further r decreasing, an infinite number of inverse period-doubling bifurcations takes place so that at $r = r_\infty + \delta$ ($\delta \to 0$) the attractor becomes consisting of the infinite (countable) number of infinitely narrow intervals, and this attractor then matches with the attractor consisting of the same (infinite) number of isolated points at $r = r_\infty - \delta$.

Calculating with the help of a pocket calculator the values of the parameter r_i, $i = 1, 2, \ldots$, at which the period-doubling bifurcations take place, Feigenbaum [30] discovered that the values r_i approximately produce a geometric progression. (Interesting that, according to the Feigenbaum paper [31], namely the absence of a possibility to use a high-speed computer helps to discover this phenomenon.) Namely, the sequential bifurcation points are subject to the law

$$\lim_{i\to\infty} \frac{r_{i+1} - r_i}{r_i - r_{i-1}} = \frac{1}{\delta}, \quad \text{or} \quad (r_\infty - r_i) \propto \delta^{-i}, \tag{2.21}$$

where $\delta = 4.6692016091\ldots$ Later, investigating (again with the help of a calculator) the another one-dimensional map $x_{n+1} = r \sin \pi x_n$, Feigenbaum discovered that the bifurcation points again produce the geometric progression, and with the same common ratio δ! Moreover, the sequence r'_i of the inverse period-doubling bifurcations follows the law (2.21) too! Thus, the results of computer modeling suggest that the law (2.21) has a general character. Indeed, with the help of the renormalization-group technique Feigenbaum proved ([32], see also [33]) that the law (2.21) is satisfied for any one-dimensional map of an interval into itself, $x_{n+1} = F(x_n)$, provided the function $F(x)$ has a single maximum within the interval where the map is defined (a more detailed description of the Feigenbaum theory may be found, e.g., in [17]).

Above we have tried to show that the maximum simplification of the model is the necessary condition for the modeling to be successful. However, at the next stage of modeling we have to turn back to reality. In other words, further we should complicate the model in order to become sure in invariability of the founded laws as well as to establish whether new effects would appear. With the help of computer modeling it has been shown that the driven pendulum model as well as a majority of other nonlinear dissipative systems show the transition to chaos according to the same scenario described above. In particular, the trajectories and Fourier spectra showed in Fig. 2.8, were calculated in the work [29] for the system of equations

$$\begin{cases} \dot{x} &= -y - z\,, \\ \dot{y} &= x + \frac{1}{5}y\,, \\ \dot{z} &= \frac{1}{5} - rz + xz\,, \end{cases} \tag{2.22}$$

where r is the external system parameter. This model was introduced by Rössler [34, 35] in order to describe dynamics of chemical reactions which take place in a vessel with mixing. The attractor for the Rössler model is shown in Fig. 2.9. From this figure it is easy to see why the existence of dynamical chaos needs $1\frac{1}{2}$ degrees of freedom, i.e. the motion of the system must take place in three-dimensional phase space as minimum. Indeed, owing to existence of the instability, the close phase trajectories diverge, i.e. the trajectory spirals away from the critical point (unstable focal point). However, first, the trajectory cannot get out from the

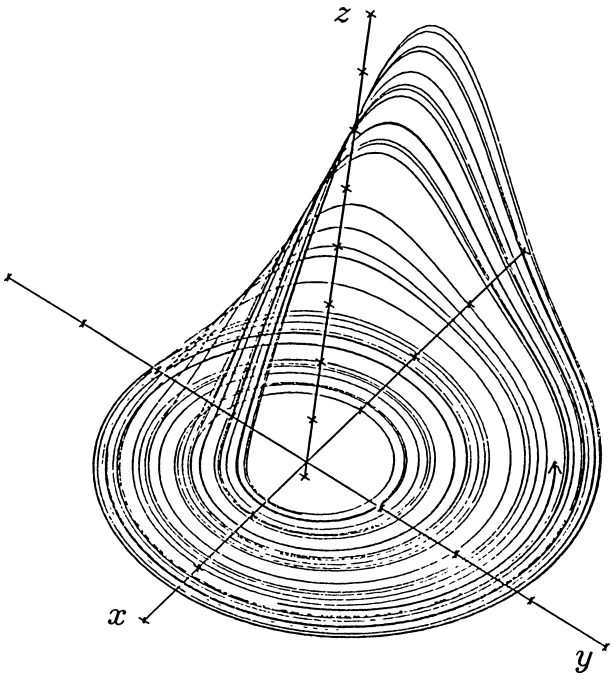

Figure 2.9: The Rössler attractor as calculated in [29].

attractor, i.e. from the bounded region of the phase space, and, second, according to Cauchy's theorem the phase trajectories cannot be intersected. Therefore, if the "untwisting" of the trajectory took place on some two-dimensional surface, then the trajectory must "go out" to a third dimension of the phase space in order to return to the vicinity of the unstable critical point.

Such a broad applicability of the Feigenbaum theory can be explained in the following way. As is known, for a dissipative system the phase volume taken by the system, must shrink as time increases. Therefore, the phase trajectories must crowd together in average. However, for the existence of an instability the nearest trajectories must diverge. These two conditions may be satisfied simultaneously, if the trajectories lying on one hypersurface, diverge, while the trajectories which belong to the orthogonal hypersurface, crowd together (see Fig. 2.10), and in average the latter exceeds the former. If the hypersurface where the phase trajectories diverge, is a two-dimensional surface (and namely such a situation takes place in a number of physical systems), than the transition to chaos in this system should be described by the Feigenbaum theory independently on the total dimensionality of the phase space.

2.5 Strange attractor

Investigation of chaotic dynamics in a dissipative system reduces mainly to looking for its strange attractor. The structure of the strange attractor is rather interesting [36, 37] (see also a review [38]). First, it occurs to be within a "sack", i.e. within a closed region of the phase space, so that each trajectory which came into the sack, cannot escape from it. Second, within the sack, only the unstable critical points exist. For example, the Rössler attractor has the single unstable focal point (see Fig. 2.9). The another well known attractor is the attractor of the Lorenz model shown in Fig. 2.11. This model was introduced by Lorenz [40] for the very simplified (three-mode) description of motion of viscous uncompressed fluid confined into a box with the heated bottom, where the Rayleigh-Bénard convention has to arise with increasing of the temperature difference between the bottom hot surface and the top cold surface. The Lorenz model is described by the

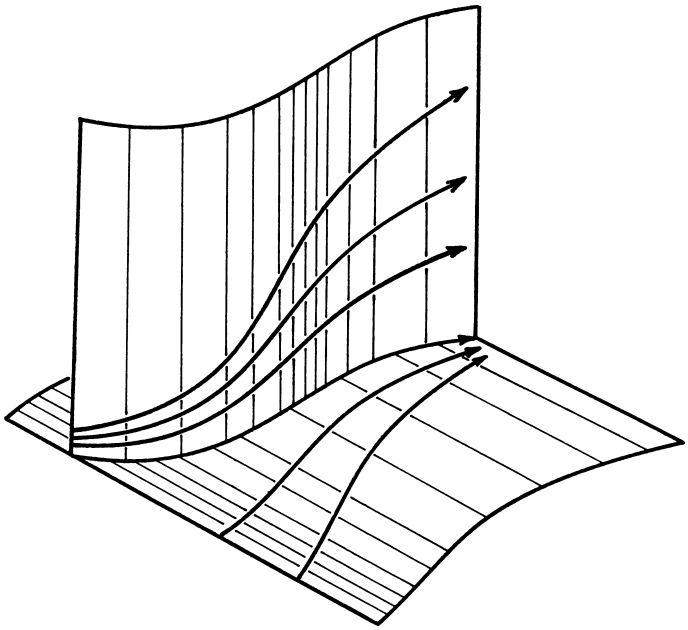

Figure 2.10: Qualitative behavior of phase trajectories on different (hyper-) surfaces.

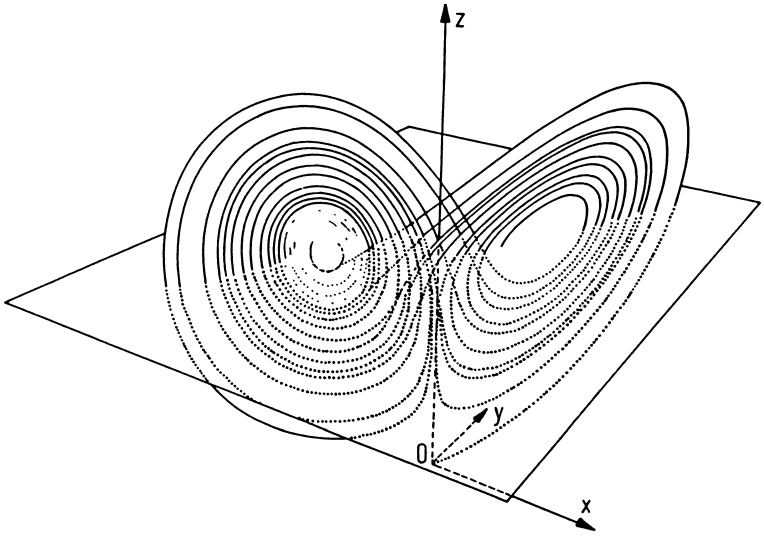

Figure 2.11: The chaotic trajectory for the Lorenz attractor at $r = 28$ as calculated by Lanford [39].

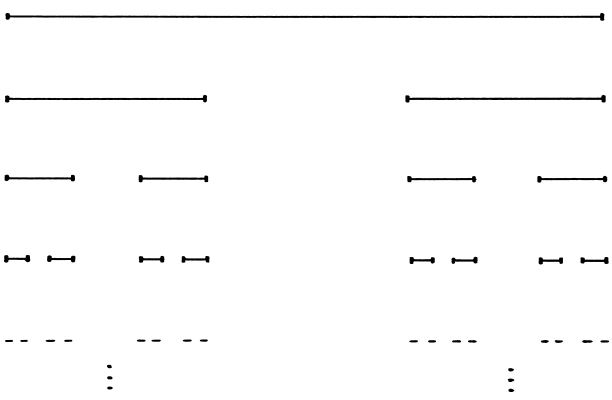

Figure 2.12: Sequent steps of construction of the "middle-thirds" Cantor set.

system of equations

$$\begin{cases} \dot{x} &= -10x + 10y\,, \\ \dot{y} &= rx - y - xz\,, \\ \dot{z} &= -\frac{8}{3}z + xy\,. \end{cases} \tag{2.23}$$

Here r is the control parameter known as the Rayleigh number. In the sack of the Lorenz model at $r > 24.74$ there exist one saddle point and two unstable focal points, so that the trajectory within the sack has to "rush about" between the unstable critical points (that reminds the children game "fifth angle"), asymptotically approaching to the strange attractor.

Because an elementary phase volume of the dissipative system should decrease with time evolution, the dimension of the strange attractor d must be lower than the dimension of the original phase space ν. Besides, at least in one direction of the phase space the strange attractor should have the structure of the Cantor set. A simplest way to explain the structure of the Cantor set, is to describe the concrete example of the "middle-thirds" set. Namely, let us take the closed unit interval, divide it into three parts and delete the middle open (i.e., excluding the end points) part as is shown in Fig. 2.12. Then let us divide each of the remaining intervals into three parts, and delete the middle parts again. Repeating this procedure by the infinite number of times, in what leaves we finally obtain the set with the structure of the Cantor set. It is characterized by the following properties:
(a) The Cantor set has the zero Lebesgue measure. Indeed, the "length" of the "middle-thirds" set is equal $1 - (\frac{1}{3} + 2\frac{1}{3^2} + 2^2\frac{1}{3^3} + \cdots) = 0$.
(b) The Cantor set is uncountable, because it contains "the same number" of points as the original unit interval. Indeed, at each step of construction of the "middle-thirds" set, it contains the uncountable number of points, therefore the limit set should have the same cardinality number as the set of irrational numbers.
(c) The Cantor set is totally disconnected, and it is *scale-invariant* (sometimes the terms "self-similar" or "auto-modeling" are used). Crude speaking, considering this set with a "microscope", we will see the same picture at any magnification. The scale invariance of the "middle-thirds" set is clear from its construction, because each next step of construction has repeated the previous one.

A fine model of the strange attractor has been proposed by Hénon [41]. The Hénon model is the two-dimensional discrete map of a rectangular into itself described by the set of equations

$$\begin{cases} x_{n+1} &= y_n + 1 - 1.4x_n^2\,, \\ y_{n+1} &= 0.3x_n\,. \end{cases} \tag{2.24}$$

Equations (2.24) were chosen in such a way that to describe approximately the Poincaré map for the Lorenz model (2.23). In constructing the strange attractor we may start from an arbitrary point and then remove few first points (about ten or hundred points). The results of numerical investigation of the model (2.24) are

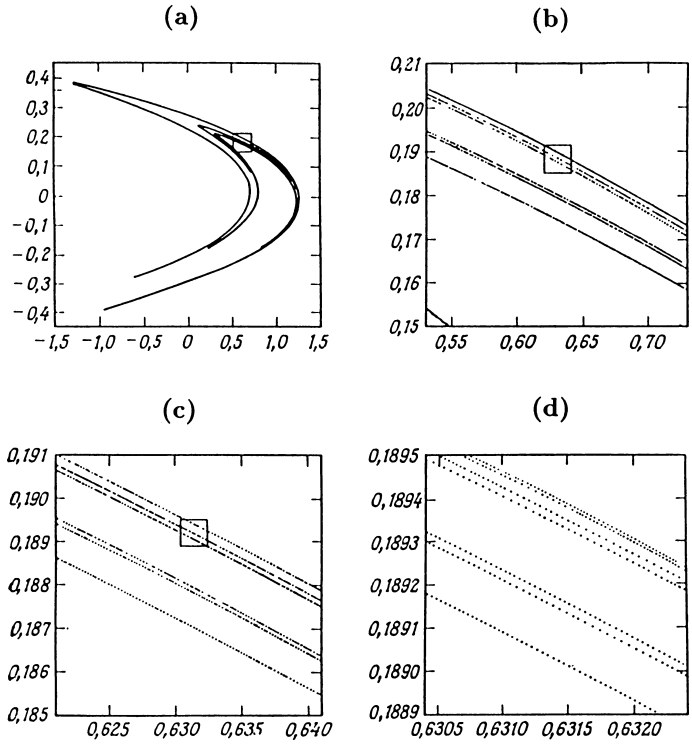

Figure 2.13: Layered structure of the Hénon attractor as calculated in [41]. (a) The whole attractor, (b) to (d) enlargements of the squares in the preceding figure.

shown in Fig. 2.13. The whole strange attractor of the Hénon model is plotted in Fig. 2.13a. Figure 2.13b scales up the part of the attractor which is inclosed in the small square in Fig. 2.13a. The small square of Fig. 2.13b is scaled up in Fig. 2.13c, and the small square of Fig. 2.13c, in Fig. 2.13d (note that in the numerical experiment the each "magnification" is to be followed by the increase of the total number of computed points in the same times). As seen, for any magnification the structure of the strange attractor remains the same: it consists of a series of continuous curves, while in the direction perpendicular to the curves, it has the structure of the Cantor set. Analogously, the strange attractor of dissipative systems described by continuous differential equations, has the structure of the Cantor set on the hypersurface where the nearest trajectories diverge (the vertical surface in Fig. 2.10) in the directions perpendicular to the trajectory.

A quantitative characteristic of the strange attractor is its *fractal*, or *Hausdorff dimension*. It is determined in the following way. In the ν-dimensional phase space of the system under consideration let us take small cubes with a size ϵ, i.e. with the volume ϵ^ν, and cover by them the strange attractor. Let $M(\epsilon)$ is the minimum number of cubes needed to cover the whole attractor. Then the fractal dimension (sometimes known as the *capacity*) is defined by the relationship [42] $M(\epsilon) \propto \epsilon^{-d}$, or

$$d = -\lim_{\epsilon \to 0} \frac{\ln M(\epsilon)}{\ln \epsilon} \,. \tag{2.25}$$

For example, for the the "middle-thirds" set the fractal dimension is equal $d = \ln 2/\ln 3 \approx 0.631$, for the logistic map at $r = r_\infty$, $d \approx 0.548$, for the Hénon attractor $d \approx 1.26$, for the Rössler attractor $d \approx 2.01$, and for the Lorenz attractor $d \approx 2.06$ [43, 44]. In computer experiments, however, the calculation of d is a rather difficult procedure, but it is relatively simple to find the Liapunov exponents λ_i, $i = 1, \cdots, \nu$ [44]. Recall that $\sum_{i=1}^{\nu} \lambda_i < 0$ for the dissipative system, and that at least one of the Liapunov exponents must be positive when an instability exists. Let λ_i are numerated in the decreasing order, $\lambda_1 > \cdots > \lambda_s > 0 > \lambda_{s+1} > \cdots > \lambda_\nu$,

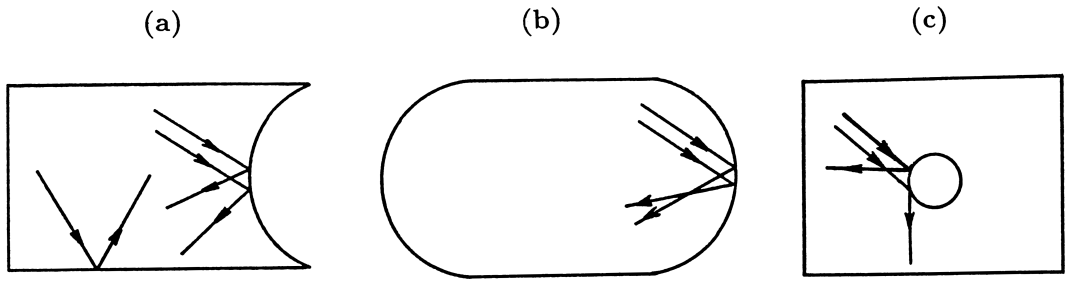

Figure 2.14: The Sinai billiards.

and let m is such an integer that

$$\lambda_1 + \lambda_2 + \cdots + \lambda_{m-1} > 0\,, \tag{2.26}$$

but

$$\lambda_1 + \lambda_2 + \cdots + \lambda_{m-1} + \lambda_m < 0\,. \tag{2.27}$$

Then let us choose the value α $(0 < \alpha < 1)$ from the condition

$$\lambda_1 + \lambda_2 + \cdots + \lambda_{m-1} + \alpha\lambda_m = 0\,. \tag{2.28}$$

According to the Kaplan-Yorke conjecture [45] which was confirmed by computer experiments, the fractal dimension d is very close to the *information dimension* d_{KY}:

$$d \approx d_{KY} = m + \alpha\,. \tag{2.29}$$

Other methods of calculation of the fractal dimension as well as characteristics of the strange attractor which are coupled with the former, can be found in [46] as well as in Part 6 of Moon's book [47].

2.6 Kolmogorov-Sinai entropy

Now let us return to conservative systems. The simplest among the conservative models with two degrees of freedom is the Sinai billiards, or stadium shown in Fig. 2.14. It describes the frictionless motion of a ball (more exactly, a disk) on the limited from all sides plane, and reflections from the boundaries are carried out according to the law "the reflection angle is equal to the incidence angle". In some cases, for example, when the billiards has the rectangular or round shape, the system is exactly integrable because it splits into two independent subsystems, each having one degree of freedom. However, if at least one of the walls is concave, Fig. 2.14a, or convex as in Fig. 2.14b, then the system dynamics becomes stochastic (according to the established terminology, in dissipative systems the word "chaotic" is commonly used, while in conservative systems the name "stochastic" is more often used). The instability which is responsible for arising of chaos, may be clarified from Fig. 2.14a: two neighboring trajectories being parallel before the collision with the concave wall, after the collision begin to diverge. It is clear that, analogously, the dynamics of the billiards with a "hole" in the middle will be stochastic too (Fig. 2.14c), as well as if instead of the "hole" we put the second immobile ball, or if two, or three, or many balls run over the billiards simultaneously. By the way, the last of the mentioned models is known as the *Lorenz gas*, and up to now it is the unique model where the validity of statistical physics laws is proved strictly mathematically [48, 49].

Simplicity of dynamics of the Sinai billiards is caused by that there are no islands of stability in this model, the stochastic trajectories occupy the whole phase space, and the "measure of chaos" does not depend on the system energy. Usually, however, the behavior of conservative systems is more complicated than that of dissipative ones. Excluding exotic cases such as the Sinai billiards, there always exist islands of stability (*invariant tori*) where the motion is regular [this statement is known as the KAM theorem

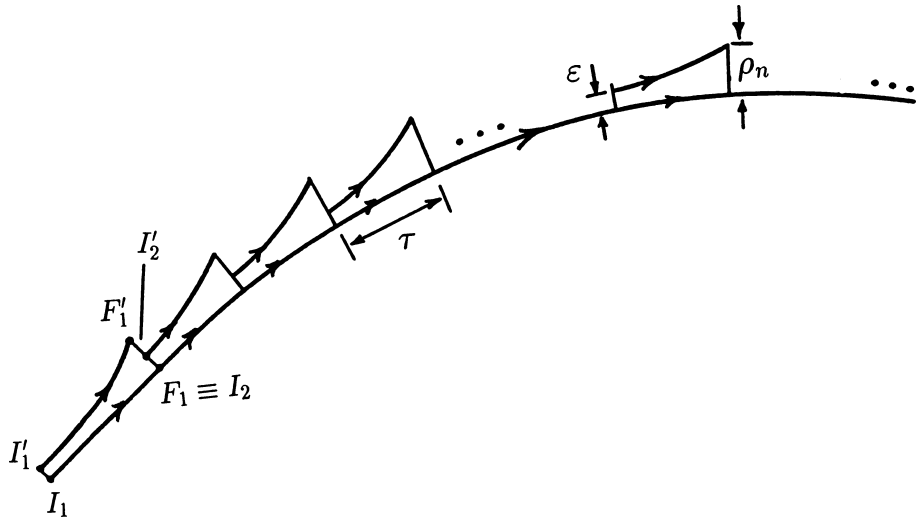

Figure 2.15: Numerical calculation of the maximal Liapunov exponent (after Benettin *et al.*[57]).

(Kolmogorov [50], Arnold [51], and Mozer [52])]. At small value of the total energy E of the system these islands occupy almost the whole available phase space, being separated from each other by thin stochastic layers. This regime is known as the *undeveloped* (or *local*, or *isolated*, or *weak*) stochastic regime. With increasing of the energy E the measure of these stochastic layers increases, and at some critical energy $E_{\rm crit}$ the neighboring stochastic layers overlap. (To estimate analytically the critical energy $E_{\rm crit}$, it is convenient to use the Chirikov criterion [53, 54, 55] known as the *resonance overlap criterion*. It is described in detail in, e.g., the books [14, 16]). For $E > E_{\rm crit}$ the *developed* (or *global*, or *connected*, or *strong*) chaos emerges in the system, and stochastic trajectories occupy a majority of the phase space, although small islands of stability continue to exist at any system energy. Thus, opposite to dissipative systems, in the conservative system the transition from the undeveloped chaos to the developed chaos is smooth. Ii is interesting that this transition is also described by the Feigenbaum theory, but with another common ratio δ of the geometric progression (for conservative systems $\delta = 8.72\ldots$).

A method of investigation of stochastic dynamics, most often used in computer experiments, is to calculate the Poincaré map, especially when the section plane is chosen "successfully". (In a general case instead the section plane we may introduce the so-called "surface without contact", i.e. a smooth surface which is intersected without touching in all its points by the phase trajectory.) The computation of correlation functions and Fourier spectra is useful too. But to estimate quantitatively the "measure of chaos" in the system, we have to calculate the so-called *Kolmogorov-Sinai* (KS) *entropy* h which characterizes the rate of diverging of neighboring trajectories. To find h numerically (see [56, 57]), we have to calculate simultaneously two trajectories, the main trajectory (called also the "reference" trajectory) and the "shifted" trajectory as is shown in Fig. 2.15. The total simulation time interval $t_{\rm max}$ is divided into N_τ small fixed intervals τ. Let the main trajectory starts at some point I_1 of the phase space. Then the point I_1' which defines the initial conditions for the "shifted" trajectory, is obtained from the point I_1 by its shift in some direction on a value ε. The shift may be done in an arbitrary direction, because the necessary direction will be installed automatically during the simulation process. Then both trajectories are calculated during the time τ. Let the main trajectory comes to the point F_1 for the time τ, and the shifted trajectory, to the point F_1' (Fig. 2.15). Denote the length of the vector $\overline{F_1F_1'}$ by ρ_1. For the next time interval, $\tau \le t \le 2\tau$, the initial points I_2 and I_2' are chosen in the following way: for the main trajectory we take $I_2 = F_1$, and for the shifted trajectory the initial point I_2' is obtained from the point I_2 by its shift on ε in the direction of the

vector $\overline{F_1F_1'}$. The described procedure is repeated by N_τ times, and then the following value is calculated:

$$\lambda(I_1, \varepsilon, N_\tau) = \frac{1}{\tau N_\tau} \sum_{i=1}^{N_\tau} \ln\left(\frac{\rho_i}{\varepsilon}\right). \tag{2.30}$$

To define the KS entropy strictly mathematically, we must also take the limits $\varepsilon, \tau \to 0$ and $t_{\max}, N_\tau \to \infty$. In computer experiment, however, numerical parameters ε, τ and N_τ should be chosen in such a way that at their variation in some limits (e.g., when they increase as well as decrease in two or three times) the value λ is to be approximately unchanged. Usually they are taken to be $\varepsilon \sim 10^{-5} \div 10^{-3}$, $\tau \sim 0.1 \div 1$ and $N_\tau \sim 10^3 \div 10^4$ in units which are natural for a given problem [56, 57, 58]. From the procedure described above it is clear that $\lambda(I_1)$ is the maximal Liapunov exponent averaged over the given trajectory. If the initial point I_1 is chosen within the limits of the island of stability, it should be $\lambda(I_1) = 0$; otherwise $\lambda(I_1) = \lambda_c > 0$ and $\lambda(I_1)$ should be independent on the choice of the initial point I_1 within the stochastic region. For the system with two degrees of freedom the KS entropy is determined from $\lambda(I_1)$ with the help of averaging over the position of the initial point within the available phase space [60], $h = \langle\lambda(I_1)\rangle_{I_1}$, or

$$h = \lambda_c \mu\,, \tag{2.31}$$

where μ is the measure of the stochastic region which may be estimated as $\mu \approx A_c/A$, A_c is the area of the region of the Poincaré map occupied by the stochastic trajectory, and A is the total area of the region which is available for the given energy of the system [58]. When the number of degrees of freedom is more than two, the KS entropy is determined by the relationship [59, 60]

$$h = \langle \sum_{\lambda_i > 0} \lambda_i \rangle, \tag{2.32}$$

so that we have to calculate all positive Liapunov exponents [61, 62]. Note, however, that the value h determined by the expression (2.31), leads to sufficiently reliable characteristic of the chaos for many-dimensional systems too.

Thus, the distance between the neighboring trajectories increases in average exponentially fast according to the law $\rho(t) \propto \exp(ht)$. If the system is exactly integrable, we have $h = 0$; otherwise $h > 0$ always. Until the chaos is undeveloped, the value h is small, but when the system energy E excesses the critical value E_{crit}, the KS entropy increases sharply (usually to a value $\sim 10^{-2}$). "Maximal" chaos with the KS entropy $h = \infty$ corresponds to the Markov process. (Note that adaptation the name "entropy" to h is connected with the fact that the KS entropy as well as the Liapunov exponents define an average loss of information on the initial conditions during the system evolution and, therefore, h is closely coupled with Shannon's information entropy; see detail in, e.g., [17]).

Describing the driven pendulum model we have noted that the motion becomes unstable if it takes place in a neighborhood of the separatrix. In the conservative system the stochastic layers arise near separatrices too; namely, they arise in the region of the intersection of two different separatrices (more rigorously, the intersection of the stable and unstable saddle manifolds). If these separatrices intersect one time at least, they must intersect an infinite number of times, and they create the so called *homoclinic structure* as shown in Fig. 2.16. The stochastic layers emerge namely in the region of the homoclinic structure. (Note that the strange attractor in the dissipative system arises in the region of attraction of the homoclinic structure too). In analytical approach, the homoclinic structure is often looked for with the help of the Melnikov method [63] (see also [64]).

2.7 Conclusion

Thus, the study of chaotic dynamics raises two main questions: first, what is the mechanism of the transition from regular to chaotic regimes of motion, and second, how to describe the system dynamics in the

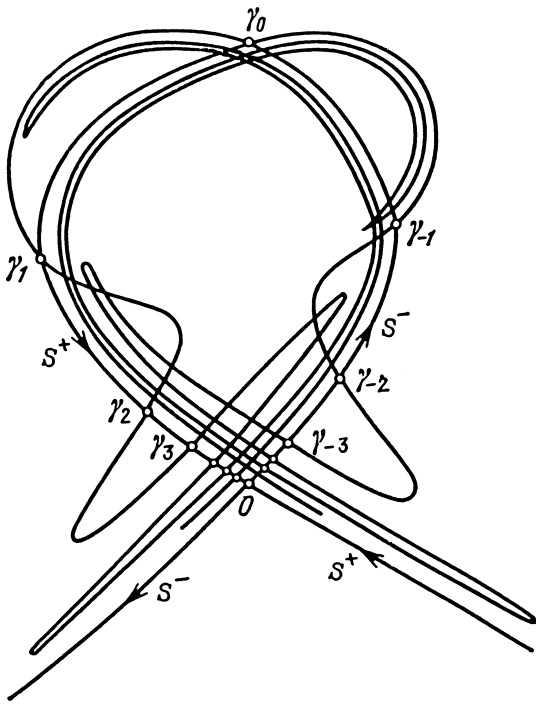

Figure 2.16: Section of the homoclinic structure. $S^{\pm}$ are the separatrices, γ_n $(n = 0, \pm 1, \ldots)$ are the intersection points.

chaotic regime. The transition to chaos usually proceeds thought the infinite sequence of period-doubling bifurcations. In computer or laboratory experiment in this case it will be enough to find few first bifurcation points only, and then the further system behavior can be predicted with the help of the Feigenbaum theory.

However, another scenarios of the route to chaos exist as well. In particular, in the mechanism of route to chaos thought *intermittency*, the regular system motion begins to be interrupted by short chaotic bursts when the control parameter r increases [65, 66, 67, 68]. At the beginning, the chaotic bursts are short and occur rarely, but with r increasing the system spends more and more time in the chaotic regime until the intervals with the regular system motion disappear at all. For example, the transition from regular to chaotic regimes through intermittency was observed in the Lorenz model (2.23) when the parameter r was increased in the interval $166 \leq r \leq 167$ [67].

Another important mechanism of the transition to chaos, the so-called "severe" (sharp? rigid?) mechanism, was observed, for example, in the Lorenz model (2.23) too with increasing of the parameter r from 1 to $24.74\ldots$ (When the parameter r decreases starting from infinity, the chaos in the Lorenz model arises according to the Feigenbaum scenario.) Namely, at small r, $1 < r < 13.926\ldots$, the system behavior is regular, and there are one saddle and two stable focal points in the system. At $r = 13.926\ldots$ the homoclinic trajectories arise (i.e., the two trajectories, the first going out and the second going into the saddle, are united into one closed trajectory), which surrounds the stable critical points. With the further increase of r, the homoclinic trajectories are transferred into unstable limit cycles, the radius of these cycles decreases, and at $r = 470/19 \approx 24.7368$ the cycles collapse into points, absorbing the stable focal points and creating the unstable focal points in the same places.

For the complete description of the system motion in the chaotic regime we have to find the corresponding homoclinic structure in the case of the conservative system, or the strange attractor for the dissipative system. In solution of this problem the computer modeling usually plays the main role. Note also that with the change of the system parameters, the structure of the chaotic manifold may change as well, i.e., the chaos $\rightarrow$ chaos transitions may take place. For example, even in the simplest model such as the logistic

map (2.9), within the region $r_\infty < r < 4$ there exist "windows" with the regular dynamics. In particular, with r variation starting from the value $r_c = 1 + \sqrt{8}$, the transition from the regular to the chaotic regime takes place, and in the case of r decreasing this transition is carried out by the intermittency mechanism.

Note also that the structure of the chaotic manifold is "stable", i.e. it remains unchanged under the action of small external perturbations, e.g., a small-amplitude external noise.

In conclusion note that in physics for a long time it has been wide-spread a mistake that any problem of classical mechanics has, at least in principle, the exact solution, and the only problem is to have a sufficiently high-speed computer for its solution. This mistake was based on the fact that in the linear approximation, indeed, any problem is exactly integrable. Of course, there exist a number of nonlinear models which are exactly integrable too (for example, even the Hénon-Heiles model (2.2) is integrable for some particular sets of the parameters such as $\omega_y = 1$ and $\mu = 1/3$, or $\omega_y = 4$ and $\mu = 16/3$, or $\mu = 2$ and arbitrary ω_y), and these models play an extremely important role in physics (see below Chapter 6). It occurs, however, that when a nonlinearity is taken into account, almost all systems are nonintegrable, i.e. they have no analytical solution in an integral form in principle, and their dynamics is chaotic. When this fact has became clear, the stochastic theory spreads over practically all branches of physics as well as of adjacent sciences. In particular, we may give the following examples (much more enormous list of applications of the stochastic theory can be found in Moon's book [47]):

1. *Statistical mechanics.* The stochastic theory put the base for the main hypotheses of statistical mechanics, namely for the ergodic hypothesis (which implies that the time average exists, and it has to coincide with the average over the phase space), and Boltzmann's hypothesis (which states that the average kinetic energy is to be equally shared among all degrees of freedom of the system). It appears that in order these hypotheses to be true, it is not necessary to have an enormous (10^{23}) number of atoms as it was assumed earlier, but it is enough to have a single atom only as in the Sinai billiards! But it does be necessary to have an instability. The instability is also responsible for the time irreversibility of statistical mechanics laws which naturally emerges when the "rough" distribution function is used, i.e. when we reject the attempts to put the initial conditions with the infinite accuracy. At the same time, however, we are losing the configurations which have an extremely small probability to occur (for example, that all gas molecules come together in one half of a volume at a given time moment).

2. *Meteorology.* The Lorenz model (2.23) appears namely in this field of science. In particular, the stochastic theory shows senseless of attempts to predict a weather for a long time period by solution of deterministic motion equations for the atmosphere. The latter approach can predict the weather for a period of $\Delta t \sim \lambda_{\max}^{-1}$ only, this period is about two weeks.

3. *Aero- and hydrodynamics.* A great hope has been put on that the stochastic theory would help to describe the transition from a laminar to the turbulent motion of liquid or gas as well as to describe characteristics of the turbulence regime itself.

4. *Radiotechnique.* When an information is transferred along a line or optical fibre, it is desirable that the noise produced by the line itself and then introduced into the signal, is reduced to a minimum. To achieve this goal, we have to choose the line parameters in such a way that the system has to be as close as possible to the exactly integrable system, i.e. the line has to be characterized by the minimal KS entropy. The corresponding example will be presented in Chapter 6.

5. *Optics.* Lasers with smoothly varying frequency have a large area of application. One of methods to construct such a laser is to make the wideband (noise) laser and then to subtract the necessary frequency with the help of a resonator. But to produce a continuous frequency spectrum, the laser should be constructed in a way that to work in the chaotic regime.

6. *Celestial mechanics.* The Hénon-Heiles model (2.2) belongs to this field of science. The stochastic mechanics explains, in particular, why it is necessary to correct the trajectory of motion of a satellite or rocket from time to time.

7. *Chemistry, biology etc.* As the last example we note that the logistic map (2.9) has been introduced by P.F. Verhulst as far as in 1845 to simulate the evolution of a population in a closed area. Namely, if we denote by x_n the relative (normalized) number of species in the n-th year, the linear on x term in the

right-hand side of Eq. (2.9) will describe an increasing of the population due to duplication of species, while the quadratic term, their ruin owing to limits of meal resources, territory, *etc.*

Chapter 3

Molecular Dynamics Method

In the present Chapter we consider the first among the two main methods of modeling used in physics, the Molecular Dynamics (MD) method. It consists in computer modeling of motion of a "sufficiently large" number of particles with a given interaction law. The word "molecular" in the name of the MD method has a historical origin; in a real system electrons, atoms, molecules, quasiparticles (e.g., solitons), planets, stars, galaxies, *etc.* may play the role of "molecules" as well. The motion of particles is usually assumed to be classical, i.e. they move according to Newton's laws. For an atomic systems, an effective interaction potential which approximately takes into account quantum corrections to the interaction law, may be used [69].

The MD method is used for investigation of systems which either have no analytical description at all, or have an analytical description but its accuracy and the region of its validity are unknown. In the latter case, the results of MD modeling are used as an "etalon", i.e., as the "obtained from the first principles exact solution" which could then be compared with the results of different approximate approaches. First of all we have to mention the systems which are far from the thermal equilibrium state. Investigation of dynamical characteristics of the thermodynamically equilibrium state by the MD modeling leads to important and interesting results too. Last, investigation of thermodynamical characteristics such as properties of phase transitions by the MD method is useful too, although in the latter case the Monte Carlo method is more suitable usually.

As has been mentioned in Introduction, computer modeling in many points looks like a laboratory experiment and, moreover, it has a series of important advantages. First, in the computer experiment we can make a snapshot and analyze the positions of all particles in the system at any given time moment. Often it is useful also to plot the trajectories of particles during some time interval. Clearly that to obtain such pictures in laboratory experiments is impossible in principle. Second, with the help of computer modeling we can investigate the system behavior under the conditions which cannot be achieved in a real experiment. For example, we may heat up a plasma to as high temperature as we want, artificially restricting the system volume using appropriate boundary conditions. Third, in the computer experiment we can investigate the systems which are inaccessible for us in a laboratory. For example, we can study star collisions, evolution of galaxies, *etc.*

As an illustration, let us describe an example where a single star collides with a double star [70]. Figure 3.1 shows star trajectories for one particular choice of initial conditions. As seen, the process is rather complicated. Depending on initial conditions (such as the target parameter, i.e. the shortest distance between the single star and the center of the double system, and the orbital phase, i.e. the relative positions of stars in the double star in the initial configuration), one of the following three scenarios may take place: **(a)** the "fly past", when the orbital phase of the double star has been changed in the collision process, but the double star remains to exist as a whole (Fig. 3.2a), **(b)** the "exchange", when the single star replaces a star in the double system (Fig. 3.2b), and **(c)** the "ionization", when all three stars occur to become uncoupled in the final state (Fig. 3.2c). The results of modeling of more than 10^3 experiments are summarized in the diagram in Fig. 3.3.

In the example described above the MD method has been used to model the motion of three particles.

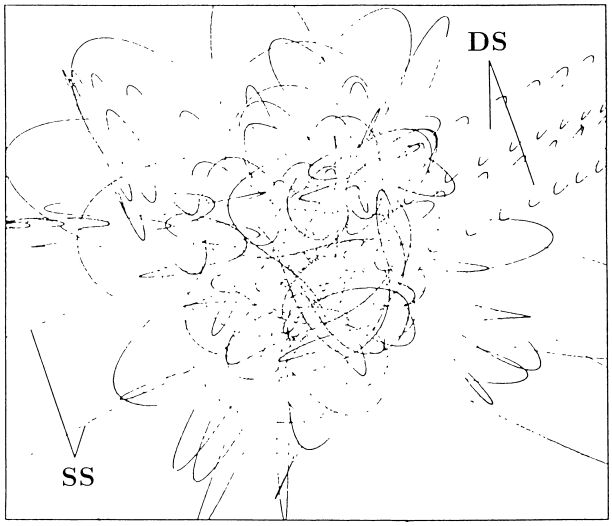

Figure 3.1: Trajectories of stars' motion for one particular choice of initial conditions as calculated in [70]. **SS** indicates the trajectory of the single star, and **DS**, that of the double system.

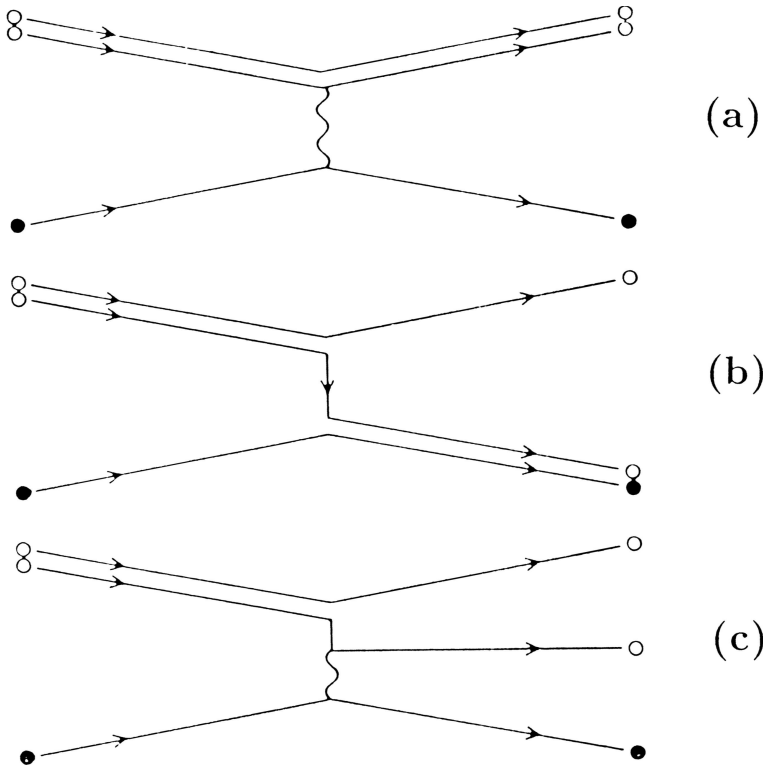

Figure 3.2: Schematic presentation of different stars' collision scenarios: (a) fly past, (b) exchange, and (c) ionization (from [70]).

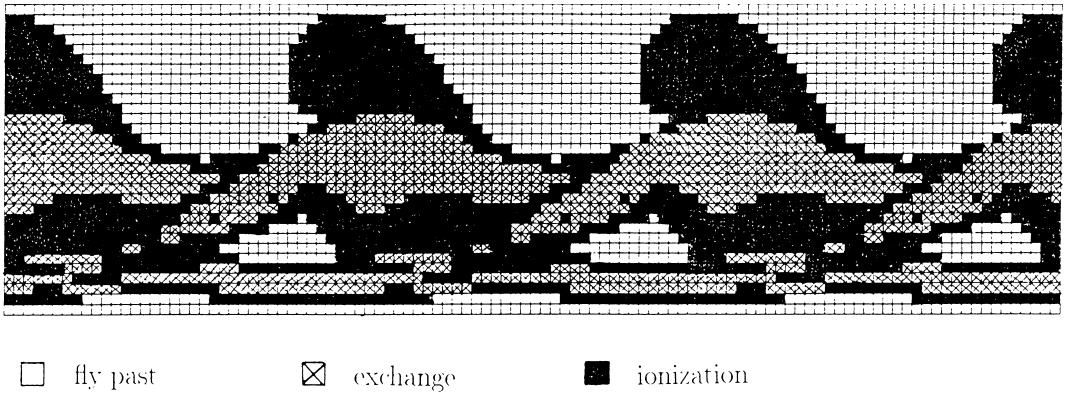

Figure 3.3: Diagram for different events at stars' collision for different initial conditions as calculated in [70].

Sometimes, as, for example, in the Henon-Heiles model (Sec. 2.1), the MD method is used to study the motion of a single particle only. In modeling of physical systems, however, a more typical is the situation when the number of particles should be as large as possible, and this number is restricted by the computer power only. In fact, the number of particles in the MD method is about $10^2 \div 10^3$, and up to 10^6 in the extreme cases. An answer on the question, is this number sufficient for that the physical situation being described correctly, first of all depends on the dimensionality of the model under consideration. For example, in modeling of one-dimensional systems, a chain of 10^3 atoms is practically undistinguishable from the infinite chain. But when we study three-dimensional systems, a cube of 10^3 atoms has the linear extension of 10 atoms, so in the latter case the model describes properties of a small cluster rather than the bulk properties of a solid state.

The MD modeling was, is, and (at least in a near future) will be a sufficiently expensive job. Usually we can make a small number of computer experiments only, and we may vary a restricted number of model parameters. Thus, the model under consideration as well as the method of its solution should be chosen with a special care. Below in Sections 3.2 to 3.4 we consider several characteristic examples of MD modeling. First, however, let us present some general principles which would be used in choosing of models and methods for their solution.

3.1 General statements

First of all we have to settle the model dimensionality. There exist one-dimensional (1D) models where atoms move along one coordinate only, two-dimensional (2D) models where atoms are aligned with a plane, and three-dimensional (3D) models. Sometimes, one- or two-dimensional models for a system of atoms inserted into an external (usually periodic) potential are used. For example, in the two-dimensional system of atoms adsorbed on a crystal surface, the substrate periodic potential plays the role of the external potential.

Of course, real physical objects are placed into three-dimensional space always. On the other hand, except some specific problems, linear extension of the system should be sufficiently large. Therefore, if it is possible, the priority should be given to one- or two-dimensional models. But such a possibility may not exist at all. The most striking examples of situations where the one-dimensional model incorrectly describes the physical situation, are the problem of validity of statistical physics laws (where the use of the 1D model leads to the Fermi-Pasta-Ulam paradox, see below Sec. 6.1), and the problem of phase transitions (one-dimensional systems do not exhibit phase transitions at all for any $T > 0$).

Because in numerical experiments the number of particles is finite always, the question on boundary conditions arises. The following three types of boundary conditions are used typically:

(a) *Free boundary conditions.* In this case there are no additional conditions for the boundary atoms of the model's volume. Note, however, that the atoms at the boundary are in special conditions because they

have neighbors from one-hand side only. Of course, the free boundary conditions can only be used when the atoms attract each other at large distances at least, because otherwise the atoms will run away from the system in opposite directions.

(b) *The fixed boundary conditions*, when the coordinates of the boundary atoms are fixed and their velocities are artificially put to zero, i.e. the boundary atoms are excluded from the MD modeling. Sometimes the following modification of this method is used (the so-called "damped boundary conditions"): additionally to pinning of the boundary atoms, some conditions are imposed on few "marked" atoms (or atomic layers) located near the boundaries. For example, while the motion of atoms in the system is subjected to Newton's laws, the motion of the "marked" atoms is described by the motion equations with a viscous friction, and the value of the friction coefficient smoothly increases with approaching to the boundary (from zero for "bulk" atoms to infinity for boundary atoms). In this case impulses or waves which go to the boundary region, are damped in this region and cannot go back to the bulk of the system. In the result, dynamical properties of such system will remind those of the infinite system where also a wave going to infinity, cannot be returned back. An another example of modified boundary conditions is the case when motion of atoms near the boundary is described by the Langevin equations (see below Chapter 4); in this case we can model a system in contact with a thermostat at a nonzero temperature.

It is important to note that in the cases of the free or fixed boundary conditions, we must exclude the boundary and "marked" atoms from the consideration if we are interesting in "bulk" system properties.

(c) May be, the most popular are the *periodic*, or *cyclic (Born-Karpman) boundary conditions*. For example, if we study a one-dimensional model, then in order to impose the periodic boundary conditions on the chain of N atoms, we have to introduce a fictitious $(N+1)$-th atom which is assumed to move analogously to the first atom in the chain. In other words, the chain is coupled into the cycle. Similarly, one gets a tori for the two-dimensional model, and the three-dimensional tori in the case of the three-dimensional model. If the atomic system is placed into a box with a fixed volume, we have to assume that when an atom crosses the boundary and leaves the box, at the same time a new atom with the same velocity is to appear on the opposite boundary of the box. The periodic boundary conditions look to be the most attracting ones, because in this case all the atoms are treated on equal foot, and all atoms may be used for calculation of "bulk" properties. However, the cyclic conditions have their own inconveniences. Namely, imposing artificially the periodicity on the system, at the same time we introduce the minimum wave vector $k_{\min} = 2\pi/L$ and the minimum frequency $\omega_{\min} = 2\pi/t_{\max}$, where $t_{\max} = L/v_{\text{sound}}$, L is the linear system size and v_{sound} is the maximum sound velocity in the system. In the result the spacial Fourier spectrum of the system becomes discrete (the total momentum may take only the values which are multiplies of $k_{\min}$), and the modeling is correct for times $t < t_{\max}$ only.

When the fixed or periodic boundary conditions are used, we also have to choose a geometry for the "box" surrounding the system. In modeling of gas or liquid, the most useful way is to take the simplest box shape, i.e. to use the square (cube) box. But when we model the system which may have a crystalline structure, the shape of the box must be matched with the symmetry of this structure. Namely, the crystal must be imposed into the box without distortions, because otherwise we will artificially introduce topological defects which can not be removed during simulation.

The next important stage of model construction is a choice of the interatomic potential. In the MD modeling it is usually assumed that the interaction is pairwise and isotropic, i.e. that the potential energy of the system can be presented in the form

$$V = \sum_{i<j} v(r_{ij})\,, \tag{3.1}$$

where r_{ij} is the distance between the atoms with the numbers i and j. Of course, sometimes this assumption does not correspond to reality. For example, two hydrogen atoms attract one another and form the hydrogen molecule, but the coupling of a third hydrogen atom to the molecule is energetically unfavorable.

Typically one of the following functions is used for $v(r)$:

(a) The "hard-core" potential has the form

$$v(r) = \begin{cases} \infty & \text{if} \quad r < a_0, \\ 0 & \text{if} \quad r > a_0, \end{cases} \tag{3.2}$$

where a_0 corresponds to the diameter of particles. Also one may use the hard-core potential with an attraction,

$$v(r) = \begin{cases} \infty & \text{if} \quad r < a_0, \\ -v_0 & \text{if} \quad a_0 < r < a_1, \\ 0 & \text{if} \quad r > a_1, \end{cases} \tag{3.3}$$

where v_0 is the well depth. The potentials (3.2) and (3.3) are the simplest used in the MD simulation.

(b) The power potential is described by an expression

$$v(r) = v_0 \left(\frac{a_0}{r}\right)^{\nu}. \tag{3.4}$$

For $\nu = 1$ or $\nu = 3$ this potential corresponds to Coulomb or dipole-dipole potentials respectively. As examples of applications of the potential (3.4) we can mention the interaction between electrons and ions in the ionized plasma, or the interaction of partially ionized atoms chemically adsorbed on a metal surface [71].

(c) The widely used Mie-Lennard-Jones (LJ) potential is described by an expression

$$v(r) = \frac{v_0}{\mu - \nu}\left[\nu\left(\frac{r_0}{r}\right)^{\mu} - \mu\left(\frac{r_0}{r}\right)^{\nu}\right]. \tag{3.5}$$

This potential has the minimum of the depth $v(r_0) = -v_0$ at $r = r_0$, and the frequency of vibrations near the minimum is

$$\omega_0 = \left(\frac{\mu\nu v_0}{m r_0^2}\right)^{1/2}, \tag{3.6}$$

where m is the mass of particles. The usual combination is $\mu = 12$ and $\nu = 6$ which describes the interaction of inert gas atoms. The first term in the right-hand side of Eq. (3.5) describes the short-range repulsion owing to overlapping of electronic shells, while the second term corresponds to the long-range van-der-Waals attraction due to mutual polarization of the atoms.

(d) The Morse potential has the form

$$v(r) = \frac{v_0}{\mu - \nu}\left[\nu e^{-\mu\gamma(r-r_0)} - \mu e^{-\nu\gamma(r-r_0)}\right], \tag{3.7}$$

where the parameters r_0 and v_0 have the same meaning as in the LJ potential, and the parameter γ is coupled with the vibrational frequency at the minimum by the relationship

$$\omega_0 = (\mu\nu v_0\gamma^2/m)^{1/2}; \tag{3.8}$$

usually γ is taken to be $\gamma \sim r_0^{-1}$. The combination $\mu = 2$ and $\nu = 1$ is used as a rule, while another combinations may be used as well. The Morse potential is typically used to describe chemical bonds in molecules.

(e) Last, in some cases it is necessary to use the real interaction potential for $v(r)$. For example, such a situation emerges in modeling of the crystalline structure of metals, because the symmetry of the lattice which corresponds to the minimum of the energy, is very sensitive to the shape of the potential $v(r)$. In this case the function $v(r)$ should be calculated (by solving, e.g., the corresponding Schrödinger equation for two atoms) and tabulated before the MD simulation. Then, during the simulation, a necessary value of $v(r)$ is being found with the help of an interpolation procedure.

The choice of the concrete potential depends, of course, on the physical problem under investigation. But for other equal conditions, the preference should be given to that function $v(r)$ which can be calculated with a shorter computer time. From this point of view the potentials **(b)** and **(c)** have the advantage before the potential **(d)**, because the operation of taking to a power is performed much more fast and accurate than the operation of taking an exponent.

The molecular dynamics method assumes the calculation of trajectories of all model particles. As a rule, the solution of motion equations is carried out by one or another modification of the Runge-Kutta method. The main requirement is, of course, the maximum accuracy of the solution. But the increase of accuracy is not a simple task. The only method of computer solution of Newtonian differential equations is to make the discretization. For this, we have to introduce a time step Δt and then to solve not the original differential equations but the difference equations. Of course, numerical errors at this stage are unavoidable. To decrease the errors we may, for example, to decrease the time step Δt. But this can be done to a certain limit only, because for too small Δt the resulting computation errors will overcome all gains in the accuracy. An another way to improve the accuracy consists in using the Runge-Kutta method of a higher order. But this approach leads to a sharp increase of the computation time, because the more complicated is the Runge-Kutta scheme, the longer will be the computer time of calculation of the right-hand sides of motion equations, i.e., the more is the number of times the function $v(r_{ij})$ for all interatomic distances r_{ij} is to be calculated. On the other hand, in the MD modeling we typically need to follow for system evolution for a sufficiently long time $t_{\max}$. Thus, the following contradiction appears: the more accurate we will calculate the trajectories of the system motion, the shorter will be the total time of modeling $t_{\max}$ and, therefore, the lower will be the accuracy of calculation of the averaged over time characteristics of the system. But namely the time-averaged values are typically the main goal of the MD modeling. A resolution of this contradiction would be in a compromise: the "technical" computational parameters (e.g., the time step Δt) should be chosen in such a way that to achieve the maximum accuracy in the *final* results of modeling (i.e., in the time-averaged characteristics), not be too warning on accuracy of calculation of the trajectory itself. (The final choice of these "technical" parameters is carried out usually during the "training" of the program.) Moreover, for a nonintegrable system, because of the existence of an instability, the precise calculation of the trajectory (i.e., the calculation which allows to return to the initial point when all velocities are reversed) is impossible in principle.

One of methods to control the accuracy of the MD program is to test its work for the cases when the exact solution of the problem is known, for example, when the nonlinear interactions are replaced by the linear ones. For conservative systems there is also the standard test consisting in checking the energy conservation. Typically in the MD calculations the total energy must be conserved with an accuracy of $0.05 \div 1\%$. (In practice, this test is realized by calculation of the total energy at given time moments, e.g., at each 100-th time step of Δt.) Note that in some schemes of solution of motion equations the energy conservation is supported artificially (e.g., see [72, 73]); in this case we need to look for other methods to test the accuracy.

To finish with the model construction, we have to choose appropriate initial conditions. The choice of the initial configuration depends on the problem under consideration. When we are investigating the system which is close to the exactly integrable one (as, for example, in solitonic problems, see below Chapter 6), it is natural to take one of the exact solutions of the corresponding integrable system (e.g., the state with one or two solitons) as the initial configuration. In investigation of statistical properties of gases or liquids we may take the configuration with random positions of atoms. When we are studying the thermodynamical properties of crystals, it would be better to take the regular arrangement of atoms in the crystal lattice at the initial time moment, but to randomize their velocities. The initial conditions should be chosen with a sufficient attention, because an inappropriate initial state may lead to a metastable configuration which will not be escaped for a reasonable modeling time.

In any case, however, we cannot take the "absolutely correct" initial configuration, because for this we have to know the exact solution of the given problem. Therefore usually we have to wait for some time during which the system will reach the "working" regime, and only after that we may start the calculation of the characteristics under investigation. For example, when we study the equilibrium system parameters, the

"waiting" time corresponds to achieving the equilibrium state. In "solitonic" problems during the waiting time the elementary excitations of the system should take their stationary ("true") shape. The waiting time depends on the problem under consideration, and it may be about $(0.25 \div 0.5)t_{\max}$.

In the next Sections 3.2 to 3.4 we present characteristic examples of the MD modeling, and in the last Sections 3.5 and 3.6 we discuss some important technical details of the application of the MD method.

3.2 Melting of two-dimensional crystals

3.2.1 Introductional remarks

The checking of validity of the main hypotheses of the statistical mechanics was among the first applications of the MD method. Namely, if we take a real (i.e., anharmonic) potential of interaction between atoms and start from an arbitrary configuration, will the thermal equilibrium state (i.e. the state where the kinetic energy is equally shared among all system degrees of freedom) be achieved with time evolution? The answer which was obtained in the modeling of the one-dimensional atomic chain with the first computer "ANIAC", suddenly appeared to be negative. This phenomenon, named the Fermi-Pasta-Ulam paradox, will be considered in a more detail below in Sec. 6.1. Note only that the reasons for the negative answer were a small capacity of early computers that did not allow to wait for achieving the equilibrium state, and the using of the one-dimensional model which is characterized by very specific properties.

The two-dimensional model already does not exhibit such a surprise. For example, let us consider the Benettin and Tenenbaum simulation [74], where the two-dimensional system of $10 \times 10 = 100$ atoms has been investigated. They have used the fixed boundary conditions so that the positions of $4 \times 9 = 36$ atoms were kept fixed at the boundary of the square box, and $N = 8 \times 8 = 64$ atoms were allowed to move. The atoms interacted via the LJ potential with the parameters adopted to describe the interaction of argon atoms. The configuration with random coordinates and velocities of atoms was taken as the initial one. Then the Newton motion equations were solved with the time step $\Delta t = 0.05$ (in the natural for the given problem system of units) during the time $t_{\max} = 2 \times 10^4$, and a time $\sim 10^2$ was given for the system relaxation. The distribution of the energy over the linear modes as well as the auto- and cross-correlation functions and the Fourier spectrum were investigated. The results show (see Fig. 3.4) that the thermal equilibrium state is achieved for a time ~ 10 provided the system energy exceeds the value $E_{\text{crit}} \approx 0.03N$, that corresponds to the temperature ~ 3K for the chosen model parameters. The existence of the critical energy E_{crit} which is nothing else that the threshold of the transition to the developed chaos (see the previous Chapter 2), is caused by a small number of atoms in the system; it has to be expected that $E_{\text{crit}} \to 0$ in the limit $N \to \infty$ [55].

When the existence of the thermal equilibrium has been proved, we may begin the investigation of properties of the equilibrium state. Recall that for an analytical description of the equilibrium state we usually should have some "small parameter". For example, in gases the kinetic energy of atoms (molecules) essentially exceeds the potential energy of their interaction, and the latter may be accounted with the help of a perturbation theory. On the other hand, in crystals the atoms take the regular lattice sites and only slightly vibrate with respect to equilibrium positions, so that the kinetic energy plays the role of a small perturbation. But in liquids where the kinetic and potential energies are of the same order, an analytical description becomes very complicated and inexact, and the MD simulation becomes to be practically the only reliable method of the investigation [75].

An another example where the computer modeling plays an important role, is the investigation of phase transitions between different phases. Recall that the phase transition (PT) may be either discontinuous (the first-order PT) or continuous (the second-order PT). From the physical viewpoint the difference between the first-order and second-order PTs consists in the following [76]. In the first-order PT which takes place with changing of some system parameter, e.g., the temperature T, at a certain value T_c the free energies of two different phases becomes equal each other. The discontinuous PT is carried out "locally" in space. For example, with T decreasing at $T \leq T_c$ nucleus (islands, drops, grains, *etc.*) of the new phase are

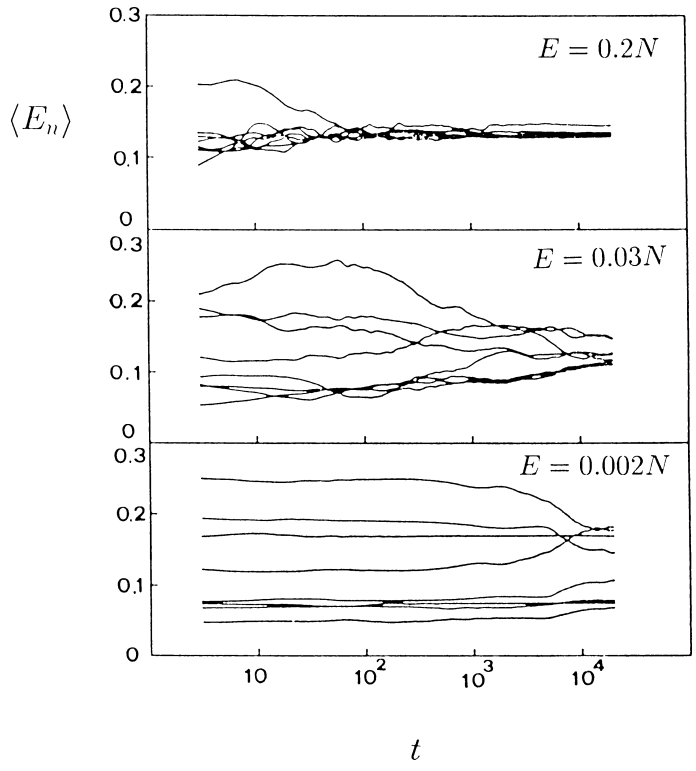

Figure 3.4: Averaged energies $\langle E_n \rangle$ of the groups of normal modes at three energies, as functions of the averaging time t. E is the total energy of the system (after [74]).

spontaneously created at random sites (or at irregularities if those exist) in the system, and when the nucleus reaches some critical size, it will grow up to macroscopic sizes (see below Chapter 5). However, at $T < T_c$ the old phase may exist as well in a metastable state (for example, the overcooled liquid or gas). Therefore, the main characteristic feature of the discontinuous phase transition that has to be observed in laboratory or computer experiment, is the existence of a hysteresis. Note that for the existence of the first-order PT, there must be an attraction (or at least an "effective" attraction, i.e. the lowering of the interatomic repulsion when the atoms gather together) between the atoms.

The second-order PT is the transition between two phases, one of which must have the lower symmetry than the another phase, i.e. the symmetry group of the first phase should be a subgroup of the symmetry group of the second phase. The continuous PT is carried out by the fluctuation mechanism: when T decreases approaching to T_c (but $T > T_c$), "domains" of the new phase are appearing from time to time due to fluctuations, and these domains have a finite (microscopic) lifetime. The closer is T to T_c, the larger are the sizes and lifetimes of these domains, and at $T = T_c$ they occupy the whole space, so that at $T < T_c$ the old phase exists already as fluctuating domains only. The second-order PT takes place in the whole space simultaneously, and the hysteresis is absent.

The classical example of the phase transition is the melting of the crystalline structure, i.e. the crystal–liquid or order–disorder transition. From experiments we know that this phase transition is the first-order PT usually. However, there are no restrictions which forbid for the melting to be the continuous PT in some systems. An interesting system where the melting is the continuous phase transition, is the two-dimensional Coulomb gas.

3.2.2 Physical model

As was first shown by Peierls [77, 78] and Landau [79] (see also [80] and [76]), the existence of the long-range order in a system is connected with the system dimensionality first of all. Let the function $u(\vec{r})$ describes the displacements of atoms situated near the point $\vec{r}$, from their regular positions in the nodes of the ideal crystal lattice. Then as the criterion of the crystalline order, we may consider the behavior of the function

$$\delta^2(\vec{r}) = \langle [u(\vec{r}) - u(0)]^2 \rangle \,, \tag{3.9}$$

which describes the mean-square displacement of atoms separated by the distance $\vec{r}$, with respect to one another. Clearly that in the crystalline phase the limit of the function $\delta^2(\vec{r})$ must be finite as $r \equiv |\vec{r}| \to \infty$, while in the disordered phase the value $\delta^2(\vec{r})$ will increase unboundedly. The function $\delta^2(\vec{r})$ can be estimated if we turn to normal modes,

$$u(\vec{r}) = \int d\vec{q}\, u_{\vec{q}} \exp(i\vec{q}\vec{r}) \,, \tag{3.10}$$

and then estimate the elastic energy of the system in the harmonic approximation as

$$E \sim \int d\vec{r} \left(\frac{\partial u(\vec{r})}{\partial \vec{r}} \right)^2 \sim \int d\vec{q}\, q^2 \mid u_{\vec{q}} \mid^2 . \tag{3.11}$$

Because each degree of freedom has to take an energy $\sim k_B T$, the following estimation follows from Eq. (3.11):

$$\mid u_{\vec{q}} \mid^2 \sim k_B T / q^2. \tag{3.12}$$

Substituting Eqs. (3.10–3.12) into Eq. (3.9), at $r \to \infty$ we approximately obtain (see details in [80]) that

$$\delta^2(r) \sim T \int_{2\pi/r} d\vec{q}\, q^{-2}, \tag{3.13}$$

where we have to take into account the lower integration limit only.

In the three-dimensional case, where $d\vec{q} \propto q^2\, dq$, the integral (3.13) is finite, and a crystalline order may exist in the system (of course, when the temperature is low enough). On the other hand, in the one-dimensional system, where $d\vec{q} = dq$, the integral (3.13) linearly diverges with r at $r \to \infty$, $\delta^2(r) \propto Tr$. This means that in one-dimensional systems at any $T \neq 0$ the disordered phase can exist only, and the PT critical temperature is $T_c = 0$ (note that the reasons given above assume the short-range interatomic interaction).

In the intermediate case of the two-dimensional system, where $d\vec{q} \propto q\, dq$, the function $\delta^2(r)$ rises with r at $r \to \infty$ too, $\delta^2(r) \propto T \ln r$. But now this growth is logarithmic, i.e., it is not so fast as in the one-dimensional model or in the disordered phase. Thus, although the two-dimensional system cannot exhibit the real crystalline structure, a more or less ordered phase must exist at low temperatures. This phase got the name *floating phase*, or the phase with quasi-long-range order [81].

There is a number of physical systems which may be considered approximately as two-dimensional ones. For example, it may be a film of atoms adsorbed on the crystal surface in the case when we can neglect by the substrate potential relief. As other examples we may mention the system of electrons "levitating" over the surface of liquid helium, or the electrons within the inversion layer in the silicon MOS transistor.

Thus, in the two-dimensional system at a low temperature it should exist the floating phase. On the other hand, at high temperatures the disordered phase has to exist, because it has a higher entropy than the ordered phase. Naturally it arises the question on the temperature and the mechanism of melting of the floating phase. The theory of this phase transition has been proposed by Kosterlitz and Thouless [82]. It is interesting that the two-dimensional melting was first observed in the computer simulation with the help of the MD method (namely, the P^3M method, see below Sec. 3.5) [83], and only after this, the verification in the laboratory experiment has been followed [84]. Below we briefly discuss the Kosterlitz-Thouless (KT) melting mechanism, and then describe in more detail the MD simulation results of Bedanov *et al.* [85].

According to the KT theory, the two-dimensional system melts by the dislocation mechanism. Namely, a pair of dislocations with opposite signs may be created in the ordered phase. From dislocation theory

it is known (e.g., see [86]) that in the two-dimensional system the two dislocations with opposite signs (i.e., with the opposite directions of Burger's vector) attract each other according to the logarithmic law, so that the creation of the pair increases the system energy on a value $\sim \varepsilon_0 \ln r$, where ε_0 is some energy constant and r is the distance between the dislocation centers. On the other hand, the dislocation pair can be placed in the lattice in $\sim (L/a_0)^2$ ways, where L is the linear system extension and a_0 is the mean interatomic distance. Thus, the creation of the dislocation pair increases the system entropy on a value $\Delta S \sim k_B \ln(L/a_0)$. Therefore, the creation of the pair with its consequent dissociation (i.e., the separation of the dislocations for a distance $r \sim L$) changes the free energy of the system on the value

$$\Delta F \sim \varepsilon_0 \ln L - k_B T \ln L \,. \tag{3.14}$$

From Eq. (3.14) we see that at $T > T_c \sim \varepsilon_0/k_B$ the creation of free dislocations becomes energetically favorable, so that they have to arise abundantly and, consequently, to destroy the order in the system.

In a more careful consideration of the dislocation melting mechanism we have to take into account that the crystalline phase is to be characterized by two order parameters, namely by the *translational order parameter* and the *orientational order parameter*. The translational order assumes the periodicity in atomic arrangement, which is exhibited, for example, in the existence of clear X-ray diffraction picture. It is described by the correlation function

$$G_r^{(q)}(r) \propto \langle \rho^{(q)}(0)\rho^{(q)}(r)\rangle \,, \tag{3.15}$$

where $\rho^{(q)}(r) = \exp[i\vec{q}\vec{u}(r)]$ is the Fourier transform of the density function. The orientational order is connected with the constancy of the crystallographic directions, so that the vectors $\vec{e}_{kk'}$ which connect the atom k (occupying the lattice node with the coordinate $\vec{r_k}$) with the neighboring atoms k', have a tendency to be parallel or to create fixed angles with analogous vectors in other places of the crystal. The orientational order is described by the correlation function

$$G_\varphi(r) \propto \langle \psi(0)\psi(r)\rangle \,, \tag{3.16}$$

where

$$\psi(r) = \frac{1}{N}\sum_{k=1}^{N} \delta(r - |r_k|)\frac{1}{\nu_k}\sum_{k'=1}^{\nu_k} \exp(i\nu\theta_{kk'}) \,, \tag{3.17}$$

$\theta_{kk'}$ is the angle between some fixed axes and the vector $\vec{e}_{kk'}$, ν is the number of the nearest neighboring nodes in the lattice (for the triangular lattice $\nu = 6$), and ν_k is the number of the nearest neighbors for the given atom k. In the disordered phase the correlation functions decrease exponentially with r,

$$G_r(r) \propto \exp(-r/\xi_r), \;\; G_\varphi(r) \propto \exp(-r/\xi_\varphi), \;\; r \to \infty, \tag{3.18}$$

while in the floating phase they decrease according to a power law (recall that in the true crystalline phase the correlation functions oscillate and tend to a nonzero limit at $r \to \infty$).

At the ordinary melting of the three-dimensional crystal occurring through the first-order phase transition, the both orders disappear simultaneously. But in the floating phase of the two-dimensional crystal, these orders are destroyed by defects of different types: the translational order is destroyed by the creation of dislocation pairs with their consequent dissociation (an isolated dislocation is shown in Fig. 3.5a), while the orientational order is violated owing to the creation and dissociation of disclination pairs (an example of the isolated disclination is presented in Fig. 3.5b). As seen from Fig.3.5a, the isolated dislocation can be considered as the coupled pair of disclinations with opposite signs. The KT theory predicts that the melting of the floating phase with the increase of temperature may proceed in two stages: first (at $T = T_{cr}$) the translational order is destroyed, and then (at $T = T_{c\varphi}$) the orientational order is destroyed, and both phase transitions are continuous. Thus, within the temperature interval $T_{cr} < T < T_{c\varphi}$ it should exist a specific phase which has been named the *hexatic* (anisotropic liquid) phase, where the translational order is already destroyed and the correlation function $G_r(r)$ decays exponentially with r, while the orientational order still exists so that the function $G_\varphi(r)$ decays with r according to the power law.

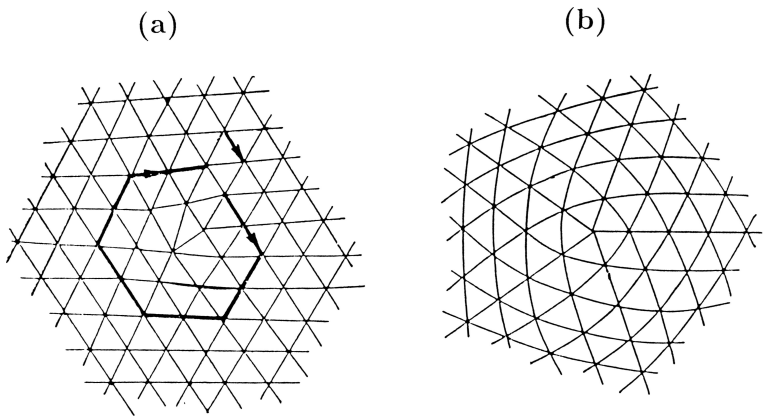

Figure 3.5: Topological defects in a two-dimensional crystal: (a) the isolated dislocation, and (b) the isolated disclination (from [87]).

3.2.3 Numerical algorithm

To check the predictions of the KT theory, Bedanov *et al.* [85] considered the two-dimensional system of electrons. The motion equations have the following form:

$$m\frac{d^2\vec{r}_k}{dt^2} = -\sum_{j\,(j\neq k)} \frac{\partial V(\vec{r_j} - \vec{r_k})}{\partial \vec{r_k}}\,, \tag{3.19}$$

where $V(\vec{r}) = e^2/\mid \vec{r}\mid$, m is the electron mass, and e is its charge.

First of all let us demonstrate the choice of units which are natural for the given problem. As the unit of length we take the value a_0 determined by the relationship

$$\pi a_0^2 = 1/n\,, \tag{3.20}$$

where n is the two-dimensional concentration of electrons, and as the unit of time we take $t_0 = a_0/v_0$, where v_0 is determined by the equation

$$\frac{1}{2}mv_0^2 = \frac{1}{2}k_BT\,. \tag{3.21}$$

Then it appears that the motion equations involve only a single dimensionless parameter

$$\Gamma = e^2\sqrt{\pi n}/k_BT\,. \tag{3.22}$$

In a result, we have to investigate the phase transition with the variation of the single parameter Γ, and then we will be able to describe the phase transition with variation of the temperature T (for a fixed n) as well as the phase transition with variation of the concentration n (for a fixed T).

In the computer experiment it is more convenient to vary the temperature T. This may be made in the following way: at each step of the solution of the motion equations we rescale the velocities of all atoms,

$$v_i \to v_i(1+\varepsilon)\,, \tag{3.23}$$

where ε is some small constant ($\varepsilon > 0$ for increasing of temperature and $\varepsilon < 0$ for its decreasing) which is to be chosen at "training" of the program. This procedure continues until the necessary value of T is reached (usually it takes $\sim 10^2$ time steps). Recall that according the definition, the value $\frac{1}{2}k_BT$ has to be equal to the average kinetic energy per one degree of freedom of the system.

Because the electrons repel each other, we expect that the floating phase has the structure of the triangular lattice. At the simulation [85] the system was imbedded into the rectangular box, and the periodic

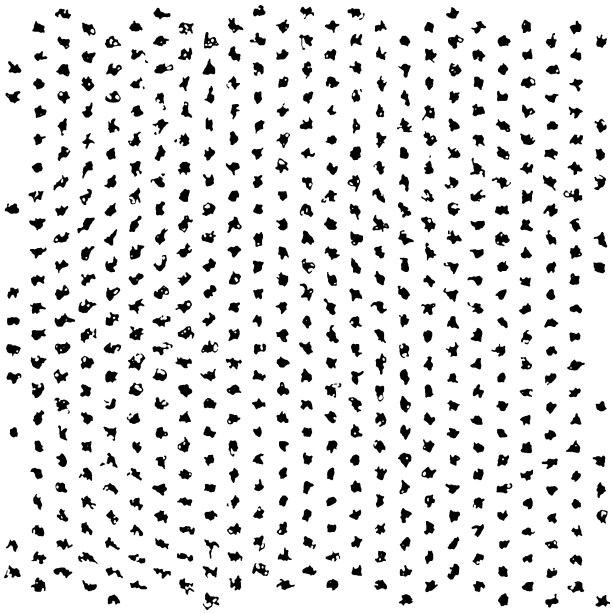

Figure 3.6: Two-dimensional crystal in the floating phase (after Bedanov *et al.* [85]).

boundary conditions were used. In order that the system is to be closed into tori without distortions of the triangular lattice, a certain condition for a possible number of atoms should be satisfied. In the work [85] the authors took $N = 2(12 \times 21) = 504$. For the initial configuration the atoms were placed to regular sites of the triangular lattice, while their velocities were set in a random way. Then the system was allowed to relax to equilibrium for a time $\sim 10^3$ time steps, and during the next $\sim 10^3$ time steps different system characteristics such as the correlation functions, were calculated. Then the temperature was changed and, after the relaxation to a new equilibrium state, the calculation was repeated.

3.2.4 Results of simulation

The simulation results of Bedanov *et al.* [85] are presented in Figs. 3.6 to 3.8. For example, Fig. 3.6 demonstrates a typical arrangement of atoms for the floating phase, and Fig. 3.7, a configuration for the disordered phase. The melting was observed in the interval $145 < \Gamma < 159$. This result is in agreement with the experimental results for electrons over the surface of liquid helium, where the melting was observed at $\Gamma = 137 \pm 15$ [84]. Snapshot pictures directly demonstrate the creation of topological defects and their evolution, i.e., the transformation of coupled dislocation pairs to free dislocations (disclination pairs) and then to free disclinations as the temperature increases. To reveal the hexatic phase, the correlation functions (3.15) and (3.16) were calculated, and then the long-distance tails of the correlation functions were approximated with the help of the mean-square method by simple functions, the exponential functions (3.18) or the power function $G_{r,\varphi}(r) \propto r^{-\nu}$. The phase transition should be followed by the change of the asymptotic behavior of the correlation functions so that the exponent ν or the correlation length ξ have to change. Simulation results presented in Fig. 3.8 show that it does exist the temperature interval where the orientational order still exists but the translational order is already destroyed. Thus, the MD modeling validates the KT theory of dislocation melting for the two-dimensional Coulomb system.

In conclusion let us briefly discuss other interesting results of computer simulation of the melting process. First, the two–stage topological melting has been observed for the two-dimensional Coulomb system only. When atoms in the two-dimensional system interact according to the dipole-dipole or LJ law, the MD simulation shows one–stage melting of the floating phase by the first-order phase transition [85]. If the two-dimensional system is subjected to an external periodic potential, phases with usual (long-ranged)

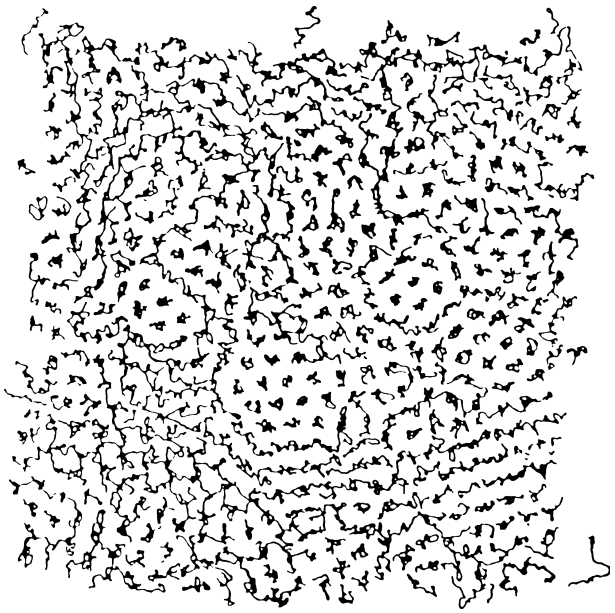

Figure 3.7: Two-dimensional liquid (after Bedanov *et al.* [85]).

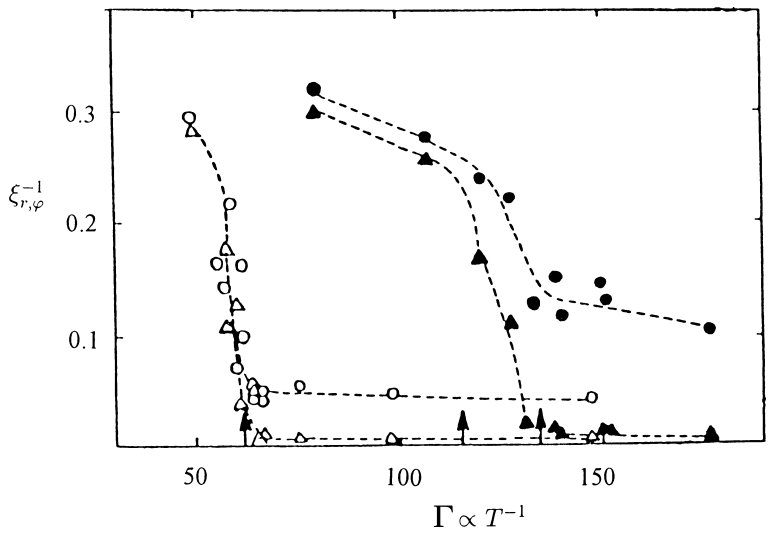

Figure 3.8: Inverse correlation lengthes ξ_r^{-1} (triangles) and ξ_φ^{-1} (circles) versus the dimensionless parameter Γ for the Coulomb (full) and dipole (open) two-dimensional systems (after Bedanov *et al.* [85]).

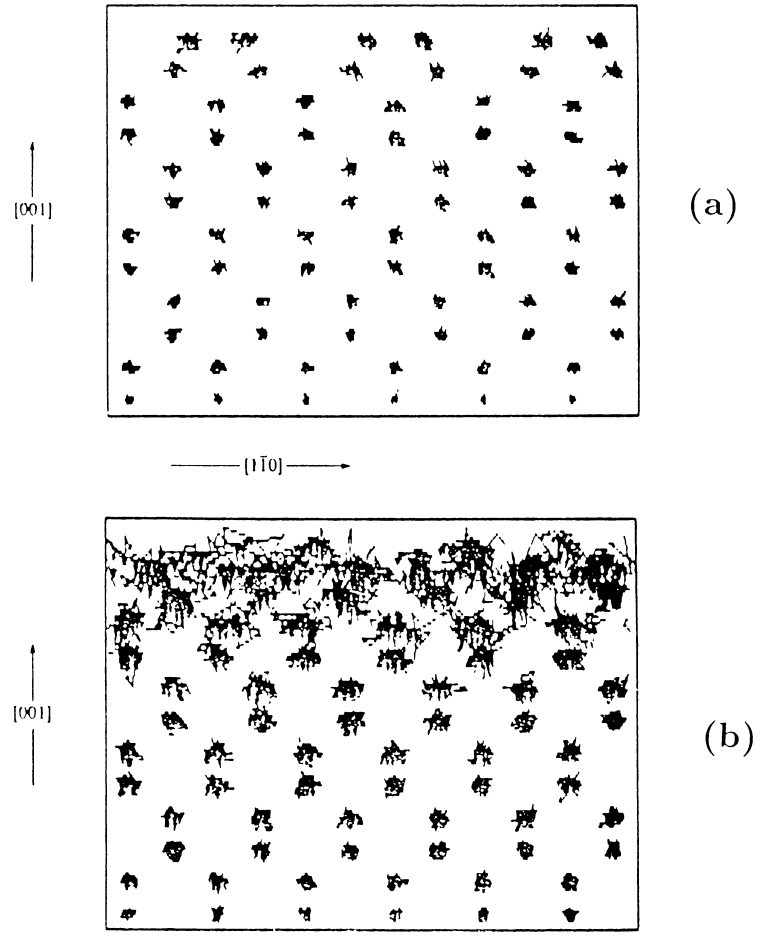

Figure 3.9: Surface melting. Atomic trajectories of the MD simulation for the **Si**(100) surface for (a) $T =$ 1003 K, and (b) $T = 1683$ K (after Abraham [88]).

crystalline order may exist in the system. When the amplitude of the external potential is high enough (as it is, for example, in films of gas atoms chemically adsorbed on a crystal surface), it is better to use the lattice-gas model which is equivalent to the Ising model with an external magnetic field. The Ising model is to be investigated by the Monte Carlo method described in Chapter 5. Otherwise, when the amplitude of the external potential is small, "solitonic" structures (see Chapter 6) may exist in the system, and the latter may be in the floating phase and to melt by the KT mechanism. A more detailed discussion of these questions can be found, e.g., in the monograph [87].

Finally, melting of a three-dimensional crystal exhibits interesting features too. According to the simplest melting scenario, an amplitude of vibrations of crystal atoms increases as the temperature rises, and when this amplitude achieves some critical value, i.e., when the amplitude becomes so high that the crystalline order is to be destroyed, the crystal melts (the so-called Lindeman criterion). However, a real crystal always has a surface. Because the surface atoms have a lower number of nearest neighbors than the bulk atoms, the amplitude of vibrations for the surface atoms should be larger than that for the bulk atoms. This reason makes the base for the assumption that the surface layers may melt before the bulk ones, i.e., that a think surface layer has to melt at a temperature T_c' which is lower than the bulk melting temperature T_c, and the width of the melted surface layer should increase to infinity as T approaches to T_c. The MD simulation [88] indeed directly demonstrates such surface melting (see Fig. 3.9). Laboratory experiments also validate that in some systems the melting process indeed starts from the surface.

3.3 Surface diffusion

3.3.1 Introductional remarks

The MD method is of a large value in investigations of kinetic and dynamic characteristics of systems in the thermal equilibrium state. In a general case these characteristics are expressed through different spacial and time correlation functions of the system. Correlation functions contain the complete information on the equilibrium state and, besides, they describe the linear response of the system to an infinitesimal external perturbation.

Many of correlation functions can be directly measured in spectroscopy experiments. All spectroscopy methods are based on the irradiation of the sample with some particles (photons, electrons, positrons, neutrons, ions, atoms, molecules) and the measuring the energy and angular distribution of scattered particles. Different spectroscopy methods help to get a rich information on the system under investigation. For example, the positions of peaks (more rigorously, hollows) in the energy distribution give the frequencies (energies) of elementary excitations (quasiparticles) in the system, while the width of the corresponding peak as well as its shape contains the information on the quasiparticle dynamics. However, an extraction of this information is often a rather difficult problem, and typically it is done with the help of comparison of experimental spectra with those predicted theoretically (e.g., see review papers [89] and [90]). For example, the line width in the vibrational spectrum is caused by damping of a given mode, i.e. by the rate η of the energy exchange between the mode and other elementary excitations of the system. The measuring of η is very important because the value of η determines in an essential degree the system dynamics. But additionally to η, two more mechanisms lead to contributions into the line width. There are the so-called *inhomogeneous broadening* η_x and *dephasing broadening* η_t. The first contribution η_x is coupled with a spacial disorder in the system, i.e., with the existence of defects of the crystalline structure, impurities, *etc.* The dephasing broadening η_t is caused by a temporal chaos of the system, i.e. by its nonintegrability. Note that in some systems the ratio η_t/η may achieve so large values as 10^2. The extraction of different contributions into the line width, is in need of the corresponding theory. In developing of this theory the MD simulation plays a significant role. As examples we can mention the works [69] and [91].

Different transport coefficients such as the diffusion coefficients are important characteristics of the system too. The diffusion plays a large role in many processes which are important from the practical viewpoint, for example, in crystal growth or in heterogeneous catalytic reactions, and also in microelectronics, *etc.* [92, 93]. There are three different diffusion coefficients: the self-diffusion coefficient, the collective diffusion coefficient (coupled with the system conductivity by Einstein's relationship), and the chemical diffusion coefficient (the latter appears in the first Fick law). All diffusion coefficients are defined through the velocity correlation functions. In particular, the self-diffusion coefficient is defined as

$$D_s = \frac{1}{\nu}\frac{1}{N}\sum_{i=1}^{N}\int_0^\infty dt\,\langle \vec{v}_i(0)\vec{v}_i(t)\rangle\,, \tag{3.24}$$

where ν is the system dimensionality ($\nu = 2$ for the surface diffusion). D_s describes the Brownian motion of a given particle,

$$\langle[\vec{r}_i(t) - \vec{r}_i(0)]^2\rangle \simeq 2\nu D_s t \;\text{ at }\; t \to \infty. \tag{3.25}$$

Below we consider one concrete example of investigation of the self-diffusion by the MD method.

3.3.2 Diffusion of Na atom adsorbed on the surface of Na crystal

De Lorenzi and Jacucci [94] have investigated by the MD method the self-diffusion of the **Na** atom adsorbed on the surface of the **Na** metal crystal. The b.c.c. **Na** crystal was modeled as a parallelepiped of N atoms. The number N depends on the surface under investigation. De Lorenzi and Jacucci [94] have modeled three crystallographic faces: the (112) face which has a furrowed (channelled) structure, was modeled by a slab of $N = 288$ atoms (12 planes, 3×8 atoms in each plane), the densely packed (compact) (110) face was modeled

by $N = 240$ atoms (8 planes, 30 atoms in each plane), and for the comparatively rare (loosely packed) (100) face it was taken $N = 250$ atoms (10 planes, 25 atoms in each plane). The periodic boundary conditions were imposed on the side planes of the parallelepiped, while the top surface and the bottom surface were left free (the "working" surfaces).

In order to model the crystal with the b.c.c. lattice, the authors had to use the realistic potential for the interaction between the **Na** atoms, because with the 12-6 LJ potential the sample relaxed to the f.c.c. lattice. The potential $v(r)$ was calculated preliminary with the help of the pseudopotential technique, and then it was tabulated with the step $0.00123\,a_0$ (a_0=4.30 Å is the lattice constant of the **Na** crystal). The real potential $v(r)$ differs from the LJ potential by that the former is nonmonotonic — it has a small maximum at the distance $r \sim 2a_0$ (the so-called Friedel oscillations).

The question on the initial conditions was solved in an original way. Namely, one atom was taken off from the bottom surface and placed on the top surface of the parallelepiped. Due to this trick two problems were solved simultaneously in a single simulation run, namely the problem of diffusion of the extra atom (adatom) on the top surface, and the problem of diffusion of the vacancy in the bottom surface.

The choice of the temperature T in computer simulation of such a type is determined by two factors. On the one hand, at low T the adatom will move too slowly, and for a realistic simulation time we will not achieve the asymptotic (3.25) and, therefore, will not have the necessary number of simulation runs in order to find a reliable value of D_s. On the other hand, at high T the probability of fluctuationally–driven generation of new adatom–vacancy pairs is high, and at $T \geq T_c'$ the surface will be melted at all. According to the work [94], the choice $T = 0.4\,T_c$ is the optimal one.

During the simulation the initial time interval was given for relaxation to the equilibrium state. During this initial time period the Newton motion equations were replaced by the Langevin equations (see the next Chapter 4) in order to accelerate the relaxation and to avoid the situation when the system catches itself in a metastable state. At the same time during the initial time interval the temperature was slowly increased from zero to the given value $T = 0.4\,T_c$.

The trajectories of the adatom motion obtained in the MD simulation, are presented in Figs. 3.10 to 3.12. On the furrowed (112) face (see Fig. 3.10) the diffusion is strongly anisotropic, the adatom moves along the channels only. Sometimes, however, the so-called "exchange" diffusion takes place, when the adatom A "push out" the surface atom B from its regular position in the row of surface atoms, and occupies the created free site. In the new configuration we again have the single adatom (B) which, however, is now in the nearest neighboring furrow. Thus, in this way an effective diffusion across the furrows takes place. The diffusion across the furrows is about ten times slower than the diffusion along the furrows. Comparison of the results for the (110) and (100) faces (see Figs. 3.11 and 3.12 respectively) shows that on a more "crumbly" (100) face the adatom moves essentially slower than on the smooth (110) face. The qualitative picture of surface diffusion obtained in the simulation, is in agreement with the experimental results obtained for the diffusion of atoms adsorbed on the metal surfaces with the help of the field ion microscopy (e.g., see [92]).

Note that the computer simulation [94] demonstrates also "long jumps" (i.e., the jumps of the adatom for distances of a several lattice constants at once without stops in the intermediate adsorption sites) which, however, were not observed in laboratory experiments. The existence of the long jumps in the simulation is probably caused by the too low value of the effective damping in the model used by the authors [94]. Indeed, they took into account a single friction mechanism (i.e., the mechanism of energy exchange between the moving adatom and the substrate) only, namely, the phonon friction which corresponds to the adatom energy losses owing to creation of vibrations of substrate atoms. But in the case of the metal substrate the "electron-hole" friction mechanism, i.e., the energy losses due to creation of electron-hole pairs in the substrate, plays an essential role as well [90]. With accounting of this additional friction, the long jumps have to become improbable.

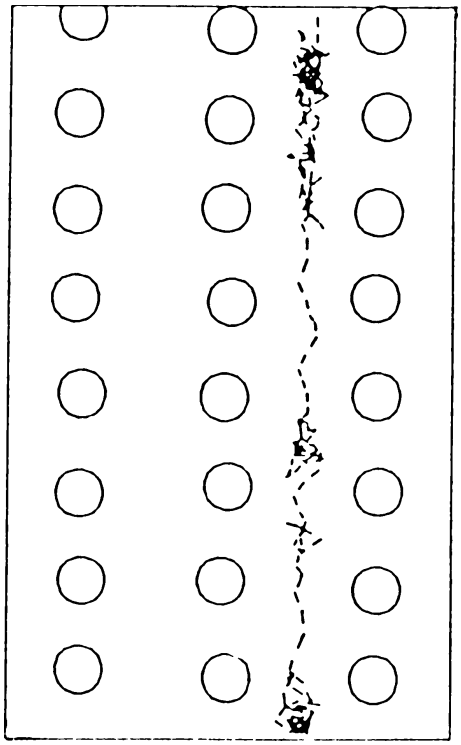

Figure 3.10: Adatom strolling along a channel on the b.c.c. (112) surface, in a 1000 time steps (∼20 ps) run as calculated by De Lorenzi and Jacucci [94].

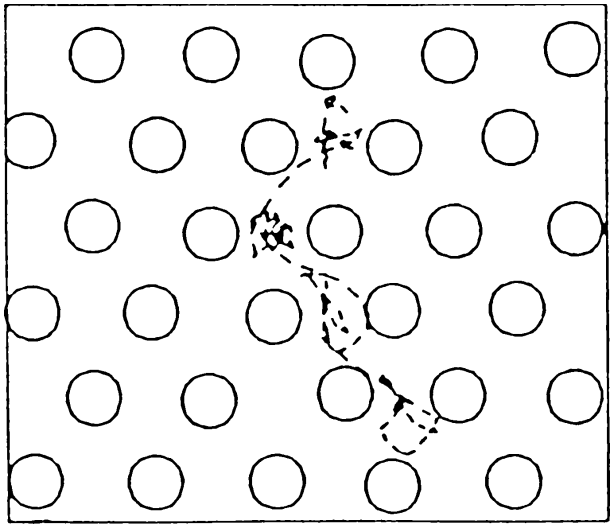

Figure 3.11: Adatom migration on the b.c.c. (110) surface. The positions of the surface atoms at the beginning of the interval are shown as open circles (from [94]).

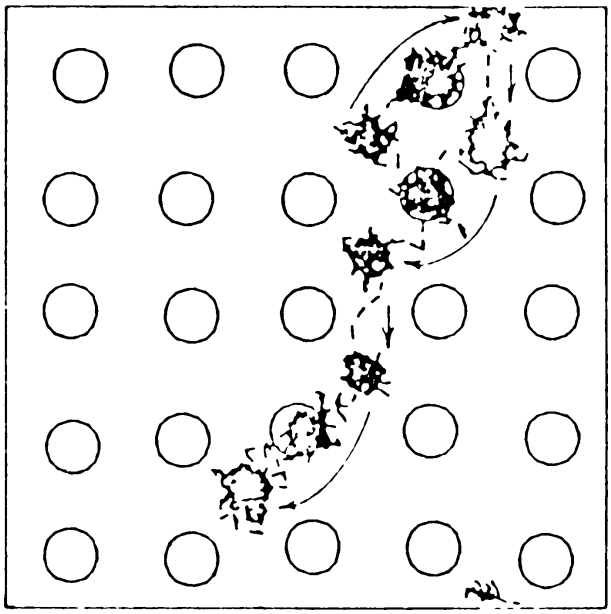

Figure 3.12: Adatom migration on the b.c.c. (100) surface, during a run of 100 ps. The positions of all the surface atoms at the beginning of the run are represented with circles; only the trajectories of the adatom originally present on the surface and of the surface atoms directly interested by its migration are represented. Short straight arrows indicate direct jumps of the adatom from one equilibrium site on the surface to another. Longer curved arrows indicate exchange events, in which the adatom changes identity (from [94]).

3.4 Epitaxial crystal growth

One more class of systems where the MD simulation plays a large role, is the class of systems far from the thermal equilibrium state. First, the MD simulation helps to consider the process of relaxation to equilibrium. Second, we can investigate the systems which are in a steady state, i.e., in the stationary but nonequilibrium state. Nonequilibrium state may be supported by a flux of energy or particles from outside. As an example, in the present section we consider the second from the mentioned processes, namely the epitaxial crystal growth. In this process certain atoms are placed into the defined sites of the substrate, so that in the result we obtain, e.g., a chip with given characteristics. But in order to control this process and to grow the crystal with the given parameters, we have to answer to the following questions: **(a)** What should be the velocity of arriving atoms? **(b)** At what angle should we direct the incoming beam? **(c)** What should be the substrate temperature?

Indeed, if the energy of the arriving atom and the substrate temperature are small, the atom which comes to the surface at a random site, will be "pinned" at this site because the atom will have no enough energy to migrate along the surface. In the result, the obtained structure will be disordered, and the growing crystal will be amorphous. On the other hand, the increase of the energy of the incoming beam will lead either to creation of defects in the substrate or to sputtering of the latter. Thus, in order to grow a more or less perfect crystal, we have to find an optimal regime at which the atom which comes to the surface, will migrate along the surface for some time looking for a "suitable" site with the lowest potential energy.

Note that the incoming atomic beam can be created by two methods. First, we can evaporate a drop of the given substance. In this way we get the atomic beam with thermal velocities. Second, we can sputter the source bombarding it by high-energy ions. In this case we obtain the atomic beam with high velocities. Thus, it arises the question, which method to choose. To answer it, Müller [95] has investigated with the help of the MD simulation the influence of the energy of the incoming atomic beam on the growth of thin films. He

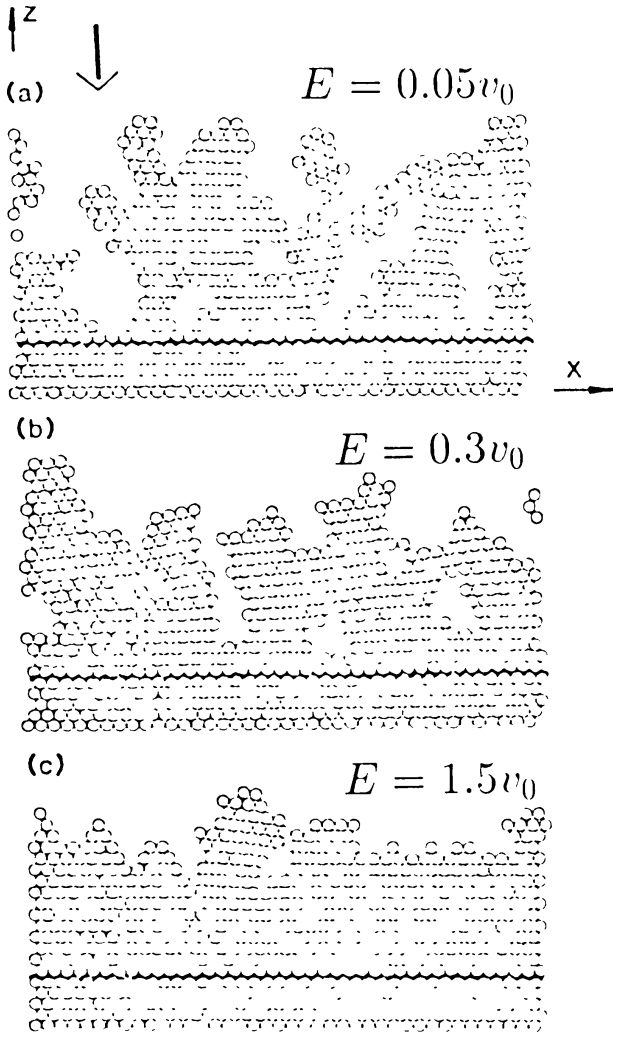

Figure 3.13: Microstructure of the film after deposition of 550 atoms as calculated by Müller [95] for different energies of arriving atoms: (a) $E = 0.05v_0$, (b) $E = 0.3v_0$, and (c) $E = 1.5v_0$ (v_0 is the parameter of the LJ potential).

considered a two-dimensional system with a "glass" shape, i.e., the system was restricted by the inequalities $0 < x < L$ and $0 < y < \infty$, where L is the "glass" width. It was taken $L = 40a_0$ so that along the x axis one may place up to 40 atoms in one row. On the side walls of the "glass" the periodic boundary conditions were imposed. For the interaction potential between the atoms it was taken the Lennard-Jones potential (3.5). As the initial configuration it was taken the perfect crystal consisting of five atomic layers placed on the "glass" bottom. The first (lowest) atomic layer was kept immobile (the fixed boundary conditions). As for the atoms from the second layer, it was assumed that, besides the forces from the neighboring atoms, the viscous friction force acts additionally on these atoms. In this way an energy has been taken away from the system. The motion of other atoms was subjected to the Newton motion equations.

On the substrate constructed in this way, from the "glass" top the atoms with a given velocity (energy) directed normal to the crystal surface, and a random x coordinate were injected one by one. The crystals growing in the simulation, are shown in Fig. 3.13 for different incident kinetic energies of the incoming beam. As seen, at small energy of the arriving atoms (Fig. 3.13a) the microstructure of the growing crystal is "crumbly", while at high energies the obtaining structure is more or less closely packed (Figs. 3.13b and 3.13c). In the next Fig. 3.14 the average density of the growing crystal versus the incident kinetic energy

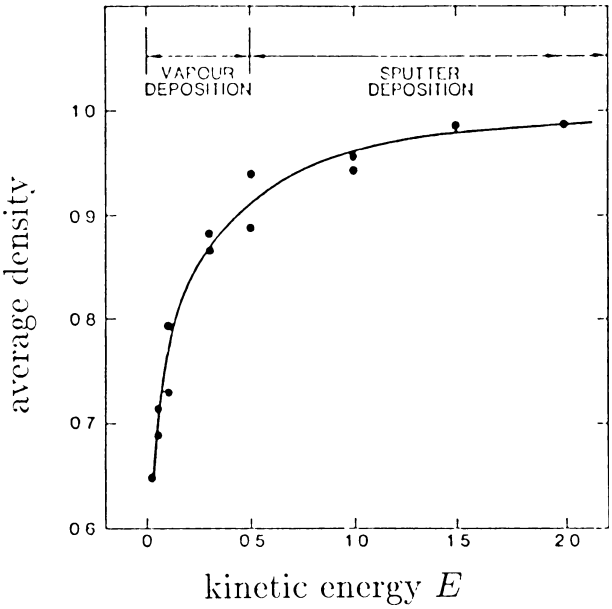

Figure 3.14: Mean density of the growing film versus the kinetic energy of incident atoms (after [95]).

of the incoming atomic beam is presented. As seen, the growth of "good" crystals is achieved only when one uses the beam with high atomic velocities, so the beam from the sputtering sources is to be preferable. Besides, it was found the average distance which the adsorbed atom should "run" in order to get a "true" position in the crystalline lattice — it occurs to be about $2a_0$.

3.5 PP, PM and P^{3}M methods

The "core" of the MD method is the solution of the Newton motion equations for as many particles as possible. It is clear, however, that with the increase of the number of particles N the computation time raises catastrophically. In the present section we briefly describe special methods which help to reduce the computer time.

In a general case, each particle interacts with all other particles in the system. Therefore for the "direct" accounting of all interactions the computation time will be proportional to $N(N-1) \approx N^2$. Moreover, when we use the periodic boundary conditions, the situation becomes completely catastrophical: although the number of motion equations is N only, each particle interacts with not only the $N-1$ other particles, but with their "images" (see Fig. 3.15) including its own "image", i.e., with an infinite number of particles! An elegant procedure to sum the forces acting on the given particle from the infinite number of neighboring particles, was proposed by Ewald [96] (see also [97] and, e.g., [98]). Practically, however, the "direct" accounting of all interactions is not used.

The situation simplifies when the interaction potential is short-ranged as, for example, for the 12-6 LJ potential (3.5) or the Morse potential (3.7). In this case we can neglect by interactions of particles separated by large enough distances. To do this, it is introduced the so-called "truncating radius" r^*, and it is taken into account the interaction of the given particle with those N^* particles only which occur within the circumference (or the sphere in the three-dimensional case) of the radius r^* with the center on the given particle. A value of r^* is typically of the order of $(2 \div 3)a_0$, where a_0 is the mean interparticle distance. Note that during the training of the program we should check either the increasing of r^* changes essentially the simulation results (in particular, this is very important in investigation of first-order phase transitions, see [99]). Owing to fluctuations of the local density of particles, the number of particles which are within the sphere of the fixed radius r^*, may vary (fluctuate) in time. That leads to some distortion of the computation

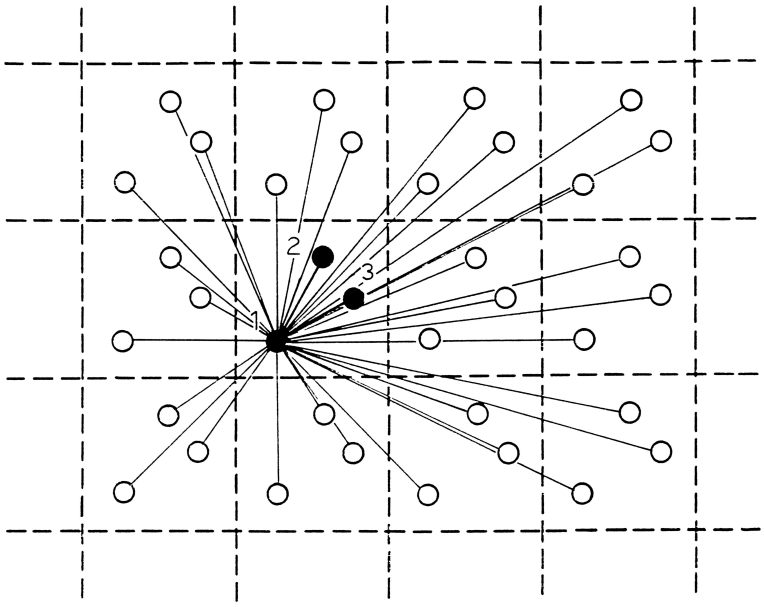

Figure 3.15: Atomic images emerged because of the periodic boundary conditions.

results, because the potential energy changes jump-like with N^* variation. In practical realization of this method, a special "coupled lists" program is used, where the numbers of the neighboring particles for the each given particle are stored [100]. The method described above, is known as the **PP** (particle-particle) method. Clearly that the computation time in the **PP** method is proportional to NN^*.

In the case of long-range (Coulomb) interparticle interactions the using of the **PP** method becomes unrealistic even for one-dimensional systems. In this case, however, we can use the **PM** (particle-mesh) method. The idea of the method consists in the following. Let us split the whole system into small cells (squares or cubes) with a side r_m and take $r_m \gg r_0$, so that each cell will contain several particles. Let us consider the forces acting on the given particle from the particles belonging to some sufficiently removed cell. Because these forces are approximately of the same value, their sum can be approximately substituted by a force acting on the given particle from some effective particle placed into the center of the removed cell. Practically this procedure reduces to calculation of the electrostatic potential $\varphi(r,t)$ (or the gravitation potential in astronomical problems) with the help of the solution of Poisson's equation $\Delta\varphi(r,t) = 4\pi\rho(r,t)$ over the mesh with the cell size r_m, and then to determination of the force acting on the k-th particle with the charge e as

$$\vec{F}_k(t) = -e\vec{\nabla}\varphi(r,t) \mid_{r=r_k} . \tag{3.26}$$

The algorithm of the solution is the following:
(a) The number of particles in each cell and, consequently, the cell charges $\rho_i(t)$ are calculated (the so-called procedure of "distribution of charge between the mesh cells").
(b) The discrete Poisson equation is solved. For example, in the one-dimensional case this equation has the following form:

$$\varphi_{i-1}(t) - 2\varphi_i(t) + \varphi_{i+1}(t) = 4\pi\rho_i(t) . \tag{3.27}$$

Different methods of solution of such equations are described, in particular, in the book [7].
(c) Using an interpolation procedure, the potential $\varphi(r,t)$ is found for the points where the particles are situated, and then the forces (3.26) are calculated.
(d) The motion equations are solved by the Runge-Kutta method.

The **PM** method is used mainly for solution of the plasma physics problems. The computation time in the **PM** method is proportional to N_M^ν, where $N_M = L/r_m$, and ν is the dimensionality of the system (ν=1,2

or 3). The deficiency of the **PM** method consists in large errors for the forces acting between short-ranged particles.

The quality features of both methods are accumulated in the hybrid $\mathbf{P^3M}$ method ($\mathbf{P^3M}$=**PP**+**PM**). In this method the force $F_k(t)$ acting on the particle, is split into two contributions,

$$F_k(t) = F_k^{(\mathrm{short})}(t) + F_k^{(\mathrm{long})}(t), \tag{3.28}$$

where $F_k^{(\mathrm{short})}(t)$ is the "short-range" contribution caused by interactions of closely situated particles, while $F_k^{(\mathrm{long})}(t)$ is the "long-range" force acting from the more removed particles. Then the force $F_k^{(\mathrm{short})}(t)$ is calculated exactly by the **PP** method, while the force $F_k^{(\mathrm{long})}(t)$ is found with the help of the **PM** method. Comparing with the direct use of the **PP** method, the computation time in the $\mathbf{P^3M}$ method is shorter in several orders of magnitude. However, the corresponding computer programs are too complicated [7].

3.6 Canonical ensemble

To conclude this Chapter devoted to the MD method, let us discuss in more detail the adaptation of the method to statistical physics problems. First, let us recall that in the MD method the time average (but not the average over ensembles) is calculated, so that the existence of a developed chaos in the system is the necessary condition which must be fulfilled. Second, note that in the standard variant of the MD method we keep fixed the number of atoms, the total system energy and its volume, i.e., we are considering the *microcanonical ensemble*. In thermodynamics, however, the *canonical ensemble* is more often used, when the temperature (the average kinetic energy of atoms) is fixed instead of the energy.

The simplest way (see [101]) of extension of the MD method to the canonical ensemble is to rescale the velocities v_l at each time step in such a way that the condition

$$\sum_{l=1}^{N} \frac{1}{2} m_l v_l^2 = \frac{1}{2} \nu N k_B T \tag{3.29}$$

is fulfilled. However, the phase trajectories calculated in this way are discontinuous.

Another method of introduction of the canonical ensemble (see [102, 103, 104]) is to impose the so-called *nonholonomic constraints*, i.e., to add to the system of Newtonian equations the additional condition

$$\sum_{l=1}^{N} \frac{1}{2} m_l v_l^2 = \frac{1}{2} (\nu N - 1) k_B T \,. \tag{3.30}$$

The imposing of the constraints leads to appearing of an additional term in the motion equations of the Hamiltonian system, so that the resulting set of equations takes the form

$$\left\{ \begin{array}{rcl} dq_l/dt &=& p_l/m_l \,, \\ dp_l/dt &=& -\partial V/\partial q_l - \alpha p_l \,, \end{array} \right. \tag{3.31}$$

where

$$\alpha(t) = -\frac{\sum_l (\partial V/\partial q_l)(p_l/m_l)}{\sum_l p_l^2/m_l} \,. \tag{3.32}$$

The method described above has the definiens that the temperature in this method is strictly fixed. Therefore, all fluctuations of the temperature are artificially suppressed, while in the canonical ensemble the temperature always fluctuates, and $\delta T \propto 1/\sqrt{N}$. In a result, the calculated distribution function has an error proportional to $N^{-1/2}$.

For calculation of the canonical ensemble by the MD method it may be used also the Nosé method [105], which assumes the introduction of an additional degree of freedom $s(t)$, $p_s(t)$ in such a way that the

distribution function, projected onto the physical system from the extended system of N particles plus the variable s, is to be depended exponentially on the system energy (as it just must be for the canonical ensemble). The additional variables $s(t)$ and $p_s(t)$ are called usually as "demons". These "demons" should act in such a way that to keep fixed the system temperature. Namely, let us introduce new variables $\tilde{q}$, $\tilde{p}$, $\tilde{s}$, $\tilde{p_s}$, and $\tilde{t}$ (called the "virtual variables") related with the real variables by the equations

$$\left\{\begin{array}{lcl} q_l &=& \tilde{q_l}\,, \\ s &=& \tilde{s}\,, \\ p_l &=& \tilde{p_l}/\tilde{s}\,, \\ p_s &=& \tilde{p_s}/\tilde{s}\,, \\ t &=& \int^{\tilde{t}} d\tilde{t}\,/\tilde{s}\,. \end{array}\right. \tag{3.33}$$

Let us postulate the system Hamiltonian for the extended system in the virtual variables in the form

$$\tilde{H}(\tilde{q},\tilde{p},\tilde{s},\tilde{p_s}) = \sum_{l=1}^{N} \frac{\tilde{p_l}^2}{2m_l\tilde{s}^2} + V(\tilde{q}) + \frac{\tilde{p_s}^2}{2Q} + gk_BT\ln\tilde{s}\,, \tag{3.34}$$

where the term $\tilde{p_s}^2/2Q$ corresponds to the "demon" kinetic energy, $gk_BT\ln\tilde{s}$ describes its potential energy, Q is the "demon" mass, and g is the integer constant chosen below.

Hamiltonian (3.34) leads to the following motion equations:

$$\left\{\begin{array}{lcll} d\tilde{q_l}/d\tilde{t} &=& \partial\tilde{H}/\partial\tilde{p_l} &=\ \tilde{p_l}/m_l\tilde{s}^2\,, \\ d\tilde{p_l}/d\tilde{t} &=& -\partial\tilde{H}/\partial\tilde{q_l} &=\ -\partial V/\partial\tilde{q_l}\,, \\ d\tilde{s}/d\tilde{t} &=& \partial\tilde{H}/\partial\tilde{p_s} &=\ \tilde{p_s}/Q\,, \\ d\tilde{p_s}/d\tilde{t} &=& -\partial\tilde{H}/\partial\tilde{s} &=\ \left(\sum_l \tilde{p_l}^2/m_l\tilde{s}^2 - gk_BT\right)/\tilde{s}\,. \end{array}\right. \tag{3.35}$$

Then the average over the microcanonical ensemble for the extended system with the "demons",

$$\langle A\rangle_{\tilde{t}} \propto \int d\tilde{\Gamma}_{\nu N+1}\, A(q,p,s,p_s)\,\delta(E-\tilde{H})\,, \tag{3.36}$$

calculated by the standard MD method for the set of equations (3.35), with accounting of the relationship

$$d\tilde{\Gamma}_{\nu N+1} = d\Gamma_{\nu N}\, ds\, dp_s\,, \tag{3.37}$$

for the case of $g = \nu N + 1$ may be rewritten in the form

$$\langle A\rangle_{\tilde{t}} \propto \int d\Gamma_{\nu N}\, A(p,q) e^{-H/k_BT}, \tag{3.38}$$

so that it just coincides with the average over the canonical ensemble for the original model.

Numerical integration of the system of equations (3.35) is too complicated, because the virtual time $\tilde{t}$ is nonlinearly coupled with the real time t. Therefore it is more convenient to return to the real variables, that leads to the following system of equations [105]:

$$\left\{\begin{array}{lcl} dq_l/dt &=& p_l/m_l\,, \\ dp_l/dt &=& -\partial V/\partial q_l - sp_sp_l/Q\,, \\ ds/dt &=& s^2p_l/Q\,, \\ dp_s/dt &=& \left(\sum_l p_s^2/m_l - gk_BT\right)/s - sp_s^2/Q\,, \end{array}\right. \tag{3.39}$$

which should be solved with the help of a computer. Note that in the latter case the average $\langle A\rangle$ is to be calculated with the relation

$$\langle A\rangle = \frac{\langle s^{-1}A(p,q)\rangle_t}{\langle s^{-1}\rangle_t}\,, \tag{3.40}$$

and that we have to set $g = 3N$ here. As for the "demon" mass Q, in its choosing we have to come of the facts that at $Q \to 0$ the "demon" separates from the system, while at $Q \to \infty$ its motion becomes too slow. Therefore we have to take an optimal value for Q such that the frequency of the $s(t)$ variation be of the same order as a characteristic system frequency.

Unfortunately, the Nosé method is imperfect too, because it has some additional conserved quantities (for example, the conservation of the total momentum) which disturb the distribution function. Therefore, for the correct simulation of the canonical ensemble we have to use the MD method based on the solution of Langevin equations.

Besides the NVE (microcanonical) and NVT (canonical) ensembles, there exist modifications of the MD method which allow to study the NPE ensemble or the NPT ensemble as well, where the system volume may change, while the pressure is kept constant (see [105, 1]).

Chapter 4

Stochastic Equations

4.1 Langevin equation

To explain the idea of introduction of the Langevin equation, let us consider the concrete example taken from surface physics: motion of an atom (adatom) adsorbed on a crystal surface. The role of the substrate in this case may be reduced to the following four factors.

(a) First of all, the surface atoms of the substrate produce some potential $V(\vec{r})$ which plays the role of the external potential for the adatom. At the same time we have to take into account that the coordinates of substrate atoms are not rigidly fixed, the surface atoms can "adjust" themselves to the adatom position by shifting in such a way that to lower the total potential energy of the system (the so-called "polaronic effect"). Such "adiabatic" potential $V(\vec{r})$ may be introduced in a rigorous way only when the adatom velocity is much lower than a characteristic velocity of substrate atoms. In the opposite case we have to take into account the retarded effects of the substrate response, and the potential $V(\vec{r})$ will depend on the whole previous trajectory of the adatom.

(b) When $T \neq 0$, the substrate atoms oscillate near their equilibrium positions. Owing to this, the potential produced by the substrate atoms is "smoothed", the corresponding "effective" potential $V_{\rm eff}(\vec{r})$ differs from the "bare" potential $V(\vec{r})$ by the Debye-Waller factor.

(c) There always exists an energy exchange between the adatom and the substrate, in a result of which the moving adatom may lose a part of its kinetic energy for excitation of various substrate degrees of freedom (phonons, plasmons, electron-hole pairs, *etc.*). This effect may be approximately described with the help of a friction force $\vec{F}^{f}(t)$ applying to the adatom. When the adatom velocity is small, this force may be approximated by the linear expression,

$$\vec{F}^{f}(t) = -\,m\,\eta\,\dot{\vec{r}}_a(t)\,, \tag{4.1}$$

where η is the viscous friction coefficient.

(d) Besides the energy flux from the adatom to the substrate, at $T \neq 0$ there always exists the reverse flux from the substrate to the adatom. Namely, the surface atoms of the substrate due to their thermal vibrations "push" the adatom in random directions. This effect may be modeled by introduction of some random force $\vec{F}^{r}(t)$ acting on the adatom. Evidently, the following equation should be satisfied,

$$\langle \vec{F}^{r}(t) \rangle = 0\,, \tag{4.2}$$

because in the opposite case we may redefine the "regular" force

$$\vec{F}(t) = -\left[\frac{\partial V_{\rm eff}(\vec{r})}{\partial \vec{r}}\right]_{\vec{r}=\vec{r}_a(t)}\,. \tag{4.3}$$

In the simplest approximation the random force is assumed to correspond to the Gaussian white noise, so that its correlator is taken to be

$$\langle F^{r}_{\alpha}(t)\,F^{r}_{\beta}(t')\rangle = A\,\delta_{\alpha\beta}\,\delta(t-t')\,, \tag{4.4}$$

where α and β are the Cartesian coordinates, and

$$A = 2\,\eta\, m\, k_B T. \tag{4.5}$$

The coefficient A is determined uniquely by the condition that the system must tend to thermal equilibrium as $t \to \infty$ (the so-called *fluctuation-dissipation theorem*, see [76]).

In a general case, the friction force may be non-Markovian as well as nonlinear with respect to particle's velocity. For example, the non-Markovian friction force $\vec{\mathcal{F}}^f(t)$ is introduced with the help of the following expression,

$$\mathcal{F}^f_\alpha(t) = -\,m \sum_{\beta=1}^{\nu} \int_0^\infty d\tau\, \mathcal{N}_{\alpha\beta}(\tau)\, \dot{r}_\beta(t-\tau)\,, \tag{4.6}$$

where $\mathcal{N}_{\alpha\beta}(\tau)$ is known as the response tensor, or the memory kernel. In this case the random force $\vec{\mathcal{F}}^r$ will be non-Gaussian: according to the fluctuation-dissipation theorem, its correlator should be equal to

$$\langle \mathcal{F}^r_\alpha(t)\, \mathcal{F}^r_\beta(t')\rangle = \mathcal{N}_{\alpha\beta}(|t-t'|)\, m\, k_B T\,. \tag{4.7}$$

Equations (4.6) and (4.7) reduce to Eqs. (4.1), (4.4) and (4.5), if we put

$$\mathcal{N}_{\alpha\beta}(t) = 2\,\eta\,\delta_{\alpha\beta}\,\delta(t)\,. \tag{4.8}$$

Note also that in a general case the friction coefficient depends on the adatom coordinate, $\eta \to \eta\,(\vec{r}_a(t))$, so that in the case of using of a constant friction coefficient we should assume that η is some effective friction averaged over the trajectory of adatom motion.

Of course, in order to calculate the functions $V_{\text{eff}}(\vec{r})$ and $\mathcal{N}_{\alpha\beta}(\tau)$, we must solve the full system of Newtonian motion equations for the adatom and all substrate atoms with the help of the MD method as it was described above in Sec. 3.3. Often, however, these functions may be properly modeled. For example, for the adatom problem described above, $V_{\text{eff}}(\vec{r})$ is a periodic function along the surface which is characterized by the symmetry determined by the corresponding crystallographic face of the substrate (if the surface is not reconstructed). The amplitude of the function $V_{\text{eff}}(\vec{r})$ determines the activation energy of diffusion of a single adatom, ε_a, while the shape of $V_{\text{eff}}(\vec{r})$ may be approximated if the frequencies ω_α of the adatom vibrations at the minimum of the substrate potential well are known. Besides, the value of the friction coefficient η is coupled with the width of vibrational line Γ. Thus, all necessary parameters in Eqs. (4.1) to (4.3) may be properly approximated if we know the parameters ε_a, ω_α and Γ from independent experiments.

So, instead of the consideration of motion of the whole system "adatom plus substrate", we may consider the motion of the singled out subsystem, for example, the adatom only as we have done above. Thus, the adatom motion is modeled by the following equations,

$$\begin{cases} m\,\ddot{\vec{r}}_a(t) + m\,\eta\,\dot{\vec{r}}_a(t) + \partial V_{\text{eff}}(\vec{r}_a)/\partial \vec{r}_a = \vec{F}^r(t)\,, \\ \langle \vec{F}^r(t)\rangle = 0\,, \\ \langle F^r_\alpha(t)\, F^r_\beta(t')\rangle = 2\,\eta\, m\, k_B T\,\delta_{\alpha\beta}\,\delta(t-t')\,. \end{cases} \tag{4.9}$$

The set of equations (4.9) is called the *Langevin equation*. It must be emphasized from the very beginning that this equation is the *stochastic equation*, not an ordinary differential equation (see below Sec. 4.4).

In some particular cases, Eq. (4.9) has an analytical solution. For example, when the external potential is absent, $V_{\text{eff}}(\vec{r}) \equiv 0$, the solution of Eq. (4.9) is

$$\vec{v}(t) = \vec{v}(t_0)\, e^{-\eta(t-t_0)} + \frac{1}{m}\int_{t_0}^{t} dt'\, e^{-\eta(t-t')}\,\vec{F}^r(t')\,, \tag{4.10}$$

where t_0 is the initial time moment.

The memory on the initial velocity is lost for a time $\sim \eta^{-1}$, therefore at $t - t_0 \gg \eta^{-1}$ the first term in the right-hand side of Eq. (4.10) may be neglected. Then, the velocity autocorrelation function may be written in the following way,

$$\begin{aligned} K_{\alpha\beta}(\tau) &\equiv \langle v_\alpha(t+\tau)\, v_\beta(t)\rangle \\ &= m^{-2} \int_{-\infty}^{t+\tau} dt' \int_{-\infty}^{t} dt''\, e^{-\eta(t+\tau-t'+t-t'')} \langle F^r_\alpha(t')\, F^r_\beta(t'')\rangle \\ &= \delta_{\alpha\beta}\,(2\,\eta\, k_B T/m) \left(\int_{-\infty}^{t} dt''\, e^{2\eta t''}\right) e^{-\eta(2t+\tau)} \\ &= \delta_{\alpha\beta}\,(k_B T/m)\, e^{-\eta\tau}, \end{aligned} \tag{4.11}$$

where we assumed $\tau > 0$. Thus, the correlation function decays exponentially, and it holds the relationship $\frac{1}{2} m \langle \vec{v}^2 \rangle = \nu \frac{1}{2} k_B T$.

From Eq. (4.11) we can easily find the mean-square displacement of the particle,

$$\begin{aligned} \langle [\vec{r}(t) - \vec{r}(t_0)]^2 \rangle &= \textstyle\sum_{\alpha=1}^{\nu} \langle \left(\int_{t_0}^{t} dt'\, v_\alpha(t') \right)^2 \rangle \\ &= \nu \int_{t_0}^{t} dt' \int_{t_0}^{t} dt'' \langle v_\alpha(t')\, v_\alpha(t'')\rangle \\ &= 2\nu \int_{t_0}^{t} dt' \int_{t_0}^{t'} dt'' K_{\alpha\alpha}(t' - t'') \\ &= 2\nu (k_B T/m\,\eta) t\,. \end{aligned} \tag{4.12}$$

Thus, the Langevin equation for the particle subjected to a viscous medium describes its Brownian motion with the diffusion coefficient

$$D = D_f \equiv k_B T/m\,\eta\,. \tag{4.13}$$

Besides, the solution of the Langevin equation for the harmonic potential $V_{\text{eff}}(\vec{r}) \propto \vec{r}^2$ is known too. Namely, for the one-dimensional equation

$$\ddot{x}(t) + \eta\, \dot{x}(t) + \omega_0^2\, x(t) = \frac{1}{m}\, F^r(t) \tag{4.14}$$

for times $t \gg \eta^{-1}$, when a memory on the initial state is lost, the solution has the following form,

$$x(t) = \int^{t} dt'\, R(t - t')\, F^r(t')\,, \tag{4.15}$$

where

$$R(t) = \frac{1}{m\,\omega_*} \exp\left(-\frac{1}{2}\,\eta\, t\right) \sin(\omega_* t) \tag{4.16}$$

and $\omega_*^2 = \omega_0^2 - \eta^2/4$. If $\omega_0 < \eta/2$, we have to make the substitutions $\omega_* = i\,\omega'_*$ and $\sin \omega_* t = i \sinh \omega'_* t$ in Eq. (4.16), where $\omega'_* = \eta^2/4 - \omega_0^2$. With the help of Eqs. (4.15) and (4.16) we may calculate the correlation functions of coordinates or velocities. The Fourier transform of the correlation function gives the following expression:

$$K(\omega) = \int_{-\infty}^{\infty} dt\, e^{-i\omega t} K(t) \propto \frac{(\omega_0^2 - \omega^2) - i\eta\,\omega}{(\omega_0^2 - \omega^2)^2 + (\eta\,\omega)^2}\,. \tag{4.17}$$

Thus, Eq. (4.14) leads to the Lorentz shape of the vibrational line with the half-width $\Gamma = \eta$ (the so-called "full width at half maximum", or FWHM). Recall that in real physical systems the inhomogeneous broadening (leading to the Gaussian line shape) as well as the broadening owing to dephasing effects are presented additionally.

In a general case the Langevin equation (4.9) may be solved by numerical methods.

4.2 Numerical solution of the Langevin equation

In order to solve the Langevin equation with a computer, first of all we need of a subroutine that produces the normal random numbers, i.e. the sequence $\{\xi_n\}$ of random numbers with Gaussian distribution,

$$\mathrm{Prob}\,(\xi_n) = (2\pi)^{-1/2} \exp(-\xi_n^2/2)\,, \tag{4.18}$$

which also satisfy the conditions

$$\langle \xi_n \rangle = 0 \quad \text{and} \quad \langle \xi_n\, \xi_m \rangle = \delta_{nm}\,. \tag{4.19}$$

Such random numbers may be obtained from the standard random numbers $\{\gamma_m\}$, i.e., the random numbers uniformly distributed in the interval (0,1), with the help of the formula

$$\xi_n = \left(\sum_{m=1}^{12} \gamma_m \right) - 6\,, \tag{4.20}$$

i.e. we obtain one normal random number from twelve standard random numbers (see details below in Sec. 5.1.2).

The further integration of the stochastic equation is carried out practically in the same way as it does in the case of an ordinary differential equation [106]. The only difference is that we have to choose a Runge-Kutta scheme with the constant time step Δt, and then at each time step $t_n = n\,\Delta t$ we should add the random force

$$F_n^r = B\,\xi_n \tag{4.21}$$

to the sum of the regular force $F_n = F(t_n)$ and the friction force $F_n^f = F^f(t_n)$. The coefficient B in Eq. (4.21) is defined by the integration of Eq. (4.4),

$$\int_{-\infty}^{\infty} dt\, \langle F^r(t)\, F^r(t') \rangle = 2\,\eta\, m\, k_B T\,; \tag{4.22}$$

with accounting of Eqs. (4.19) and (4.21) that gives the expression

$$B = (2\eta m k_B T/\,\Delta t)^{1/2}. \tag{4.23}$$

The appearance of the time step Δt in the expression for the random force is caused by the fact that the Langevin equation is the stochastic one. Emphasize that Δt in Eq. (4.23) corresponds to the time between the consequential calls of the subroutine which calculates the forces. Therefore, if we use the forth-order RK method (which calls the "force subroutine" 4 times in one step) with the time step τ, then we must take $\Delta t = \tau/4$ in Eq. (4.23). Note also that the described method works as well if the damping coefficient η depends on the coordinates of moving particles. However, when η depends on the velocities, the problem becomes more involved [193].

To control the work of the computer program, one may test, for example, the condition of the equally distribution of the kinetic energy among all degrees of freedom in the stationary regime.

In a result of the numerical solution of the Langevin equation we obtain one particular realization of the temporal trajectory of system motion. When we are studying the stationary state of the system, we have to remove some initial part of the trajectory, and then to use the remainder for calculation of different averages similarly as it was done in the MD method. Note that the equilibrium solution of the Langevin equation does not depend on the friction coefficient η, thus the value of η may be chosen appropriately in order to accelerate the approaching to equilibrium (usually $\eta \sim \omega_0$ is taken). In investigation of nonstationary (e.g., relaxation) processes, the whole trajectory is used, and sometimes we have to make the averaging over the initial states of the system. Besides, to get reliable results, it is always useful to carry out also the averaging over an ensemble of different realizations of the trajectory, i.e. to average over the set of trajectories calculated with different sequences of random numbers.

4.3 Generalized Langevin equation

In a general case, the idea of simulation with the help of stochastic equations consists in the separation of the whole system into two parts, the "main" one and the "thermostat". A part of the whole system that is the most interesting one for the problem under investigation, is singled out as the main part, and its evolution is modeled with the help of rigorous solution of motion equations. The remainder of the whole system is related to the thermostat, and it is assumed that the influence of the thermostat on the main part reduces to creation of some external potential, and to arising of the energy exchange (dissipation of the kinetic energy) and the corresponding random force. Of course, we have to be sure that a "developed chaos" really exist in the remainder which we have modeled as the thermostat. A rigorous introducing of the Langevin equation is possible only in the case when there exists a hierarchy of times in the system. This means that a characteristic time τ_A of motion of the singled out subsystem (e.g., the adatom in the example described above) must essentially exceed the characteristic time τ_S of the thermostat (e.g., the substrate). As the time τ_A of the main subsystem we may take, for example, the period of its characteristic oscillations, while as the time τ_S of the thermostat, the time of decay of temporal fluctuations. In the case $\tau_A \gg \tau_S$ the fluctuations in the thermostat are fast, and the action of the thermostat on the main subsystem may be approximately described by the δ-correlated random force. Otherwise, in the case of $\tau_A < \tau_S$, the employing of Gaussian random force is not rigorous, and we have to use Eqs. (4.6) to (4.8) corresponded to non-Markovian dynamics.

The simplest way to carry out the described program is to assume that the motion of atoms situated at the boundary between the main subsystem and the thermostat, undergoes the Langevin motion equations instead of Newtonian ones. A more rigorous approach developed by Adelman and Doll [107] and Tully [108], is based on the Zwanzig-Mori-Kubo technique [109, 110, 111]. Namely, let the whole system consists of N atoms, and N_A atoms belong to the singled out subsystem **A**, while N_S atoms, to the thermostat **S**. Let $\vec{x}$ denote the array (vector) of coordinates of all atoms, $\vec{x}_A$ denote the coordinates of the main subsystem, and $\vec{x}_S$ denote the thermostat coordinates. Introduce also the projection operators $\mathcal{P}$ and $\mathcal{Q}$ which separate the atoms of the main subsystem and the thermostat respectively, i.e. $\vec{x}_A = \mathcal{P}\,\vec{x}$ and $\vec{x}_S = \mathcal{Q}\,\vec{x}$. To reduce notations, let us introduce the displacements $\vec{u}$ of atoms from the equilibrium positions $\vec{x}^{(0)}$, the latter correspond to the absolute minimum of the potential energy $V(\vec{x})$ of the system. Then, the rigorous motion equations may be written in the following form,

$$\ddot{\vec{u}} = \vec{\mathcal{F}}(\vec{x}) = -\overleftrightarrow{\Omega}^2(\vec{x}^{(0)})\,\vec{u} + \delta\vec{\mathcal{F}}(\vec{x})\,, \tag{4.24}$$

where $\vec{\mathcal{F}} = -\partial V/\partial\vec{x}$, and $\overleftrightarrow{\Omega}^2$ is the dynamical matrix which describes the system evolution in the harmonic approximation. If we neglect all nonlinear effects in the thermostat, Eq. (4.24) may be rewritten in the following form,

$$\ddot{\vec{u}}_A = -\overleftrightarrow{\Omega}^2_A\,\vec{u}_A - \overleftrightarrow{\Omega}^2_{AS}\,\vec{u}_S + \delta\vec{\mathcal{F}}_A(\vec{x}_A, \vec{x}_S^{(0)})\,, \tag{4.25}$$

$$\ddot{\vec{u}}_S = -\overleftrightarrow{\Omega}^2_S\,\vec{u}_S - \overleftrightarrow{\Omega}^2_{SA}\,\vec{u}_A\,, \tag{4.26}$$

where $\overleftrightarrow{\Omega}^2_A = \mathcal{P}\,\overleftrightarrow{\Omega}^2\,\mathcal{P}$, $\overleftrightarrow{\Omega}^2_S = \mathcal{Q}\,\overleftrightarrow{\Omega}^2\,\mathcal{Q}$, $\overleftrightarrow{\Omega}^2_{AS} = \mathcal{P}\,\overleftrightarrow{\Omega}^2\,\mathcal{Q}$, and $\overleftrightarrow{\Omega}^2_{SA} = \mathcal{Q}\,\overleftrightarrow{\Omega}^2\,\mathcal{P}$.

A formal solution of the linear equation (4.26) is

$$\begin{aligned} u_S(t) = {} & \cos(\Omega_S t)\,u_S(0) + \Omega_S^{-1}\sin(\Omega_S t)\,\dot{u}_S(0) \\ & -\Omega_S^2\,\Omega_{SA}^2\,u_A(t) + \Omega_S^{-2}\cos(\Omega_S t)\,\Omega_{SA}^2\,u_A(0) \\ & + \int^t dt'\,\Omega_S^{-2}\cos[\Omega_S(t-t')]\,\Omega_{SA}^2\,\dot{u}_A(t')\,. \end{aligned} \tag{4.27}$$

The substitution of Eq. (4.27) into Eq. (4.25) gives the *generalized Langevin equation* (GLE) which describes the evolution of N_A atoms of the singled out subsystem only:

$$\ddot{\vec{u}}_A(t) = -\overleftrightarrow{\Omega}^2_{\text{eff}}\,\vec{u}_A(t) - \overleftrightarrow{\mathcal{N}}(t)\,\vec{u}_A(0) - \int^t dt'\;\overleftrightarrow{\mathcal{N}}(t-t')\,\dot{\vec{u}}_A(t)$$

$$+\vec{\mathcal{F}}^r(t)+\delta\vec{\mathcal{F}}_A\left(\vec{x}_A(t),\vec{x}_S^{(0)}\right), \tag{4.28}$$

where

$$\overleftrightarrow{\Omega}^2_{\rm eff}=\overleftrightarrow{\Omega}^2_A-\overleftrightarrow{\mathcal{N}}(0)\;=\overleftrightarrow{\Omega}^2_A-\overleftrightarrow{\Omega}^2_{AS}\overleftrightarrow{\Omega}^{-2}_S\overleftrightarrow{\Omega}^2_{SA}\,, \tag{4.29}$$

$$\overleftrightarrow{\mathcal{N}}(t)\;=\overleftrightarrow{\Omega}^2_{AS}\cos\left(\overleftrightarrow{\Omega}_S t\right)\overleftrightarrow{\Omega}^{-2}_S\overleftrightarrow{\Omega}^2_{SA}\,, \tag{4.30}$$

$$\vec{\mathcal{F}}^r(t)=-\overleftrightarrow{\Omega}^2_{AS}\cos\left(\overleftrightarrow{\Omega}_S t\right)\vec{u}_S(0)-\overleftrightarrow{\Omega}^2_{AS}\sin\left(\overleftrightarrow{\Omega}_S t\right)\dot{\vec{u}}_S(0)\,, \tag{4.31}$$

and, as it should be according to the fluctuation-dissipation theorem,

$$\langle\vec{\mathcal{F}}^r(t)\,\vec{\mathcal{F}}^r(0)\rangle=k_BT\,\overleftrightarrow{\mathcal{N}}(t)\,. \tag{4.32}$$

The generalized Langevin equation (4.28) is the matrix equation for N_A atoms only. But this equation includes the response matrix $\overleftrightarrow{\mathcal{N}}(t)$, and to calculate the latter, the full system of N motion equations must be solved. The idea of simulation with the help of the GLE consists in the following trick: let us take artificially for $\overleftrightarrow{\mathcal{N}}(t)$ some simple expression which (we hope) describes the thermostat dynamics quite well. For example, we may assume that the thermostat consists of a set of Brownian harmonic oscillators, that leads analogously to Eqs. (4.14–4.16) to the approximation

$$\overleftrightarrow{\mathcal{N}}(t)=\overleftrightarrow{\mathcal{N}}_0\exp\left(-\frac{1}{2}\,\overleftrightarrow{\eta}\,t\right)\left[\cos\left(\overleftrightarrow{\omega}_* t\right)+\frac{1}{2}\,\overleftrightarrow{\eta}\,\overleftrightarrow{\omega}^{-1}_*\sin\left(\overleftrightarrow{\omega}_* t\right)\right]. \tag{4.33}$$

At the same time, $\overleftrightarrow{\mathcal{N}}(t)$ satisfies the differential equation

$$\ddot{\overleftrightarrow{\mathcal{N}}}(t)=-\overleftrightarrow{\omega}^2_0\overleftrightarrow{\mathcal{N}}(t)-\overleftrightarrow{\eta}\,\dot{\overleftrightarrow{\mathcal{N}}}(t)\,, \tag{4.34}$$

where

$$\overleftrightarrow{\omega}^2_0=\overleftrightarrow{\omega}^2_*+\frac{1}{4}\,\overleftrightarrow{\eta}^2\,. \tag{4.35}$$

The approximation (4.33) depends on three matrices, $\overleftrightarrow{\mathcal{N}}_0$, $\overleftrightarrow{\eta}$, and $\overleftrightarrow{\omega}_*$, each of them has the dimension $\nu N_A\times\nu N_A$. They should be chosen in such a way that, for example, to describe correctly the vibrational spectrum of the thermostat.

The employing of the non-Markovian random force (4.32) in computations is not convenient. Using Eq. (4.34), the set of equations (4.28), (4.29), (4.32) and (4.33) may be reduced to the set of Markovian stochastic equations but, however, already for $2N_A$ particles (N_A atoms of the main subsystem and N_A "fictitious" atoms with the coordinates $\vec{s}(t)$):

$$\ddot{\vec{u}}_A(t)=\vec{\mathcal{F}}_A(\vec{x}_A(t),\vec{x}_S^{(0)})+\overleftrightarrow{\mathcal{N}}^{1/2}_0\overleftrightarrow{\omega}_0\,\vec{s}(t)\,, \tag{4.36}$$

$$\ddot{\vec{s}}(t)=\overleftrightarrow{\mathcal{N}}^{1/2}_0\overleftrightarrow{\omega}_0\,\vec{u}_A(t)-\overleftrightarrow{\omega}^2_0\,\vec{s}(t)-\overleftrightarrow{\eta}\,\dot{\vec{s}}(t)+\vec{F}^r(t)\,, \tag{4.37}$$

where the random force $\vec{F}^r(t)$ is now Markovian,

$$\langle\vec{F}^r(0)\,\vec{F}^r(t)\rangle=2\,k_BT\,\overleftrightarrow{\eta}\;\delta(t)\,. \tag{4.38}$$

Methods of numerical solution of the system of stochastic equations (4.36) to (4.38) are described, e.g., in the works [112] and [113].

Note also that the separation of the whole system into the main part and the thermostat may be changed during the system evolution. For example, if we included into the main part the adatom together with a few nearest neighboring atoms of the substrate, then during the adatom diffusion on the surface, its surrounding will change, i.e. some surface atoms will "pass out" into the thermostat, while the others that occur to be in the nearest surrounding of the adatom at the given instant, should "pass into" from the thermostat to the main subsystem. The method of accounting of such an effect was described by Tully [108].

4.4 Fokker-Planck-Kramers equation

Emphasize ones more that the Langevin equation is the stochastic one, it differs from an ordinary differential equation by the existence of random force. The stochastic equation may be reduced to a differential equation, but the latter will be the differential equation with partial derivatives called the *Fokker-Planck* (FP) *equation*. In the present section we deduce the FP equation from the Langevin one (for a more detailed description of the theory of stochastic processes see, for example, the monograph [114]).

Temporarily, let us turn from the continuous time t to the discrete time variable $t_n = n\,\Delta t$. Then, the stochastic equation for some measurable quantity $q(t)$ in a general case may be written in the following way,

$$\Delta q(t_n) \equiv q(t_n) - q(t_{n-1}) = K\left(q(t)\right)\,\Delta t + G\left(q(t)\right)\,\Delta w(t_n)\,, \tag{4.39}$$

where the first term in the right-hand side of Eq. (4.39) called the *drift term* describes the action of the regular force, and the second term called the *diffusion term*, the action of the random force. According to its definition, the quantity $\Delta w(t_n)$ must satisfy the following two conditions,

$$\begin{cases} \langle \Delta w(t_n) \rangle = 0\,, \\ \langle \Delta w(t_n)\,\Delta w(t_m) \rangle = \delta_{nm}\,\Delta t\,, \end{cases} \tag{4.40}$$

where δ_{nm} is Kronecker's delta symbol.

Note, however, that the functions K and G in Eq. (4.39) depend on the coordinates which in turn are the functions of time. Therefore, the question arises, at which time instant t they should be calculated. There may be two variants:

(a) In the so-called *Ito's calculus*, one takes $t = t_{n-1}$. In this case the calculation of different averages is significantly simplified, because, for example,

$$\langle G\left(q(t_{n-1})\right)\,\Delta w(t_n) \rangle = \langle G\left(q(t)\right) \rangle\,\langle \Delta w(t) \rangle = 0\,. \tag{4.41}$$

(b) In the *Stratonovich calculus* the time instant t is taken at the middle between the n-th and $(n-1)$-th steps, $t = \frac{1}{2}(t_n + t_{n-1})$. This calculus is convenient by that differentiation and integration operations are carried out with the same rules as in the ordinary differential and integral calculi.

Returning to the continuous time, $\Delta t \to dt$, Eqs. (4.39) and (4.40) are rewritten in the following form,

$$\begin{cases} dq(t) = K(q)\,dt + G(q)\,dw\,, & \text{(a)} \\ \langle dw \rangle = 0\,, & \text{(b)} \\ \langle dw(t)\,dw(t) \rangle = dt\,. & \text{(c)} \end{cases} \tag{4.42}$$

From Eq. (4.42c) one can see that the stochastic equation differs from a differential one by existence of the diffusion term such that $\langle dw \rangle \sim \sqrt{dt}$.

When we solved the Langevin equation, we calculated the trajectory of system motion which then was used for calculation of temporal averages. For stochastic processes, however, it is more natural to use a distribution function $f(q,t|q_0,t_0)$ which is the conditional probability that the system is in the state q at the instant t, if at the previous time moment t_0 it was in the state q_0. Let us deduce now an equation for the distribution function $f(q,t|q_0,t_0)$. For this, we take an arbitrary function $u(q)$ which depends on the coordinates, and write down its average (over the ensemble) quantity $\langle u \rangle$ at time t as

$$\langle u \rangle = \int dq\,u(q)\,f(q,t|q_0,t_0)\,. \tag{4.43}$$

Then the derivative of $u(q)$ with respect to time is equal to

$$\frac{d}{dt}\langle u \rangle = \int dq\,u(q)\,\frac{\partial f(q,t|q_0,t_0)}{\partial t}\,. \tag{4.44}$$

On the other hand, the differential of the function $u(q)$ with an accuracy up to dt, with the help of Eq. (4.42a) may be written in the form

$$\begin{aligned} du &= (\partial u/\partial q)\, dq + \tfrac{1}{2}(\partial^2 u/\partial q^2)\, dq\, dq \\ &= (\partial u/\partial q)\, [K(q)\, dt + G(q)\, dw(t)] + \tfrac{1}{2}(\partial^2 u/\partial q^2)\, G^2(q)\, dw(t)\, dw(t)\,, \end{aligned} \tag{4.45}$$

where the appearance of the second derivatives is caused by the fact that in the stochastic equation $\langle dw\rangle \sim \sqrt{dt}$.

Averaging Eq. (4.45) over time with the help of the Ito calculus, dividing both sides of the equation by dt, and taking into account Eqs. (4.42b,c), we obtain

$$\frac{\langle du\rangle}{dt} = \left\langle \frac{\partial u}{\partial q}\, K(q) \right\rangle + \frac{1}{2}\left\langle \frac{\partial^2 u}{\partial q^2}\, G^2(q) \right\rangle . \tag{4.46}$$

The first term in the right-hand side of Eq. (4.46) may be rewritten as

$$\begin{aligned} \left\langle \frac{\partial u}{\partial q}\, K(q) \right\rangle &= \int dq\, f(q,t|q_0,t_0)\, K(q)\, \frac{\partial u}{\partial q} \\ &= -\int dq\, u(q)\, \frac{\partial}{\partial q}[K(q)\, f(q,t|q_0,t_0)]\,, \end{aligned} \tag{4.47}$$

where we also have carried out the integration by parts. Analogously, the second term in the right-hand side of Eq. (4.46) may be transformed, if we carry out the integration by parts two times. Then, comparing the obtained expression for $\langle du\rangle/dt$ with Eq. (4.44) and taking into account that the function $u(q)$ is an arbitrary one, we obtain that the distribution function $f(q,t|\cdots)$ must satisfy the following equation,

$$\frac{\partial f(q,t|\cdots)}{\partial t} = -\frac{\partial}{\partial q}[K(q)\, f(q,t|\cdots)] + \frac{1}{2}\frac{\partial^2}{\partial q\, \partial q}[G^2(q)\, f(q,t|\cdots)]\,. \tag{4.48}$$

Eq. (4.48) is called the *Fokker-Planck* (FP) *equation*.

The described deduction may easily be generalized for a system of many variables, $q \to \vec{q} \equiv \{q_l\}$. The stochastic equation in this case takes the following form,

$$\begin{cases} dq_l(t) = K_l(q)\, dt + \sum_m G_{lm}(q)\, dw_m(t)\,, \\ \langle dw_m(t)\rangle = 0\,, \\ \langle dw_l(t)\, dw_m(t)\rangle = \delta_{lm}\, dt\,, \end{cases} \tag{4.49}$$

and the Fokker-Planck equation, the form

$$\begin{aligned} \frac{\partial f(\vec{q},t|\cdots)}{\partial t} &= -\sum_l \frac{\partial}{\partial q_l}[K_l(\vec{q})\, f(\vec{q},t|\cdots)] \\ &+ \frac{1}{2}\sum_{kl} \frac{\partial^2}{\partial q_k \partial q_l}\left(\sum_m G_{km}\, G_{lm}\, f(\vec{q},t|\cdots)\right) . \end{aligned} \tag{4.50}$$

As an example, let us consider the Langevin equation (4.9). For a single degree of freedom $x(t)$, Eq. (4.9) may be rewritten as the set of equations

$$\begin{cases} \dot{x} = v\,, \\ \dot{v} = -\eta\, v - V'_{\rm eff}(x)/m + F^r(t)\,. \end{cases} \tag{4.51}$$

Putting $q_1 = x$ and $q_2 = v$, one can see that the system (4.51) takes the form of the stochastic equations (4.49) with the following parameters,

$$\begin{aligned} K_1 = v\,,\ \ K_2 = -\eta\, v - V'_{\rm eff}(x)/m\,, \\ G_{11} = 0\,,\ \ G_{12} = 0\,,\ \ G_{21} = 0\,, \\ G_{22} = (2\,\eta\, k_B T/m)^{1/2}. \end{aligned} \tag{4.52}$$

Thus, in this case the FP equation for the function $f(x, v, t)$ has the form

$$\frac{\partial f}{\partial t} + v\,\frac{\partial f}{\partial x} - \frac{1}{m}\,V'_{\rm eff}(x)\,\frac{\partial f}{\partial v} = \eta\,\frac{\partial}{\partial v}\left[\left(v + \frac{k_B T}{m}\,\frac{\partial}{\partial v}\right)\,f\right]. \tag{4.53}$$

Equation (4.53) is known as the *Fokker-Planck-Kramers* (FPK) *equation.* The generalization for the system with many degrees of freedom in the present case looks as

$$\frac{\partial f}{\partial t} + \vec{v}\,\frac{\partial f}{\partial \vec{x}} - \frac{1}{m}\,\frac{\partial V}{\partial \vec{x}}\,\frac{\partial f}{\partial \vec{v}} = \frac{\partial}{\partial \vec{v}}\left[\overleftrightarrow{\eta}\,\left(\vec{v} + \frac{k_B T}{m}\,\frac{\partial}{\partial \vec{v}}\right)\,f\right]. \tag{4.54}$$

Note that if the left-hand side of Eq. (4.54) is put to zero,

$$\frac{\partial f}{\partial t} + \vec{v}\,\frac{\partial f}{\partial \vec{x}} - \frac{1}{m}\,\frac{\partial V}{\partial \vec{x}}\,\frac{\partial f}{\partial \vec{v}} = 0\,, \tag{4.55}$$

then we obtain nothing else that the Liouville equation, but for the "singled out" subsystem only which undergoes the Newtonian motion equations. Thus, the right-hand side of Eq. (4.54) describes the influence of the thermostat on the "main" subsystem.

In the case of strong friction, $\eta \gg \omega_0$, where ω_0 is a characteristic frequency of system vibrations, we may neglect by the first (inertial) term in the Langevin equation. Then, comparing the equation

$$m\,\eta\,\dot{x} + V'_{\rm eff}(x) = F^r(t) \tag{4.56}$$

with the stochastic equation (4.42), we see that now $K(x) = -V'_{\rm eff}(x)/m\eta$ and $G = (2\,k_B T/m\eta)^{1/2}$, so that the FP equation takes the form

$$\frac{\partial f(x,t)}{\partial t} = \frac{\partial}{\partial x}\left(\frac{V'_{\rm eff}(x)}{m\,\eta}\,f(x,t)\right) + \left(\frac{\partial^2}{\partial x^2}\,f(x,t)\right)\frac{k_B T}{m\,\eta}\,. \tag{4.57}$$

Equation (4.57) is called the *Smoluchowsky equation.*

Thus, on the one hand, the stochastic dynamics may be investigated with the help of solution of the Langevin equation using the random number generator and calculating as large as possible number of trajectories. On the other hand, we may solve the FPK equation, i.e. the equation with partial derivatives. Both the approaches are mathematically equivalent, and they must lead to the same final results.

For solution of the FP equation, a number of approximate methods have been developed which are described, e.g., in Risken's monograph [115]. As an example, in the next section we describe the results of numerical solution of the FPK equation.

4.5 Kramers theory

4.5.1 The problem

Perhaps, the most known application of stochastic equations is connected with the theory of thermally activated processes firstly developed by Kramers [116].

Let $V(x_1, \ldots, x_N)$ denote the potential energy of the system with N degrees of freedom, and let the function $V(x_1, \ldots, x_N)$ has minima (the absolute minimum as well as relative, or local minima), maxima and saddle points. If this system is in contact with a thermostat having a temperature T, then it will stay mainly near one of the minima of V, the absolute minimum (the ground state of the system) or a local minimum (a metastable state). However, when $T \neq 0$, then, owing to action of random (fluctuating) forces, the system may pass from a given minimum **A** to a nearest neighboring minimum **B** as shown in Fig. 4.1. Thus, it emerges the question about the probability R of such a transition per unit of time.

Approximately we may suppose that the transition from **A** to **B** is carried out along the adiabatic trajectory. To construct it, first we have to find the saddle trajectory, i.e. the trajectory which connects the

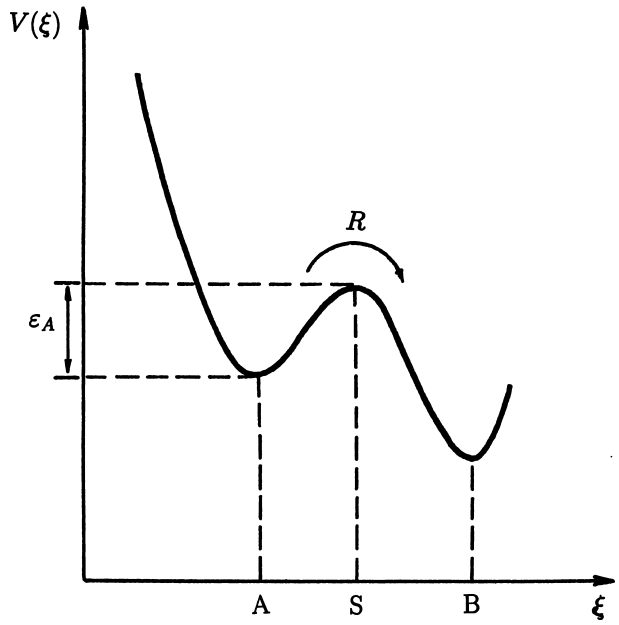

Figure 4.1: Escape of an atom from a potential well.

points of the minima **A** and **B** on the surface $y = V(x_1, \ldots, x_N)$ in the $(N+1)$-dimensional space passing through a nearest saddle point **S**. The saddle trajectory is determined by solution of the equation

$$\frac{\partial \vec{x}}{\partial \tau} = -\frac{\partial V}{\partial \vec{x}} \tag{4.58}$$

(where τ is a parameter varying along the trajectory), i.e. it is the curve of steepest descent, and it describes non-inertial (i.e. with the infinitely large friction) motion of the system from the saddle point **S** to the nearest minima **A** and **B**. If there are several saddle trajectories, as the adiabatic trajectory we have to select that saddle trajectory which is characterized by the minimum energy barrier. Motion along the adiabatic trajectory will occur with the lowest energy losses and, therefore, it will take place with the highest probability. Along the adiabatic trajectory we may introduce a coordinate ξ (which is a curved coordinate in a general case), so that the potential energy becomes the function of a single variable, $V(\xi)$, and the problem reduces to one-dimensional one.

Thus, the problem under investigation is reduced approximately to solution of the one-dimensional Langevin equation or the corresponding FPK equation for the potential $V(\xi)$ schematically shown in Fig. 4.1. As an initial condition we may assume that at the initial instant $t = 0$ the system is near the minimum **A** in the "thermal equilibrium" state, i.e. that

$$f(\xi, \dot{\xi}, t = 0) = \frac{m\,\omega_A}{2\,\pi\,k_B T} \exp\left(- \frac{m\,\dot{\xi}^2/2 + V_A(\xi)}{k_B T} \right), \tag{4.59}$$

where m is the reduced mass, and the function

$$V_A(\xi) = \frac{1}{2}\, m\,\omega_A^2 (\xi - \xi_A)^2 \tag{4.60}$$

approximates $V(\xi)$ near the minimum point $\xi = \xi_A$. To define the probability R of the transition **A**→**B** per time unit, we have to calculate the flux of particles through the saddle point $\xi = \xi_S$ from the left-hand to the right-hand side,

$$j(t) = \int_0^\infty d\dot{\xi}\, \dot{\xi}\, f(\xi_S, \dot{\xi}, t)\,. \tag{4.61}$$

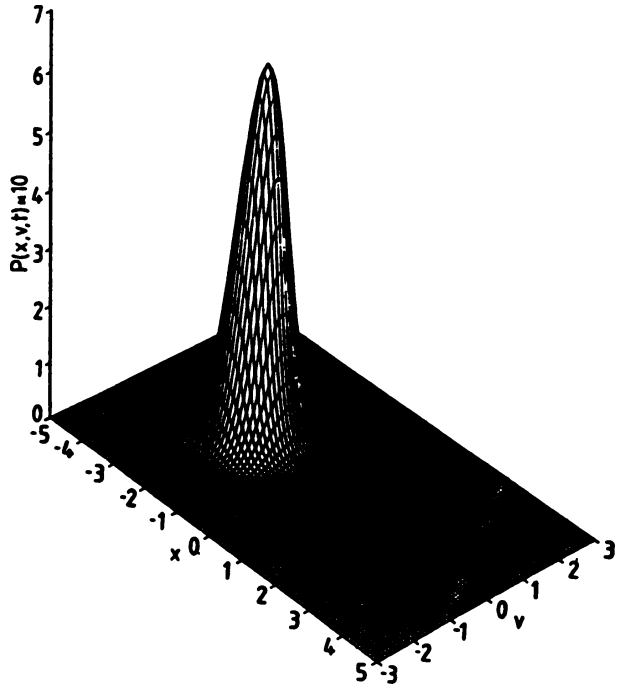

Figure 4.2: The initial state probability density function $P(x, v, t)$ at $t = 0$, the Maxwell-Boltzmann distribution with $\varepsilon_A/k_BT = 4$ and corresponding to the equilibrium state if only the left minimum of the potential $V(x)$ in Fig. 4.1 were present (from [119]).

Then, taking into account the normalization of the function (4.59), we find

$$R \approx j(t) \quad \text{for} \quad t \sim \omega_A^{-1}. \tag{4.62}$$

The approximate solution obtained by Kramers for the case of $\varepsilon_A \equiv V(\xi_S) - V(\xi_A) \gg k_BT$, has the Arrhenius form,

$$R = R_0 \exp(-\varepsilon_A/k_BT)\,, \tag{4.63}$$

where

$$R_0 \approx \begin{cases} (\omega_A/2\pi)(\eta/\eta^*) & \text{for } \eta \ll \eta^* \equiv (\omega_A/2\pi)(k_BT/\varepsilon_A)\,, \quad \text{(a)} \\ (\omega_A/2\pi) & \text{for } \eta^* \ll \eta < \omega_S\,, \quad \text{(b)} \\ (\omega_0/2\pi)(\omega_S/\eta) & \text{for } \eta > \omega_S\,, \quad \text{(c)} \end{cases} \tag{4.64}$$

and the value ω_S determines the curvature of the barrier near the saddle point, $V(\xi) \approx -\frac{1}{2}m\,\omega_S^2(\xi - \xi_S)^2$ at $\xi \approx \xi_S$. Thus, the rate of escape from the potential well reaches the maximum and approximately does not depend on the friction η in the case of intermediate friction (4.64b). Note that this case corresponds to the region where the Eyring absolute rate theory [117] may be applied, and that namely this case corresponds to a majority of real situations in physics and chemistry. In the case of small friction, $\eta \ll \eta^*$, the rate of the process is slowed, because in order to overcome the barrier, the system must (fluctuationally) get from the thermostat an energy which exceeds the activation barrier ε_A, and this process is carried out during too long time in the case of small η. On the other hand, in the case of strong friction, $\eta > \omega_S$, the rate of the escape from the potential well is small due to a small velocity of the system motion in the configuration space when it overcomes the saddle point.

However, the solution obtained by Kramers is an approximate one. It was checked and made more precise in a number of works (e.g., see the survey [118]). Below we describe the results of one of such works, where the computer simulation was used.

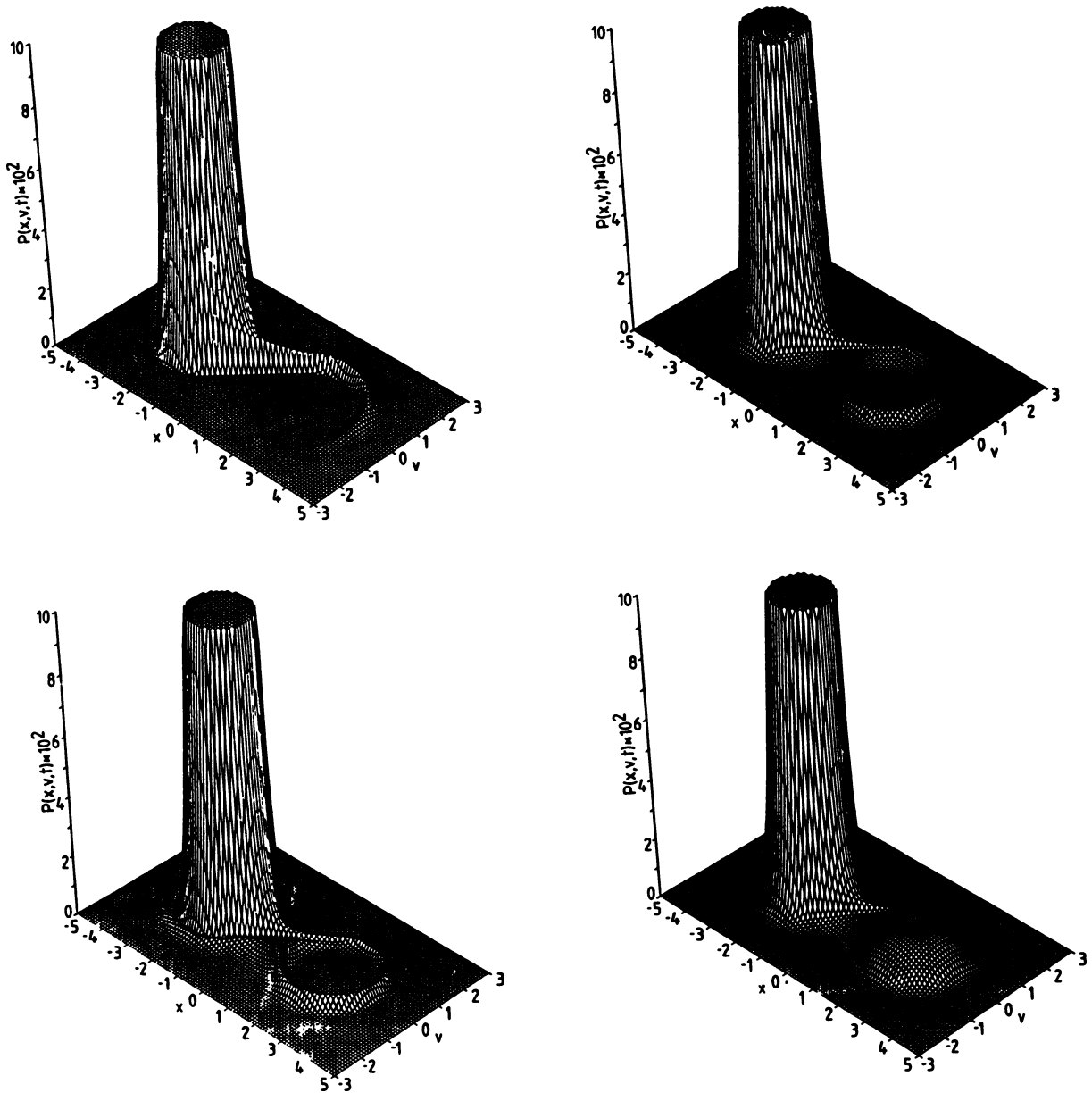

Figure 4.3: The time evolution of the probability density function $P(x, v, t)$ in the phase space as determined from the Fokker-Planck-Kramers equation for the potential (4.65) and from the initial state in Fig. 4.2 as calculated by Cartling [119]. The damping constant $\eta = 0.01$ and the sequence of times is: (a) 4, (b) 10, (c) 40, and (d) 200.

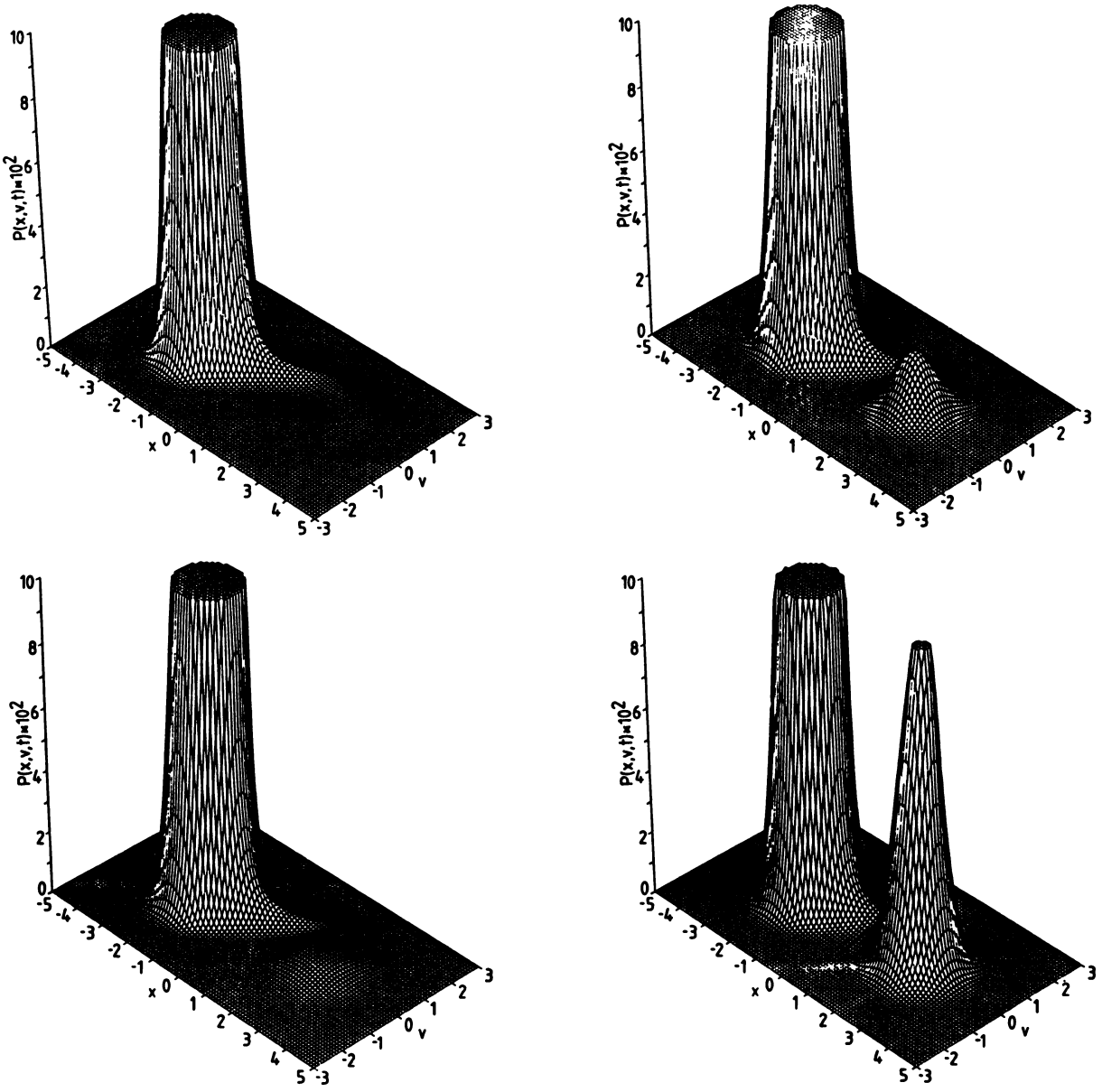

Figure 4.4: As in Fig. 4.3, but for the damping constant $\eta = 0.4$ and the sequence of times: (a) 4, (b) 10, (c) 20, and (d) 100 (after [119]).

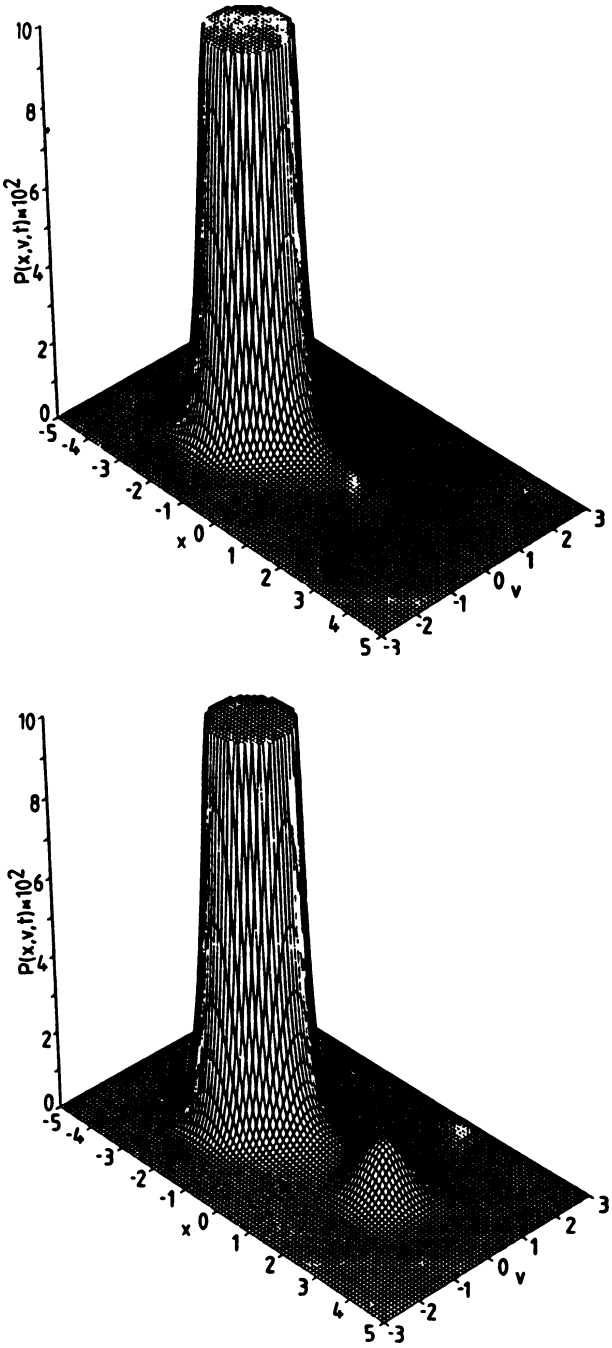

Figure 4.5: As in Fig. 4.3, but for the damping constant $\eta = 10.0$ and the sequence of times: (a) 20 and (b) 100 (after [119]).

4.5.2 Numerical solution of the FPK equation

To check the Kramers theory, Cartling [119] solved numerically the FPK equation for the potential

$$V(x) = \begin{cases} \frac{1}{2}(x+2)^2 & \text{if } x < -1, \\ 1 - \frac{1}{2}x^2 & \text{if } |x| \le 1, \\ \frac{1}{2}(x-2)^2 & \text{if } x > 1, \end{cases} \tag{4.65}$$

i.e. it was chosen $m = 1$, $x_A = -2$, $x_B = 2$, $x_S = 0$, $\varepsilon_A = 1$, and $\omega_A = \omega_S = 1$. The region of variation of the variables was restricted to the rectangle $|x| \le 5$, $|v| \le 3$, and at the border of the rectangle, the function $f(x, v, t)$ was put to zero. The transformation of the differential FPK equation to the difference equation was carried out by standard numerical method with the help of introducing a mesh (e.g., see [120]). The mesh step was taken as $h_x = h_v = 0.02$, and the time step was chosen as $\Delta t = 0.025$. As an initial condition, it was taken the Maxwell-Boltzmann distribution (4.59) shown in Fig. 4.2. Evolution of the function $f(x, v, t)$ for different friction coefficients η (the small friction $\eta = 0.01$, the intermediate friction $\eta = 0.4$, and the strong friction $\eta = 10$) for the temperature $k_B T = 0.25$ is shown in Figs. 4.3 to 4.5. To determine the rate R, there were calculated the probabilities to find the particle at each of the minima, at the left minimum Q_L and the right one Q_R,

$$Q_L(t) = \int_{-\infty}^{\infty} dv \int_{-\infty}^{0} dx\, f(x, v, t), \qquad Q_R(t) = \int_{-\infty}^{\infty} dv \int_{0}^{\infty} dx\, f(x, v, t), \tag{4.66}$$

so that $Q_L(t) + Q_R(t) = 1$, and then the transition rates $R_{L\to R}$ and $R_{R\to L}$ were found with the help of the kinetic equation,

$$\frac{dQ_L(t)}{dt} = R_{L\to R} Q_L(t) - R_{R\to L} Q_R(t). \tag{4.67}$$

The simulation has shown that the approximate Kramers expressions (4.63), (4.64) describe the numerical results with a sufficient accuracy.

Chapter 5

Monte Carlo Method

The Monte Carlo (MC) method is associated with a whole class of numerical methods, where the using of the random number generator (RNG) plays the central role. The Monte Carlo method together with the Molecular Dynamics method is the main method of computer simulation, especially for investigation of systems in the thermal equilibrium state.

In fact, we already met with the MC method in the previous Chapter 4 when we discussed the numerical solution of the Langevin equation (more exactly, the method of solution of the Langevin equation is more logically to call as the hybrid of MD and MC methods). Recall that in the solution of the Langevin equation the RNG has been used to imitate the fluctuating force acting on the subsystem under investigation from the thermostat. This example may lead to an impression that the MC method is an imitation one, i.e. that it gives some approximate description of a real system. But this impression is wrong. All approximations are connected with the model under consideration only, while the MC method gives the mathematically rigorous solution of the model with any preliminary stated accuracy.

In a general case all MC methods reduce to the same scheme. Namely, let we need to calculate the value of a variable a. To do this, we have to invent a random variable ξ such that the mathematical expectation of ξ will coincide with the value of a,

$$a = \langle \xi \rangle \equiv M\{\xi\}\,, \tag{5.1}$$

where

$$M\{\xi\} = \frac{1}{N}\sum_{n=1}^{N} \xi_n \;\text{ at }\; N \to \infty\,. \tag{5.2}$$

According to the probability theory, for a finite N the computational error of the MC method will always be proportional to $N^{-1/2}$. Consequently, the length of the sequence $\{\xi_n\}$ must be very long as a rule.

Naturally the following questions arise:

1. In what a way should ξ be chosen, or, more specifically, what should the corresponding probability distribution $P(\xi)$ be equal to?

2. How can the corresponding Markov chain, i.e. the sequence of random numbers with the given distribution $P(\xi)$, be constructed?

Actually the MC method has found a wide application with appearing fast electronic computers only. Metropolis and Ulam [121] were the first who have done this in 1949.

5.1 Generation of random numbers

5.1.1 Generation of the standard random numbers

Evidently that for exploring the MC method, first of all we are in need of a generator of random numbers. Here let us consider the so called *standard random number generator* (SRNG), and then we will show how we can get random numbers with any given probability distribution $P(\xi)$ from the standard random numbers.

The *standard random numbers* (SRNs) are the random numbers γ uniformly distributed within the unit interval (0,1). The sequence of the SRNs $\{\gamma_n\}$ must satisfy the following four requirements:

(1) Its probability distribution must be equal one,

$$P(\gamma) = 1\,. \tag{5.3}$$

(2) Its mathematical expectation should be the following:

$$M\{\gamma\} \equiv \langle\gamma\rangle = \frac{1}{2}\,. \tag{5.4}$$

(3) The dispersion has to be equal to

$$D\{\gamma\} \equiv \langle(\gamma - \langle\gamma\rangle)^2\rangle = \frac{1}{12}\,. \tag{5.5}$$

Recall that according to the definition, the dispersion for a function $A(\xi)$ of a random variable ξ, which is defined in an interval (a, b) with a probability distribution $P(\xi)$, is equal to

$$D\{A\} \equiv \langle(A - \langle A\rangle)^2\rangle = \int_a^b d\xi\, P(\xi)\,[A(\xi) - M\{A\}]^2, \tag{5.6}$$

where

$$M\{A\} \equiv \langle A\rangle = \int_a^b d\xi\, P(\xi)\, A(\xi)\,. \tag{5.7}$$

(4) There must be no correlation between different elements of the sequence $\{\gamma_n\}$, i.e., it should be fulfilled the relationship

$$\left\langle \left(\gamma_i - \frac{1}{2}\right)\left(\gamma_k - \frac{1}{2}\right)\right\rangle = \left\langle \gamma_i - \frac{1}{2}\right\rangle\left\langle \gamma_k - \frac{1}{2}\right\rangle = 0 \tag{5.8}$$

for any $i \neq k$.

In what a way can we get random numbers with a computer? The decisive factor here is the fact that any application of the MC method to physical problems requires a very long sequence of random numbers. Consequently, such methods of obtaining the random numbers (RNs) as, for example, their getting from a preliminary prepared table, or the setting of the RNs with an external noise generator, occur to be ineffective (a large table cannot be stored, and for the noise generator it is too difficult to control the quality of the obtained RNs).

The only realistic method turns to be the calculation of the RNs with the recurrence first-order formula

$$\gamma_{n+1} = f(\gamma_n)\,. \tag{5.9}$$

Clearly that in this way we can get only pseudo-random numbers but not the true random numbers. The sequence $\{\gamma_n\}$ calculated with Eq. (5.9), always has a finite period simply in consequence of that the set of different numbers available with a computer, is finite. Therefore, if some number in the sequence $\{\gamma_n\}$ is been met in the second time, then according to Eq. (5.9) all next numbers will be repeated too. Thus, the problem is to find such a function $f(\gamma)$ that the period of the obtained sequence $\{\gamma_n\}$ will be as long as possible.

In order to choose an appropriate function $f(\gamma)$ in Eq. (5.9), let us combine the consequent random numbers $\gamma_1, \gamma_2, \ldots$ in pairs (γ_1, γ_2), (γ_3, γ_4), $\ldots$, and let us consider these pairs as coordinates of points in a unit square (see Fig. 5.1). Clearly that these points must uniformly fill in the whole square, i.e., the graph of the function $y = f(x)$ should fill in the square as dense as possible. On the other hand, the function $f(x)$ has to be maximally simple (for example, linear), because we need a very large amount of RNs. To satisfy these two requirements, we may take the function

$$f(x) = gx \,(\mathrm{mod}\, 1) \tag{5.10}$$

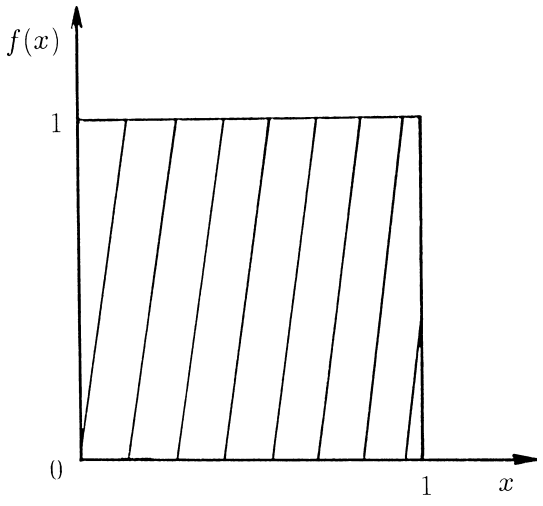

Figure 5.1: The function $f(x) = gx \,(\mathrm{mod}\, 1)$ for the random number generator.

with a sufficiently large g ($g \gg 1$) as is shown in Fig. 5.1. Recall that we met already Eq. (5.10) in Chapter 2 devoted to the stochastic theory. Namely, Eq. (5.10) for $g = 2$ is nothing else that the logistic map (2.9) for the $r = 4$ case, when its strange attractor takes the whole interval (0,1). In fact, instead of Eq. (5.10) it is more convenient to use the following recurrence formula for integer numbers m_n (D. Lemer):

$$m_{n+1} = gm_n \,(\mathrm{mod}\, M)\,. \tag{5.11}$$

Here $\mathrm{mod}\, M$ is the operation of taking of the integer remainder from the division by M, g and M have to be relatively simple, and the inequality $g < \sqrt{M}$ should be satisfied. The needed standard random numbers γ are then obtained with the equation

$$\gamma_n = m_n/M\,. \tag{5.12}$$

The described method is known as the *congruential method*, or the *residual method*. Note that the period of the sequence $\{\gamma_n\}$ obtained in this way, cannot exceed M, and that the sequence itself depends on the choice of the initial value m_0. Usually for the 32-bit computers one takes $M = 2^{31} - 1 = 2147483647$ (the so-called Mersenn's number) and $g = 7^5 = 16807$ (another choice is $g = 48271$ or $g = 69621$ [185]), for the 36-bit computers, $M = 2^{36}$ and $g = 5^{15}$, for the 40-bit computers, $M = 2^{40}$ and $g = 5^{17}$, and for the 49-bit computers, $M = 2^{48}$ and $g = 11^{13}$. In the case of $g \ll M$, however, the algorithm (5.11) is not too good — if one meets with a small γ_n value, the values of the following random numbers will be small too. Because the multiplication and especially the division are comparative slow computer operations, it is better to write a subroutine for the random number generator using computer instructions. Concrete programs for RNGs as well as the bibliography on this question may be found, e.g., in Heerman's book [1] (see also [9, 185, 186]).

The main requirements to the SRNG in the MC method are the following:

(a) The random number generator should work with high speed, and the period of the resulting sequence must be sufficiently long. For this, it is desirable to use the 32-bit computer (or, better, the 64-bit computer), and the double precision would be used.

(b) The each RNG must be tested before its exploration, i.e., the validity of the conditions (5.3), (5.4), (5.5), and (5.8) must be verified. This point is especially actual in using the RNG developed for a computer of one type, on the computer of an another type. At this, it is desirable to test not only the whole sequence $\{\gamma_n\}$, but also its separate sections (consisting, for example, of 2000 sequent elements), as well as to test, whether the period of the random sequence is sufficiently long for the problem under investigation. The detailed description of different tests for checking the quality of RNs may be found, e.g., in the book of Sobol' [122]. However, it must be noted that any test cannot guarantee the getting of "good" random numbers, although it can save from the using of wittingly "bad" sequences at least.

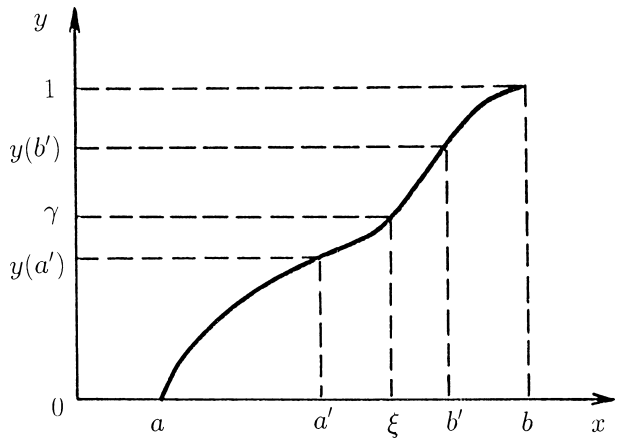

Figure 5.2: The transformation of the standard random numbers γ into the random numbers ξ with a given probability distribution $P(\xi)$.

(c) To have a more reliable results, it is desirable to carry out the MC simulation with different SRNGs. For example, some runs may be carried out not only with the "main" SRNG, but also with an another SRNG, and then the results have to be compared.

5.1.2 Generation of random numbers with a given distribution

Often in the MC simulation we need a sequence of random numbers $\{\xi_n\}$ defined in an interval (a, b) with a given probability distribution $P(\xi)$. Let us prove that the values of ξ can be found from the equation

$$\int_a^{\xi} P(x)\,dx = \gamma\,, \tag{5.13}$$

i.e., to get a next random number ξ, we have to take the next standard random number γ and then to solve the equation (5.13).

To prove, let us consider the function

$$y(x) = \int_a^x P(x)\,dx\,. \tag{5.14}$$

According to the definition of the probability distribution, we have

$$P(x) > 0 \;\text{ and }\; \int_a^b P(x)\,dx = 1\,. \tag{5.15}$$

Thus, it follows that $y(a) = 0$, $y(b) = 1$, and $y'(x) = P(x) > 0$, i.e., the function $y(x)$ monotonically increases from 0 to 1 as shown in Fig. 5.2. Consequently, any line $y = \gamma$, where $0 < \gamma < 1$, intersects the graph of the function $y = f(x)$ in one and only one point (the abscissa of which is just taken as ξ), and Eq. (5.13) always has a single solution only.

Now let take an arbitrary interval (a', b') inserted into the interval (a, b). Points from the interval $a' < x < b'$ are in one-to-one correspondence with the ordinates of the function $y(x)$, and these ordinates satisfy the inequality $y(a') < y < y(b')$. Therefore, if ξ belongs to the interval $a' < x < b'$, the corresponding γ is to belong to the interval $y(a') < y < y(b')$ and vice versa (see Fig. 5.2). Thus,

$$\text{Prob}\,\{a' < \xi < b'\} = \text{Prob}\,\{y(a') < \gamma < y(b')\}\,. \tag{5.16}$$

But because γ is uniformly distributed within the interval (0,1), we have

$$\text{Prob}\,\{y(a') < \gamma < y(b')\} = y(b') - y(a') = \int_{a'}^{b'} P(x)\,dx\,. \tag{5.17}$$

From Eqs. (5.16) and (5.17) it follows that

$$\text{Prob}\,\{a' < \xi < b'\} = \int_{a'}^{b'} P(x)\,dx\,, \tag{5.18}$$

and Eq. (5.18) just means that the random value ξ which is the root of Eq. (5.13), has the probability distribution $P(x)$.

Let us consider several specific examples of obtaining of RNs with a given $P(x)$.

(1) The random number ξ is uniformly distributed within an interval (a, b), if its distribution function is constant in this interval,

$$P(x) = \frac{1}{b-a}\,,\ \ a < x < b\,. \tag{5.19}$$

In this case the integral (5.13) is trivial, and it follows that

$$\xi = a + (b-a)\gamma\,. \tag{5.20}$$

(2) Clearly that the three of random numbers $(\gamma_1, \gamma_2, \gamma_3)$ correspond to the coordinates of points which uniformly fill in the unit cube. On the other hand, if we are in need of random points uniformly filled in the sphere of a radius R, we have to pass to spherical coordinates r, φ and θ (radius, azimuthal and polar angles respectively). Then each random point within the sphere is defined by three of random points (r, φ, θ), which are solutions of the equations

$$\frac{1}{\frac{1}{3}R^3}\int_0^r dr\, r^2 = \gamma_1\,, \tag{5.21}$$

$$\frac{1}{2\pi}\int_0^\varphi d\varphi = \gamma_2\,, \tag{5.22}$$

$$\frac{1}{2}\int_0^\theta d\theta\,\sin\theta = 1 - \gamma_3\,. \tag{5.23}$$

From Eqs. (5.21) to (5.23) it follows that

$$r = R\gamma_1^{1/3}, \tag{5.24}$$

$$\varphi = 2\pi\gamma_2\,, \tag{5.25}$$

$$\theta = \arccos(2\gamma_3 - 1)\,. \tag{5.26}$$

(3) One of important problems where the MC method is used, is the modeling of a random free path for a particle moving in a scattered or dissipating medium. In this case we are in need of random numbers ξ defined in the infinite interval $(0, \infty)$ with the probability distribution

$$P(x) = \frac{1}{\lambda}\exp\left(-\frac{x}{\lambda}\right), \tag{5.27}$$

where $\lambda = M\{\xi\}$ is the mean free path. The solution of Eq. (5.13) in this case yields the relationship $\xi = -\lambda\ln(1-\gamma)$, or, that is the same,

$$\xi = -\lambda\ln\gamma\,, \tag{5.28}$$

because γ and $(1-\gamma)$ are distributed in the interval (0,1) in the same way. The generalization of the distribution (5.27) is the so called *gamma-distribution*, which is characterized by the distribution function

$$P(x) = \frac{1}{(n-1)!}\,x^{n-1}e^{-x}, \tag{5.29}$$

where $n \geq 1$ is an integer. In this case the random numbers can be obtained with the formula [122]

$$\xi = -\ln(\gamma_1\gamma_2\ldots\gamma_n)\,. \tag{5.30}$$

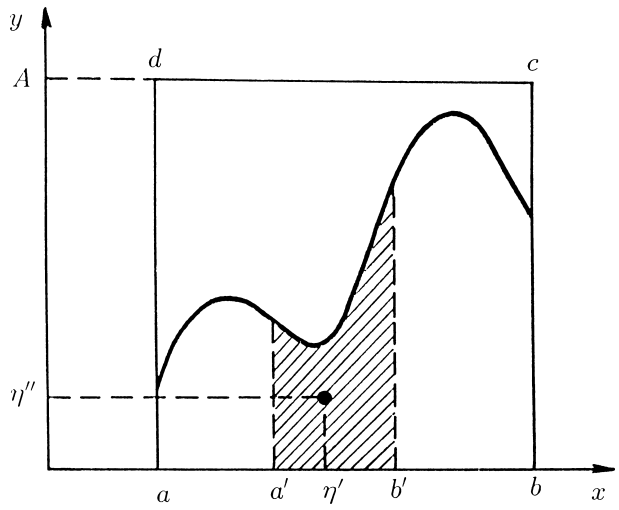

Figure 5.3: Illustration of Neumann's method.

(4) Now let us consider a situation when the integral of $P(x)$ cannot be expressed through elementary functions, or when the function $P(x)$ was determined in an experiment, so that it exists in a table form only. If the random value ξ is defined in a finite interval (a, b) and its distribution function is bounded, $P(x) \leq A$, the raffling of ξ may be done by Neumann's method in the following way:
(i) choose two sequent values γ' and γ'' from the standard random number sequence $\{\gamma_n\}$, and construct the random point $\Gamma(\eta', \eta'')$ with the coordinates $\eta' = a + \gamma'(b-a)$ and $\eta'' = \gamma'' A$;
(ii) if the point Γ occurs to be below the curve $y = P(x)$, we put $\xi = \eta'$; otherwise, when the point Γ lies above the curve $y = P(x)$, the pair (η', η'') is to be thrown out, and we are taking the next pair (γ', γ'').

Indeed, the random points Γ are uniformly distributed within the rectangular "*abcd*" which has the area $A(b-a)$ (see Fig. 5.3). The probability that the point Γ will occur to be below the curve $y = P(x)$ and will not be thrown out, is equal to the ratio of areas

$$\frac{\int_a^b P(x)\,dx}{A\,(b-a)} = \frac{1}{A\,(b-a)}\,. \tag{5.31}$$

But the probability that the point Γ will occur to be below the curve $y = P(x)$ in the interval $a' < x < b'$, is equal to the following ratio of areas:

$$\frac{\int_{a'}^{b'} P(x)\,dx}{A(b-a)}\,. \tag{5.32}$$

Consequently, among all values of ξ which have been picked out, a part of those getting into the interval (a', b'), is equal to the ratio

$$\frac{\int_{a'}^{b'} P(x)\,dx}{A(b-a)} \Big/ \frac{1}{A(b-a)} = \int_{a'}^{b'} P(x)\,dx\,, \tag{5.33}$$

q.e.d. Note that the sequence $\{\xi_n\}$ obtained in this way, is approximately four times shorter that the original sequence $\{\gamma_n\}$.

(5) Very often we need the so called *normal random numbers* ξ which are defined in the interval $(-\infty, +\infty)$ with the probability distribution

$$P(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \tag{5.34}$$

so that $M\{\xi\} = 0$ and $D\{\xi\} = 1$. To produce them, we may use the central limit theorem of the probability theory which states that if $\{\xi_n^{(1)}\}$, $\{\xi_n^{(2)}\}$, ... are independent sequences of random numbers with the same

distribution function, then the sum of N such random numbers, $\xi_n = \sum_{i=1}^{N} \xi_n^{(i)}$, is to produce at $N \to \infty$ the sequence $\{\xi_n\}$ of normal random numbers, and

$$M\{\xi\} = NM\{\xi^{(i)}\} \text{ and } D\{\xi\} = ND\{\xi^{(i)}\}\,. \tag{5.35}$$

Therefore, the normal random numbers can be obtained from the standard random numbers with the relationship

$$\xi \approx \sqrt{\frac{3}{N}} \sum_{n=1}^{N} (2\gamma_n - 1)\,, \tag{5.36}$$

where $N \gg 1$. For practical use one usually put $N = 12$, so that Eq. (5.36) takes the form

$$\xi \approx \left(\sum_{n=1}^{12} \gamma_n \right) - 6\,. \tag{5.37}$$

In this case the obtained sequence $\{\xi_n\}$ is in twelve times shorter than the original sequence $\{\gamma_n\}$. Recall that namely such random numbers have been used in solution of the Langevin equation.

From the normal random numbers ξ with the help of a simple renormalization

$$\tilde{\xi} = a + \sigma\xi \tag{5.38}$$

we can get the *Gauss random numbers* $\tilde{\xi}$ distributed with the probability function

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-a)^2}{2\sigma^2} \right), \tag{5.39}$$

for which $M\{\tilde{\xi}\} = a$ and $D\{\tilde{\xi}\} = \sigma^2$.

(6) However, the described above method, Eq. (5.37), is not the best one — to get one normal random number, we have to use twelve standard random numbers. Instead [194], one may use the Box-Muller transformation [195], which allows us to transform uniformly distributed random variables to a new set of random variables with a Gaussian distribution. Starting with two independent standard random numbers $\gamma_{1,2}$ and making the transformation $\xi_1 = \sqrt{-2\ln\gamma_1}\cos(2\pi\gamma_2)$, $\xi_2 = \sqrt{-2\ln\gamma_1}\sin(2\pi\gamma_2)$, we obtain two new independent random numbers $\xi_{1,2}$ which have a Gaussian distribution with zero mean and a standard deviation of one. For practical use, the polar form of the Box-Muller transformation is both faster and more robust numerically [196]. The algorithm of it is the following [194]:

```
float x1, x2, w, y1, y2;
do {    x1 = 2.0 * ranf() - 1.0;
        x2 = 2.0 * ranf() - 1.0;
        w = x1 * x1 + x2 * x2;
   } while ( w >= 1.0 );
w = sqrt( ( -2.0 * ln( w ) ) / w );
y1 = x1 * w;
y2 = x2 * w;
```

where `ranf()` is the routine to obtain a standard random number.

(7) Finally, sometimes we need discrete random numbers. For example, such a situation emerges in step-by-step simulation of some process, when at a given step with a probability p (at $\xi = 0$) the process will go in one way, and with a probability $1-p$ (at $\xi = 1$), in the another way. The discrete random numbers ξ_n can be obtained from the SRNs γ_n by the rule

$$\xi_n = \begin{cases} 0 \text{ if } \gamma_n \le p\,, \\ 1 \text{ if } \gamma_n > p\,. \end{cases} \tag{5.40}$$

A more detailed discussion of the questions mentioned in this section may be found, e.g., in Ref. [122].

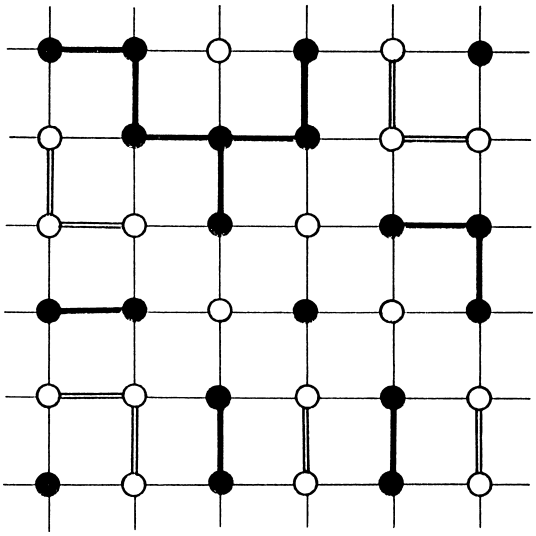

Figure 5.4: Square lattice for the nodal problem of the percolation theory. The sites united into "black" clusters are connected by thick lines, and those united into "white" clusters, by double lines.

5.2 Some applications of the MC method

5.2.1 Percolation theory

Let us consider a square lattice and suppose that each its site is either occupied by an atom with a probability p, or it is free with the probability $(1-p)$ as is shown in Fig. 5.4. Let us assume that two occupied sites are *coupled* to one another, if they are either nearest neighbors (NN), or if they can be connected by a chain of the NN occupied sites. Let us call by the ("black") *cluster* the set of occupied sites in which any two sites are coupled (similarly one can define the "white" cluster consisting of free sites, see Fig. 5.4). Clearly that at $p \to 0$ the average cluster size tends to zero, while in the limit $p \to 1$ the cluster will occupy the whole lattice. Naturally it arises the question, at what value $p = p_c^s$ does the infinite cluster emerge in the first time? If we suppose that the "bonds" are conducting, and we solder electrodes to the opposite lattice sides, then a current between the electrodes will pass when $p > p_c^s$ only. The value p_c has thus been called by the *percolation threshold*.

The problem of the same class emerges when we are considering a metallic mesh where the each section (i.e., a piece connecting the nearest neighboring sites) has a resistance r. Let us again solder the electrodes to the opposite mesh sides, and measure the system conductivity σ. If now we begin to break the sections with the probability $1-p$, than the conductivity σ will decrease, and at $p \le p_c^b$ the conductivity will fall down to zero.

The former among the described above problems is known as the *node problem*, and the latter, as the *bond problem*. According to Hammereli's theorem, for the same lattice the inequality $p_c^b \le p_c^s$ is fulfilled. Intuitively it is clear, because in order to blockade one site, we have to break all the bonds which come together to this site.

Note that the bond problem can be transformed into the node problem but for an another lattice called the *covered lattice*. The covered lattice is constructed in the following way: put the site of the covered lattice into the middle point of the each bond of the original lattice, and then connect two sites of the covered lattice to each other, if and only if the corresponding bonds of the original lattice come to the same site. The resulting covered lattice may occur to be more complicated than the original one. For example, the covered lattice for the square lattice cannot be placed on a plane.

The described problems are called the *percolation problems*. Analytical results in the percolation theory are not numerous, and the most natural way of their solution is the MC method. For this, we have to

Table 5.1: Percolation thresholds for lattices of different kinds (analytical results are denoted by asterisk; $p_0 = 2\sin(\pi/18) \approx 0.347296$) (from [124] and [188])

spatial dimensionality	lattice	p_c^b (bond problem)	p_c^s (node problem)
2	triangular	p_0^*	0.5^*
	square	0.5	0.592746
	honeycomb	$(1-p_0)^*$	0.70
3	simple cubic	0.24881	0.3117 ± 0.0003
	b.c.c.	0.18	0.25
	f.c.c.	0.12	0.20
	diamond	0.39	0.43

construct an array of the necessary dimensionality, and to numerate the array elements analogously to the sites of the lattice. To each array element, the value 0 or 1 (e.g., the free or occupied site respectively) is given with the help of the generator of discrete random numbers. The obtained array is then analyzed to find the clusters (the corresponding algorithms may be found, e.g., in Appendix to the Müller-Krumbhaar paper [123]).

Apart from the calculation of the percolation threshold p_c, it is interesting to study the average cluster size as a function of $p_c - p$ at $p < p_c$, $\langle N \rangle = f_N(p_c - p)$, as well as to investigate the dependence of the system conductivity on $p - p_c$ at $p > p_c$, $\sigma = f_\sigma(p - p_c)$. Numerical results show that $f_N(x)$ and $f_\sigma(x)$ are power functions, so that the percolation processes belong to the class of critical phenomena [187].

In addition to the square lattice, one may study the triangular or hexagonal lattices as well as three-dimensional lattices of different symmetry. Some results for the percolation thresholds obtained by the MC method, are summarized in Table 5.2.1. Besides, the problem may be generalized by accounting of more removed bonds, or by filling the lattice by atoms of different kinds, *etc.*

Percolation models have found a wide application in description of different physical phenomena, such as **(1)** jump conductivity of an amorphous semiconductor, **(2)** destroying of the spontaneous magnetization of a ferromagnetic, when it is doped by nonmagnetic impurities, **(3)** formation of a polymer gel, **(4)** penetration of gas molecules thought a pore filter, *etc.* A more detailed discussion of percolation problems may be found, e.g., in Refs. [124, 125, 188].

5.2.2 Penetration of neutrons through an absorbing plate

Now let us consider a standard example from the neutron physics, namely the problem of shielding of radiation. For the sake of concreteness let take the simplest variant of the problem, the penetration of neutrons through a plane.

Let the neutron beam is forwarded normal to a uniform infinite plane of a width h. When a neutron collides with an atom in the plane, it may either be absorbed with a probability p_c, or may be scattered with the probability $p_s = 1 - p_c$. For the sake of simplicity we assume that the neutron energy remains unchanged at the scattering, and that any direction of the neutron after the scattering is equiprobable (the latter is sometimes true for substances consisting of heavy atoms). Figure 5.5 demonstrates possible scenarios of the neutron motion: (a) the neutron passes through the plane, (b) the neutron is absorbed in the plane, and (c) the neutron is reflected from the plane. Let suppose also that between the scattering events the neutron moves freely, and let its motion is characterized by a mean path length λ_0, so that the random path lengthes λ are distributed with the probability density $P(\lambda) = \lambda_0^{-1} \exp(-\lambda/\lambda_0)$.

Let we have to calculate the probability P^+ that the neutron passes through the plane, the probability P^- of the neutron reflection, and the probability P^c of the neutron absorption in the plane. To describe the corresponding computation algorithm, first we have to introduce the following variables. Let N denote the number of incoming neutrons, N^- denote the number of neutrons reflected from the plane, N^+ denote the number of neutrons passed through the plane, N^c denote the number of neutrons absorbed in the plane, and x

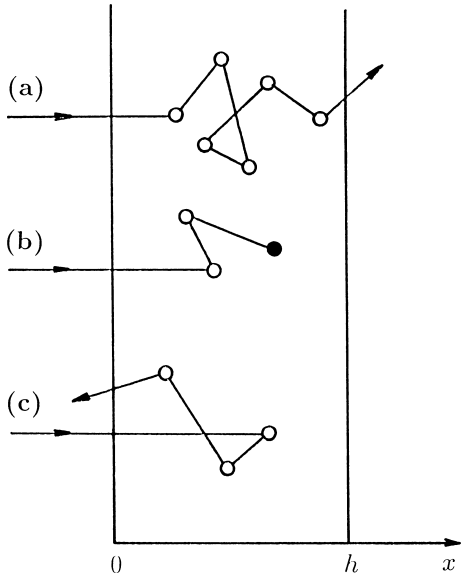

Figure 5.5: Penetration of neutrons through a plane: (a) passing through, (b) absorption in, and (c) reflection from the plane.

denote the horizontal coordinate of the neutron. At the beginning we have to put $N = N^+ = N^- = N^c = 0$. The further step-by-step operation of the program is the following.

Step 1. Suppose that the neutron comes to the plane from the left-hand side (see Fig. 5.5) so that we put $x := 0$ and

$$N := N + 1 \tag{5.41}$$

(here the symbol ":=" designates the assign operator. Recall that in the standard FORTRAN the symbol "=" is used as the assign operator). Put also $\mu := 1$ for the variable μ which will be needed at next steps.

Step 2. With the help of the standard random number generator let us "raffle" the free path length λ:

$$\lambda = -\lambda_0 \ln \gamma \,. \tag{5.42}$$

Step 3. Calculate the new neutron coordinate after its free run for the distance λ:

$$x := x + \lambda\mu \,. \tag{5.43}$$

Step 4. Test the condition that the neutron has passed through the plane,

$$x \geq h \,. \tag{5.44}$$

If this condition is fulfilled, the calculation of the trajectory of the given neutron is finished, and the counter of the passed neutrons is increased by one,

$$N^+ := N^+ + 1 \,. \tag{5.45}$$

In the opposite case we have to test the reflection condition

$$x \leq 0 \,. \tag{5.46}$$

If this condition is fulfilled, the calculation of the trajectory is finished too, and the counter of the reflected neutrons is increased by one,

$$N^- := N^- + 1 \,. \tag{5.47}$$

Last, if both conditions (5.44) and (5.46) are not fulfilled, i.e. if $0 < x < h$, then, consequently, the neutron must undergo a scattering in the plane, and we have to raffle its "fate" at the scattering. For this, let us take the next standard random number γ and test the absorption condition

$$\gamma < p_c\,. \tag{5.48}$$

If the latter inequality is fulfilled, the calculation of the trajectory is finished again, and the counter of the absorbed neutrons is increased by one,

$$N^c := N^c + 1\,. \tag{5.49}$$

For the all three cases when the calculation of the given trajectory has been finished, we have to check whether the number of incoming neutrons N reaches the preliminary designed number $N_{\max}$, and either to stop the computation, or to return to the step **1**.

Step 5. If the neutron was not gone out from the plane as well as it was not absorbed, we assume that the neutron undergoes the scattering at the point with the coordinate x. Then we have to raffle a new direction of its velocity. Because we have assumed in our model that any direction of the scattering is equiprobable, the cosine of the scattering angle, μ, must be uniformly distributed in the interval $(-1, 1)$. From this it follows that the value of μ should be raffled by the formula

$$\mu = 2\gamma - 1\,. \tag{5.50}$$

Then we have to return to the step **2** (it must be noted that each value γ must be used not more than one times, so that for the calculation of one section of the trajectory we are in need of three different SRNs).

When all $N_{\max}$ trajectories have been calculated, the necessary probabilities are defined approximately as

$$P^+ \approx N^+/N, \quad P^- \approx N^-/N, \quad \text{and} \quad P^c \approx N^c/N. \tag{5.51}$$

The computation method described above is quite natural but, unfortunately, it is imperfect. In particular, for a "thick" plane it is too difficult to obtain a reliable value for P^+, because a majority of neutrons is absorbed in the plane, and the resulting value of N^+ occurs to be too small. In this case we may use other variants of the MC method, for example, the so-called *balance method*. Let us suppose that at the initial time moment a "pack" consisting of a large number w of neutrons, penetrates into the plane, so that the operator (5.41) is to be replaced by $N := N + w$, and then the "pack" moves as a whole along the same trajectory. Consequently, when the "pack" leaves the plane, we have to use the operators $N^\pm := N^\pm + w$ instead the operators (5.45) or (5.47) respectively. When the "pack" collides with a plane's atom, the number of neutrons absorbed from the "pack" is equal wp_c, while the number of scattered neutrons is equal wp_s in average. Therefore, instead of raffling the absorption process we may simply add the value wp_c to the counter of absorbing neutrons, $N^c := N^c + wp_c$, and then we may continue to follow for the motion of the scattered "pack" which is assumed to be scattered as a whole into the same direction becoming "thinner" in the p_s^{-1} times, i.e. $w := wp_s$. The rest of the algorithm described above remains the same.

The gain in accuracy in the balance method is achieved owing to the fact that now all the trajectories are going out from the plane, no trajectories stop in the plane (a rigorous prove may be found in the Sobol' book [122]). The value w is called usually as the neutron "weight", and instead of the "pack" consisting of w neutrons, one says about a single neutron with the weight w. The initial weight w_0 at the step **1** is usually taken as one. This is not in a contradiction with the reasons of a "large pack", because it can be seen that all w values obtained at the calculation of a given trajectory, contain the initial value w_0 as the common multiplier.

Note that for walls of a complicated shape the MC method is in fact the unique method of calculating the probabilities $P^\pm$.

5.2.3 Calculation of definite integrals

Let us now consider the problem of approximate calculation of definite integrals. Let we have a function $y = g(x)$ defined in the interval (a, b), and let this function is bounded so that we can choose the number A such that

$$A > y_{max} \equiv \max\, g(x)\,. \tag{5.52}$$

Let we have to calculate the integral

$$I = \int_a^b dx\, g(x)\,. \tag{5.53}$$

Because the computation of definite integrals is equivalent to the calculation of areas, we may proceed in the following way. Consider the rectangular *abcd* (Fig. 5.3) and fill it uniformly by points with random coordinates. Let the total number of points is M, and the number of points under the curve $y = g(x)$, is M_I. Then the area under the curve may be approximately calculated as

$$I \approx (M_I/M)(b-a)A\,. \tag{5.54}$$

However, the described computation method is too primitive and leads to large errors, especially in the cases when the function $g(x)$ has, for example, a sharp maximum within the interval (a, b). For a more accurate solution of the problem under investigation the so-called *preferable choice method* is used usually.

Let us take a random variable ξ defined in the interval (a, b) with a probability distribution $P(\xi)$, and with the help of this variable let us calculate a new random variable η according to the formula

$$\eta = g(\xi)/P(\xi)\,. \tag{5.55}$$

Then one can see that

$$M\{\eta\} = \int_a^b dx\, P(x)\,[g(x)/P(x)] = I, \tag{5.56}$$

and this gives the method of computation of I:

$$I \approx \frac{1}{N}\sum_{n=1}^{N}\eta_n = \frac{1}{N}\sum_{n=1}^{N}\frac{g(\xi_n)}{P(\xi_n)}\,. \tag{5.57}$$

Intuitively it is clear that the lower is the dispersion of the random variable η, the higher will be the accuracy of the described method. Indeed, with the help of the central limit theorem of the probability theory it can be proved (e.g., see [122]) that with a very high probability (equal 0.997) the error of the approximation (5.57) does not exceed the value $3\sqrt{D\{\eta\}/N}$. On the other hand, we can lower $D\{\eta\}$ with the help of a suitable choice of the primary random variable ξ. For this, let us calculate the dispersion of the random variable η:

$$D\{\eta\} = M\{\eta^2\} - I^2 = \int_a^b dx\,[g^2(x)/P(x)] \;-\; I^2. \tag{5.58}$$

Let prove that this expression achieves its minimum when $P(x)$ is proportional to $|g(x)|$. To find the minimum of the dispersion $D\{\eta\}$ given by Eq. (5.58) for any possible choice of $P(x)$, let us use the well known inequality

$$\left[\int_a^b dx\,|u(x)v(x)|\right]^2 \le \left[\int_a^b dx\,u^2(x)\right]\left[\int_a^b dx\,v^2(x)\right]. \tag{5.59}$$

Putting $u(x) = g(x)/\sqrt{P(x)}$ and $v(x) = \sqrt{P(x)}$, from the inequality (5.59) we get

$$\left[\int_a^b dx\,|g(x)|\right]^2 \le \left[\int_a^b dx\,\frac{g^2(x)}{P(x)}\right]\left[\int_a^b dx\,P(x)\right] = \int_a^b dx\,\frac{g^2(x)}{P(x)}\,. \tag{5.60}$$

Equations (5.58) and (5.60) mean that

$$D\{\eta\} \geq \left[\int_a^b dx\,|g(x)|\right]^2 - I^2. \tag{5.61}$$

It leaves to prove that the lowest boundary is achieved when $P(x)$ is proportional to $|g(x)|$. Let

$$P_g(x) = \frac{|g(x)|}{\int_a^b dx\,|g(x)|}\,. \tag{5.62}$$

It is not difficult to calculate that for the probability density (5.62) we have

$$\int_a^b dx\,\left[g^2(x)/P_g(x)\right] = \left[\int_a^b dx\,|g(x)|\right]^2, \tag{5.63}$$

and the dispersion $D\{\eta\}$ is indeed equal to the right-hand side of Eq. (5.61).

Note that it is practically impossible to find the "best" density (5.62), because for this we have to calculate the integral $\int_a^b |g(x)|\,dx$. The calculation of the latter integral is the problem equivalent in its complexity to the original one. Besides, we should not choose too complicated functions $P(x)$, because the raffling of the ξ values becomes a very complicated procedure. But the reasons, described above, can and must be used in choosing of $P(x)$.

For a practical sake, one-dimensional integrals (5.53) are not been calculated by the MC method, there exist more accurate quadrature formulas for these purposes. But when we are meeting with multi-dimensional integrals, the situation changes: quadrature formulas become extremely complicated, while the MC method may be applied almost without changes. Already for integrals over the three-dimensional space, an accuracy of the MC method becomes comparable with the accuracy of quadrature formulas, and in cases of integrals over a space of dimensionality more than three, the MC method becomes the unique numerical method of computation of definite integrals. In particular, the MC method may be used in calculation of path integrals which are widely used in statistical physics [126], quantum physics [127] and quantum field theory [128].

5.3 The MC method in statistical physics

5.3.1 Idea of the method

When the system is in the thermodynamically equilibrium state, then, according to Boltzmann's hypothesis, its dynamics is stochastic, i.e. it undergoes to statistical laws of the probability theory. Consequently, it is natural to expect that the MC method should be the most suitable one in order to study the equilibrium state. Indeed, in the previous Chapter 4 we have used already the MC method in numerical solution of the Langevin equation. Using the stationary solution of this equation for calculation of time averages, we can study thermodynamical characteristics of the canonical ensemble. Note that a value of the friction coefficient η in this case may be chosen arbitrary (usually one takes $\eta \sim 1$), because the equilibrium characteristics do not depend on η.

On the other hand, instead of the averages over time, we may calculate the averages over the phase space,

$$\langle A\rangle = \int dq\,A(q)\,P_{\rm eq}(q)\,. \tag{5.64}$$

Here $A(q)$ is a measurable physical quantity, q is the coordinate in the $2\nu N_p$-dimensional phase space of the system under investigation (ν =1, 2 or 3 is the spacial dimensionality and N_p is the number of particles), and $P_{\rm eq}(q)$ is the equilibrium Maxwell-Boltzmann distribution,

$$P_{\rm eq}(q) = Z^{-1}\exp[-\beta E(q)]\,, \tag{5.65}$$

where $\beta = (k_B T)^{-1}$, $E(q)$ is the total system energy, and Z is the partition function,

$$Z = \int dq \, \exp[-\beta E(q)] \,, \tag{5.66}$$

which ensures the normalization of the probability distribution (5.65).

Usually the measurable $A(q)$ depends only on the coordinates of particles, and the total energy $E(q)$ may be presented as a sum of kinetic and potential energies. In this case the integrals over momenta in Eqs. (5.64) and (5.66) can be taken analytically and, moreover, they are mutually cancelled in the expression for $\langle A \rangle$. Therefore in what follows we will let $E(q)$ in Eqs. (5.65) and (5.66) denote the potential energy $V(q)$ only, and the integration in Eqs. (5.64) and (5.66) is to be carried out only over the spacial coordinates of particles.

The integrals (5.64) and (5.66) are the νN_p-dimensional integrals over the configurational space of the system under investigation. They can be found numerically by the unique method, namely by the MC method. As has been described in the previous Sec. 5.2.3, for this we have to construct a random sequence $\{\xi_n\}$ of points in the νN_p-dimensional configurational space such that the probability distribution density of this sequence, $P(q)$, is to be close as possible to the under-integral functions in Eqs. (5.64) and (5.66). Then we can calculate $\langle A \rangle$ with the formula

$$\langle A \rangle = \lim_{N \to \infty} \frac{N^{-1} \sum_{n=1}^{N} P^{-1}(\xi_n) A(\xi_n) \exp[-\beta V(\xi_n)]}{N^{-1} \sum_{n=1}^{N} P^{-1}(\xi_n) \exp[-\beta V(\xi_n)]} \,. \tag{5.67}$$

Taking into account that the functions $A(q)$ are smooth usually, the choice $P(q) = P_{\rm eq}(q)$ has to be the optimal one. In this case Eq. (5.67) reduces to the arithmetic average,

$$\langle A \rangle = \frac{1}{N} \sum_{n=1}^{N} A(\xi_n), \quad N \to \infty \,. \tag{5.68}$$

Thus, the problem under investigation reduces to a search of the sequence of random points in the phase space, ξ_1, ξ_2, ..., distributed with the probability density $P_{\rm eq}(q)$ which has to coincide with the Boltzmann distribution. Because each point in the phase space corresponds to a given state of the system, the random points ξ_n predominantly fill those part of the phase space, where the probability to find the system is high. That is to say that the microstates ξ_n are to be sampled according to their importance. The method of construction of the random sequence $\{\xi_n\}$ with the distribution density $P_{\rm eq}(q)$, based on the procedure of random walk of a point in the configurational space of the system, was firstly proposed by Metropolis *et al.* [129]. It consists in the following.

Let us turn to a discrete time $t_n = n \, \Delta t$ (further we may put $\Delta t \to 0$) and introduce some distribution density $P(q, t_n)$ which corresponds to the probability that the system is in the point q of the configuration space at the time t_n. Introduce also the value $W(q \to q')$ which describes the conditional probability of the transition of the system from the point q to the point q' per unit of time. Then, we can write the following evolution equation for the function $P(q, t_n)$:

$$\begin{aligned} \Delta P(q, t_n) &\equiv P(q, t_n) - P(q, t_{n-1}) \\ &= \textstyle\sum_{q'} W(q' \to q) \, P(q', t_{n-1}) \, \Delta t \, - \, \Big(\sum_{q'} W(q \to q') \, \Delta t \Big) \, P(q, t_{n-1}) \,, \end{aligned} \tag{5.69}$$

which is known as the *master equation*. The first term in the right-hand side of Eq. (5.69) rises $P(q, t_n)$ owing to transitions of the system to the state q from other points q' of the phase space. The rate of the increasing is equal to the probability of the transition for time Δt, $W(q' \to q) \, \Delta t$, multiplied by the probability that the system was in the state q', $P(q', t_{n-1})$. Analogously, the second term of Eq. (5.69) describes the transitions from the state q to other points q' of the phase space. In the limit $\Delta t \to 0$ Eq. (5.69) takes the form

$$\frac{\partial P(q, t)}{\partial t} = \sum_{q'} \left[W(q' \to q) \, P(q', t) \, - \, W(q \to q') \, P(q, t) \right] . \tag{5.70}$$

It can be proved that this equation has a stationary solution $P_0(q)$, and that any initial probability distribution $P(q, t = 0)$ approaches to the $P_0(q)$ as $t \to \infty$, provided the following two conditions are satisfied (the proof may be found, e.g., in [130]): **(1)** $W \geq 0$, and **(2)** the *ergodicity condition*, or the *connectivity condition*, according to which it must exist a possibility of the transition from any one point of the phase space to any other point for a finite number of steps.

Then, for the stationary distribution, Eq. (5.70) takes the form

$$\sum_{q'} \left[W(q' \to q)\, P_0(q') \; - \; W(q \to q')\, P_0(q) \right] = 0\,. \tag{5.71}$$

The simplest way to satisfy this equation is to put zero each term of the sum (5.71), i.e.

$$W(q' \to q)\, P_0(q') \; - \; W(q \to q')\, P_0(q) = 0\,. \tag{5.72}$$

Equation (5.72) is called the *detailed balance condition*.

In order that $P_0(q)$ is to coincide with the necessary function $P_{\rm eq}(q)$ (i.e., with Boltzmann's distribution (5.65)), we have to choose the transition probabilities $W(q \to q')$ which satisfy the condition

$$\frac{W(q \to q')}{W(q' \to q)} = \frac{P_{\rm eq}(q')}{P_{\rm eq}(q)} = \exp\left(-\frac{\Delta V}{k_B T}\right), \tag{5.73}$$

where $\Delta V = V(q') - V(q)$. Then the necessary random sequence of states of the system, $\xi_1, \ldots, \xi_N$, distributed with the probability density $P_{\rm eq}(q)$, may be obtained as the stationary solution of the master equation (5.69), i.e. as the sequence of states obtained with the help of this equation when the system has achieved the equilibrium state.

The rates $W(\ldots)$ in Eq. (5.73) may be chosen in different ways. The simplest and the most convenient is the choice

$$W = \begin{cases} \exp(-\Delta V/k_B T) & \text{if } \Delta V > 0\,, \\ 1 & \text{if } \Delta V < 0\,. \end{cases} \tag{5.74}$$

Thus, the transition to a state with a lower potential energy occurs with the probability one, while the transition to a state with a higher potential energy, with the probability $\exp(-\Delta V/k_B T)$ which is always lower than one.

5.3.2 Calculation scheme

The Metropolis procedure may be realized with the help of the following algorithm.

First let us consider the canonical ensemble consisting of N_p particles imbedded into some fixed volume at a constant temperature T. Let choose an initial configuration ξ_0, and initiate the counter of states n, $n := 0$.

Step A1. Put $n := n + 1$. If $n > N$, where N is the preliminary stated length of the sequence ξ_n, the calculation is finished.

Step A2*. With the help of the standard random number generator, select the j-th particle by the formula $j := \mathrm{int}\,(\gamma N_p)$.

Step A3*. In a random way, choose the vector $\Delta\vec{r}$ which gives the direction and the length of the displacement, and displace the j-th particle on $\Delta\vec{r}$ (the direction may be raffled by Eqs. (5.25) and (5.26), and the length, with the help of Gauss random numbers). In the result we get a trial configuration ξ^*.

Step A4. Calculate the change of the potential energy, ΔV, caused by the displacement of the randomly selected particle j on the random vector $\Delta\vec{r}$, $\Delta V = V(\vec{r}_j + \Delta\vec{r}) - V(\vec{r}_j)$.

Step A5. If in the result of this displacement the potential energy decreases, $\Delta V < 0$, the trial configuration is accepted as the next element of the sequence, $\xi_n := \xi^*$.

Step A6. In the opposite case, when the potential energy increases, $\Delta V > 0$, take a next standard random number γ, which has to be compared with $\exp(-\Delta V/k_BT)$. In this case it may be two variants: **(1)** if $\gamma \leq \exp(-\Delta V/k_BT)$, we put $\xi_n := \xi^*$ again; **(2)** but if $\gamma > \exp(-\Delta V/k_BT)$, then the trial configuration ξ^* should be throw out as an "unsuccessful" one, and the original configuration is duplicated, $\xi_n := \xi_{n-1}$.

Step A7. Return to the step **A1**.

The Metropolis procedure can be generalized for the case of grand canonical ensemble, where the chemical potential μ is fixed, while the number of particles may vary. Recall that in this case the averages are determined by the expression

$$\langle A \rangle = \sum_{N_p=0}^{\infty} \int dp\, dq\, A(q, N_p)\, P_{\rm eq}(p, q, N_p)\,, \tag{5.75}$$

where $P_{\rm eq}(p, q, N_p)$ is defined by the formula

$$P_{\rm eq}(p, q, N_p) = \frac{1}{\Xi N_p!} \exp\left\{-\beta\left[E(p, q, N_p) - \mu N_p\right]\right\}\,, \tag{5.76}$$

and the grand partition function is equal to

$$\Xi = \sum_{N_p=0}^{\infty} \frac{1}{N_p!} \int dp\, dq\, \exp\left\{-\beta\left[E(p, q, N_p) - \mu N_p\right]\right\}\,. \tag{5.77}$$

Repeating the arguments which have led to the expression (5.73), one can obtain that with accounting of the integration over momenta, the transition probabilities now must satisfy the condition

$$\frac{W(\{q, N_p\} \to \{q', N_p'\})}{W(\{q', N_p'\} \to \{q, N_p\})} = \frac{N_p!}{N_p'!} a^{N_p' - N_p} \exp(-\beta\, \Delta V)\,, \tag{5.78}$$

where

$$\Delta V = V(q', N_p') - V(q, N_p)\,, \tag{5.79}$$

$$a = (m k_B T/2\pi\hbar^2)^{\nu/2} \exp(\beta\mu)\,, \tag{5.80}$$

ν=1, 2 or 3 is the system dimensionality, and $\hbar$ is the Planck constant.

The algorithm is modified in the following way.

Step B1. Coincides with the step **A1**.

Step B2. Select in a random way one among the following three operations:
(a) to leave the number of particles unchanged (if the SRN γ occurs to be within the interval $0 < \gamma < 1/3$),
(b) to increase N_p by one (if $1/3 \leq \gamma < 2/3$), or
(c) to decrease N_p by one (if $2/3 \leq \gamma < 1$).

Step B3a. In the case of the variant **B2(a)**, we have to carry out the steps **A2** to **A6**, and then to return to the step **B1** instead of the **A1**.

Step B3b. In the case of the variant **B2(b)**, we have to choose in a random way the coordinates for a new particle, and to add it to the system, $N_p := N_p + 1$, obtaining in such a way a trial configuration ξ^*.

Step B3c. In the case of the variant **B2(c)**, we have to select in a random way the number $j = \mathrm{int}\,(\gamma N_p)$, and to take out the j-th particle from the system, $N_p := N_p - 1$, thus obtaining the configuration ξ^*.

Step B4. Calculate the change of the potential energy for the transition from the old configuration ξ_{n-1} to the trial configuration ξ^*, $\Delta V = V(\xi^*) - V(\xi_{n-1})$.

Step B5. Calculate the value $b = [a/(N_p + 1)] \exp(-\beta\, \Delta V)$ in the case of creation of a particle (the variant **(b)**), or the value $b = (N_p/a) \exp(-\beta\, \Delta V)$ for the case of the destruction of the particle (the variant **(c)**). If $b \geq 1$, the trial configuration is accepted, $\xi_n := \xi^*$.

Step B6. But if $b < 1$, we call the next standard random number γ and compare it with b. If $\gamma \leq b$, then we again put $\xi_n := \xi^*$. But if $\gamma > b$, the configuration ξ^* is to be throw out, and we put $\xi_n := \xi_{n-1}$.

Step B7. Return to the step **B1**.

Note that depending on the model under investigation, a concrete realization of the steps **A2*** and **A3*** marked by asterisk, may differ from those proposed above. Note also that in principle with the help of the MC method we may consider the microcanonical ensemble as well, if we introduce a "demon" which causes artificially the random point (the configuration of the system) to walk in the phase space closely to the surface of constant energy. Besides, the MC method may be modified in order to consider the NPT ensemble too. The corresponding algorithms are described, e.g., in [131] and in Heerman's book [1].

With the help of the algorithms described above, the sequence of configurations (or points in the phase space) $\xi_0, \xi_1, \xi_2, \ldots, \xi_n, \ldots, \xi_N$ is constructed. This sequence creates the Markov chain, which approaches to the stationary distribution $P_{\rm eq}(q)$ for large n. Because for the calculation of equilibrium characteristics of the system we must use a part of the Markov chain $\{\xi_n\}$, which corresponds to the stationary distribution, an initial part of the sequence should be thrown out. When the sequence $\{\xi_n\}$ is been found, any average can be calculated using Eq. (5.68). A length of the sequence should be such that $N/N_p \sim 10^3 \div 10^6$; for $N \sim 10^4 N_p$ an error in calculation of averages is about 1% [132]. Because the raffling of each configuration is in need of a several standard random numbers, the period of the sequence $\{\gamma_n\}$ produced by the SRNG, must be very large, it has to exceed the value $(10^4 \div 10^7)N_p$. The MC method allows us to simulate systems consisting of an essentially larger number of particles (larger than $500^3 \approx 10^8$) than the MD method.

The Markov chain constructed with the algorithm described above, is not too satisfactory, because the consequent configurations differ one from another by the position of one particle as maximum, while the other particles remain at the same positions. Therefore, a real difference of consequent configurations is not too large, and such Markov chain is not very suitable for calculation of thermal characteristics. In order to improve this deficiency, one may construct a new Markov chain $\{\tilde{\xi}_n\}$, selecting from the original chain $\{\xi_n\}$ the each m-th configuration, i.e.

$$\tilde{\xi}_1 = \xi_{i_0},\ \tilde{\xi}_2 = \xi_{i_0+m},\ \tilde{\xi}_3 = \xi_{i_0+2m},\ \ldots\,, \tag{5.81}$$

where the number i_0 corresponds to the configuration when the chain has achieved the equilibrium state. It is easy to see that if we take $m > N_p$, then the sequent states in the chain $\{\tilde{\xi}_n\}$ with a high probability will differ one from another by displacements of all particles.

The another deficiency of the described above method is connected with the fact that at each step we are trying to shift a single particle only, i.e., a collective motion of particles is completely ignored. In a result, it exists a probability that the system may be locked in a metastable state, and the escape from this state may be possible with a simultaneous jump of few particles (this is especially dangerous at an "unsuccessful" choice of the initial configuration ξ_0). To improve this deficiency, we may complicate the algorithm, for example, to include the possibility of simultaneous displacements of two particles at the steps **A2**–**A3**. A more powerful method based on Darwing's laws, will be described below in Sec. 5.6.

Unfortunately, with the help of the Metropolis procedure we can calculate only derivatives of the free energy with respect to thermodynamic variables (for example, the specific heat $c_V = (\partial F/\partial T)_V$, or the magnetization $M = (\partial F/\partial h)_T$, where h is the magnetic field). But this method does not allow to calculate directly the partition function Z and, consequently, the free energy $F = -k_B T \ln Z$ and the entropy $S = (E - F)/k_B T$. To improve the situation, one may try to find the entropy by another methods, for example, using its definition

$$S = k_B \ln \mathcal{N}, \tag{5.82}$$

where $\mathcal{N}$ is the number of ways to place the system in the phase space [133]. The another method is to calculate the value $A(a) = \partial F/\partial a$ for all values of the variable a from 0 to a_0, and then to find F with the help of numerical integration,

$$F(a_0) = \int_0^{a_0} da\, A(a) \,+\, F(0)\,. \tag{5.83}$$

Here, however, we must know the value $F(0)$.

A more detailed description of applications of the MC method in physics may be found, e.g., in the books edited by Binder [132, 134, 135].

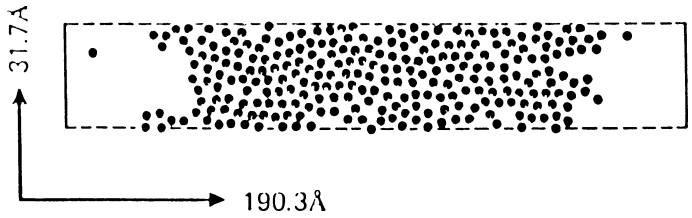

Figure 5.6: Snapshot pictures of the x-y atomic positions of the 256 Lennard-Jones argon atoms for the 10^7-th configuration in the Monte Carlo simulation for the two-dimensional liquid-vapor interface. The horizontal broken lines of the computation box depict periodic boundary conditions and the vertical lines depict hard walls (after [88]).

5.4 Modeling of phase transitions

The investigation of phase transitions is, perhaps, the main application of the MC method. Because it is described in a number of review articles and monographs (e.g., see books [132, 134, 135, 136]), the further description will be rather brief.

For the MC studies of different states of the system (such as gas, liquid and crystalline states) and phase transitions (PT) between these states, the model is choosing with the same principles as in the MD method described above in Sec. 3.1. Namely, we have to choose the dimensionality of the system, then to take the law of interaction between the particles, and to put the boundary conditions. Because the number of particles in the MC method may essentially exceed the number of particles accepted in the MD method, the MC simulation may be successfully employed for three-dimensional systems as well. Usually in this case the periodic boundary conditions are imposed. Otherwise, if we are using the free or fixed boundary conditions, it emerges the problem how to distinguish the body and surface contributions. For this, we may make several computations with different numbers of particles N_p. Indeed, for the homogeneous (one-phase) state of the system, the dependence of some intensive characteristic A versus N_p may be presented approximately in the form

$$A(N_p) \approx a_v N_p + a_s N_p^{2/3}. \tag{5.84}$$

Then, constructing the graph of the dependence $A(N_p)/N_p$ versus $N_p^{-1/3}$, we can find the body contribution a_v from the point of intersection of the dependence with the ordinate axis, and the surface contribution a_s from the tangent of the inclination angle. Note, however, that the procedure of separation of the contributions is essentially complicated near the point of the second-order phase transition, where extensions of fluctuationally arising domains start to exceed the system size, and the value a_v begins to depend on N_p.

5.4.1 Continuum models

For studying of continuum models (i.e. the models where the particles can move continuously in space), the Metropolis algorithm described in the previous section, is explored usually. The average displacement $\langle \Delta \vec{r} \rangle$ for the shift of a particle per one MC step, is selected during the program training in such a way that to obtain the maximally fast approach to the equilibrium state. In this case it is useful to compare the results of the MC modeling with those of the MD simulation of the same model.

Summary of the results on the MC modeling of homogeneous systems as well as the phase transitions in these systems, may be found in many works, in particular, in the paper of Levesque *et al.* (see Chapter 2 in the book [132]); see also the bibliography in Abraham's survey [88]. Note only that when the interparticle interaction is modeled by the hard-core potential (3.2), the system exhibits only a single first-order phase transition, the gas–liquid PT. But already a crude inclusion of attraction with the help of the step potential (3.3), leads to the existence of both phase transitions, the gas–liquid and liquid–solid PTs.

As an example, let us consider the Abraham results [88], where the system of $N_p = 256$ atoms interacting by the 12–6 Lennard-Jones potential (3.5) were studied. The system was imbedded into the rectangular box (see Fig. 5.6), two boundaries of which (parallel to the $0Y$ axis) were the elastically reflected walls, while on the boundaries parallel to the $0X$ axis, the periodic boundary conditions were imposed. Such a geometry has allowed to get a planar (more exactly, a linear for the two-dimensional system) boundary of separation of two phases, the gas and liquid ones (if we impose the same boundary conditions in both directions, the separation boundary should be a circle). It was obtained the sequence of MC configurations of a length $N \sim 10^5 N_p$. One of the configurations is shown in Fig. 5.6. One can see a sharp boundary of separation of two phases. However, the surface of separation of the phases occurs to be "rough" (unlike from a usually smooth separation surface in three-dimensional systems); a width of the transition layer is about $(3 \div 4)r_0$ (r_0 is the parameter of the LJ potential; for **Ar** atoms r_0=3.822Å).

5.4.2 Lattice models

Perhaps, the most successful applications of the MC method are connected with lattice models. Recall that these models assume that particles may occupy only sites of some (one-, two-, or three-dimensional) spatial lattice characterized by a given symmetry. At the same time, each particle may carry several (usually discrete) internal degrees of freedom, for example, a spin in the case of magnetics or a kind of the atom in modeling of binary alloys. Consequently, with each site j of the lattice we have to associate one or several variables which take, for example, one among two values, $n_j = 0$ (the free site) or $n_j = 1$ (the site occupied by an atom).

The potential energy in the lattice model consists of one-particle and many-particle terms. The one-particle term describes the particle's energy in the given lattice site, and it may include a dependence on internal degrees of freedom. If the lattice has a complex elementary cell, different sites in the cell may be characterized by different one-particle energies. When disordered systems are modeled, the one-particle energy may take random values.

The interparticle interaction is described with the help of introduction of the parameters v_{ij}, which are equal the energy of interaction of the particles occupied the sites i and j (this energy may also include a dependence on internal degrees of freedom). Emphasize that if the occupation number n_j is allowed to take only two values 0 and 1, then the lattice model supposes a short-range interparticle repulsion with a potential of the hard-core type from the very beginning. In the simplest case one assumes that $v_{ij} \neq 0$ only for the particles which occupy the nearest neighboring sites. Often, however, we have to take into account the interaction of more removed neighbors, as well as the anisotropy of interaction. Additionally to the pairwise interaction v_{ij}, sometimes it is necessary to include a "trio" interaction v_{ijk}.

The term corresponding to the kinetic energy of the system, is absent in the lattice models. As the result, wave processes (phonons, spin waves, *etc.*) are lost when the lattice models are used.

In time evolution of the system, the particles may "jump" to nearest neighboring or more removed free sites of the lattice. When the system is in contact with some reservoir of particles, the destruction of some particles and the creation of new particles (at the free lattice sites) may take place. Besides, the particles may change the value of their internal degrees of freedom. It is assumed, however, that such acts are carried out "instantly" in time, and that they are characterized by some probabilities (of the jump, destruction, creation, *etc.*) per unit of time.

We have already met with applications of the lattice models in Sec. 5.2.1, where we discussed the percolation problems. In that case it has been assumed $v_{ij} = 0$, and in the result of simulation we have obtained the geometry of connected clusters.

An interesting problem is the investigation of random walks of particles in the lattice. If we "mark" one of the particles, for example, with the help of introduction of some (fictitious) internal degree of freedom, then in this way we can calculate the self-diffusion coefficient D_s.

Two-dimensional (2D) variant of the lattice model is widely used in modeling of adsorption systems.

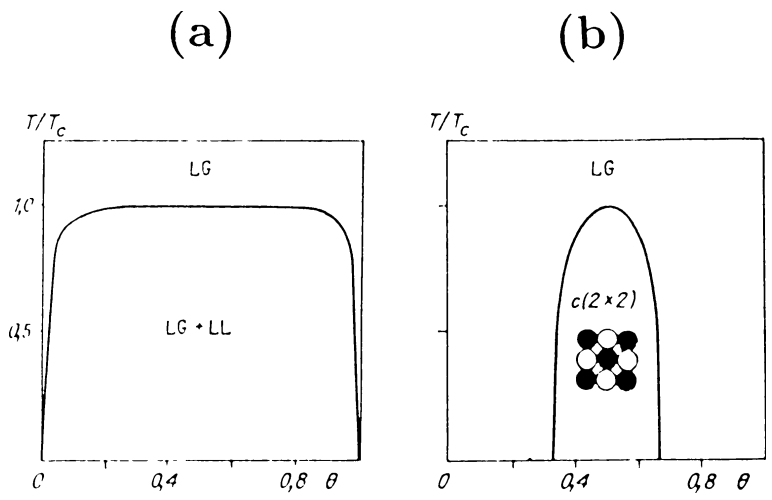

Figure 5.7: Phase diagram for the lattice-gas model with the square lattice and interaction of nearest neighbors only: (a) attraction, and (b) repulsion of atoms. LG is the lattice–gas phase, LL is the lattice–liquid phase, and c(2x2) is the ordered phase. $T_c \approx 0.567|v_{ij}|$ (after [139]).

The system Hamiltonian in this case has a form

$$H = \sum_j (\varepsilon_{(j)} - \mu) n_j + \frac{1}{2} \sum_{i \neq j} v_{ij} n_i n_j + \frac{1}{6} \sum_{i \neq j \neq k} v_{ijk} n_i n_j n_k + \cdots \tag{5.85}$$

Here the occupation number n_j takes the values 0 or 1, $\varepsilon_{(j)}$ is the one-particle energy, and μ is the chemical potential of the particle's reservoir, the latter is to be taken into account only in simulation of the adlayer which is in contact with a surrounding gas phase.

When the adlayer is isolated from the gas, the number of particles $N_p = \sum_j \langle n_j \rangle$ is to be fixed, and for the elementary MC step we must take the jump of an atom to the nearest neighboring (or a more removed) free site, $(n_i = 0,\, n_j = 1) \rightleftarrows (n_i = 1,\, n_j = 0)$ (the so-called *Kawasaki dynamics* [137]). On the other hand, when we are studying the adlayer which is in contact with a gas phase, the elementary MC steps must include adsorption and desorption processes, $(n_j = 0) \rightleftarrows (n_j = 1)$ (the so-called *Glauber dynamics* [138]).

From the sequence of configurations $\{\xi_n\}$ we can calculate the correlation function,

$$G(i-j) = \langle (n_i - \langle n_i \rangle)(n_j - \langle n_j \rangle) \rangle \,, \tag{5.86}$$

so that we can determine the phase state of the adlayer and construct the phase diagram of the system. Unlike from the solely two-dimensional system (i.e. the film adsorbed on "smooth" substrate without a potential relief), the 2D system subjected to the external periodic substrate potential (recall that the lattice model corresponds to the extreme case of infinitely high amplitude of the substrate relief) may be in a "true" crystalline phase, i.e. in the phase with nondecaying (oscillating) correlation function (5.86) at a temperature $T \neq 0$. For description of ordered phases one introduces the so-called *order parameter* which is a tensor function in a general case (e.g., see [76] or [87]).

As the simplest example, let us consider the square lattice with interaction of nearest neighbors only. If the atoms attract each other, $v_{ij} < 0$, the system exhibits the first-order PT. The phase diagram of the system is shown in Fig. 5.7a, where the full curve shows the dependence of the critical temperature $T_{\rm crit}$ on the "coverage" $\theta = N_p/N_L$ (N_p is the number of adatoms and N_L is the total number of the lattice sites). At $T > T_{\rm crit}(\theta)$ the adlayer is in the lattice-gas (LG) phase, while at $T < T_{\rm crit}(\theta)$ there coexist simultaneously two phases, the LG phase and the lattice–liquid (LL) phase. Note that the most direct proof, that a phase transition studied with the help of the computer experiment corresponds to the first-order PT, is, perhaps, the observation of a hysteresis at slow increasing and decreasing of the temperature during the simulation.

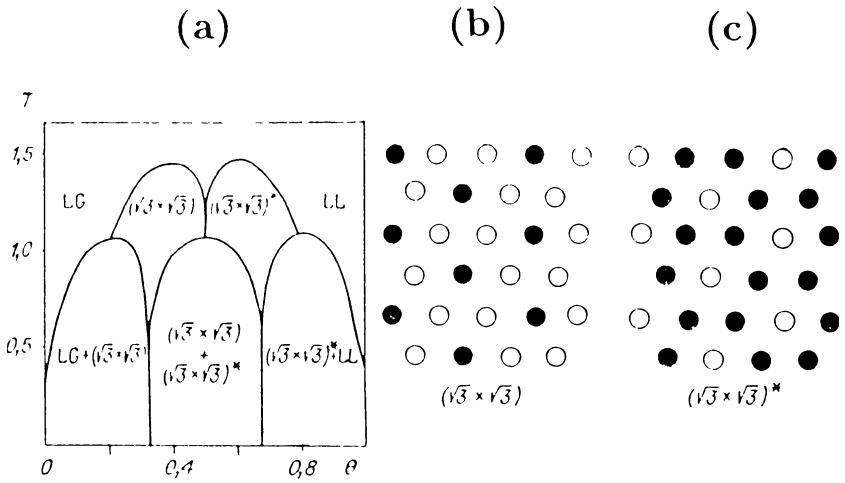

Figure 5.8: (a) Phase diagram for the lattice-gas model on the triangular lattice for repulsive interaction of nearest and next-nearest neighbors, (b) the structure $(\sqrt{3} \times \sqrt{3})$, and (c) the structure $(\sqrt{3} \times \sqrt{3})^*$ (black and white circles correspond to occupied and free lattice sites respectively) (after [143]).

It has to be noted also that the determination of the PT point itself is a rather difficult problem because of finiteness of the system under simulation (e.g., see [136, 140, 141] for more details).

Otherwise, if the atoms repel each other, $v_{ij} > 0$, the system exhibits the second-order PT of the "order–disorder" type. The corresponding phase diagram calculated by Binder and Landau [139], is presented in Fig. 5.7b. In this case at $T < T_{\rm crit}(\theta)$ the whole adlayer is in the ordered state with the c(2x2) structure (see Fig. 5.7b). Emphasize that the ground state of the c(2x2) structure is doubly degenerated: from one ground-state configuration we may obtain the another (equivalent) ground-state configuration by the shift of all atoms by one lattice constant. At $T = 0$ one of these states is realized in the adlayer, but at $T \neq 0$ the domains corresponding to the another ground state configuration, are created fluctuationally. With T increasing, the number of such domains rises, and at $T = T_{\rm crit}(\theta)$ the domains are completely mixed so that the ordered phase becomes totally destroyed. Recall that close to the point of the second-order PT, the correlation radius tends to infinity, and the phase transition takes place simultaneously in the whole system. Evidently that owing to the finiteness of the system under simulation, a region of the phase diagram near the PT points, is "smeared", and this may lead to essential errors in determination of $T_{\rm crit}(\theta)$.

The phase diagrams described above, correspond to the case of the canonical ensemble, where the number of adatoms N_p is fixed. On the other hand, when the adlayer is in contact with a gas, first we have to find the average number of adatoms, $\langle N_p \rangle$, which corresponds to a given value of the chemical potential of the gas (or to the gas pressure), and then to use the phase diagrams presented above.

Recall that the described model has an analytical solution in a single point only, namely at $\theta = \frac{1}{2}$ (the famous Onsager solution [142] of the two-dimensional Ising model for zero magnetic field).

As a more complicated example, Fig. 5.8 demonstrates the phase diagram calculated by Landau [143], for the triangular lattice with accounting the repulsive interactions of nearest and next-nearest neighbors. An interesting peculiarity of this phase diagram is a degeneration (the mixture of ordered phases) near the $\theta = \frac{1}{2}$ point.

When the model takes into account pairwise interatomic interactions only, the Hamiltonian (5.85) is invariant with respect to the substitution "occupied site" $\rightleftarrows$ "free site" [so that $\theta \rightleftarrows (1 - \theta)$], and the phase diagram is symmetric with respect to the $\theta = \frac{1}{2}$ line. However, the accounting of trio interactions ($v_{ijk} \neq 0$) destroys this symmetry.

The comparison of the phase diagram calculated by the MC method with that measured experimentally, helps to extract an important information on the interaction between adatoms. Note that the employing of analytical approaches for the construction of phase diagrams, is very complicated owing to the fact that

fluctuation effects in 2D systems are too strong and, consequently, approximate theories of the mean-field type are inapplicable. Note also that the description of a submonolayer adsorbed film with the help of the lattice model is justified, if the following two conditions are satisfied simultaneously. First, it should be fulfilled the inequality

$$k_B T \ll \varepsilon_a \,, \tag{5.87}$$

where ε_a is the activation energy for the transition of an adatom to a neighboring lattice site. When the inequality (5.87) is satisfied, the adatoms are situated in the lattice sites (at the minima of the substrate potential relief) for an almost the whole time, and from time to time only they are jumping to another sites. Second, it should be satisfied the condition

$$|v_{ij}| \ll \varepsilon_a \,, \tag{5.88}$$

because in the opposite case we can not neglect by displacements of adatoms from the relief minima caused by interatomic interactions. These two conditions are satisfied, for example, for gas atoms or molecules (such as **O**, **H**, **CO**, **NO**, *etc.*) chemically adsorbed on a metal substrate.

In investigation of multilayer adsorbed films, as well as in studying of clean crystal surfaces, it is often used an another variant of the lattice model, the so-called *solid on solid* (SOS) model introduced by Kossel. This model assumes that each site of the two-dimensional lattice is associated with the discrete variable h_j ($0 \le h_j < \infty$ for the simulation of adlayers, and $-\infty < h_j < +\infty$ for the simulation of clean surfaces) which has a meaning of a height of the column of adatoms growing over a given lattice site. The one-particle contribution into the system potential energy is assumed to be proportional to the column height h_j, $\sum_j \varepsilon_{(h_j)} h_j$, and for accounting of interatomic interaction parallel the surface it is assumed that the interaction energy of two nearest neighboring columns of highes h_i and h_j is determined by the difference of their highes. Thus, the Hamiltonian of the SOS model has the form

$$H = \sum_j (\varepsilon_{(h_j)} - \mu) h_j + \frac{1}{2}\phi \sum_{\langle i \neq j \rangle} \mid h_i - h_j \mid^m, \tag{5.89}$$

where ϕ is the parameter of interaction of the nearest neighboring columns (in the case of their attraction we have to put $\phi > 0$), and the exponent m is usually taken to be $m = 1$. The sum in the second term of the right-hand side of Eq. (5.89) is carried out over the nearest neighboring sites only. Notice that when the interval of h_j variation is restricted by the values 0 and 1 only, the Hamiltonian (5.89) reduces to the Hamiltonian (5.85) considered above. The special feature of the SOS model is that it forbids the existence of "overhanging" structures, i.e. there must be no bubbles in the adlayer. As the elementary MC acts in the SOS model we have to take the adsorption–desorption acts, $(h_j) \rightleftarrows (h_j \pm 1)$, as well as the acts of diffusion along the surface, $(h_i,\, h_j) \rightleftarrows (h_i - 1,\, h_j + 1)$.

The SOS model may be used for studying of the roughing of the surface with increasing of temperature. Examples of the MC configurations obtained in this case, are presented in Fig. 5.9 taken from [132]. More often, however, the SOS model is used to study the crystal growth (see Sec. 5.5).

The most known among the lattice models is the Ising model, where all sites are assumed to be occupied by particles having an internal degree of freedom s_j, called usually by spin, and s_j is assumed to take, for example, two values $s_j = \pm 1$ (along or opposite the external magnetic field h). The Hamiltonian of the Ising model has the form

$$H = -h \sum_j s_j + \frac{1}{2} \sum_{i \neq j} J_{ij} s_i s_j \,, \tag{5.90}$$

where J_{ij} is called the "exchange integral". For modeling of ferromagnetics we have to put $J_{ij} < 0$, and for the case of antiferromagnetics, $J_{ij} > 0$. Recall that the two-dimensional version of the Ising model for the $h = 0$ case with accounting of interaction of nearest neighbors only, has the analytical solution [142].

It is easy to see that the substitution $s_j = 1 - 2n_j$ transformers the Ising model into the lattice-gas model (5.85). As the elementary step in the Ising model, a spin in a randomly chosen site j is upset usually, $(s_j = +1) \rightleftarrows (s_j = -1)$ (in the lattice-gas model this corresponds to the case of an adlayer in contact with a gas).

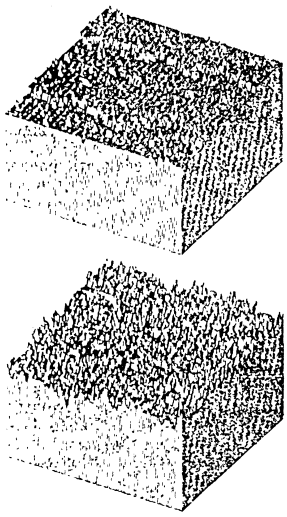

Figure 5.9: Simulation of crystal roughening in the framework of the Kossel (solid-on-solid) model. (a) At low temperature the surface is smooth, and (b) at high temperature the surface is rough (from [132]).

Note ones more that in the Ising model at $h = 0$ there exists an effective violation of ergodicity so that at low temperatures (lower than the critical temperature $T_{\rm crit}$) the ground state of the system is doubly degenerated. For example, in the case of ferromagnetics the total magnetization $M = \sum_j \langle s_j \rangle$ may be either positive or negative. Because of finiteness of the system investigated in computer simulation, there always exists a nonzero probability of the transition from one ground state to the another ground state. The probability of this transition is determined by an activation barrier, the height of which is proportional to $N_L^{(\nu-1)/\nu}$ [144]. Therefore, when we are calculating the total magnetization, we have to accumulate for averaging the absolute values of M.

In conclusion, let us recall also the Geisenberg model, where with each lattice site it is associated a continuous variable, for example, a vector of unit length $\vec{s}_j$, which may be directed arbitrary. In this case the model Hamiltonian takes the form

$$H = -\vec{h} \sum_j \vec{s}_j + \frac{1}{2} J \sum_{\langle i \neq j \rangle} \vec{s}_i \vec{s}_j \,, \tag{5.91}$$

where $\vec{h}$ is the vector corresponding to the external magnetic field.

The modeling of magnetic systems with Ising- or Geisenberg-type Hamiltonians is a very wide topic, which lies beyond the scope of the present course. For a more detailed discussion of this problem we refer a reader to special sources (e.g., see Landau's article in the book [139]).

5.5 Relaxation dynamics

The Metropolis technique described above in Sec. 5.3, was intended to investigate the thermodynamically equilibrium state of a system. However, the approach to the equilibrium state lies in the foundation of this technique from the very beginning. Consequently, the master equation (5.69) may be interpreted as the equation which describes "time evolution" of the system. However, the dynamics obtained in this way, is not true, but the *Markovian dynamics*, i.e. the dynamics "without a memory of the past". Such a dynamics is called the *relaxation dynamics*. In more detail, the Metropolis procedure uses the detailed balance condition (5.72) (known also as the *microcanonical reversibility condition*), which is usually not satisfied in real dynamics. For the same reasons, inertional effects fall out the description, and in the result, wave processes are lost in the relaxation dynamics.

As the temporal coordinate in the relaxation dynamics, the Monte Carlo step per particle is used which may be connected with the real time approximately only. Recall that for the correct description of the equilibrium state, the transition probabilities per unit time, W_{ij}, must satisfy the unique condition, the detailed balance condition (5.72) which fixes only the ratio of the forward and backward transitions $\xi_{n-1} \leftrightarrow \xi_n$. On the other hand, a rate of dynamical processes in the system is determined not only by the ratio W_{ij}/W_{ji}, but by the values of individual probabilities W_{ij} as well. According to the Kramers theory (Sec. 4.5), the probability W_{ij} itself depends exponentially on the activation energy for the given transition, which in turn depends on both configurations ξ_{n-1} and ξ_n (for example, on the positions and states of the neighboring atoms interacting with the "jumping" atom). The incorporation of these reasons into consideration significantly increases the number of model parameters. However, in order to connect a given MC step with the real time, we have to take into account the reasons described above and to modify the Metropolis algorithm in an appropriate way.

Nevertheless, the relaxation dynamics is very interesting, because it exhibits many characteristic features of true dynamics of the system. Besides, the relaxation dynamics may be carried out for an essentially longer time interval as well as for systems consisting of a much larger number of particles than the MD simulation of the real dynamics.

The main idea of using of the Metropolis procedure in studying the system dynamics, is to take as the initial configuration a state ξ_0 which is realistic from the physical viewpoint. For example, in investigation of the submonolayer adsorbed film, we may take for ξ_0 the configuration where the atoms occupy the left–half part of the lattice only. Then the sequence of configurations $\{\xi_n\}$ obtained with the Metropolis algorithm, will describe the system evolution to the equilibrium state, i.e. the "smoothing" of the initial step in atomic concentration to the state where the atoms are uniformly distributed over the surface [145]. Fixing the MC configurations through the equal time intervals, calculating the concentration of adatoms as a function of the spacial coordinate, and processing the simulation data with the help of the Motana-Boltzmann technique, we can find the chemical diffusion coefficient D_c as a function of the concentration of adatoms analogously as it is done in laboratory experiments on surface diffusion [92]. Besides, we may analyze in detail the evolution of the diffusion front which usually is characterized by a fractal structure [189].

As the next example, let us note the using of the MC simulation in investigation of dynamics of phase transitions, for example, the crystallization of a liquid when it is cooling below the critical temperature $T_{\rm crit}$ of the crystal–liquid phase transition. In this case the Metropolis procedure is carried out for a temperature T lower than $T_{\rm crit}$, while as the initial configuration, we have to choose a configuration corresponding to the thermal equilibrium state of the system at $T = T_{\rm ini} > T_{\rm crit}$ (such a configuration may be obtained from the preliminary performed MC simulation for the $T = T_{\rm ini}$ case or, more crude, simply by taking the random atomic coordinates, the latter corresponds to the $T_{\rm ini} = \infty$ case). Then, generating the sequence of the MC configurations $\{\xi_n\}$, we may follow in detail how, owing to thermal fluctuations, small "drops" or "clusters" of the new phase emerge on the background of the old (unstable at the simulation temperature) phase, and how these drops either growth to macroscopic sizes (if they exceed a "critical" size) and coagulate, or disappear [146]. The results of such a modeling may be used to test different theoretical predictions. Note, however, that if the model uses a cutting radius r^* for the interatomic potential (as it was done in the particle-particle MD method), the simulation results for the dynamics of the first-order phase transition occur to be very sensitive to a choice of the value r^* [99]. Besides of such "homogeneous" nucleation, a large interest has the simulation of the "heterogeneous" nucleation, i.e. the nucleation at various irregularities of the system such as impurities, surfaces, interfaces, *etc.*

When the procedure described above, is carried out for a temperature $T \ll T_{\rm crit}$, we may simulate a process of phase separation at fast cooling of, e.g., a liquid alloy. For example, Fig. 5.10 demonstrates snapshot pictures for the two-dimensional lattice model of a binary A-B alloy [147], where a process of crystalline growth is seen clearly. Such computer experiments are very important for the metallurgy.

The simulation technique for continuous phase transitions is the same. For example, Salid and Binder [148] investigated the submonolayer film with the coverage $\theta = 1/2$, adsorbed on the centered rectangular (rhomboidal) lattice. This model describes an adsorption on the (110) face of the b.c.c. crystal, for example,

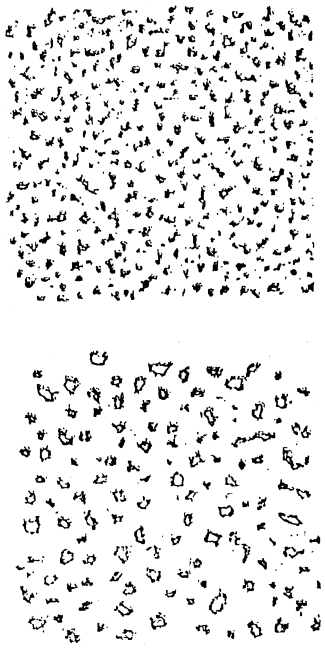

Figure 5.10: Snapshot pictures for the A-B alloy modeled by the 200x200 lattice-gas system at $\langle n_B \rangle = 0.2$ for (a) $t = 500$ and (b) $t = 8925$ Monte Carlo steps per atom after quench to the temperature $T = 0.6T_{\text{crit}}$. Black points correspond to the B-atoms, while the A-atoms are not shown (after Rao *et al.* [147]).

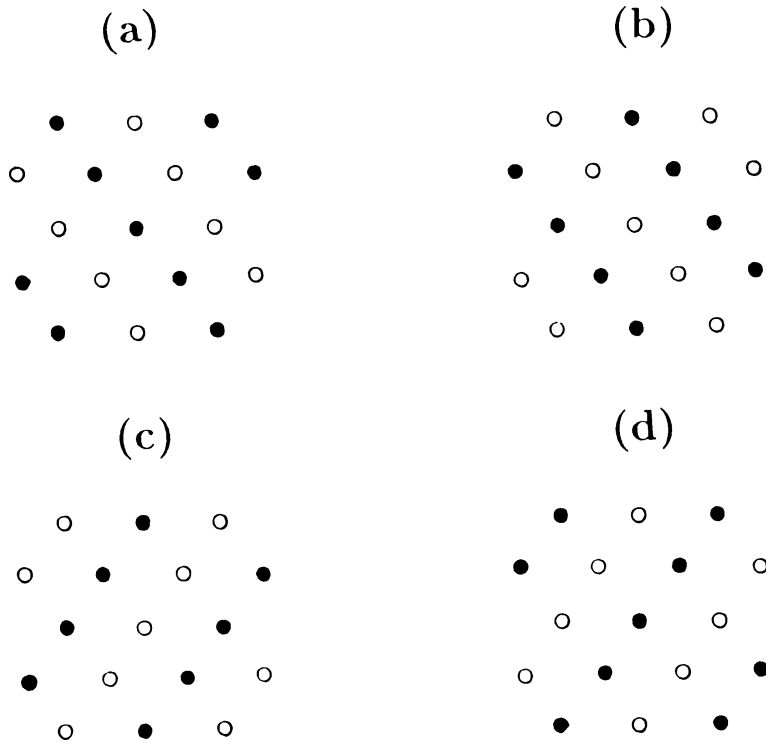

Figure 5.11: Four different realizations of the p(2x1) structure on the square lattice.

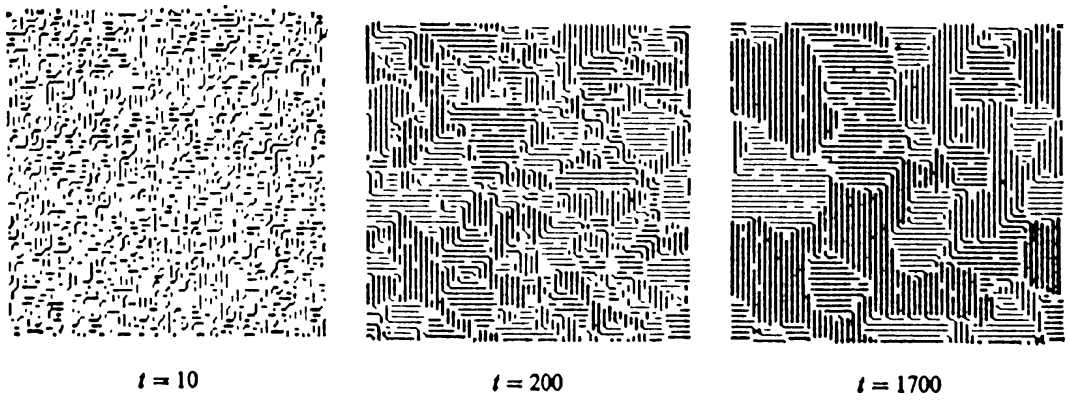

Figure 5.12: Snapshot simulation pictures for the time evolution of the quenched lattice gas model. The symmetry of the model corresponds to the p(2x1) **O-W**(110) adsorption system. Differently shaded regions correspond to the p(2x1) domains with different realizations of the p(2x1) structure as is shown in Fig. 5.11. The denoted "time" t corresponds to the number of the Monte Carlo step per atom (after [148]).

the **O-W**(110) adsystem. The $T = 0$ ground state of this system corresponds to the p(2x1) structure, and this ground state is four–fold degenerated as shown in Fig. 5.11. The results of the MC simulation at quench of the system to a temperature $T < T_{\rm crit}$ shown in Fig. 5.12, clearly demonstrate the dynamics of fluctuationally–driven creation and further evolution of domains of the p(2x1) structure.

It is important to note that the relaxation dynamics is very sensitive to the concrete realization of elementary MC steps. For example, when we investigate an adlayer with the help of the two-dimensional lattice-gas model with Kawasaki dynamics, the equilibrium state does not depend on the distance of the atomic jump in a single MC act. But the system dynamics obtained with the jumps to the nearest neighboring sites only, differs essentially from the dynamics which assumes additionally the possibility of "long jumps" to more removed lattice sites.

The relaxation dynamics is widely used to study the heterogeneous crystal growth. For example, the growth of the first layer on some substrate may be investigated with the help of the two-dimensional lattice model with the Hamiltonian (5.85), while the growth of three-dimensional crystals, with the help of the SOS model (5.89). In both cases we have to take into account the contact of the system with the particle reservoir (gas over the crystal surface). As the initial configuration it is natural to take an empty lattice, and as the elementary MC steps, the adsorption–desorption processes as well as the diffusion jumps. Recall that with the SOS model we can simulate the growth of the perfect crystal only, because the creation of bubbles (vacancies) is forbidden in this model. Such a model is not too realistic at temperatures close to the melting temperature. But for simulation of the crystal growth from a beam or a solution at strongly nonequilibrium conditions, the SOS model is quite satisfactory.

It is important to note that if in the SOS model we take for the chemical potential μ of the particle's reservoir (gas) a value, which is not equal to the value of the chemical potential μ_s characteristic for the equilibrium state of an ideal crystal at the given temperature, then the growth process at $\mu > \mu_s$ (or the evaporation of the crystal if $\mu < \mu_s$) will continue for an infinitely long time, because the model reservoir is "inexhaustible". In this case the sequence of the MC configurations $\{\xi_n\}$ approaches some stationary (but nonequilibrium) state as $n \to \infty$. Because new atoms income to the surface at random sites, the shape of the growing crystal is rough always (the so-called *kinetic rough*). A degree of this rough is high, when the rate of adsorption–desorption processes significantly exceeds the rate of diffusion, because the diffusion processes try to "smooth" surface defects.

From experiments it is known that the heterogeneous crystal growth may be carried out by one among three different mechanisms: first, by the *Frank–van der Merwe* (layer-by-layer) mechanism, second, by the *Vollmer–Weber* mechanism (when separate islands of three-dimensional crystals start to grow from the

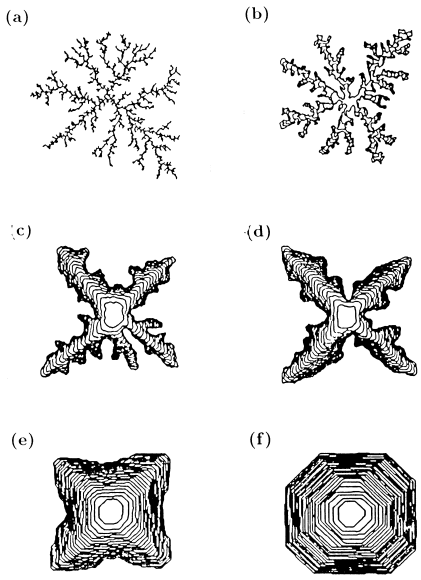

Figure 5.13: Diffusion limited growth of two-dimensional crystals for different diffusion times n_τ. (a) $n_\tau = 0$, 3600 atoms, (b) $n_\tau = 10$, 11000 atoms, (c) $n_\tau = 50$, 19000 atoms, (d) $n_\tau = 100$, 23000 atoms, (e) $n_\tau = 1000$, 23000 atoms, and (f) $n_\tau = 10000$, 27000 atoms aggregated into the crystal (after [152]).

very beginning), and third, by the *Stranski–Krastanov* mechanism (when the first layer grows by the first mechanism, while the second and consequent layers, by the second mechanism). All these scenarios may be studied with the MC simulation, if we take into account in the Hamiltonian (5.89) of the SOS model, that the one-particle energies $\varepsilon_{(h)}$ and the the parameters ϕ for the interatomic interactions, may take one values in the first layer, and another ones in the second and consequent layers [149].

Finally, choosing an appropriate initial configuration ξ_0, we may model the crystal growth over the screw dislocation, or on the stepped (vicinal) surface, *etc.* (e.g., see Müller-Krumbhaar's article in the book [132]).

In the simulation process we may support artificially an unstable state, for example, by injecting constantly new particles into the system as it was described above. As an example, let us consider also the so-called *diffusion limited crystal growth* (e.g., see [150, 151]), and, for a change, consider a case when the surface of the growing crystal is not flat.

Above (in Sec. 3.4) in the discussion of the simulation of the crystal growth by the MD method, we have noted that the energy of incoming atoms may essentially exceed the average thermal energy of the substrate atoms. In this case the incoming atom gives the surplus of energy into the substrate for some "thermalization time" τ. During the time τ the atom has an abnormally high mobility, and it may do several jumps on the surface even at zero substrate temperature. Now we demonstrate how this process may be modeled with the MC method [152].

Let us take the 2D lattice model and place in its center a "nucleus" (a single immobile atom), to which the incoming atoms can "stick". Then, let us inject atoms one by one into the lattice (for example, from the left-hand side of the lattice), and allow for the atoms to walk randomly over the lattice until they meet with the nucleus. Then, let us suppose that after the meeting with the nucleus, the atom can carry out on the nucleus' surface a limited number of random walks only, e.g., n_τ walks, and after that the atom must be irreversibly aggregated on the visiting site having the maximum number of nearest neighbors. If $n_\tau \gg 1$, then with a high probability during the diffusion over the surface of the growing crystal, the atom will find a site with the lowest potential energy (i.e., with the maximum number of neighbors), so that the atom will join itself to a "true" site of the growing crystal. In a result, the growing crystal is the "ideal" isometric

octagonal crystal bounded by equally developed (10) and (11) faces as is shown in Fig. 5.13f. Otherwise, if $n_\tau = 0$, the incoming atom "sticks" to a random place on the surface, and in the result the random fractal pattern with the fractal dimension $d = 1.70$ is growing (see Fig. 5.13a). Finally, at intermediate values of the allowed number of random walks n_τ, the so-called *dendritic growth* is observed (for example, see the $n_\tau = 100$ case in Fig. 5.13d) [190]. In the latter case, it is growing a clear anisotropic patter with protrusions, which are then transformed into well defined branches, and latter produce new protrusions, *etc.*

5.6 Accelerating relaxation to equilibrium

As was shown in the previous Sec. 5.5, the Metropolis procedure imitates successfully the system evolution to the equilibrium state. In real life, however, a single "attempt" of the transition to a new configuration takes a time $\sim 10^{-13}$ second, and all particles carry out these attempts simultaneously. Consequently, for one second a real system makes $\sim 10^{13} N_p$ attempts. On the other hand, in the computer experiment the particles try to pass to a new site one by one, and the number of these attempts per one second does not exceed $10^5 \div 10^8$. Therefore, "physical" times of approach to equilibrium in laboratory and computer experiments differ in $(10^5 \div 10^8)N_p$ times! Thus, if the time of approach to equilibrium in a laboratory experiment is, for example, about few seconds, in the computer experiment we should wait for the achieving of equilibrium for few years as minimum.

One may indicate four classes of systems characterized by slow relaxation to equilibrium.

(a) All models where some conservation laws must be satisfied, exhibit the so-called *hydrodynamical slowing* (e.g., see [153]).

(b) All systems near the point of the second-order phase transition, exhibit the so-called *critical slowing.*

(c) The so-called *metastable slowing* is observed at first-order phase transitions, because for formation of a new phase, first islands (drops) of the new phase with critical sizes should be created, and this process typically include overcoming of an activation barrier.

(d) Some systems (such as glasses, overcooling alloys, *etc.*) have a number of metastable states characterized by an energy which is only slightly above the energy of the ground state. During the cooling, these systems with a very high probability are locked in one of the metastable states, and then the system may get out from the metastable state with simultaneous motion of many particles only.

Evidently that in all these cases, the Metropolis algorithm is in need of a modification. An original method of acceleration of the relaxation to equilibrium in the MC simulation, based on Darwin's evolution laws, was proposed by Shumway and Sethna [154]. The idea of the method consists in using of "adaptable" Monte Carlo steps, which are modified during the system evolution in such a way that to achieve the most fast approach to the equilibrium.

Recall briefly the standard Metropolis procedure. At each step of construction of the sequence $\{\xi_n\}$ of the MC configurations, we have constructed from the old configuration ξ_{n-1} the trial configuration ξ^*, and then we have been decided, either to take ξ^* as the new configuration ξ_n, or to throw it out. At this step, we have selected randomly only one variant among a set of variants of the ξ^* construction (for example, we have chosen randomly the particle's number, the length and the direction of the displacement). Let the set of different variants of construction of ξ^* from ξ_{n-1} constitutes some set ("population") $\mathcal{K}$ consisting of a finite number K of elements. Some of these variants (elements of the set $\mathcal{K}$) approach the system to the equilibrium, others are wittingly unacceptable. The idea of the Shumway and Sethna method is to select successful variants and throw out the unsuccessful ones and, moreover, just during the simulation to modify purposefully the procedure of the ξ^* construction with the aim to lower the number of unsuccessful attempts.

For this, let associate some "health" coefficient h_k, where $k = 1, 2, \ldots, K$, with each "member" of the "population" $\mathcal{K}$ (i.e., with each variant of the ξ^* construction). At the beginning of the construction of the sequence $\{\xi_n\}$, we put $h_k := 1$ for all k. Then, during the simulation we will change the coefficients h_k, increasing them for the "successful" variants of the ξ^* construction, and decreasing for the "unsuccessful" members of the population $\mathcal{K}$. Namely, let at the n-th step of construction of the sequence $\{\xi_n\}$, we have

chosen with the help of the random number generator one of the members of the population $\mathcal{K}$, for example, with the number k_0 ($1 \le k_0 \le K$), as the variant of the ξ^* construction. Then, after the construction and testing the trial configuration ξ^*, we change the health coefficient for the k_0-th element in the following way: if the transition $\xi_{n-1} \to \xi^*$ lowers the system potential energy, $\Delta V < 0$, we put

$$h_{k_0} := \alpha h_{k_0} + \beta |\Delta V| \,, \tag{5.92}$$

where α and β are some positive numerical coefficients, and $\alpha < 1$. Otherwise, if the attempt of the transition leads to increasing of the potential energy, $\Delta V > 0$, we simply put

$$h_{k_0} := \alpha h_{k_0} \,. \tag{5.93}$$

Clearly that if some of the variants of the ξ^* construction will lead to increasing of the system energy for the each time, its health coefficient will constantly decrease.

Now let us introduce some "critical health level" $h_{\rm crit}$ ($0 < h_{\rm crit} < 1$). In the case when the value h_k for some variant k will become lower than the critical level, $h_k < h_{\rm crit}$, let throw out the k-th variant from the set $\mathcal{K}$ of the ξ^* construction, and instead of it let us introduce some new variant of the ξ^* construction. For this, let choose two new elements k and k' from the set $\mathcal{K}$ in a random way, but taking into account their health coefficients (i.e., an element with a higher h_k value should be chosen with a higher probability). Then, using the k and k' variants, let us combine a new variant of the ξ^* construction, and include this variant into the set $\mathcal{K}$ instead of the removing one. The concrete procedure of construction of the new element depends on the model under investigation. For example, if the variants k' and k'' describe the displacement of one of the system particles on distances $\Delta r'$ and $\Delta r''$ correspondingly, then as the new variant we may take the displacement of the particle on the distance $\Delta r' + \Delta r''$ or $\Delta r' - \Delta r''$. Or, if the procedure of the ξ^* construction includes a simultaneous shift of few particles, and the variants k' and k'' correspond to shifts of n' and n'' particles respectively, then the new variant may correspond to the simultaneous shift of $n' + n''$ particles, *etc.* Clearly that the health coefficient of the new variant must be put to one.

During the carrying out of the procedure described above, the "unsuccessful" elements of the population $\mathcal{K}$ will be throw out ("died off"), while the "successful" elements will "survive", their health coefficients will increase and, moreover, they will be "reproduced", creating new elements ("offsprings") instead of the "died off". However, this procedure has a nonzero probability that at some step the population $\mathcal{K}$ may loss of diversity and collapse. To avoid such a scenario, from time to time we may introduce "mutations" into the population $\mathcal{K}$, for example, multiplying a randomly chosen element by some random number. A rate of producing of new "mutations" has to be not too slow (because this will not save the population against the collapse) as well as not too high (because in the latter case the useful role of the health coefficients will be lost). An optimum may correspond to the rate of "mutation" producing, which is of the same order as the rate of creation of "offsprings".

The concrete values for the coefficients α, β and $h_{\rm crit}$ may be chosen at the training of the computer program. Shumway and Sethna [154] applied the described method to study the Frenkel-Kontorova model (see below Sec. 6.4.2), where with temperature decreasing the system is locked in a glass-like state. Using the method described above, they achieved the acceleration of the approach to equilibrium (comparing with the standard Metropolis algorithm) in $10^4 \div 10^6$ times. During the simulation, the maximal distance of atomic displacements per one MC step were risen from $0.1a_0$ at the beginning to $0.6a_0$ at the end of simulation, and the number of simultaneously shifted atoms were increased from one to few dozens. Note, however, that the sequence $\{\xi_n\}$ calculated by the described above method, is not Markovian.

5.7 Comparison of the MD and MK methods

In conclusion let us compare the applicability of the two main methods of computer simulation to statistical physics problems.

Molecular dynamics method	**Monte Carlo method**
1. One may check, whether the system under investigation is integrable or non-integrable, and in the latter case one can test, does the "developed chaos" exists in the system, i.e., may the thermal equilibrium be achieved.	**1.** The existence of a contact with a thermal bath, which guaranties that the thermal equilibrium will always be achieved, is postulated.
2. Temporal averages are calculated.	**2.** Configurational averages (the averages over the ensemble) are calculated.
3. The method may be used for investigation of the far from equilibrium systems, as well as for studying non-stationary processes.	**3.** The applicability of the method is strictly verified for the thermodynamically equilibrium state of the system only.
4. The method gives the true dynamics of the system, including the description of wave processes.	**4.** The method gives the relaxation dynamics of the system, which only approximately imitates the approach to the equilibrium. All vibrational processes are lost.
5. The method is usually realized with a complex computational algorithm, needs a long processor time on a powerful computer, and can simulate a system consisting of a comparative small number of particles.	**5.** The method may be realized with a comparative simple and universal algorithm, allows to model systems of a large number of particles during an essentially longer time interval.

From this comparison one can see that each of the methods has its own dignities as well as deficiencies. Therefore, the choice of a suitable method first of all depends on the problem under investigation.

Chapter 6

Integrable Systems

Up to now we mainly tried to explain the reasons leading to stochastic dynamics of various physical systems, and described different methods of investigation of the chaotic equilibrium state with the help of a computer. Now we consider the opposite class of systems, namely the systems where the nonlinearity leads not to a chaos, but to the regular motion.

6.1 Fermi-Pasta-Ulam paradox

Perhaps, during all development of physics up to recent times, when a wide employing of computer modeling has been started, the harmonic approximation for solution of physical problems was used as the main approach. Namely, the potential energy of the system was expanded into Taylor series with respect to small displacements of atoms from their equilibrium positions up to square terms, so that the force acting on a given atom occurs to be a linear function of the displacements. In this case the system evolution is governed by a system of linear equations, and the latter always has an exact solution at least in principle. In a general case this solution can be expressed as a sum over independent normal modes, i.e. the system is described by a set of harmonic oscillators.

When the total energy of the system increases, the amplitude of vibrations rises, and this leads to an anharmonicity of the vibrations as well as to a nonlinear interaction of different modes. In other words, the set of harmonic oscillators may be interpreted as a set of quasiparticles (phonons, magnons, plasmons, *etc.*). Then, with increasing of the system total energy, the number of these quasiparticles increases too, and they begin to interact between themselves. In a classical approach these anharmonic (or nonlinear) effects are assumed to be small, and they are taken into account with the help of one or another variant of the perturbation theory. However, it turns out that, as a rule, such a perturbation series diverges and, consequently, the corresponding physical system is nonintegrable usually. In the result, practically in all systems a chaos emerges, and the statistical mechanics laws start to work leading the system to thermal equilibrium.

However, observing nature and life surrounding us, we see that real systems are often very far from the chaotic equilibrium state. This phenomenon is caused by two reasons. First, a majority of systems are open, i.e. an energy and/or particles constantly enter into the system and/or are taken away from it, and that causes the system to be in a nonequilibrium state. Second, in some "special" systems the nonlinearity leads not to a chaos, but, on the contrary, to an "ordering" of their state. In the result of co-action of both reasons, in some systems the processes, which are reverse to chaotization ones, the so-called *self-organizing processes*, take place. Such processes are studied by a new science called *sinergetics* [155, 156]. And, although a number of self-organized systems is rather small, their role in science is extremely high, because namely such systems surround us and, moreover, we present an example of such a system ourselves.

In the present Chapter we give a brief introduction to the theory of "special" nonlinear systems, where the existence of nonlinearity results in appearing of specific "solitonic" solutions. In developing of this

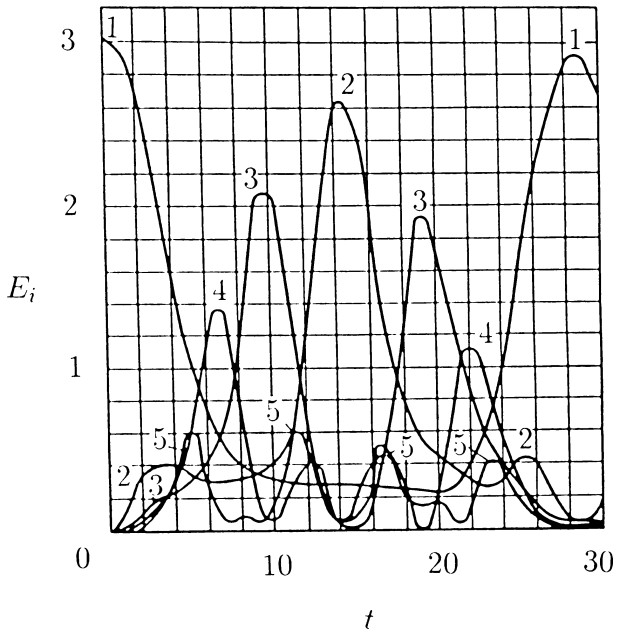

Figure 6.1: Fermi-Pasta-Ulam paradox. Energies of different linear modes E_i versus time t (after [157]).

theory the computer modeling played one of main roles as it did also in developing of the stochastic theory discussed in the previous chapters of this book.

Fermi, Pasta and Ulam [157] were the first who have met with such a problem. They tried to answer on the following question stated long time ago: why do the laws of statistical physics and thermodynamics operate, i.e. why does the kinetic energy is equally shared among the system degrees of freedom? The problem has being solved with the help of one of the first computers "MANIAC" by the method which now is known as the molecular dynamics method.

Fermi *et al.* [157] have considered a chain consisting of N_p atoms (it was taken $N_p = 32$ or 64), both chain's ends (i.e., the atoms with the numbers $n = 0$ and $n = N_p + 1$) were kept fixed. The anharmonic interaction (between the nearest neighbors only) was accounted with the help of two types of potentials, the potential with cubic anharmonicity,

$$v(x) = \frac{1}{2}g(x - a_0)^2 + \frac{1}{3}\alpha(x - a_0)^3, \tag{6.1}$$

and the potential with quadric anharmonicity,

$$v(x) = \frac{1}{2}g(x - a_0)^2 + \frac{1}{4}\beta(x - a_0)^4, \tag{6.2}$$

where g is the elastic constant, a_0 is the mean interatomic distance, and α and β are small coefficients. At $t = 0$ the chain was taken to be in rest, but the initial atomic displacements $u_n = x_n - na_0$ were taken as the following,

$$u_n(0) \propto \sin\left[\pi n/(N_p + 1)\right], \tag{6.3}$$

i.e. only the lowest linear mode was excited. It was expected that for a some time the thermalization should take place, i.e. the initial energy would be equally shared among all modes. In computer experiment, however, nothing similar to the thermodynamical behavior have been observed. An example of variation of the energies E_i versus time t for different linear modes is shown in Fig. 6.1. As seen, a part of the initial energy is transferred from E_1 to E_2, then to E_3 and so on, but after a certain time almost the whole energy comes back to the first mode. Moreover, the displacement of each atom returns to the initial value too. This recurrence phenomenon was called the *Fermi-Pasta-Ulam* (FPU) *paradox*.

To anticipate, we have to note that such a result was obtained because of a small capacity of early computers. In fact, the system described above is nonintegrable, so that a chaos will be arised after a long enough time interval [158]–[161]. Namely, the recurrence state does not exactly coincide with the initial state, and with each recurrence period the measure of chaos slowly increases. But the thermalization time is very long, so that the simulation of such problems need a more high-speed computer. That means that the given system is "close" to the exactly integrable one.

The attempts to resolve the FPU paradox have led to a drastic acceleration of the rate of development of the theory of nonlinear systems and to discovery of exactly integrable nonlinear systems exhibiting solitonic solutions. Zabusky and Kruskal [162] were the first who have done this with the help of computer simulation.

6.2 Solitons

6.2.1 Boussinesq's equation

First let us deduce the motion equation for the FPU chain. To do this, let us take the potential energy of the infinite atomic chain in the following form,

$$V = \sum_n v\left(u_{n+1} - u_n\right), \tag{6.4}$$

so that in accordance with the second Newton law we obtain the motion equations

$$m_a \ddot{u}_n = -\frac{\partial V}{\partial u_n} = v'(u_{n+1} - u_n) - v'(u_n - u_{n-1})\,. \tag{6.5}$$

Further we put $m_a = 1$ for the atomic mass.

Let us expand the interaction potential $v(\Delta u)$ into Taylor series and restrict ourselves by accounting of cubic anharmonicity only as, e.g., in Eq. (6.1),

$$v'(\Delta u) \approx v'(0) + v''(0)\,\Delta u + \frac{1}{2} v'''(0)\,\Delta u^2. \tag{6.6}$$

Now let us consider long-wave disturbances only, that allows to use the so-called continuum approximation. Namely, if we suppose that

$$|u_{n+1} - u_n| \ll a_0\,, \tag{6.7}$$

we may turn of from the discrete variable n to a continuous variable x ($x = na_0$), $u_n(t) \to u(x,t)$. Then, it is easy to deduce the following expressions,

$$u_{n\pm1} = u_n \pm a_0 u' + \frac{1}{2} a_0^2 u'' \pm \frac{1}{3!} a_0^3 u''' + \frac{1}{4!} a_0^4 u'''' + \cdots, \tag{6.8}$$

$$u_{n+1} - u_n = a_0 u' + \frac{1}{2} a_0^2 u'' + \frac{1}{3!} a_0^3 u''' + \frac{1}{4!} a_0^4 u'''' + \cdots, \tag{6.9}$$

$$u_n - u_{n-1} = a_0 u' - \frac{1}{2} a_0^2 u'' + \frac{1}{3!} a_0^3 u''' - \frac{1}{4!} a_0^4 u'''' + \cdots, \tag{6.10}$$

$$(u_{n+1} - u_n)^2 = a_0^2 u'^2 + a_0^3 u' u'' + \cdots, \tag{6.11}$$

$$(u_n - u_{n-1})^2 = a_0^2 u'^2 - a_0^3 u' u'' + \cdots\,. \tag{6.12}$$

Further, we will exploit the following notations commonly used for nonlinear equations:

$$\frac{\partial u(x,t)}{\partial x} \equiv u_x\,, \quad \frac{\partial u(x,t)}{\partial t} \equiv u_t\,, \quad \frac{\partial^2 u(x,t)}{\partial^2 x} \equiv u_{xx}\,, \quad etc. \tag{6.13}$$

Substituting Eqs. (6.6), (6.8–6.12) into Eq. (6.5), we obtain the equation

$$u_{tt} = s^2 u_{xx} + B^2 u_{xxxx} - \Lambda u_x u_{xx} \,, \tag{6.14}$$

where s is the sound velocity in the chain,

$$s^2 = a_0^2 v''(0) \,, \tag{6.15}$$

the constant B characterizes the dispersion of linear waves,

$$B^2 = \frac{1}{12} a_0^4 v''(0) \,, \tag{6.16}$$

the coefficient Λ describes the nonlinearity of interaction,

$$\Lambda = -a_0^3 v'''(0) \,, \tag{6.17}$$

and usually $v'''(0) < 0$ so that $\Lambda > 0$.

Next, we have to turn to dimensionless variables. For this, let us introduce new variables

$$\tilde{t} = t/t_0 \,, \;\; \tilde{x} = x/x_0 \,, \;\; \text{and} \;\; \tilde{u} = u/u_0 \,, \tag{6.18}$$

and rewrite Eq. (6.14) in the form

$$\tilde{u}_{\tilde{t}\tilde{t}} - \frac{s^2 t_0^2}{x_0^2} \, \tilde{u}_{\tilde{x}\tilde{x}} - \frac{B^2 t_0^2}{x_0^4} \, \tilde{u}_{\tilde{x}\tilde{x}\tilde{x}\tilde{x}} + \frac{\Lambda u_0 t_0^2}{x_0^3} \, \tilde{u}_{\tilde{x}} \tilde{u}_{\tilde{x}\tilde{x}} = 0 \,. \tag{6.19}$$

It is natural to choose the values t_0, x_0 and u_0 in the following way:

$$x_0 = s t_0 = B/s \,, \;\; t_0 = B/s^2, \;\; \text{and} \;\; u_0 = sB/\Lambda \,. \tag{6.20}$$

Then Eq. (6.19) takes the canonical form,

$$u_{tt} - u_{xx} - u_{xxxx} + u_x u_{xx} = 0 \,, \tag{6.21}$$

where we have omitted all tildes in order to reduce the notations. Equation (6.21) is just the *Boussinesq equation* which we were looking for.

6.2.2 Competition of dispersion and nonlinearity

First let us investigate Eq. (6.21) on a qualitative level. If we will keep in Eq. (6.21) the two first terms only, we obtain the simplest wave equation,

$$u_{tt} - u_{xx} = 0 \,. \tag{6.22}$$

This equation is solved with the help of Fourier transform, and the solutions correspond to the plane waves,

$$u(x,t) \propto \exp\left[i\,\omega(k) t - i\,kx\right] , \tag{6.23}$$

with the wave length $\lambda = 2\pi/|k|$ and the frequency $\omega(k) = |k|$. These waves may propagate in both directions, and the phase velocity of all waves is the same, $s = 1$ in the chosen system of units.

Now let us take into account the third term in Eq. (6.21), i.e. let us consider the equation

$$u_{tt} - u_{xx} - u_{xxxx} = 0 \,. \tag{6.24}$$

This is the linear equation too, and its solutions are again the plane waves (6.23), but now with the dispersion law

$$\omega(k) = \pm k \sqrt{1 - k^2} . \tag{6.25}$$

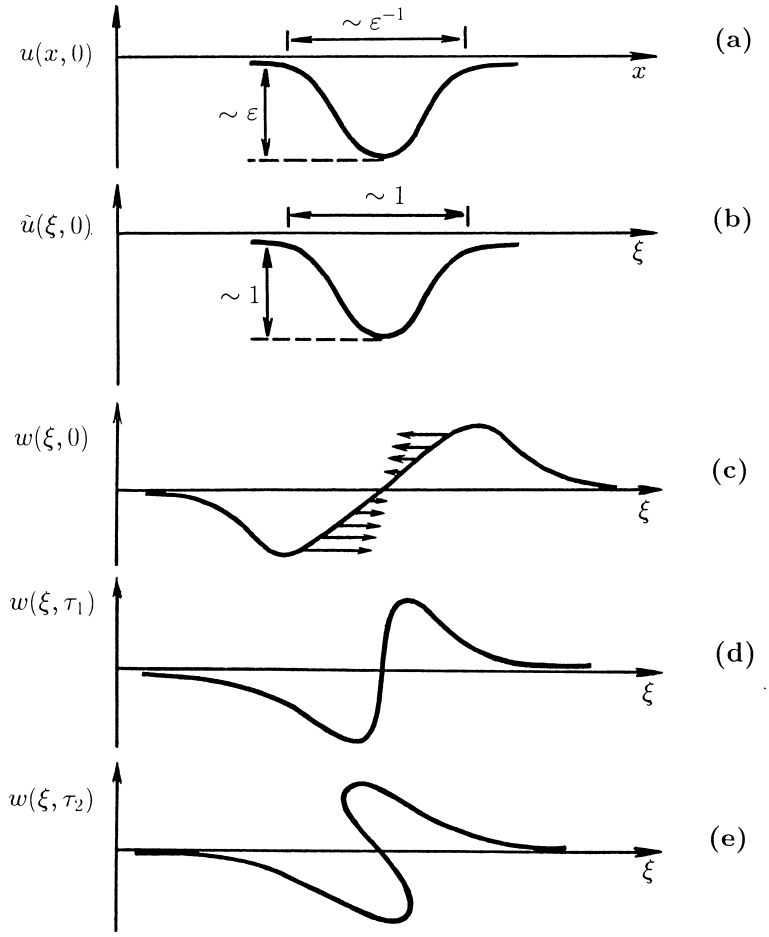

Figure 6.2: Evolution of the localized region of compression ($\tau_2 > \tau_1 > 0$) (from [164]).

Thus, now the phase velocity $v_{\rm ph}$ of the waves depends on the wave momentum,

$$v_{\rm ph} = \frac{\omega(k)}{k} = \pm\sqrt{1-k^2}. \tag{6.26}$$

Consequently, a lower is the wave length, the lower is its velocity. Therefore, if at the initial time moment we have created a localized wave packet constructing of the harmonics (6.23) with different k, then the packet will spread (delocalize) during its evolution. Namely due to this fact the dispersed linear systems are inappropriate for transmission of information.

Now let us investigate a role of the last term in Eq. (6.21) which is connected with the chain's nonlinearity. For this, let us neglect for a moment by the dispersion, and consider the equation

$$u_{tt} - u_{xx} + u_x u_{xx} = 0\,. \tag{6.27}$$

For the sake of simplicity, let us consider the waves propagating in one direction only, for instance, from the left-hand to the right-hand side. In this case it is convenient to turn to the moving reference frame, the center of which moves with the sound speed $s = 1$, i.e. to change from the coordinate x to the new coordinate $\xi \propto x - st$.

Let us suppose that at the initial time moment we had a wave packet with a small amplitude of the order ε ($\varepsilon \ll 1$), which was localized in a wide region of extension $\sim \varepsilon^{-1}$ as is shown in Fig. 6.2a. To describe the evolution of the packet, it is convenient to turn to the so-called Gardner variables introduced according to the following relations,

$$x \to \xi = \varepsilon\,(x - t)\,, \tag{6.28}$$

$$t \to \tau = \frac{1}{2}\,\varepsilon^3 t\,, \tag{6.29}$$

$$u \to \tilde{u} = \frac{1}{\varepsilon}\,u\,. \tag{6.30}$$

In these new variables the packet will be localized in a region of extension ~ 1, and it will have an amplitude ~ 1 (see Fig. 6.2b). Then, according to standard differential rules, we have

$$\frac{\partial}{\partial x}\cdots = \frac{\partial \xi}{\partial x}\frac{\partial}{\partial \xi}\cdots + \frac{\partial \tau}{\partial x}\frac{\partial}{\partial \tau}\cdots, \tag{6.31}$$

$$\frac{\partial}{\partial t}\cdots = \frac{\partial \xi}{\partial t}\frac{\partial}{\partial \xi}\cdots + \frac{\partial \tau}{\partial t}\frac{\partial}{\partial \tau}\cdots, \tag{6.32}$$

and from Eqs. (6.28) and (6.29) it follows that

$$\frac{\partial \xi}{\partial x} = \varepsilon\,,\ \ \frac{\partial \tau}{\partial x} = 0\,,\ \ \frac{\partial \xi}{\partial t} = -\varepsilon\,,\ \ \frac{\partial \tau}{\partial t} = \frac{1}{2}\varepsilon^3. \tag{6.33}$$

Substituting Eqs. (6.28–6.33) into Eq. (6.27) and keeping the terms up to $\sim \varepsilon^5$ inclusively, we get the equation

$$\tilde{u}_{\xi\tau} - \tilde{u}_\xi \tilde{u}_{\xi\xi} = 0\,. \tag{6.34}$$

Introducing the variable

$$w = \tilde{u}_\xi\,, \tag{6.35}$$

which describes the chain's deformation in the moving frame, we finally obtain the following equation for the variable w:

$$w_\tau - w w_\xi = 0\,. \tag{6.36}$$

Emphasize ones more that Eq. (6.36) approximately describes the evolution of the wave packet of small amplitude which translates in one spacial direction only. The procedure of turning from the original equation (6.27) to the approximate equation (6.36) is called the method of coordinate extension, or the many–scale method. This method is described in more detail in, e.g., the book [163], where, in particular, it is explained why we have to introduce the coefficient ε^3 in Eq. (6.29).

Using the direct substitution, we may check that in an implicit form the solution of Eq. (6.36) is given by the expression

$$w(\xi, \tau) = f(\xi + w\tau)\,, \tag{6.37}$$

where $f(z)$ is an arbitrary function. For example, if $f(z) = az$, so that at the initial time moment the deformation w had a linear section,

$$w(\xi, 0) = a\xi\,, \tag{6.38}$$

then during the evolution the slope of this section will increase,

$$w(\xi, \tau) = a\xi/(1 - a\tau)\,. \tag{6.39}$$

For a qualitative analysis of evolution of a perturbation with an arbitrary shape, let us suppose that the variables ξ and τ depend on some parameter α, so that $\xi = \xi(\alpha)$ and $\tau = \tau(\alpha)$. Then we may write

$$\frac{dw}{d\alpha} = \frac{\partial w}{\partial \xi}\frac{d\xi}{d\alpha} + \frac{\partial w}{\partial \tau}\frac{d\tau}{d\alpha}\,. \tag{6.40}$$

Let us follow now for evolution of points of a constant profile. For this, we have to find a solution of the equation

$$w(\tilde{\xi}, \tilde{\tau}) = \mathrm{Const}\,. \tag{6.41}$$

Substituting Eq. (6.41) into Eq. (6.40) and using Eq. (6.36), we get

$$\frac{d\tilde{\xi}}{d\tilde{\tau}} = \frac{\xi_\alpha}{\tau_\alpha} = -\frac{w_\tau}{w_\xi} = -w\,. \tag{6.42}$$

Consequently, evolution of a nonlinear wave is reduced to that the points on the graph of the solution moves horizontally with a speed equal to the value of the deformation taken with the opposite sign. In a result, the profile of translating wave is modified as time increases. For example, if at the initial instant a localized region of compression existed in the chain as shown in Fig. 6.2a, then during the evolution this region is being narrowed (Fig. 6.2d) and, after some time, the front of the wave will be overturned (Fig. 6.2e) [164]. At the same time, the linear section at the center of the wave will behave in accordance with Eq. (6.39).

Note that if the parameter of the interaction's nonlinearity has the another sign, i.e. if $v'''(0) > 0$, then in Eqs. (6.21), (6.27), (6.34) and (6.36) we have to change the sign in the front of the nonlinear term. In this case the theory predicts the analogous behavior for a localized region of the chain's extension.

Now we may predict the behavior of a solution of Boussinesq's equation (6.21). Namely, the dispersion term leads to smoothing of a wave packet which was localized at the beginning, while the nonlinear term, on the contrary, leads to the narrowing of the region of chain's compression and to increasing (sharping) of the wave front. Thus, we expect that during the evolution of the localized compressed region, a balance between these two processes would be achieved, and in a result some localized pulse with a permanent shape will translate. Computer modeling of Zabusky and Kruskal [162] confirmed this prediction and also showed that although the wave may become very sharp, it will never be overturned if the dispersion term is taken into account.

6.2.3 Waves of stationary shape

If we apply the method of coordinate extension to the full Boussinesq equation (6.21), then instead of Eq. (6.36) we obtain the equation

$$w_\tau \mp 6ww_\xi + w_{\xi\xi\xi} = 0\,. \tag{6.43}$$

Equation (6.43) is called the *Korteweg-de Vries* (KdV) *equation* [165]. The sign "minus" ("plus") in the front of the nonlinear term emerges at the description of waves of compression (expansion) in the case of $v'''(0) < 0$ $(v'''(0) > 0)$, when the deformations $w(\xi, \tau)$ as a function of ξ at a fixed time τ has a shape with one or several hollows (humps). The numerical factor 6 before the nonlinear term is introduced for convenience, and it may be changed by scaling the variable w.

Equation (6.43) is the equation with partial derivatives. Let us try to find one of its particular solutions, namely the solution which may be presented in the form

$$w(\xi, \tau) = \Phi(z)\,, \quad z = \xi - c\tau\,. \tag{6.44}$$

The solutions of the type (6.44) is called the *wave of stationary shape* (WSS), and it describes a wave which propagates without changing its shape. Note that the speed c in Eq. (6.44) corresponds to the WSS's speed in moving reference frame, so that in the rest frame the WSS's speed is equal to $(s+c)$, where s is the sound velocity in the chain (for our system of units $s = 1$).

Substituting Eq. (6.44) into Eq. (6.43), we obtain an ordinary differential equation (further for the sake of definiteness we will consider only the upper sign in Eq. (6.43)):

$$-c\Phi' - 6\Phi\Phi' + \Phi''' = 0\,, \tag{6.45}$$

which may be rewritten in the form

$$(-c\Phi - 3\Phi^2 + \Phi'')' = 0\,. \tag{6.46}$$

Integration of Eq. (6.46) yields

$$\Phi'' = c\Phi + 3\Phi^2 + C_1\,. \tag{6.47}$$

Multiplying Eq. (6.47) by Φ' and then integrating it, we get

$$\frac{1}{2}\Phi'^2 = \frac{1}{2}c\Phi^2 + \Phi^3 + C_1\Phi + C_2\,. \tag{6.48}$$

Equation (6.48) may be rewritten in the form

$$\Phi'(z) \equiv \frac{d\Phi}{dz} = \pm\sqrt{c\Phi^2 + 2(\Phi^3 + C_1\Phi + C_2)}\,, \tag{6.49}$$

and after that it may be integrated ones more:

$$z = z_0 \pm \int d\Phi \Big/ \sqrt{c\Phi^2 + 2(\Phi^3 + C_1\Phi + C_2)}\,. \tag{6.50}$$

Here C_1, C_2 and z_0 are the integration constants. The integral in Eq. (6.50) is the *elliptic integral*. Equation (6.50) presents the WSS solution in an implicit form. In order to get the explicit solution, we have to find the corresponding inverse function (such functions are called the elliptic functions too).

Now let us choose the integration constants in such a way that Eq. (6.44) would describe the spatially localized wave packet. For this, let us impose additionally the following boundary conditions,

$$\Phi(z),\ \Phi'(z) \to 0 \ \text{ at } \ z \to \pm\infty\,, \tag{6.51}$$

that leads to the values $C_1 = C_2 = 0$ (the constant z_0 called the phase of the WSS, may be easily removed by an appropriate choice of the beginning of coordinates). Then the function (6.44), (6.51) will describe the so-called *solitary* WSS (SWSS). In this case the integral (6.50) may be expressed through elementary functions, and the solution has the form

$$\Phi(z) = -w_0\,\mathrm{sech}^2\left(z\sqrt{w_0/2}\right), \tag{6.52}$$

where w_0 is the amplitude of the SWSS which is coupled with its speed c by the relationship

$$c = 2w_0\,. \tag{6.53}$$

Emphasize that $c > 0$, i.e., the SWSS of the KdV equation propagates with the faster-than-sound speed, and a higher is the amplitude of a single translating pulse, the higher is its speed.

Note that the original Boussinesq equation (6.21) has a solution of a SWSS form too:

$$u(x,t) = -6\sqrt{c^2-1}\left\{1 + \tanh\left[\frac{1}{2}\sqrt{c^2-1}\,(x-ct)\right]\right\}. \tag{6.54}$$

The displacement field for the SWSS of the Boussinesq equation is shown in Fig. 6.3a, and the deformation which follows the SWSS, in Fig. 6.3b. The deformation within the SWSS is described by the expression

$$w(x,t) \equiv \frac{\partial u(x,t)}{\partial x} = -3(c^2-1)\,\mathrm{sech}^2\left[\frac{1}{2}\sqrt{c^2-1}\,(x-ct)\right]. \tag{6.55}$$

As seen, the deformation is bell–shaped with a width $\sim (c^2-1)^{-1/2}$ and an amplitude $\sim (c^2-1)$ similarly to the SWSS (6.52) of the KdV equation. As for the field of displacements itself, it undergoes a jump $\Delta u(t) \equiv u(+\infty,t) - u(-\infty,t) = -12\sqrt{c^2-1}$. Contrary to the SWSS of the KdV equation, the SWSS (6.54) may propagate from the left-hand to the right-hand side ($1 < c < \infty$) as well as in the opposite direction ($-\infty < c < -1$).

Finally, for the FPU chain of atoms with the quadric anharmonicity (6.2), the method described above leads to the so-called *modified Korteweg-de Vries equation*,

$$w_\tau + 6w^2 w_\xi + w_{\xi\xi\xi} = 0\,, \tag{6.56}$$

which has the SWSS-type solution too:

$$w(\xi,\tau) = \pm\sqrt{c}\,\mathrm{sech}\left[\sqrt{c}\,(\xi - c\tau)\right],\ \ c > 0\,. \tag{6.57}$$

Emphasize that the modified Korteweg-de Vries equation allows both types of the SWSS's, the compression waves as well as the expansion ones.

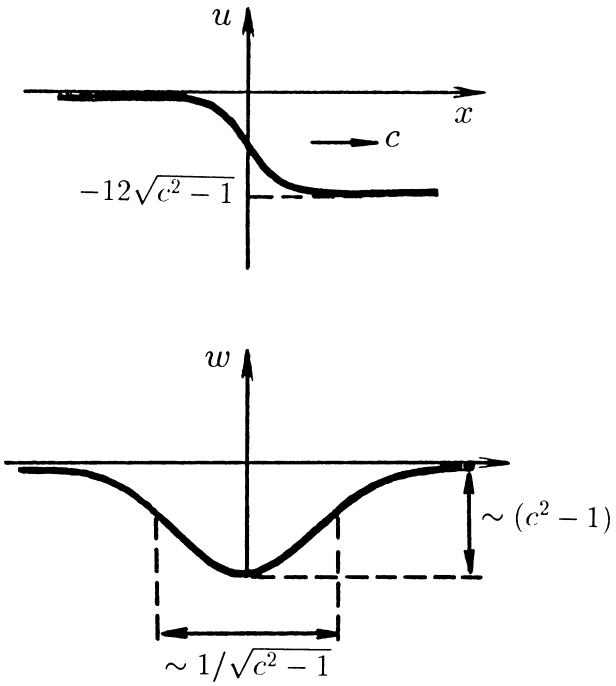

Figure 6.3: (a) Displacements $u(x,t)$ and (b) deformations $w(x,t) \equiv \partial u(x,t)/\partial x$ of the wave of stationary shape at a fixed time moment t for Boussinesq's equation (6.21) (c is the wave velocity) (from [164]).

6.2.4 Atomic chain in an external periodic potential

Let us consider now another characteristic example — a chain of interacting atoms, where each of the atoms is subjected to some external potential $v_{\rm sub}(x)$. Let for the sake of simplicity the interaction between the atoms is harmonic,

$$v(\Delta x) = \frac{1}{2}\, g\,(\Delta x - a_0)^2, \;\; \Delta x = x_{n+1} - x_n\,, \tag{6.58}$$

and let the external potential depends only on the displacement $u_n = x_n - x_n^{(0)}$ of a given atom with respect to its equilibrium position $x_n^{(0)} = na_0$ in a free chain. Assuming also that $v'_{\rm sub}\left(x_n^{(0)}\right) = 0$, let us write the potential $v_{\rm sub}(x)$ in the form

$$v_{\rm sub}(x_n) = \omega_0^2\, v_s(u_n)\,, \tag{6.59}$$

where the function $v_s(u)$ is normalized in such a way that $v_s''(0) = 1$, so that ω_0 corresponds to the frequency of vibrations of an isolated atom in the external potential. Then the motion equations take the form ($m_a = 1$):

$$\ddot{u}_n = g\,(u_{n+1} + u_{n-1} - 2u_n) - \omega_0^2\, v_s'(u_n)\,. \tag{6.60}$$

When all displacements are small, $|u_n| \ll a_0$, Eq. (6.60) may be linearized, $v'(u_n) \approx u_n$. In this case it has solutions in a form of plane waves, $u_n(t) \propto \exp\left[i\omega(k)t - ikna_0\right]$, with an optical dispersion law,

$$\omega(k) = \sqrt{\omega_0^2 + 2g\,(1 - \cos a_0 k)} \approx \sqrt{\omega_0^2 + s^2k^2}, \tag{6.61}$$

where $s = a_0\sqrt{g}$ is the sound velocity in the free chain.

Otherwise, if only the relative displacements are small, $|u_{n+1} - u_n| \ll a_0$, we may use the continuum approximation described above, $u_n(t) \to u(x,t)$. Substituting Eqs. (6.8–6.10) into Eq. (6.60), in this case we obtain the equation

$$u_{tt} - s_0^2\, u_{xx} + \omega_0^2\, v_s'(u) = 0\,. \tag{6.62}$$

In the linear case, when $v_s'(u) = u$, Eq. (6.62) is reduced to the so-called *Klein-Gordon equation*.

Analogously to the method used above, let us try to find one of particular solutions of Eq. (6.62), namely the solution of the WSS form,

$$u(x,t) = \Phi(z)\,, \quad z = x - ct\,. \tag{6.63}$$

In the present case, substituting Eq. (6.63) into Eq. (6.62) and then making the simple transformations,

$$\gamma^2\Phi'' = g\,v_s'(\Phi)\,, \tag{6.64}$$

$$\frac{1}{2}\gamma^2\Phi'^2 = g\,[v_s'(\Phi) + C_1]\,, \tag{6.65}$$

$$\Phi' \equiv \frac{d\Phi}{dz} = \pm\frac{1}{\gamma d}\sqrt{2[v_s'(\Phi) + C_1]}\,, \tag{6.66}$$

where

$$\gamma = \sqrt{1 - c^2/s^2} \tag{6.67}$$

and

$$d = s/\omega_0\,, \tag{6.68}$$

leads to the implicit solution in the form of the integral

$$z = z_0 \pm \gamma d \int d\Phi/\sqrt{2[v_s'(\Phi) + C_1]}\,. \tag{6.69}$$

Further, let us assume that the potential $v_{\rm sub}(x)$ has at least two equivalent absolute minima. Then the ground state of the system will be doubly degenerated, because two ground states will be possible: the state GS_1, where all atoms are situated in one minimum, and the state GS_2, where the atoms occupy the another minimum. Impose now on the function (6.63) the following boundary conditions: let the left-hand section of the chain (at $z \to -\infty$) is in the GS_1, while the right-hand section (at $z \to +\infty$), in the GS_2 (to satisfy these conditions, we have to choose the integration constants and the integration limits in Eq. (6.69) in an appropriate way). In a result, we obtain a solitary WSS called the *kink*. Contrary to the SWSS of the KdV equation, the kink is topologically stable, so that it can exist in an immobile state too.

The following three variants of choice of the external potential are used most often:

(a) the model with the piecewise parabolic potential

$$v_s(u) = \frac{1}{2}\,(|u| - 1)^2\,; \tag{6.70}$$

in this case the kink has the following shape:

$$\Phi(z) = \pm\,[1 - \exp\,(-|z|/\gamma d)]\,\operatorname{sign} z\,; \tag{6.71}$$

(b) the ϕ^4-model, where

$$v_s(u) = \frac{1}{4}\,(u^2 - 1)^2, \tag{6.72}$$

$$\Phi(z) = \pm\tanh\,(z/\gamma d\sqrt{2})\,; \tag{6.73}$$

(c) the *sine-Gordon* (SG) *model*, where

$$v_s(u) = 1 - \cos u\,, \tag{6.74}$$

$$\Phi(z) = 4\tan^{-1}\,\exp\,(\pm z/\gamma d) + 2\pi n\,, \tag{6.75}$$

n being an integer. The signs + or − in Eqs. (6.71), (6.73) and (6.75) describe the kink (which links GS_1 with GS_2) or the antikink (that links GS_2 with GS_1) correspondingly.

As one can see, in all cases the kink is localized within the region of extension of the order d (d is called the kink width), and the kink can move with the subsonic velocity only, $|c| < s$. When the kink moves, its width is equal to γd, i.e. the faster is the kink, the more narrow is it (the "relativistic" narrowing).

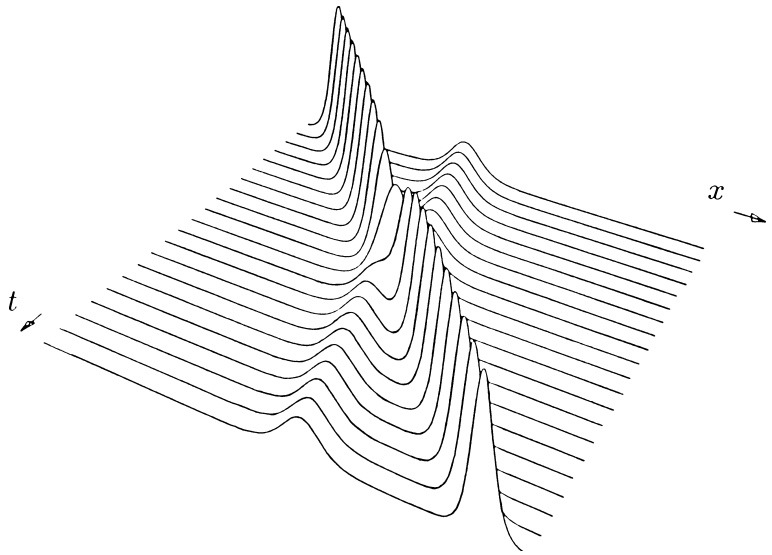

Figure 6.4: Collision of two waves of the stationary shape for the Korteveg-de-Vris equation (from [163]).

6.2.5 Solitonic solutions

Thus, some partial differential equations admit particular solutions of the SWSS form. Such a solution describes the spatially localized perturbation which propagates through the system without changing the shape. But this solution describes an isolated SWSS, i.e. it corresponds to the case when there are no other perturbations in the system except the SWSS. From the physical viewpoint, however, the question about a behavior of the SWSS in a general case, i.e. when other perturbations are present in the system, appears.

A stability of the SWSS with respect to small perturbations may be clarified with the help of the standard perturbation technique. For this, we have to substitute the expression

$$u(x,t) = \Phi(x - ct) + \delta u(x,t) \tag{6.76}$$

into the original equation with partial derivatives, and then to linearize the obtained equation in small perturbation $\delta u(x,t)$. In a result we get the linear equation for the function $\delta u(x,t)$, which has the same form as the Schrödinger equation for a fictitious particle moving in the potential determined by the function $\Phi(x)$. The SWSS stability is then determined by the eigenvalues λ_k of this Schrödinger equation: if even one of the eigenvalues is negative, then the given SWSS is unstable. For the examples described above, it occurs that $\lambda_k \geq 0$, and it exists only a single zero eigenvalue which corresponds to the so-called *Goldstone mode* which describes the possibility of a spacial displacement of the SWSS as a whole for an arbitrary distance.

It is very important, however, to investigate the stability of the SWSS not only with respect to small perturbations, but also for a more general case, for example, to check the SWSS stability when it collides with a large-amplitude wave packet or with another SWSS. The simplest and the most reliable method of solution of this problem is to use computer simulation. For this, we have to take an appropriate initial state, and then to investigate its evolution with the help of the MD method. Characteristic examples of such experiments are presented in Figs. 6.4 to 6.6. Each wave profile in these figures describes the graph of the function $u(x,t)$ on x for some fixed time t. As a whole, the figures correspond to the superpositions of a number of such profiles for the increasing sequence of times separated by equal intervals.

Figure 6.4 demonstrates the collision of two SWSSs of the KdV equation, when one SWSS overtakes the another SWSS, the latter having the smaller amplitude (and, therefore, the smaller speed), and passes throughout it. Figure 6.5a shows the collision of two SWSSs of the SG equation with different polarities (the kink-antikink collision), while Fig. 6.5b, the collision of two kinks of the same polarity. As seen, in all these cases the solitary waves of stationary shape come out from the collision region, completely preserving

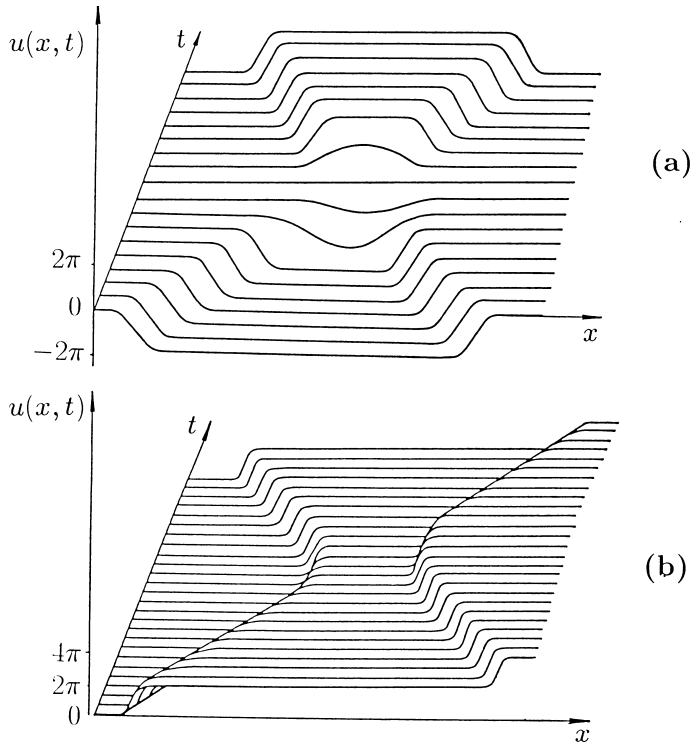

Figure 6.5: Collision of kinks for the sine-Gordon equation: (a) kink-antikink collision, (b) collision of two kinks of the same polarity (from [164]).

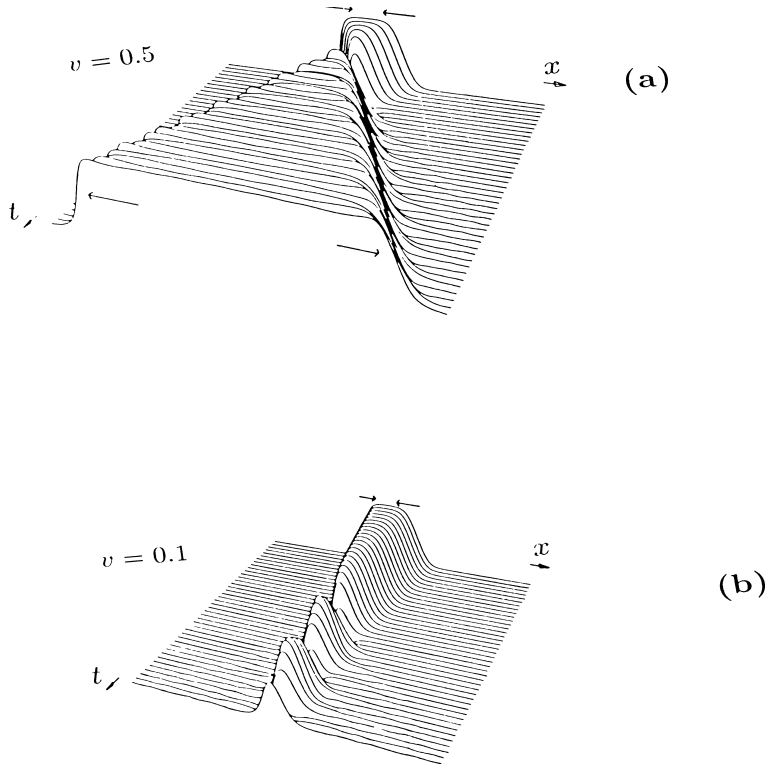

Figure 6.6: Collision of kinks for the ϕ^4-model: (a) collision of fast kinks, and (b) slow kinks' collision (v is the difference of kinks' velocities) (from [163]).

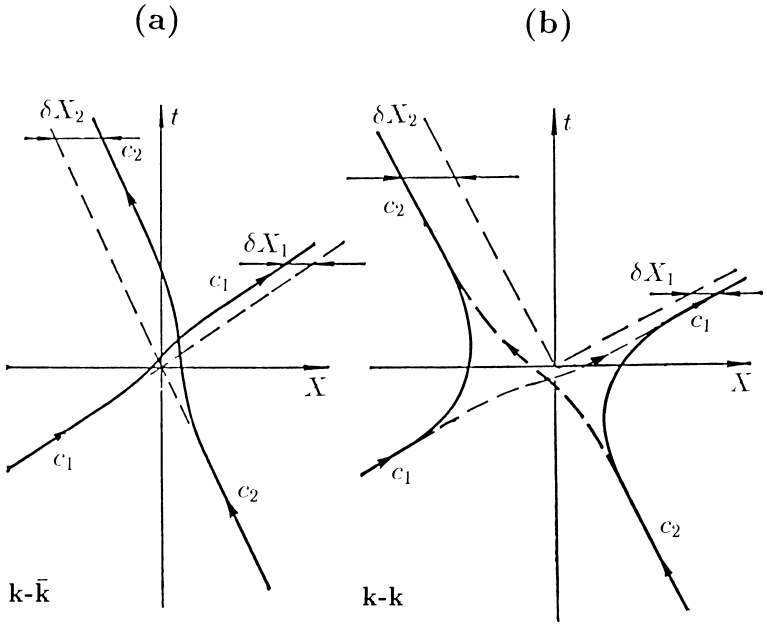

Figure 6.7: The kink center coordinates versus time for (a) the kink-antikink collision, and (b) the kink-kink collision in the SG model (schematically) (from [164]).

their identity (i.e. the shapes, speeds, *etc.*), so that these SWSSs behave like real particles. Zabusky and Kruskal [162] have named such SWSSs by *solitons*. Each soliton may be assigned with some energy, speed, effective mass (and also with the topological charge $\sigma = \pm 1$ in order to distinguish the kink from the antikink in the case of the SG model), which remain unchanged at the collisions. At the same time, the soliton trajectories on the (x,t)-plane before and after the collision do not coincide. This fact indicates that the solitons during the collision undergo a phase shift, i.e. the soliton centers occur to be slightly shifted from their positions for the case when the interaction of solitons would be completely absent. For example, Fig. 6.7b describes the coordinates of centers of two colliding SG kinks of the same polarity. This process may be interpreted as the "elastic" collision of two quasiparticles which repel each other and exchange by their momenta during the collision. The another process, the collision of the kink and the antikink of the SG model shown in Fig. 6.7a, may be interpreted as the "elastic" interaction of two quasiparticles which effectively attract each other. During the collision, they "accelerate" and "pass through" one another "shifting" in the direction to each other. Note that we may think that during the collision the solitons pass through each other, as well as we may think that the solitons exchange by their identities; both the interpretations are true. The only important fact is that in all the cases the solitons completely preserve their characteristics.

However, the collision scenario described above, is observed not for all SWSSs. For example, Fig. 6.6 shows the kink-antikink collision for the ϕ^4-model. When the relative velocity of the colliding kinks is high, the kinks pass through each other analogously to the SG model (Fig. 6.6a). But, opposite to the SG model, now the collision is "inelastic", a part of the kink kinetic energy is converted into the radiation, so that after the collision it appears a "ripple" (oscillating tails) behind the kinks. On the other hand, when two slow kinks collide (Fig. 6.6b), they annihilate during the collision, i.e. they lose such amount of the kinetic energy that they cannot go away for an infinite distance, and they create a coupled oscillating state (the so-called *breather*) which continues to radiate waves (phonons) and finally disappears. Therefore, the behavior of the SWSSs of the ϕ^4-model is not completely the same as those of the SG solitons, although they remind each other in many features.

6.3 Exactly integrable systems

Computer simulation shows a surprising stability of SWSSs for some nonlinear systems. This fact has been served as a prompt that these systems may be completely integrable, i.e. that their solution may be found in an analytical form. Indeed, soon it has been found by Gardner *et al.* [166] and Zahkarov and Shabat [167] that under very weak restrictions (namely, in a class of functions which approach sufficiently fast to a finite limit as $|x| \to \infty$), a general solution of some nonlinear equations may be found with the help of the *inverse scattering transform* method which is the canonical transformation to generalized variables of the action–angle type. These variables correspond to possible types of elementary excitations of the system ("nonlinear normal modes"), namely phonons and solitons, which have an infinite lifetime, and preserve their identity in collisions with other excitations. A solution of the problem with an arbitrary initial condition may be presented with the help of the extended Fourier transform as a "set" of phonons and solitons, the interaction of which is asymptotically reduced to simple additive phase shifts. Because the systems described by partial differential equations are the systems with an infinite degrees of freedom, they must have the infinite set of conservation laws (integrals of motion) in order to be completely integrable. At the same time, contributions of different nonlinear normal modes into the Hamiltonian (as well as into other integrals of motion) are completely separated.

The number of exactly integrable nonlinear equations is relatively small. Among the examples described above, the KdV equation (as well as the modified KdV equation) and the SG equation are the exactly integrable ones. Besides, the exactly integrable *nonlinear Schrödinger* (NLS) *equation* found a wide application in physics too. It is very surprising that problems from various branches of physics such as hydrodynamics, plasma physics, nonlinear optics, low temperature physics, solid state physics, surface physics, elementary particles physics, astrophysics, biophysics, *etc.* are very often reduced to one among namely these three equations.

Below we briefly describe the area of application of the KdV, NLS and SG equations. The detailed description of methods of solution of these equations may be found, e.g., in the books [163, 168, 169].

6.3.1 Korteweg-de Vries equation

The soliton of the KdV equation (6.43) is given by the expression (6.52). The shape of the KdV soliton is described by a single parameter only, by its speed c. The soliton amplitude is proportional to c, and the spacial localization, to $1/\sqrt{c}$, so that the faster is the KdV soliton, the "higher" and "narrower" it is. The KdV soliton is the dynamical one, i.e. it can exist with a faster-than-sound velocity c only; it delocalizes and disappears as c approaches to the sound velocity $s = 1$.

The structure of solutions of the modified KdV equation (6.56) is more rich. First, it has solutions of the type of one–parameter solitons (6.57) which are similar to the KdV solitons. Contrary to the latter, however, the modified KdV solitons may have two different signs describing a local compression or extension of the system. Besides, the modified KdV equation admits solutions of the form of two–parameter solitons, the so-called *breathers*. The latter has the following shape [164],

$$w(\xi,\tau) = 2\frac{\partial}{\partial\xi}\tan^{-1}\left\{\frac{\kappa}{k}\frac{\sin(k\xi-\omega\tau)}{\cos[\kappa(\xi-c\tau)]}\right\}, \tag{6.77}$$

where

$$\kappa = \sqrt{(k^3-\omega)/3k} \tag{6.78}$$

and

$$c = 3k^2 - \kappa^2. \tag{6.79}$$

The solution (6.77) corresponds to the localized (in the region $\sim \kappa^{-1}$) perturbation with the amplitude $\sim \kappa$ moving with the speed c. If we turn to the frame moving with the speed c,

$$\tilde{\xi} = \xi - c\tau\,, \tag{6.80}$$

and put

$$k\xi - \omega\tau = k\tilde{\xi} - \Omega\tau\,,\ \ \Omega = \omega - kc\,, \tag{6.81}$$

then one can see that the breather oscillates with the frequency Ω. In the limit $\kappa \to 0$ the solution (6.77) is transformed into weakly localized small–amplitude linear waves, while in the limit $\Omega \ll 1$ the breather (6.77) may be interpreted as a coupled state of two one–parameter solitons (6.57) with opposite signs which oscillate with respect to the common center of mass with the frequency Ω and the amplitude $\sim -\ln|\Omega|$ in the moving frame.

The KdV equation has been deduced by Korteweg and de Vries [165] in order to describe long surface waves on shallow water without friction effects. However, from the deduction of the KdV equation given above, one can see that the KdV equation (or its modifications) will always emerge in description of propagation of a small (but finite) amplitude plane wave in a homogeneous medium with dispersion and nonlinearity. The variable w describes in this case the deviation of some physical quantity such as height, velocity, density, *etc.* from the corresponding equilibrium value. This class of problems include the ion–acoustic and hydro–magnetic waves in plasma [170], the Rossbi waves in homogeneous rotated liquid [163], *etc.*

6.3.2 Nonlinear Schrödinger equation

The nonlinear Schrödinger (NLS) equation

$$i\psi_t + \psi_{xx} - \omega_0\psi + g|\psi|^2\psi = 0 \tag{6.82}$$

is the exactly integrable equation too. This equation is called in such a way because its form coincides with the quantum Schrödinger equation for the potential $\omega_0 - g|\psi|^2$, so that the constant $g \quad (g > 0)$ may be interpreted as the "intensity of attraction".

In Eq. (6.82) the function $\psi(x,t)$ takes complex values. If we represent it in the following form [164],

$$\psi(x,t) = \Phi(x,t)\ \exp\left[i\varphi(x,t)\right], \tag{6.83}$$

where $\Phi(x,t)$ and $\varphi(x,t)$ are real functions ($\Phi(x,t) > 0$), then the NLS equation is reduced to the set of two real equations,

$$\begin{cases} \Phi_{xx} - (\omega_0 + \varphi_t + \varphi_x^2 - g\Phi^2)\,\Phi = 0\,, \\ (\Phi^2)_t + 2\,(\Phi^2\varphi_x)_x = 0\,. \end{cases} \tag{6.84}$$

The second equation of the set (6.84) may be interpreted as the local conservation law for the quantity

$$N = \int \Phi^2\,dx \equiv \int |\psi|^2\,dx\,, \tag{6.85}$$

which has a meaning of the number of particles in non-ideal bose gas described by the self-consistent equation (6.82). In this case $\Phi^2 = |\psi|^2$ is the density of particles, and $J = 2\Phi^2\varphi_x$ is the density of the particles' flux.

The NLS soliton has the following shape,

$$\psi(x,t) = A\,\frac{\exp\left(ikx - i\omega t\right)}{\cosh\left[\kappa(x - ct)\right]}\,, \tag{6.86}$$

where the soliton amplitude is equal to

$$A = \kappa\sqrt{2/g}\,, \tag{6.87}$$

and the "filling" of the soliton is determined by the parameters

$$k = \frac{1}{2}\,c \tag{6.88}$$

and

$$\omega = \omega_0 + k^2 - \kappa^2. \tag{6.89}$$

The NLS soliton (6.86) is the dynamical two–parameter soliton which describes the localized (in the region $\sim \kappa^{-1}$) perturbation with the amplitude A given by Eq. (6.87), translating with the speed c, and at the same time oscillating with the frequency $\Omega = \omega - \frac{1}{2}\,c^2$. The soliton (6.86) corresponds to the density of the particles' flux $J = c\,|\psi|^2$. In the limit $\kappa \to 0$ the amplitude of the soliton (6.86) decreases and it delocalizes transforming into a linear wave. In the opposite, extremely nonlinear case of $\kappa \to \infty$, the soliton solution transforms into a moving singularity,

$$\psi(x,t) \to \pi\sqrt{2/g}\;\delta(x-ct)\;\exp(-i\Omega t)\,. \tag{6.90}$$

The NLS equation and its modifications play a significant role in the theory of wave propagation in physical systems with dispersion which admit solutions in the form of harmonic wave packets,

$$\psi(x,t) = a(x,t)\;\exp\{i[kx - \omega(x)t]\}\,. \tag{6.91}$$

The change of the amplitude $a(x,t)$ of such spacial–temporal sinusoidal oscillations occurs owing to nonlinear effects, namely due to the reverse action of higher harmonics created by nonlinear terms, on the primary wave.

The well known example of the wave packet (6.91) is, e.g., the amplitude–modulated radio waves which carry out the usual radio communication. Another important area of application, where a carrier wave is modulated by a slowly varying envelope pulse, corresponds to the coherent laser optics. For example, a coherent optical pulse of one nanosecond duration in, say, the region of blue light, has about 10^6 vibrations of the carrier in the envelope pulse. In nonlinear optics there exist methods which allow to obtain the evolution equation for the slowly–varying amplitude and phase of the envelope from the original motion equations. For this, one typically neglects by high-order derivatives for the envelope as well as by higher harmonics of the carrier wave in the original motion equations. These ideas may be applied to a number of areas such as plasma physics, hydrodynamics and optics; the key idea is to find a slowly nonlinear expansion for vibrations of the system under consideration (see detail in [163]).

Above it was noted that dispersive linear systems are not suitable for the transmission of information, because localized pulses in these systems are smoothed. But if we will use a nonlinear system, for example, if we will transmit signals through a fibre with the parameters chosen in such a way that this system will be described by the NLS equation, and instead the pulses we will use the NLS solitons, then we get an ideal system for transmission of information (Hasegawa and Tappert [171], Mollenauer *et al.* [172]).

Also, we should note very interesting applications of the NLS equation for description of biological molecules (the so-called *Davydov solitons*) [173].

6.3.3 Sine-Gordon equation

The sine-Gordon (SG) equation

$$u_{tt} - u_{xx} + \sin u = 0 \tag{6.92}$$

is the exactly integrable equation too. The SG equation admits two types of solitons: one–parameter solitons (kinks) and two–parameter solitons (breathers). The SG kink moving with the velocity c ($|c| < 1$), is described by the expression

$$u(x,t) = 4\tan^{-1}\exp\left[\sigma(x-ct)/\gamma\right], \tag{6.93}$$

where $\sigma = \pm 1$ is the topological charge of the kink, and $\gamma = \sqrt{1-c^2}$ describes the Lorentz narrowing of the moving kink. The SG kink is the topological soliton. The principal difference of topological solitons from dynamical ones consists in the following [164]. The properties of dynamical solitons are determined by the type of dispersion and the sign (and character) of nonlinear terms of the equation, i.e. by local properties of differential equations. On the other hand, the existence of topological solitons is determined only by the

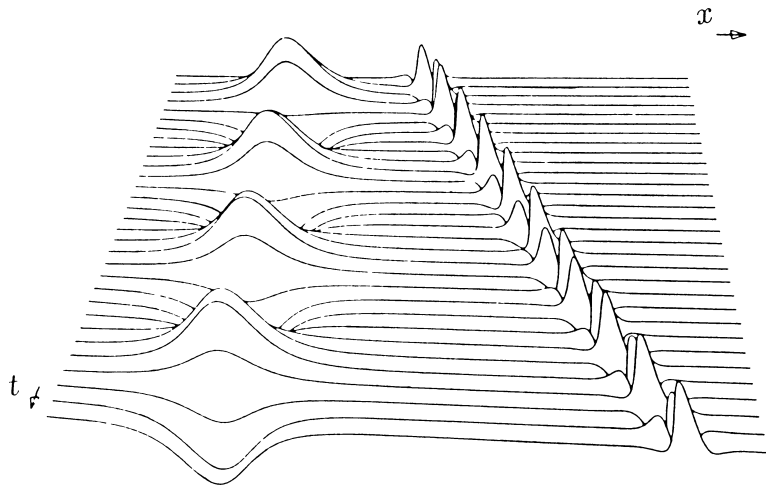

Figure 6.8: Two breathers (immobile and moving ones) for the SG equation (from [163]).

character of the ground state, namely by the existence of the discrete degeneration of the GS, while the type of nonlinearity is not essential. A single topological soliton may be inserted into the system (or be removed from it) only through a free end of the system (which lies at an infinity). In the "bulk" of the system the topological solitons may appear or disappear (annihilate) by kink-antikink pairs only.

The SG breather moving with the velocity c ($|c| < 1$), is described by the expression

$$u(x,t) = 4\tan^{-1}\left\{\frac{\sqrt{1-\nu^2}}{\nu}\ \operatorname{sech}\left[\frac{\sqrt{1-\nu^2}}{\gamma}(x-ct)\right]\sin(\omega t - kx)\right\}, \tag{6.94}$$

where the parameter ν gives the frequency of internal vibrations of the breather,

$$0 < \nu < 1\,, \tag{6.95}$$

while the "filling" of breather is determined by the parameters

$$\omega = \nu/\gamma \tag{6.96}$$

and

$$k = c\omega\,. \tag{6.97}$$

The SG breather is the dynamical soliton which may be considered as a quasiparticle with an internal degree of freedom (see Fig. 6.8). It is similar to the breather of the modified KdV equation or to the NLS soliton. In the case $(1-\nu) \ll 1$ the SG breather is the high–frequency small–amplitude one; in the limit $\nu \to 1$ it may be interpreted as a coupled state of two phonons which attract one another owing to anharmonic effects. In the opposite case of $\nu \ll 1$ the frequency of internal breather's vibrations is small, and it may be considered as a coupled state of kink and antikink vibrating with respect to the common center of mass with the amplitude $\Delta x \sim -\ln|\Omega|$ and the frequency $\Omega = \omega - ck = \gamma\nu$. In the limit $\nu \to 0$ the amplitude of vibrations logarithmically increases, and the breather breaks down into free kink and antikink.

The area of application of the SG equation in physics is large. First of all, the SG equation naturally emerges in systems which may be described with the help of the Frenkel-Kontorova (FK) model [174]. The classical FK model describes a harmonic chain of atoms subjected to the external sinusoidal potential. As has been shown in Sec. 6.2.4, motion equations for such a model reduce to the SG equation. With the help of the FK model we may describe dislocations in solids, and also crowdions (crowdion describes a structure which emerges when an extra atom is inserted into a closely packed atomic row in a metal) [175].

The FK model is widely used in investigation of adsorption systems [87]. In this case the adsorbed atoms play the role of particles, while the external potential is created by the surface atoms of the substrate. One among interesting physical phenomena described by the FK model in adsystems, is the transition from commensurate to incommensurate structure of adatoms.

Starting from the classical work of Frank and van der Merwe [176], the FK model is used to describe the epitaxial crystal growth. Because in a general case the lattice periods of the substrate and the growing crystal do not coincide, their matching at the interface boundary results in a large mismatch energy. To lower this energy, the first layer of the growing crystal adjusts to the substrate, and in order to compensate the difference of the lattice periods, the so-called *misfit dislocations* are created in this layer, and the latter are described by the FK model.

The FK model and its modifications are widely used to describe the proton transport in biological molecules [173].

The next application of the SG equation lies in the description of superionic conductors, where the conductivity is carried out by charge density waves.

The SG model is used also to describe Bloch's walls which separate domains in ferromagnetics (e.g., see [163]).

Interesting effects emerge for ultra–short coherent optical pulses (duration of $10^{-9} \div 10^{-12}$ sec) traversing a quantum two-level system with a large relaxation time. In this case the leading edge of the pulse excites the system atoms, while the trailing edge regenerates the pulse energy owing to stimulated radiation. Dynamics of such a process is described by the SG equation [177].

The SG equation is used also to describe the Josephson junction, i.e. the sandwich consisting of two layers of superconducting metals separated by thin dielectric tunnel barrier. In this case a physical object corresponding to soliton is the quantum of the magnetic flux called *fluxon*. Fluxons are surprisedly stable, they can be kept, be displaced in a given direction, and they can interact with electronic devices. Therefore, in principle the fluxon may be used as a bit in electronic information systems, and the operations with fluxons may be carried out with an extremely high speed and low energy expenses (e.g., see [178]).

Last, note the using of the SG equation in nonlinear field theory [179], where kinks owing to their topological stability claim to a role of fundamental elementary particles.

6.3.4 Toda chain

The exactly integrable KdV, NLS and SG equations describe continuous systems. On the other hand, real physical systems are discrete usually. As a rule, the accounting of discreteness effects destroys the complete integrability of the system. However, there exist two models, which describe a discrete atomic chain and, at the same time, they are exactly integrable (note that there are a number of other discrete exactly integrable systems which, unfortunately, have no direct correspondence to physical objects).

First, the motion equation for an atomic chain where the potential energy of interaction between the atoms is described by the law $\propto x^{-2}$, is exactly integrable [180]. Unfortunately, there are no solitonic solutions in this model.

A more interesting is the second exactly integrable model, the so-called *Toda chain* [181]. In this model only the adjacent atoms interact between themselves, and the interaction is carried out with the help of the exponential law

$$v(x) = \frac{a}{b}\,\exp\left[-b(x-a_0)\right] + a\,(x-a_0)\,. \tag{6.98}$$

The graph of this potential is shown in Fig. 6.9. If we expand $v(x)$ into Taylor series in small $(x-a_0)$, we obtain

$$v_{\rm int}(x) \approx \text{Const} + \frac{1}{2}\,(ab)\,(x-a_0)^2\left[1-\frac{1}{3}\,b\,(x-a_0)\right]. \tag{6.99}$$

Thus, at sufficiently small deviations of the interatomic distances from the equilibrium distance a_0, $|x-a_0| \ll b^{-1}$, the Toda potential is close to the harmonic one with the elastic constant $g = ab$. For higher deviations,

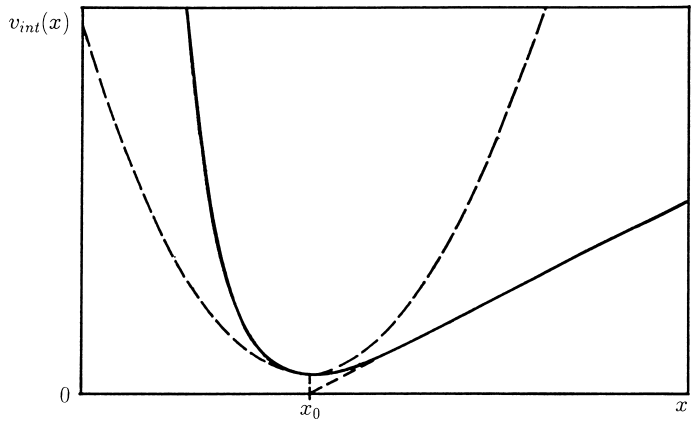

Figure 6.9: The Toda potential.

a measure of nonlinearity of interaction is determined by the parameter b. On the other hand, in the limit $b \to \infty$ the Toda potential reduces to the hard-core potential (3.2) with the core diameter a_0.

Dynamics of the atomic chain is described by the following set of equations ($m_a = 1$),

$$\ddot{u}_n = v'(u_{n+1} - u_n + a_0) - v'(u_n - u_{n-1} + a_0)\,, \tag{6.100}$$

where u_n is the displacement of the n-th atom from the equilibrium position $x_n^{(0)} = na_0$. It is more convenient to write the solution of Eq. (6.100) through the relative displacements

$$w_n \equiv u_{n+1} - u_n\,, \tag{6.101}$$

which are the discrete analog of the deformation $w = \partial u/\partial x$ used above in the long-wave limit.

The Toda soliton has the following shape,

$$w_n(t) = -\frac{1}{b}\,\ln\left\{1 + \sinh^2(\kappa a_0)\,\mathrm{sech}^2[\kappa(na_0 - ct)]\right\}, \tag{6.102}$$

where the parameter κ is coupled with the soliton velocity c by the relationship

$$c = s\,\frac{\sinh(\kappa a_0)}{\kappa a_0}\,, \tag{6.103}$$

and $s = a_0\sqrt{g/m_a}$ is the sound velocity in the chain. From Eq. (6.102) one can see that the Toda soliton is the one–parameter dynamical soliton localized within a region $\sim \kappa^{-1}$ and translating with a faster–than–sound speed. The atomic displacements u_n for the Toda soliton are equal to [164]

$$u_n = \frac{1}{b}\,\ln\left[\frac{1 + \exp\left[2\kappa((n-1)a_0 - ct)\right]}{1 + \exp\left[2\kappa(na_0 - ct)\right]}\right]. \tag{6.104}$$

From this equation it follows that the Toda soliton describes a local compression of the chain which is characterized by a jump of displacements

$$\Delta u \equiv u(+\infty) - u(-\infty) = -2\kappa a_0/b\,. \tag{6.105}$$

The limit $\kappa a_0 \ll 1$ corresponds to small–amplitude Toda solitons; in this case the motion equation (6.100) may be reduced to the KdV equation, and the Toda soliton is reduced to the KdV soliton. In the opposite case of $\kappa a_0 \gg 1$, the region of localization of the Toda soliton is $\kappa^{-1} \ll a_0$, so that at each instant the Toda soliton actually consists of only one atom. Thus, the atoms move in a relay–race fashion, and the compressed bond number consequently changes during the soliton motion.

The Toda solitons are stable not only with respect to collisions with each other and phonons, but also with respect to collisions with impurities. This property yields a possibility to suppose that the Toda solitons play a significant role in thermal conductivity of crystals. Besides, the Toda model may be used to describe long transmission lines constructed of nonlinear capacitors and inductances.

It is interesting to mention the important role of system dimensionality for its integrability. One-dimensional chain of hard balls which corresponds to the limiting case of the Toda model, is the completely integrable system. On the other hand, two-dimensional system of hard balls corresponds to Lorentz-gas model described above in Chapter 2, and this model is characterized by the "maximum chaos" (i.e., there are no islands of stability in this model).

6.4 Systems close to integrable

6.4.1 General remarks

Evidently that behavior of real physical systems may be described by exactly integrable equations approximately only. Often, however, a physical system occurs to be close to the exactly integrable one. In this case it is convenient to describe its behavior in terms of the corresponding exactly integrable system. Namely, it is convenient to present the system state as a superposition of nonlinear normal modes, phonons and solitons. Now, however, these modes are not independent, but they interact between themselves, and the quasiparticles (solitons) themselves are "soft" ("deformable") and may have a finite lifetime. Besides, the nonintegrable systems are characterized by nonzero Kolmogorov-Sinai entropy, so that there always exists some "small" chaos in these systems.

When a deviation of the system from the exactly integrable one is "small", it may be taken into account with the help of one or another variant of the perturbation theory. Unfortunately, to proceed beyond the second-order expansion of the perturbation theory is too difficult usually. Therefore, it remains open the question on a reliability of the results obtained with the help of the perturbation theory. Moreover, in many cases the perturbation can not be considered as a small one, so that the employing the perturbation technique becomes problematic at all. In these cases the computer simulation becomes practically the unique reliable method of investigation.

Methods of computer simulation of systems close to integrable ones, in principle are not different from those described in the previous chapters. If the original physical system is discrete (so that it corresponds to a set of interacting particles), it may be investigated by standard methods of molecular dynamics. When the model under investigation is described by partial differential equations, a natural method of its solution is the converting to difference equations (the description of the corresponding methods as well as the bibliography on this question may be found, for example, in the last chapter of the book [163]). Note that in construction of the model, one has to give a care to the choice of boundary conditions, because they play an essential role in soliton behavior near the system boundary. For example, in the SG model the kink is elastically reflected by the fixed chain's end (and the kink repels from the fixed end). On the other hand, when the chain's end is free, the SG kink is attracted to the end, and during the elastic reflection from the free end, the kink is converted into the antikink. When the periodic boundary conditions are used, the kink will run over the circle, making a whole run for a time $T = Na_0/c$. Therefore, when we want to get rid of the influence of boundary conditions, the most preferable would be the using of damped boundary conditions.

The main questions that emerge in investigation of systems close to integrable ones, are the following.

First, we have to find "nonlinear normal modes", i.e. to find a shape of the SWSS. For this, we have to guess the initial configuration which is the most close to the "true" SWSS for the given system (by taking, for example, the soliton's shape for the most close integrable system and "correcting" it with the help of the perturbation theory), and then we have to follow for the relaxation of the SWSS to the stationary shape. At the same time we are solving the questions about the stability of the SWSS and the character of its motion, i.e. either the motion is freely (without friction) or the SWSS motion is permanently followed by radiation of phonons, which leads to decreasing of the SWSS's speed and its stop or disappearing.

The next important question is about the character of interaction of the given mode with other modes (phonons and other SWSSs). Usually the following scenarios may be observed: **(a)** collisions are quasielastic (i.e., at the collision of quasiparticles a part of their kinetic energy is transferred into phonon radiation), **(b)** collisions lead to creation of new quasiparticles (for example, the creation of a breather at the soliton-antisoliton collision), and **(c)** in the result of the collision the primary quasiparticles are totally destroyed.

A more complex is the question about the evolution of the initial state which corresponds to excitation of a few nonlinear modes (i.e., the search of many-soliton solutions). Finally, the investigation of evolution of some important initial states has a large interest from the practical viewpoint, for example, the decay of the initial rectangular pulse into a set of stable normal modes (SWSSs and phonons). Also a large interest has the study of the SWSS motion in a "thermalized" chain, i.e. the situation when in the initial state there exist, additionally to the SWSS, phonons distributed according to the Maxwell–Boltzmann law.

Let us list now main perturbations which most often occur in modeling of physical objects.

(1) At deduction of Boussinesq's equation and then the KdV equation (see Sec. 6.2) we have used the expansion into Taylor series, and then we have neglected by terms with higher derivatives. Naturally it arises a question of the influence of these neglected terms on the solution of the original problem.

(2) All exactly integrable systems considered in Sec. 6.3 were infinite, while real physical objects are usually finite. Except the something artificial case of using the periodic boundary conditions, the transition from an infinite system to the finite one destroys, as a rule, the exact integrability of the system.

(3) Real physical systems always contain various defects, for example, irregularities of the ideal structure, impurity atoms, *etc.* Naturally it emerges the question of the interaction of phonons and SWSSs with the defects. This problem may be reduced to motion equations with variable coefficients. In some cases (for example, in investigation of disordered systems) the model may be described by nonlinear equations with random coefficients.

(4) A large practical interest has the investigation of systems subjected to some external field. If this field is "weak", it leads to polarization of the soliton, i.e. to small deviation of its shape and, consequently, its parameters too. "Strong" external fields may lead to creation of an instability, for example, to spontaneous creation of new solitons in the system. A more complex is the question of action of variable fields, which may lead either to chaotization of system dynamics, or, on the contrary, to its self–organization.

(5) As a rule, physical systems are not isolated, but they are in contact with some "thermostat". In this case, two questions appear. On the one hand, we have to clarify the influence of dissipation on the soliton dynamics. On the other hand, it emerges a problem about the soliton behavior under the action of a random force. The accounting of both factors leads to equations of the Langevin type, which describe the system behavior at a finite temperature.

(6) Finally, physical systems are discrete usually. Except the Toda chain, the accounting of discreteness effects destroys the exact integrability of the system. Below we consider in more detail one interesting example of such type, namely the Frenkel-Kontorova model which is the discrete analog of the SG model.

6.4.2 The Frenkel-Kontorova model and its generalizations

Recall (see Sec. 6.2.4) that the Frenkel-Kontorova (FK) model describes a linear chain of atoms interacting with the help of the potential $v_{\rm int}(x)$ and subjected to the external periodic potential $v_{\rm sub}(x)$. In the standard FK model it is assumed that only the nearest neighboring atoms interact, and that this interaction is described by the harmonic potential

$$v_{\rm int}(x) = \frac{1}{2}\, g\,(x - a_0)^2, \tag{6.106}$$

and also that the external potential is sinusoidal,

$$v_{\rm sub}(x) = \frac{1}{2}\,\varepsilon_s(1 - \cos 2\pi x/a_s)\,. \tag{6.107}$$

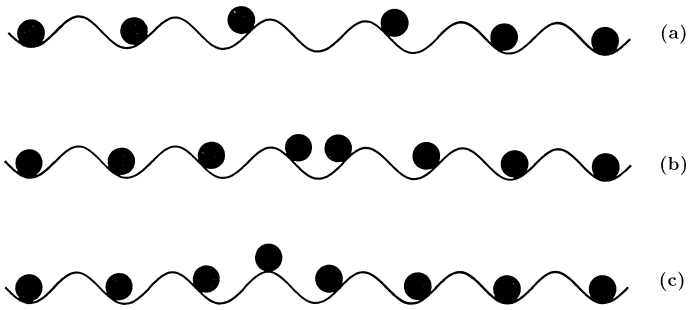

Figure 6.10: Structure of kinks in the Frenkel-Kontorova model: (a) the antikink configuration, (b) the kink at the minimum of the Peierls-Nabarro potential, and (c) the saddle kink configuration.

If, besides, the period of the chain a_0 and that of the external potential a_s are close to one another, $a_0 \approx a_s$ (or $a_0 \approx qa_s$, where q is an integer), then in the case of $ga_0^2 \gg \varepsilon_s$ ("strong" springs) we may use the continuum approximation, where the system motion equations reduce to the SG equation. Recall that nonlinear normal modes of the SG equation correspond to phonons, kinks (topological solitons) and breathers (dynamical solitons).

The main difference of the FK model from the exactly integrable SG model is connected with the discreteness of the former. It occurs, however, that it is still convenient to describe dynamics of the FK model in terms of the SG quasiparticles, i.e. phonons, kinks and breathers. The role of discreteness effects in this case reduces to the following factors.

The structure of phonon modes in the FK model remains practically the same as it was in the SG model, the only difference appears in the dispersion law for large wave vectors k (see Eq. (6.61)).

Kinks in the FK model are topologically stable objects too, and they are characterized by an infinite lifetime. But the kink shape is slightly changed (the FK kink is narrower than the SG kink). Besides, the free motion of the SG kink is changed to the motion of the FK kink in the external periodic potential of the height $\varepsilon_{\rm PN}$ (the so-called Peierls-Nabarro (PN) potential) [182]. Namely, the SG system is characterized by the continuous symmetry group of translations, while the FK model is characterized by the discrete group, i.e. the FK model is invariant under displacement of all atoms only for discrete distances multiple of a_s. In a result the shape and parameters of the isolated FK kink depend on the coordinate of its center. In particular, there exist discrete kink positions where the kink potential energy is minimal (the corresponding kink and antikink configurations are shown in Fig. 6.10a and Fig. 6.10b respectively). To move the kink over the distance a_s to the nearest equivalent state, we have to pass through the "saddle" configuration shown in Fig. 6.10c, overcoming the activation barrier $\varepsilon_{\rm PN}$. In the result, a moving with a velocity c kink will "breathe", i.e. its shape will slightly oscillate with the frequency ω_c ($\omega_c = 2\pi/t_c$, $t_c = a_s/c$) leading to phonon radiation. Evidently that the intensity of this radiation is maximal, when the frequency ω_c or its higher harmonics comes into resonance with phonon frequencies of the chain. Therefore, during the motion of the FK kink in the PN relief, the kink losses its energy for the phonon radiation, and after some time it will be "pinned" at one of minima of the PN relief. Then the kink will vibrate with the frequency ω_{PN} still radiating phonons, and finally it will stop. From the mathematical point of view, the discreteness of the chain removes the Goldstone zero-frequency mode of the SG model.

The collisions of the FK kinks are quasielastic, i.e. phonons are always radiated at the collisions. If the losses of the kinetic energy of the colliding kink and antikink are sufficiently large, then in the collision process they create their coupling state, i.e. breather.

Finally, the dynamical soliton, or breather, in the FK model has a finite lifetime, because it constantly radiates phonons thus losing its energy, and finally it disappears. However, contrary to quasilocal vibrations in the linear chain with impurities, the decay of the breather's amplitude is described by a power law instead of the exponential one.

The applications of the FK model were described above in Sec. 6.3.3. Note, however, that in order to describe real physical objects, the FK model is in need of generalizations. Below we briefly list the important from the physical viewpoint generalizations of the standard FK model (in more detail these questions are considered in, e.g., the review article [183]).

(1) An interesting generalization of the FK model consists in consideration of incommensurate systems, i.e. the case when the ratio a_0/a_s is an irrational number. In this case the investigation of stationary states is of the main interest. The equations describing the stationary states correspond to the Chirikov map, which, together with the logistic map (Sec. 2.3), is the classical model exhibiting the transition from regular to chaotic behavior. Namely, with decreasing of the parameter g (the spring elasticity) at a fixed ratio a_0/a_s the so-called Aubry transition takes place, i.e. the transition from regular trajectories of the Chirikov map to chaotic ones. Simultaneously the nature of the ground state of the FK model changes (the so-called "sliding mode" disappears).

(2) In modeling of real systems, the external periodic potential $v_{\rm sub}(x)$ is usually non-sinusoidal. This leads to changing of kink parameters, in particular, to changing of the shape and amplitude of the PN relief. For example, in some cases the minimum of the kink potential energy is achieved not for the configuration of Fig. 6.10b, but for the configuration shown in Fig. 6.10c or even for a more complicated configuration. For some shape of the function $v_{\rm sub}(x)$ the kink may have one or more the so-called "shape modes", i.e. the shape of the kink may vibrate while its center remains immobile. If the potential $v_{\rm sub}(x)$ has, besides the absolute minima, additional local minima, then the kink is characterized by a complex structure (the kink consists of "sub-kinks"). Finally, if the function $v_{\rm sub}(x)$ has more than one absolute minimum over the period a_s, then the model admits the existence of kinks of different types. In the latter case, a consequence of kinks must satisfy some topological constrains.

(3) Evidently that the interatomic interaction in a real chain may be described by the harmonic law (6.106) approximately only. The main effect which emerges in the FK model with anharmonic interatomic interaction, is the violation of the kink-antikink symmetry, so that now the kink and the antikink are characterized by different parameters.

(4) Various generalizations of the FK model to more than one spacial dimension are very important also. First, we may still consider one-dimensional atomic chain, but to assume that the atoms may move not only along the chain but also in the transverse direction. Second, we may consider two-dimensional (2D) model, which describes two-dimensional array of atoms. In the latter case there exist two modifications of the model. In the *scalar* 2D FK model it is assumed that only a single variable (which describes, for example, the atomic displacements in a single direction only) is coupled with each atom. This model may be interpreted as the system of parallel FK chains which interact between themselves. On the other hand, in the *vector* 2D FK model two variables (which describes, for example, the atomic displacements in both directions) are coupled with each atom. Note that 2D models may differ also by a type of the 2D lattice formed by the minima of the external potential (e.g., square, triangular, *etc.* lattices).

(5) At last, it is interesting to study the effects of finiteness of the system, the role of defects, external fields, contact with a thermostat, *etc.*

Note that the kink describes the system configuration with the inserted extra atom (Fig. 6.10b), while the antikink configuration describes the chain with the extra vacancy (Fig. 6.10a). The amplitude of the PN relief $\varepsilon_{\rm PN}$ is always lower than the amplitude of the original external potential $\varepsilon_{\rm sub}$, so that kinks move along the chain more freely than isolated atoms (i.e., kink's motion is characterized by a lower activation energy). Because kinks transfer mass during their motion along the chain, it is clear that namely the kinks are responsible for diffusivity and conductivity of the system.

Bibliography

[1] Heerman, D.W., 1986, *"Computer Simulations Methods in Theoretical Physics"* (Springer-Verlag, Berlin).

[2] Gould, H. and J. Tobochnik, 1988, *"An Introduction to Computer Simulation Methods. Applications to Physical Systems"* (Addison-Wesley Publishing Company).

[3] Korn, G.A. and T.M. Korn, 1968, *"Mathematical Handbook for Scientists and Engineers"* (McGraw-Hill Book Company).

[4] Voevodin, V.V. and Yu.A. Kuznetsov, 1984, *"Matrices and Calculations"* (Nauka, Moscow) (in Russian; English translation:...).

[5] Turchak, L.I., 1987, *"Bases of Numerical Methods"* (Nauka, Moscow) (in Russian; English translation:...).

[6] Samarsky, A.A. and A.V. Gulin, 1989, *"Numerical Methods"* (Nauka, Moscow) (in Russian; English translation:...).

[7] Hockney, R.W. and J.W. Eastwood, 1981, *"Computer Simulation Using Particles"* (McGraw-Hill Book Company).

[8] Forsythe, G.E., M.A. Malcolm, and C.B. Moler, 1977, *"Computer Methods for Mathematical Computations"* (Prentice-Hall, Inc).

[9] W.H. Press, B.P. Flannery, S.A. Teukolsky, and W.T. Vetterling, *"Numerical Recipes: The Art of Scientific Computing"*, 2007 (Cambridge University Press, Cambridge) (http://www.nr.com).

[10] Landau, L.D. and E.M. Lifshitz, 1965, *"Mechanics"* (Nauka, Moscow) (in Russian; English translation:...).

[11] Arnold, V.I. and A. Avez, 1968, *"Ergodic Problems of Classical Mechanics"* (Benjamin, New York).

[12] Sinai, Ya.G., 1981, Priroda, No.3, p.72-80 (in Russian).

[13] Chirikov, B.V., 1982, Priroda, No.7, p.15-25 (in Russian).

[14] Lichtenberg, A.J. and M.A. Liberman, 1983, *"Regular and Stochastic Motion"* (Springer-Verlag, New York).

[15] Rabinovich, M.I. and D.I. Trubetskov, 1984, *"Introduction into Theory of Vibrations and Waves"* (Nauka, Moscow) (in Russian; English translation:...).

[16] Zaslavsky, G.M., 1984, *"Stochasticity of Dynamical Systems"* (Nauka, Moscow) (in Russian; English translation:...).

[17] Schuster, H.G., 1984, *"Deterministic Chaos"* (Physik-Verlag, Weinheim).

[18] Arnold, V.I., 1989, *"Mathematical Methods of Classical Mechanics"* (Nauka, Moscow) (in Russian; English translation of the previous edition: 1979, Springer-Verlag, New York).

[19] Hénon, M. and C. Heiles, 1964, Astron. J. **69**, 73.

[20] Testa, J., J. Pérez, and C. Jeffries, 1982, Phys. Rev. Lett. **48**, 714.

[21] Ruelle, D. and F. Takens, 1971, Commun. Math. Phys. **20**, 167.

[22] Inoue, M. and H. Koga, 1982, Prog. Theor. Phys. **68**, 2184.

[23] Gwinn, E.G. and R.M. Westervelt, 1985, Phys. Rev. Lett. **54**, 1613.

[24] Hockett, K. and P.J. Holmes, 1985, *"Josephson Junction, Annulus Maps, Birkhoff Attractors, Horseshoes and Rotation Sets"* (Center for Applied Math Report, Cornell University).

[25] Kerr, W.C., M.B. Williams, A.R. Bishop, K. Fesser, and P.S. Lomdahl, 1985, Z. Phys. B**59**, 103.

[26] Metropolis, M., M.L. Stein, and P.R. Stein, 1973, J. Combinatorial Theory **15**A, 25.

[27] May, R.M., 1976, Nature **261**, 459.

[28] Grossmann, S. and S. Thomae, 1977, Z. Naturforsch. **32**A, 1353.

[29] Crutchfield, J.P., D. Farmer, N. Packard, R. Shaw, G. Jones, and R.J. Donnelly, 1980, Phys. Lett. **76**A, 1.

[30] Feigenbaum, M.J., 1978, J. Stat. Phys. **19**, 25.

[31] Feigenbaum, M.J., 1980, Los Alamos Science, Summer, 4.

[32] Feigenbaum, M.J., 1979, J. Stat. Phys. **21**, 669.

[33] Vul, E.B., Ya.G. Sinai, and K.M. Khanin, 1984, Uspechi Mat. Nauk (Russ. Math. Surveys) **39**, 3.

[34] Rössler, O.E., 1976, Phys. Lett. **57**A, 397.

[35] Rössler, O.E., 1976, Z. Naturforsch. **31**A, 1168.

[36] Ruelle, D., 1980, Math. Intelligencer **2**, 126.

[37] Eckman, J.P., 1981, Rev. Mod. Phys. **53**, 643.

[38] Eckman, J.P. and D. Ruelle, 1985, Rev. Mod. Phys. **57**, 617.

[39] Lanford, O.E., 1977, *"Turbulence Seminar"*, In: *"Lecture Notes in Mathematics"* **615**, 114, Eds. P. Bernard and T. Rativ (Springer, Heidelberg-New York).

[40] Lorenz, E.N., 1963, J. Atmos. Sci. **20**, 130.

[41] Hénon, M., 1976, Comm. Math. Phys. **50**, 69.

[42] Kolmogorov, A.N. and V.M. Tikhomirov, 1959, Uspechi Mat. Nauk (Russ. Math. Surveys) **14**, 3.

[43] Grassberger, P. and I. Proccacia, 1983, Phys. Rev. Lett. **50**, 346.

[44] Wolf, A., J.B. Swift, H.L. Swinney, and J.A. Vasano, 1985, Physica **16**D, 285.

[45] Kaplan, J. and J. Yorke, 1979, In: *"Functional Differential Equations and Approximation of Fixed Points"*, Eds. H. O. Peitgen and H. O. Walther (Springer, Heidelberg-New York).

[46] Farmer, J.D., E. Ott, and J.A. Yorke, 1983, Physica **7**D, 153.

[47] Moon, F.C., 1987, *"Chaotic Vibrations"* (John Wiley & Sons, Inc).

[48] Sinai, Ya.G., 1963, Docl. Acad. Nauk SSSR **153**, 1261.

[49] Sinai, Ya.G., 1970, Uspechi Mat. Nauk (Russ. Math. Surveys) **25**, 2.

[50] Kolmogorov, A.N., 1954, Docl. Acad. Nauk SSSR **98**, 527.

[51] Arnold, V.I., 1963, Uspechi Mat. Nauk (Russ. Math. Surveys) **18**, 5.

[52] Mozer, J., 1967, Math. Ann. **169**, 163.

[53] Chirikov, B.V., 1959, At. Ener. **6**, 630.

[54] Chirikov, B.V., 1960, Plasma Phys. (J.N.E.Pt C) **1**, 253.

[55] Chirikov, B.V., 1979, Phys. Rep. **52**, 265.

[56] Casartelly, M., E. Diana, L. Galgani, and A. Scotti, 1976, Phys. Rev. A**13**, 1921.

[57] Benettin, G., L. Galgani, and J.M. Strelcyn, 1976, Phys. Rev. A**14**, 2338.

[58] Vorob'ev, P.A. and G.M. Zaslavsky, 1987, Zh. Eksp. Teor. Fiz. **92**, 1564 [Sov. Phys. JETP **??**, ???].

[59] Kolmogorov, A.N., 1959, Dokl. Akad. Nauk SSSR **124**, 754.

[60] Sinai, Ya.G., 1966, Izvestiya AN SSSR, Ser. Mat. **30**, 15.

[61] Shimada, I. and T. Nagashima, 1979, Prog. Theor. Phys. **61**, 1605.

[62] Benettin, G., L. Galgani, A. Giorgilli, and J.M. Strelcyn, 1980, Meccanica **15**, 21.

[63] Melnikov, V.K., 1963, Trans. Moscow Math. Soc. **12**, 1.

[64] Guckenheimer, J. and P.J. Holmes, 1983, *"Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields"* (Springer-Verlag, New York).

[65] Pomeau, Y. and P. Manneville, 1979, In: *"Intrinsic Stochasticity in Plasmas"*, Eds. G. Laval and D. Gresillon (Ed. de Physique, Orsay).

[66] Manneville, P. and Y. Pomeau, 1979, Phys. Lett. **75**A, 1.

[67] Manneville, P. and Y. Pomeau, 1980, Physica **1**D, 219.

[68] Pomeau, Y. and P. Manneville, 1980, Comm. Math. Phys. **74**, 189.

[69] Tully, J.C., Y.J. Chabal, K. Raghavachari, J. M. Bowman, and R. R. Lucchese, 1985, Phys. Rev. B**31**, 1184.

[70] Hut, P. and G.J. Sussman, 1987, Scientific American **257**, No 4.

[71] Braun, O.M. and V.K. Medvedev, 1989, Usp. Fiz. Nauk **157**, 631 [Sov. Phys. Usp. **32**, 328].

[72] Kanatani, K., 1984, J. Comput. Phys. **53**, 181.

[73] Greenspan, D., 1984, J. Comput. Phys. **56**, 28.

[74] Benettin, G. and A. Tenenbaum 1983, Phys. Rev. A**28**, 3020.

[75] Lagar'kov, A.N. and V.M. Sergeev, 1978, Usp. Fiz. Nauk **125**, 400 [Sov. Phys. Usp. **??**, ???].

[76] Landau, L.D. and E.M. Lifshitz, 1958, *"Statistical Mechanics"* (Pergamon Press, Ltd., London).

[77] Peierls, R.E., 1934, Helv. Phys. Acta **7**, Suppl.2, 81.

[78] Peierls, R.E., 1935, Ann. Inst. H. Poincaré **5**, 177.

[79] Landau, L.D., 1937, Zh. Eksp. Teor. Fiz. **7**, 627 [Sov. Phys. JETP **??**, ???].

[80] Peierls, R., 1979, *"Surpises in Theoretical Physics"* (Princeton University Press, Princeton).

[81] Mermin, N.D., 1968, Phys. Rev. **176**, 250.

[82] Kosterlitz, J.M. and D.J. Thouless, 1973, J. Phys. C**6**, 1181.

[83] Hockney, R.W. and T.R. Brown, 1975, J. Phys. C**8**, 1813.

[84] Grimes, C.C. and G. Adams, 1979, Phys. Rev. Lett. **42**, 795.

[85] Bedanov, V.M., G.V. Gadiyak, and Yu.E. Lozovik, 1985, Zh. Eksp. Teor. Fiz. **88**, 1622 [Sov. Phys. JETP **??**, ???]; see also preprint No 29-83, Institute for Theoretical and Applied Mechanics (Novosibirsk).

[86] Hirth, J.P. and J. Lothe, 1968, *"Theory of Dislocations"* (McGraw, New York).

[87] Lyuksyutov, I.F., A.G. Naumovets, and V.L. Pokrovsky, 1988, *"Two-Dimensional Crystals"* (Naukova Dumka, Kiev) (in Russian; English translation: 1992, Academic Press, Boston).

[88] Abraham, F.F., 1984, J. Vac. Sci. Technol. B**2**, 534.

[89] Gadzuk, J.W. and A.C. Luntz, 1984, Surface Sci. **144**, 429.

[90] Braun, O.M., A.I.Volokitin, and V.P.Zhdanov, 1989, Usp. Fiz. Nauk **158**, 421 [Sov. Phys. Usp. **32**, 605].

[91] Riehl, J.P. and D.J. Diestler, 1976, J. Chem. Phys. **64**, 2593.

[92] Naumovets, A.G. and Yu.S. Vedula, 1985, Surface Sci. Reports **4**, 365.

[93] Gomer, R., 1990, Reports on Progress in Physics **53**, 917.

[94] De Lorenzi, G. and G. Jacucci, 1985, Surface Sci. **164**, 526.

[95] Müller, K.H., 1987, Surface Sci. **184**, L375.

[96] Ewald, P.P., 1921, Ann. Physik **64**, 253.

[97] Brush, S,G., H.L. Sahlin, and E. Teller, 1966, J. Chem. Phys. **45**, 2102.

[98] Ziman, J.M., 1964, *"Principles of the Theory of Solids"* (Cambridge University Press, Cambridge).

[99] Heerman, D.W., W. Klein, and D. Stauffer, 1982, Phys. Rev. Lett. **49**, 1262.

[100] Hockney, R.W., S.P. Goel, and J.W. Eastwood, 1974, J. Comput. Phys. **14**, 148.

[101] Woodcock, L.V., 1971, Chem. Phys. Lett. **10**, 257.

[102] Hoover, W.G., A.J.C. Ladd, and B. Moran, 1982, Phys. Rev. Lett. **48**, 1818.

[103] Ladd, A.J.C. and W.G. Hoover, 1983, Phys. Rev. B**28**, 1756.

[104] Evans, D.J., 1983, J. Chem. Phys. **78**, 3297.

[105] Nosé, S., 1984, J. Chem. Phys. **81**, 511.

[106] Schneider, T. and E. Stoll, 1978, Phys. Rev. B**17**, 1302.

[107] Adelman, S.A. and J.D. Doll, 1976, J. Chem. Phys. **64**, 2375.

[108] Tully, J.C., 1980, J. Chem. Phys. **73**, 1975.

[109] Zwanzig, R., 1960, J. Chem. Phys. **32**, 1173.

[110] Mori, H., 1965, Prog. Theor. Phys. **33**, 423.

[111] Kubo, R., 1966, Rep. Prog. Theor. Phys. **29**, 255.

[112] Tully, J.C., G.H. Gilmer, and M. Shugard, 1979, J. Chem. Phys. **71**, 1630.

[113] Helfand, E., 1979, Bell System Technical Journal **58**, 2289.

[114] Gardiner, C. W., 1983, *"Handbook of Stochastic Methods"* (Springer-Verlag, Berlin).

[115] Risken, H., 1984, *"The Fokker-Planck Equation. Methods of Solution and Applications"* (Springer, Berlin).

[116] Kramers, H.A., 1940, Physica **7**, 284.

[117] Glasstone, S., K. J. Laidler, and H. Eyring, 1941, *"The Theory of Rate Processes"* (..?.., New York).

[118] Hänggi, P., P. Talkner, and M. Borkovec, 1990, Rev. Mod. Phys. **62**, 251.

[119] Cartling, B., 1987, J. Chem. Phys. **87**, 2638.

[120] Mitchell, A.R. and D.F. Griffiths, 1981, *"The Finite Difference Method in Partial Differential Equations"* (Wiley, Chicester, England).

[121] Metropolis, N. and S. Ulam, 1949, J. Amer. Statistical Assoc. **44**, 335.

[122] Sobol', I.M., 1973, *"Numerical Monte Carlo Methods"* (Nauka, Moscow) (in Russian; English translation:...).

[123] Müller-Krumbhaar, H., 1979, In: *"Monte Carlo Methods in Statistical Physics"*, Ed. K. Binder (Springer-Verlag, Berlin).

[124] Efros, A.L., 1982, *"Physics and Geometry of Disorder"* (Nauka, Moscow) (in Russian; English translation: 1986, Mir, Moscow).

[125] Stauffer, D., 1985, *"Introduction to Percolation Theory"* (Taylor and Francis, Philadelphia).

[126] Feynman, R.P., 1972, *"Statistical Mechanics"* (W.A.Benjamin, Inc., Massachusetts).

[127] Feynman, R.P. and A.R. Hibbs, 1965, *"Quantum Mechanics and Path Integrals"* (McGraw-Hill Book Company, New York).

[128] Creutz, M., 1983, *"Quarks, Gluons and Lattices"* (Cambridge University Press, Cambridge).

[129] Metropolis, N., A.W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, J. Chem. Phys. **21**, 1087.

[130] Feller, W., 1950, *"An Introduction to Probability Theory and its Applications"*, Vol. I (Wiley, New York).

[131] Wood, W.W., 1968, In: *"Physics of Simple Liquids"*, Eds. H. N. W. Temperley, J. S. Rowlinson, and G. S. Rushbrooke (North-Holland, Amsterdam).

[132] Binder, K. (Ed.), 1979, *"Monte Carlo Methods in Statistical Physics"* (Springer-Verlag, Berlin).

[133] Ma, S.K., 1981, J. Stat. Phys. **26**, 221.

[134] Binder, K. (Ed.), 1987, *"Applications of the Monte Carlo Method in Statistical Physics"* (Springer-Verlag, New York).

[135] Binder, K. (Ed.), 1992, *"The Monte Carlo Methods in Condensed Matter Physics"* (Springer, Berlin).

[136] Mouritsen, O.G., 1984, *"Computer Studies of Phase Transitions and Critical Phenomena"* (Springer-Verlag, Berlin).

[137] Kawasaki, K., 1972, In: *"Phase Transitions and Critical Phenomena"*, Eds. C. Domb and M. S. Green (Academic, New York).

[138] Glauber, R.J., 1963, J. Math. Phys. **4**, 294.

[139] Binder, K. and D.P. Landau, 1976, Surface Sci. **61**, 577.

[140] Privman, V., 1990, In: *"Finite Size Scaling and Numerical Simulations of Statistical Physics"*, Ed. V. Privman (World Scientific, Singapore).

[141] Borgs, C. and W. Janke, 1992, Phys. Rev. Lett. **68**, 1738.

[142] Onsager, L., 1944, Phys. Rev. **65**, 117.

[143] Landau, D.P., 1983, Phys. Rev. B**27**, 5604.

[144] Palmer, R.G., 1982, Adv. Phys. **31**, 669.

[145] Bowker, M. and D.A. King, 1978, Surface Sci. **71**, 583.

[146] Binder, K. and D. Stauffer, 1976, Adv. Phys. **25**, 343.

[147] Rao, M., M.H. Kalos, J. Marro, and J.L. Lebowitz, 1976, Phys. Rev. B**13**, 4328.

[148] Salid, A. and K. Binder, 1984, J. Stat. Phys. **35**, 517.

[149] Venäläinen, O., J. Heiniö, and K. Kaski, 1991, Physica Scripta T**38**, 66.

[150] Witten, T.A. and L.M. Sander, 1981, Phys. Rev. Lett. **47**, 1400.

[151] Meakin, P., 1992, Physica Scripta **46**, 295.

[152] Otálora, F. and J.M. García-Ruiz, unpublished.

[153] Forster, D., 1975, *"Hydrodynamic Fluctuations, Broken Symmetry and Correlation Functions"* (W.A.Binjamin, Inc., London).

[154] Shumway, S.L. and J.P. Sethna, 1991, Phys. Rev. Lett. **67**, 995.

[155] Haken, H., 1978, *"Synergetics"* (Springer, Berlin).

[156] Haken, H., 1983, *"Advanced Synergetics"* (Springer-Verlag, Berlin).

[157] Fermi, E., J. Pasta, and S. Ulam, 1955, Los Alamos Rpt. LA-1940; see also *"Collected Papers of Enrico Fermi"* vol.**2**, p.978-988 (1965, Univ. of Chicago Press, Chicago).

[158] Izrailev, F.M. and B.V. Chirikov, 1966, Docl. Acad. Nauk SSSR **11**, 30.

[159] Zaslavsky, G.M. and R.Z. Sagdeev, 1967, Zh. Eksp. Teor. Fiz. **52**, 1083 [Sov. Phys. JETP **??**, ???].

[160] Saito, N., N. Hirotomi, and A. Ichimura, 1975, J. Phys. Soc. Jpn. **39**, 1431.

[161] Saito, N. and A. Ichimura, 1976, Prog. Theor. Phys. Suppl. **59**, 137.

[162] Zabusky, N.J. and M.D. Kruskal, 1965, Phys. Rev. Lett. **15**, 240.

[163] Dodd, R.K., J.C. Eilbeck, J.D. Gibbon, and H.C. Morris, 1982, *"Solitons and Nonlinear Wave Equations"* (Academic Press, Inc., London).

[164] Kosevich, A.M. and A.S. Kovalev, 1989, *"Introduction to Nonlinear Physical Mechanics"* (Naukova Dumka, Kiev) (in Russian; English translation: ...).

[165] Korteweg, D.J. and G. de Vries, 1895, Philos. Mag. **39**, 422.

[166] Gardner, C.S., J.M. Greene, M.D. Kruskal, and R.M. Miura, 1967, Phys. Rev. Lett. **19**, 1095.

[167] Zahkarov, V.E. and A.B. Shabat, 1971, Zh. Eksp. Teor. Fiz. **61**, 118 [Sov. Phys. JETP **34**, 62 (1972)].

[168] Bullough R. K. and P. J. Caudrey (Eds.), 1980, *"Solitons"*, Topics in Current Physics vol.**17** (Springer, Berlin).

[169] Novikov, S.P. (Ed.), 1980, *"Theory of Solitons: The Inverse Scattering Method"* (Nauka, Moscow) (in Russian; English translation:...).

[170] Kadomtsev, B.B. and V.I. Karpman, 1971, Usp. Fiz. Nauk **109**, 193 [Sov. Phys. Usp. **??**, ??].

[171] Hasegawa, A. and F. Tappert, 1973, Appl. Phys. Lett. **23**, 142.

[172] Mollenauer, L.F., R.H. Stolen, and J.P. Gordon, 1980, Phys. Rev. Lett. **45**, 1095.

[173] Davydov, A.S., 1984, *"Solitons in Molecular Systems"* (Naukova Dumka, Kiev) (in Russian; English translation: ...).

[174] Frenkel, Ya. and T. Kontorova, 1938, Phys. Z. Sowietunion **13**, 1.

[175] Kosevich, A.M., 1972, *"Fundamentals of Crystal Lattice Mechanics"* (Nauka, Moscow) (in Russian; English translation: ...).

[176] Frank, F.C. and J.H. van der Merwe, 1949, Proc. Roy. Soc. (London) A **198**, 205; 216; A **200**, 125.

[177] McCall, S.L. and E.L. Hahn, 1967, Phys. Rev. Lett. **18**, 908.

[178] Parmentier, R. D., 1978, In: *"Solitons in Action"*, Eds. K. Lonngren and A. Scott (Academic Press, New York).

[179] Skyrme, T.H.R., 1958, Proc. Roy. Soc. (London) A **247**, 260.

[180] Kulish, P. P., 1976, Teor. Mat. Fiz. **26**, 198.

[181] Toda, M., 1981, *"Theory of Nonlinear Lattices"* (Springer-Verlag, Berlin).

[182] Currie, J.F., S.E.Trullinger, A.R.Bishop, and J.A.Krumhansl, 1977, Phys. Rev. B **15**, 5567.

[183] Braun, O.M. and Yu.S. Kivshar, 1993, *"Nonlinear Dynamics of the Frenkel-Kontorova Model"*, Phys. Rep.

[184] ==========

[185] S.K. Park and K.W. Miller, Commun. ACM **31** (1988) 1192-1201.

[186] W.H. Press and S.A. Teukolsky, Computer in Physics **6** (1992) 522-524 No.5.

[187] J. Adler, Computer in Physics **8** (1994) 287-295 No.3.

[188] A. Bunde and S. Halvin, *"Percolation"*, in: *"Fractals and disordered systems"*, Eds. A. Bunde and S. Halvin (Springer, Berlin, 1991) p.51.

[189] J.-F. Gouyet, M. Rosso, and B. Sapoval, *"Fractal surfaces and interfaces"*, in: *"Fractals and disordered systems"*, Eds. A. Bunde and S. Halvin (Springer, Berlin, 1991) p.229.

[190] A. Aharony, *"Fractal growth"*, in: *"Fractals and disordered systems"*, Eds. A. Bunde and S. Halvin (Springer, Berlin, 1991) p.151.

[191] M.P. Allen and D.J. Tildesley, *"Computer simulation of liquids"* (Oxford University Press, 1987).

[192] D. Potter, *"Computational Physics"* (John Wiley & Sons, Chichester, 1973).

[193] Braun, O.M. and Ferrando, R., Phys. Rev. E **65**, 061107 (2002).

[194] see http://www.taygeta.com/random/gaussian.html

[195] Box, G.E.P. and M.E. Muller, 1958, *"A note on the generation of random normal deviates"*, Annals Math. Stat. **29**, 610-611.

[196] Rubinstein, R.Y., 1981, *"Simulation and the Monte Carlo method"* (John Wiley & Sons).