
CHAPTER 16

Statistical Mechanical Modeling and Its Application to Nanosystems

Keivan Esfarjani and G. Ali Mansoori

*Department of Chemical Engineering, University of Illinois at Chicago,
Chicago, Illinois, U.S.A.*

CONTENTS

1.	Introduction	1
2.	Thermodynamics of Small Systems	3
2.1.	Nonextensivity of Nanosystems	6
2.2.	The Gibbs Equation for Nanosystems	7
2.3.	Statistical Mechanics and Thermodynamic Property Predictions	9
3.	Intermolecular Forces and Potentials	10
3.1.	Step 1: Atomic Force Microscopy Measurement and Empirical Modeling	14
3.2.	Step 2: Theoretical Modeling	14
3.3.	Step 3: Development of Nanoparticle Potentials	17
4.	Nanoscale systems Computer-Based Simulations	17
4.1.	Classification of the Methods	17
4.2.	Monte Carlo Method and Random Numbers	19
4.3.	Molecular Dynamics Method	31
5.	Conclusions	41
	Glossary	41
	References	42

1. INTRODUCTION

A scientific and technological revolution has begun, and it is in our ability to systematically organize and manipulate matter on a bottom-up fashion starting from an atomic level as well as to design tools, machinery, and energy conversion devices in nanoscale toward the development of nanotechnology. There is also a parallel miniaturization activity to scale

down large tools, machinery, and energy conversion systems to micro and nanoscales toward the same goals [1–4]. The science of miniaturization is being developed, and the limitations in scaling down large systems to nanoscale are under investigation. Significant accomplishments in performance and changes of manufacturing paradigms are predicted to lead to several breakthroughs in the 21st century. The answer to the question of how soon will this next industrial revolution arrive depends a great deal on the intensity of scientific activities of the scientific communities in academic disciplines, national laboratories, or even entire industries all around the world.

The systems of interest to nanoscience and nanotechnology are the isolated individual nanostructures and their assemblies. Nanosystems are made of countable (limited) number of atoms or molecules. Their sizes are larger than individual molecules and smaller than microsystems. One of the characteristic features of nanosystems is their high surface-to-volume ratio. Their electronic and magnetic properties are often distinguished by quantum mechanical behavior, whereas their mechanical and thermal properties can be formulated within the framework of classical statistical mechanics of small systems, now known as nonextensive statistical mechanics, as it is discussed in this chapter. Nanosystems can contain all forms of condensed matter, be it soft or hard, organic or inorganic, or biological components. A deep understanding of the interactions between individual atoms and molecules composing the nanosystems and their statistical mechanical modeling is essential for nanotechnology.

Nanotechnology revolution will change the nature of almost every human-made object and activity. The objective of this chapter is to link the foundation of molecular-based study of substances and the basis for nanoscience and technology to produce predictive computational and modeling options for the behavior of nanosystems.

In statistical mechanical modeling of nanosystems, concepts from both classical and quantum mechanical many-body theories are employed. In doing so one is exposed to the question of whether the position and momentum of an object could simultaneously and precisely be determined [5] and how could one be able to brush aside the Heisenberg uncertainty principle (Fig. 1) to work at the atomic and molecular level, atom by atom, as is the basis of nanotechnology. The Heisenberg uncertainty principle helps determine the size of electron clouds, and hence the size of atoms. It applies only to the subatomic particles like electron, positron, photon, and so forth. It does not forbid the possibility of nanotechnology that has to do with the position and momentum of such large particles like atoms and molecules. This is because the mass of atoms and molecules is quite large, and quantum mechanical calculations by the Heisenberg uncertainty principle places almost no limit on how well atoms and molecules can be held in place [5].

This review chapter is concerned with one of the most important subjects of computational nanotechnology; namely, statistical mechanical modeling and its application to molecular systems to predict the properties of, and processes involving, nanoscale structures. The modification and implementation of statistical mechanics for nanoscale systems would lead to possibilities to study the evolution of physical, chemical, and biophysical systems on significantly reduced length, time, and energy scales. The statistical mechanical modeling and its application to nanosystems is achieved through mathematical modeling and computer-based

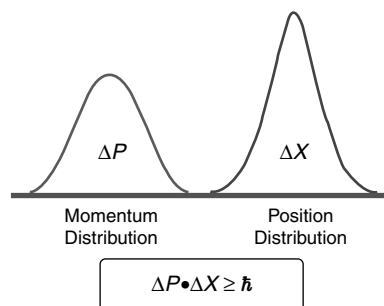


Figure 1. Heisenberg Uncertainty Principle.

molecular simulations employing concepts from statistical mechanics of large systems as well as quantum mechanical many-body theories. If performed correctly, it can provide insight into the formation, evolution, and properties of nanostructures, self-assemblies, and their relation with macroscopic behavior of matter.

Statistical mechanics of small systems is based on the fields of quantum mechanics, thermodynamics, statistics, and mathematical modeling or computer-based simulations that allow for computation and prediction of the underlying dynamics of nanosystems and processes in condensed matter physics, chemistry, materials science, biology, and genetics. It can provide deep insights into the formation, evolution, and properties of nanosystems and mechanisms of nanoprocesses. This is achieved by performing precise analytic formulation or numerical experiments on many aspects of various condensed phases. The precision of such calculations depends on the accuracy of the theory used and the interatomic and intermolecular potential energy functions at hand. Computer simulations at this scale that form the basis of computational nanoscience could allow for an understanding of the atomic and molecular scale structures, energetics, dynamics, and mechanisms underlying the physical and chemical processes that can unfold in isolated nanostructures and their assemblies under different ambient conditions.

In this review chapter, we first introduce the specificities of the thermodynamics of nanosystems, which includes an examination of laws of thermodynamics, the concept of nonextensivity of nanosystems, the Gibbs equation for nanosystems, and statistical mechanics and thermodynamic property predictions for nanosystems. Then we discuss advances in direct measurement, computation, and modeling of interparticle forces and potentials for nanostructures. Methods for computer-based simulation in nanosystems are categorized and presented later, including stochastic methods and molecular dynamics techniques.

2. THERMODYNAMICS OF SMALL SYSTEMS

Thermodynamics and statistical mechanics of large systems consisting of many particles beyond the thermodynamic limit (V goes to infinity, $N/V = \rho_N$ is constant) are well developed [6, 7]. Normally, thermodynamics describes the *most likely macroscopic behavior* of large systems (consisting of 10^{23} particles ~ 1 ccm in volume or more) under slow changes of a few macroscopic properties. The really large systems (like astrophysical objects) as well as small systems (like those of interest in nanotechnology) are excluded.

Recent developments in nanoscience and nanotechnology have brought about a great deal of interest into the extension of thermodynamics and statistical mechanics to small systems consisting of countable particles below the thermodynamic limit. Hence, if we extend thermodynamics and statistical mechanics to small systems, to remain on a firm basis we must go back to its founders and, like them, establish new formalism of thermodynamics and statistical mechanics of small systems starting from the safe grounds of mechanics. In this report, we to point out difficulties with and possibly the changes that we have to make in thermodynamics and statistical mechanics to make them applicable to small systems.

Structural characteristics in nanoscale systems is a dynamic phenomenon unlike the static equilibrium of macroscopic phases. Coexistence of phases is expected to occur over bands of temperature and pressure, rather than along just sharp points. The Gibbs phase rule loses its meaning, and many phaselike forms may occur for nanoscale systems that are unobservable in the macroscopic counterparts of those systems [8–10].

Such questions as property relations and phase transitions in small (nano) systems are subjects to be investigated and formulated provided the formulation of working equations of thermodynamics and statistical mechanics of small systems are developed. It is worth mentioning that the molecular self-assembly (bottom-up technology) that was originally proposed by Feynman [11] has its roots in phase transitions. A comprehensive understanding of thermodynamics and statistical mechanics of small systems will help bring about the possibility of formulation of alternative self-assemblies of various atomic and molecular assemblies.

The earliest work on this subject was the well-recognized contribution of T. L. Hill, published in 1963 and 1964 [8], entitled, *Thermodynamics of Small Systems*. This area of study

was recently renamed “nanothermodynamics” [9], and will be addressed in later parts of this review chapter.

To formulate the thermodynamics of small systems, one has to start evaluating thermodynamics from first principles reviewing the concepts, laws, definitions, and formulations, and to draw a set of guidelines for their applications to small systems. Historically, thermodynamics was developed for the purpose of understanding the phenomenon of converting thermal energy to mechanical energy. With the advent of technology, thermodynamics assumed a universal aspect applicable for all kinds of systems under variations of temperature, pressure, volume, and composition. The concept of irreversible thermodynamics [12, 13] was an extension of thermodynamics to dynamic systems under various external forces generating varieties of fluxes in the system. It is quite appropriate to examine the laws of thermodynamics for a small (nano) system that does not satisfy the “thermodynamic limit.”

The Zeroth Law of thermodynamics consists of the establishment of an absolute temperature scale. Temperature of a macroscopic system is a well-defined property and is a measure of the system’s thermal equilibrium (or lack of it). The temperature of a nanosystem fluctuates around an average value. The fluctuation range is a function of the system’s size and internal structure. Although the temperature of a small system is also a well-defined thermodynamic property, as for macroscopic systems, it generally has larger fluctuations with respect to time and space that would magnify as the size of the system reduces.

Because of the probabilistic aspect of statistical mechanics and thermodynamics, the theory of fluctuations has played an important role [14]. Fluctuations previously were set aside in thermodynamic property calculations of macroscopic systems because of the thermodynamic limit. However, recent advances in the fluctuation theory of statistical mechanics of macroscopic systems have been quite successful in quantifying and predicting the behavior of complex mixtures [7]. In nanosystems, as is well explained in the following statement by the US National Initiative on Nanotechnology [2], fluctuations play an important role:

“There are also many different types of time scales, ranging from 10^{-15} s to several seconds, so consideration must be given to the fact that the particles are actually undergoing fluctuations in time and to the fact that there are uneven size distributions. To provide reliable results, researchers must also consider the relative accuracy appropriate for the space and time scales that are required; however, the cost of accuracy can be high. The temporal scale goes linearly in the number of particles N , the spatial scale goes as $O(N \cdot \log N)$, yet the accuracy scale can go as high as N^7 to $N!$ with a significant prefactor.”

Fluctuations in small systems are generally appreciable in magnitude, and they are not yet quantitatively correlated to their properties. If we are going to extend applications of thermodynamics and statistical mechanics to small systems, accurate calculation of fluctuations are quite necessary.

According to the above statements, systems with countable number of atoms or molecules bring about a difficulty for defining the concept of “thermal equilibrium” in small systems. Equilibrium is defined as the condition in which there is no visible change in the thermodynamics properties of a system with respect to time. We also define thermal, mechanical, and chemical equilibriums for a macroscopic thermodynamic system. Another important concept in thermodynamics is reversibility and irreversibility. A reversible system is that when the direction of a driving force (like heat or work) is changed, the system goes back to its original condition. Thermodynamic property relations for macroscopic systems in equilibrium or when they are reversible are identical. A nanosystem may be reversible, but it may not be at equilibrium. In general, it is safe not to use the concept of equilibrium for a small system. Similar to macrosystems, a small system can be also defined as isolated, closed (also known as control mass), open (also known as control volume), adiabatic, isothermal, isometric (or isochoric, also known as constant volume), and isobaric.

The First Law of thermodynamics, which is the same as the principle of conservation of mass and energy, provides us with the relation between heat, work, mass, and energy contents of a system. The equation for the First Law of thermodynamics for a closed (controlled mass) macrosystem takes the following form,

$$dE = \delta Q_{in} + \delta W_{in} \quad (1)$$

where dE stands for the exact differential energy increase in the system resulting from the addition of inexact differential amounts of heat and work, δQ_{in} and $\delta W_{in} = -P_{ext} dV$, respectively, to the system, and P_{ext} is the external pressure. This equation is still valid for small systems. However, clear distinction or separation between the terms “heat” and “work” may be a bit foggy for certain small systems. Energy is defined in thermodynamics as “the ability of a system to perform work.” There are three kinds of energy important in thermodynamics: Kinetic energy, potential energy, and internal energy resulting from intermolecular interactions in a system. Work and heat are means of energy exchanges between a system and its surroundings or another system. Transfer of energy through work mode is a visible phenomenon in macroscopic thermodynamics. However, it is invisible in a nanosystem, but it occurs as a result of the collective motion of a microscopic assembly of particles of the system resulting in changes in energy levels of its constituting particles. Transfer of energy through heat mode is also an invisible phenomenon that occurs at the atomic and molecular level. It is caused by a change not of the energy levels but of the population of these levels. Because of fluctuations in pressure and temperature in small systems, we may not be able to define an isometric or isobaric system in nanoscale. The First Law, as we know it, is fully applicable in small systems as well as macroscopic systems.

Although the First Law, which is actually the conservation of energy, is one of the fundamental laws of physics, it alone cannot explain everything related to conversion of thermal to mechanical energy. Another law is required to understand how thermal energy can change form to mechanical energy—how the energy in a barrel of oil can move an automobile and for what maximum distance, for instance. This is the realm of the Second Law and entropy production.

The Second Law of thermodynamics was originally proposed in the nineteenth century by Lord Kelvin, who stated that heat always flows from hot to cold. Rudolph Clausius later stated that it was impossible to convert all the energy content of a system completely to work because some heat is always released to the surroundings. Kelvin and Clausius had macrosystems in mind in which fluctuations from average values are insignificant in large timescales. According to the Second Law of thermodynamics, for a closed (controlled mass) system we have [13],

$$dP_S = dS - \delta Q_{in}/T_{ext} \geq 0 \quad (2)$$

This definition of the differential entropy production in a closed (controlled mass) system, dP_S , which is originally developed for macrosystems, is still valid for nano/small systems. As it will be demonstrated later, for small systems, the term entropy, S , may assume a new statistical form because of the nonextensivity of nanosystems. Now, by joining the First and the Second Law equations for a closed system, and with the consideration of the definition of work, $\delta W_{in} = -P_{ext} dV$, the following inequality can be derived for the differential total entropy of a system:

$$dS \geq (1/T_{ext}) dE + (P_{ext}/T_{ext}) dV \quad (3)$$

This is a general inequality that seems to be valid for large as well as small systems. Of course, for a system in equilibrium, the inequality sign will be removed, and the resulting equation is the Gibbs equation. However, because of appreciable fluctuations in small systems, definition of the static equilibrium, as we know it for large systems, will be difficult, if not impossible. In nanosystems and in very short periods of time, such fluctuations are observed to be significant, violating the Second Law [15]. However, for longer time periods, nanosystems are expected to be closer to reversibility than macroscopic systems.

The Third Law of thermodynamics, also known as “the Nernst heat theorem,” states that the absolute zero temperature is unattainable. It is impossible to reach absolute zero because properties of macroscopic systems are in dynamic equilibrium and their microstates fluctuate about their average values. Likewise, the energy of a macroscopic system is an average value, and individual microstates will fluctuate around this value. As a result, the temperature can not reach absolute zero because of energy fluctuation. The nonattainability of absolute zero temperature seems also to be valid for nanosystems, because of also to fluctuations.

However, it may be possible to devise a confined nanosystem whose fluctuations are more damped and, as a result, more likely to approach closer to absolute zero temperature than macrosystems [16], or to have it become even negative, as for paramagnetic spins that are cooled off by applying a magnetic field so that entropy decreases with energy and leads to negative temperatures.

2.1. Nonextensivity of Nanosystems [17]

In the 19th century, Ludwig Boltzmann derived the Second Law by assuming that matter was composed of particulate bodies—atoms and molecules by applying Newtonian mechanics along with principles of statistics. According to Boltzmann, the Second Law of thermodynamics is probabilistic in nature. He worked on statistical mechanics, using probability to describe how the properties of atoms determine the properties of matter. In particular, he demonstrated the Second Law of thermodynamics in a statistical statement form. According to Boltzmann:

$$S = k \ln(W) \quad (4)$$

where S is the entropy of a system, $k = 1.33 \times 10^{16}$ erg/°C is the thermodynamic unit of measurement of entropy, now known as the Boltzmann constant, and W is the “probability” of the system in its mathematical sense i.e., the number of distinct ways of arranging the atoms or molecules that are consistent with the overall properties of the system.

Two important characteristics of Boltzmann entropy are [18]: its nondecrease—if no heat enters or leaves a system, its entropy cannot decrease; and its additivity—the entropy of two systems, taken together, is the sum of their separate entropies.

However, in statistical mechanics of finite (nano) systems, it is impossible to completely satisfy both of the above-mentioned characteristics [19].

Boltzmann entropy, S , as defined by Eq. (4), is for a macroscopic (large) state over a statistical ensemble. It is considered the entropy of a coarse-grained distribution, and it later was expressed by Gibbs in terms of the probability distribution of the observational states of the system, resulting in the well-known Boltzmann–Gibbs equation:

$$S = -k \sum_{i=1}^W p_i \ln p_i \quad (5)$$

The Boltzmann–Gibbs expression for entropy was derived by Gibbs from Eq. (4) in the 1870s by considering a system consisting of a large number, N , elements (molecules, organisms, etc.) classified into W classes (energy-states, species, etc.). W is the total number of such microscopic possibilities. In this equation, p_i is the probability of distribution of a set of particles i in the system.

For over a century, engineers, physicists, and chemists have used this formula for entropy to describe various macroscopic physical systems. It is the starting point of the science of statistical mechanics and thermodynamics and of the formulation of ensemble theories [20–23].

Boltzmann and Boltzmann–Gibbs entropy formulas have limitations because they are based on the assumption of additivity of entropy [24]. The very notions of extensivity (additivity) and intensivity in thermodynamics are essentially based on the requirement that the system is homogeneous, which is provided for big systems with weak interactions or, more precisely, in the thermodynamic limit, $N, V \rightarrow \infty, N/V = \text{const}$. These notions make no strict sense for inhomogeneous systems such as finite (nano) systems or systems characterized by the size of the order of correlation length [25]. This is indicative of the fact that small/nano systems are truly nonextensive, and statistical mechanics of nonextensive systems may be applied for these cases.

It has been demonstrated that the domain of validity of classical thermodynamics and Boltzmann–Gibbs statistics, Eq. (5), is restricted, and as a result, a good deal of attention has been paid to such restrictions [24–27]. This branch of science is categorized in a special part of thermodynamics, which is named “nonextensive thermodynamics.” Nonextensive

thermodynamics or thermodynamics of nonextensive systems has a formalism that is proper for the study of complex systems that do not exhibit extensivity.

To overcome difficulties of nonextensive systems, a new statistics was proposed by Tsallis [24] that has recently been modified [26]. According to Tsallis, entropy can be written by the following equation:

$$S_q = k \frac{1 - \sum_{i=1}^W p_i^q}{q - 1} \left(\sum_{i=1}^W p_i = 1; q \in \Re \right) \quad (6)$$

where k is a positive constant and W is the total number of microscopic possibilities of the system. This expression recovers the usual Boltzmann–Gibbs entropy ($S_q = k \sum_{i=1}^W p_i \ln p_i$) in the limit of $q \rightarrow 1$. The entropic index q characterizes the degree of nonextensivity reflected in the following pseudoadditivity entropy rule [24, 26],

$$S_q(A + B)/k = S_q(A)/k + S_q(B)/k + (1 - q)[S_q(A)/k][S_q(B)/k] \quad (7)$$

where A and B are two independent systems. The cases $q < 1$, $q = 1$ and $q > 1$, respectively, correspond to superadditivity (superextensivity), additivity (extensivity), and subadditivity (subextensivity). Some refer to q parameter as the entropic index or the nonextensive index.

The above expression for entropy is applied in many systems, including cosmic and nano systems. Equations (6) and (7) for entropy can be adapted to suit the physical characteristics of many nonextensive systems while preserving the fundamental property of entropy in the Second Law of thermodynamics; namely, that the entropy production of a system is positive with time in all processes [27]. Although Eq. (6) for entropy reduces to the Gibbs–Boltzmann’s expression, Eq. (5), in the case of extensivity (i.e., when the entropy of a system is merely the sum of the entropies of its subsystems). This entropy expression is considered to be much broader than the Gibbs–Boltzmann expression as it describes many nonextensive phenomena, including small systems, that are of interest in this report.

As mentioned above, the proposed general form of entropy of nonextensive systems, given by Eq. (6) and the entropic index q (intimately related to and determined by the microscopic dynamics) characterizes the degree of nonextensivity of the system. However, the entropy in Eq. (6) reduces to the usual Gibbs–Boltzmann formula, Eq. (5), in the limit $q \rightarrow 1$. For $q \neq 1$, S_q is not extensive and gives rise to a host of new and interesting effects (which would be relevant for the description of thermodynamically anomalous systems).

2.2. The Gibbs Equation for Nanosystems

For a stationary, large closed system at thermal and mechanical equilibrium with its surroundings the Second Law inequality [Eq. (3)] will be reduced to the following equation

$$dU = T dS - P dV \quad (8)$$

This is known as the Gibbs equation for a closed macroscopic system at equilibrium. Considering that all the terms appearing in this equation are system properties and the differentials are all exact differentials, it can be concluded that

$$U = U(S, V, N) \quad (9)$$

The Gibbs equation for a macroscopic open (controlled volume) system containing a pure compound will then take the following form: Because $U = U(S, V, N)$, for an open pure system, we can write

$$dU = (\partial U / \partial S)_{V, N} dS + (\partial U / \partial V)_{S, N} dV + (\partial U / \partial N)_{V, S} dN \quad (10)$$

Where,

$$\mu = (\partial U / \partial N)_{S, N} \quad (11)$$

is defined as the chemical potential. As a result

$$dU = T dS - P dV + \mu dN \quad (12)$$

This is known as the Gibbs equation for open pure systems. If we wish to derive parallel equations to the above for nanosystems, we must consider the fact that surface effects are much more appreciable in small systems compared with macroscopic systems. For small systems with nonnegligible surface effects proportional to $N^{2/3}$, another term, at the ensemble level, is added to the Gibbs equation [8, 9]. As a result, the detailed geometry and structure of the system, and the fact that there exist surface forces, edge effects, system rotation, translation, and so forth, must be taken into account. In this case, it is assumed that the following relation will hold for small closed systems [8]:

$$dU = T dS - P dV + \varepsilon dn \quad (13)$$

In this equation,

$$\varepsilon = (\partial U / \partial n)_{S, V, N} \quad (14)$$

is named the subdivision potential, and n is called the “number of noninteracting small systems,” a number that is generally much smaller than the number of particles (N) in the nanosystem. Then for small systems, instead of Eq. (9), the following equation will hold,

$$U = U(S, V, N, n) \quad (15)$$

According to Hill [8], for small systems ($N < N_{\text{Avogadro}}$) we must consider the detailed geometry and structure of the system and the fact that there exist surface forces, edge effects, and so on. In this case it is suggested that one uses the following relation for the Gibbs equation for a multicomponent small system:

$$dU = T dS - P dV + \sum_i \mu_i dN_i + \varepsilon dn \quad (16)$$

In the above equations,

$$\varepsilon = (\partial U / \partial n)_{S, V, N_i} \quad (17)$$

which is the subdivision potential in a mixture. It is a result of contributions of surface effects, edge effects, system rotation and translation, and so forth, all of which are appreciable in small systems and are negligible for macroscopic systems; the term εdn in Eq. (16) does not contribute appreciably to large systems, but the effects just mentioned are not negligible if the system under consideration consists of an assembly of a small number of molecules. In this equation, n is called the number of noninteracting smaller systems inside the nanosystem, which is generally much smaller than the number of particles (N) in the nanosystem.

Equation 16 reduces to the following format in the case of a one-component small system:

$$dU = T dS - P dV + \mu dN + \varepsilon dn \quad (18)$$

hence

$$U = U(S, V, N, n) \quad (19)$$

Equation (18) can be rearranged in the following form:

$$dS = (1/T) dU + (P/T) dV - (\mu/T) dN - (\varepsilon/T) dn \quad (20)$$

then, upon integration, we will have

$$S = (1/T)U + (P/T)V - (\mu/T)N - (\varepsilon/T)n \quad (21)$$

Through the application of nonextensive statistical mechanics, one can find a statistical mechanical expression for the subdivision potential through which it makes it possible to formulate the details of the thermodynamics of small systems. It is also possible to solve for the partition function of small systems [17]. The difference between extensive and nonextensive systems in Eq. (21) is the term $-(\varepsilon/T)n$. This term should be calculated from the difference between extensive entropy and nonextensive entropy [17]:

$$-(\varepsilon/T)n = k\left\{-[1/(1-q)]\left[1 - \sum_i p_i^q\right] + \sum_i p_i \ln p_i\right\} \quad i = 1 \rightarrow W \quad (22)$$

When parameter $q = 1$, the subdivision potential disappears as it should for macroscopic systems. It can also be shown that [17]

$$-\partial[(\varepsilon/T)]/\partial q = -k[1/(1-q)^2]\left[1 + (1-q)\sum_i p_i^q(-1 + \ln p_i)\right] \quad i = 1 \rightarrow W \quad (23)$$

Equations (22) and (23) constitute the statistical mechanical definitions of subdivision potential.

2.3. Statistical Mechanics and Thermodynamic Property Predictions

Statistical mechanical treatment of macroscopic systems consisting of an assembly of molecules starts with the Boltzmann equation, along with the use of statistical ensemble averaging techniques for the probability distribution of molecules or particles in the system. There exist various ensemble averaging techniques in statistical mechanics, including microcanonical, canonical, grand canonical and Gibbs ensembles.

It should be pointed out that the use of thermodynamic limit and of extensivity are closely interwoven with the development of classical statistical mechanics, as reported in many books published on this topic [20–23]. Hence, to extend the use of statistical mechanics to nano/nonextensive/small systems, we must go back to its foundations and establish the new formalism starting from the safe grounds of mechanics and statistics.

Wang and coworkers [26] recently have formulated the grand canonical partition function of a nonextensive system in the following form:

$$Z_q = \left\{ \sum_i^v \exp[-q\beta(\varepsilon_i - \mu N_i)] \right\}^{1/q} \quad (24)$$

For an ideal gas by considering Eq. (24) the partition function will reduce to the following form:

$$Z_q = \left\{ \frac{V}{h^3} \left(\frac{2\pi m k T}{q} \right)^{3N/2} \right\}^{1/q} \quad (25)$$

Based on the well-known equation,

$$U_q = - \left(\frac{\partial \ln Z}{\partial \beta} \right)_{N, V} \quad (26)$$

we can write the chemical potential in the following form,

$$\mu = G_q/N = [U_q + PV - TS_q]/N \quad (27)$$

Therefore, we are in the position that we can develop an equation for subdivision potential.

It is seen that for conventional (extensive) systems in which $q = 1$, we have $\varepsilon = 0$, and this is quite in agreement with the basic arguments in nonextensive thermodynamics.

With the availability of analytic expressions for the chemical potential and subdivision potential of nanosystems, it would be possible to predict thermodynamic properties and phase transitions in nanosystems. It should be pointed out that the science of phase transitions is the basic science behind molecular self-replication [28, 29].

Most of the existing theories of fluids, solids, mixtures, and phase transitions are developed for systems at an infinite thermodynamic limit using canonical or grand canonical ensembles [6, 7]. The known thermodynamic property relations, such as the virial expansion, van der Waals equation, and so forth, are well defined for large systems in thermodynamic limit conditions. Thermodynamic property relations in nanosystems are functions of the geometry and internal structure of the system under consideration. In contrast to macroscopic thermodynamics of large systems, the thermodynamic properties and property relations of small systems will be generally different in different “environments.” As a result, appropriate ensemble technique must be used for different nano/small systems to predict their behavior accurately.

Principles of phase separations/transitions are well defined and formulated in the thermodynamic limit (infinite in number of atoms or molecules), as is demonstrated for the PVT relation of pure systems depicted by Fig. 2. Considering that nanoscale systems consist of finite number of particles (intermediate in size between isolated atoms and molecules and bulk materials), principles of phase transitions need to be reformulated for nanosystems.

Small systems have an additional degree of freedom unknown to large (infinite) systems. In addition to well-known structural phase transitions such as melting or boiling, fragmentation into several large fragments and possibly also into many monomers or small clusters is characteristic for the disintegration of finite systems [14]. For example, comparing the boiling phenomena in macro and nano systems, [30] it is observed that boiling in nanosystem has a new feature; fragmentation. This is a result of the important size fluctuations and correlations that distinguish fragmentation from evaporation in large systems.

At present it is difficult, if not impossible, to come up with universal thermodynamic property relations valid for all nanosystems. However, computer simulation techniques, as discussed later in this report, may allow us to predict properties and the phase transitions behavior of small systems. To apply statistical mechanics to nanosystems, it is also necessary to measure or compute the intermolecular interactions in the systems, which is the subject of the next section of this report.

3. INTERMOLECULAR FORCES AND POTENTIALS

The most important ingredient in theoretical modeling and simulation methods is the interaction potential among the particles. To have a successful theoretical model or simulation, one needs to adopt a physically correct interparticle potential energy model. The majority of the intermolecular potential models, for example, Lennard–Jones, $\phi(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, are designed to give a statistically averaged (effective) representation of such forces

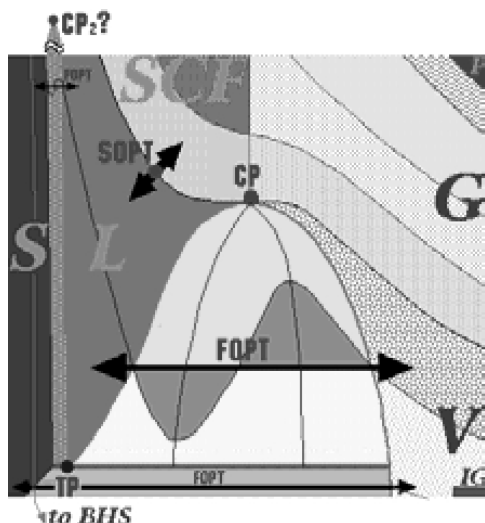


Figure 2. Pure compound phase transitions in the thermodynamic limit.

for macroscopic systems consisting of many particles. Even the ranges of accuracy of the available intermolecular potential parameters and constants are limited and are aimed at the prediction of macroscopic properties. That is probably why the application of the existing force models for the prediction of such nanocrystalline structures as diamondoids, aggregation of amphiphilic molecules into micelles and coacervates, biological membranes, and interactions between lipid bilayers and biological membranes are not quantitatively accurate. The intermolecular potential energy database for fluids and solids in macroscale are rather complete. Parameters of interaction energies between atoms and simple molecules have been calculated through such measurements as x-ray crystallography, light scattering, nuclear magnetic resonance spectroscopy, gas viscosity, thermal conductivity, diffusivity, and the virial coefficients data. Most of the present phenomenological models for interparticle forces are tuned specifically for statistical mechanical treatment of macroscopic systems. However, such information may not be sufficiently accurate in the treatment of nanosystems in which the number of particles are finite and some of the existing statistical averaging techniques fail [6, 31–35].

Exact knowledge of the nature and magnitude of intermolecular interactions is quite important for the correct prediction of the behavior of nanostructures. For example, in the process of formation of crystalline nanostructures, close-packing is achieved when each particle is touching six others that are arranged so that their centers form a hexagon (see Fig. 3). Three dimensional close-packing can be created by adding layers of close-packed planes one above the other so that each particle is in contact with three others in each of the planes directly above and below it. There are two mutually exclusive options for the positions of the particles when commencing a new layer, as indicated by the black and white interstices of Fig. 3. One may associate these options with intermolecular potential energy differences. This is an example of the role of intermolecular forces in structure. Knowledge about intermolecular interactions is very critical in designing and fabricating nanostructure-based devices such as thin film transistors, light-emitting devices, semiconductor nanorods, composites, and so forth. However, the nature and role of intermolecular interactions in nanostructures is very challenging and not well understood [32–36].

Molecular structural characteristics of macroscopic systems based on the knowledge of statistically averaged (effective) pair–interparticle interactions are already defined and formulated in the thermodynamic limit [37–41]. As a result, the available atomic and molecular interaction database for macrobodies by means of effective pair–intermolecular potentials is only an approximation. Because the number of particles are finite in nanostructures, those statistically averaging tools are not well suited for this purpose. As a result, the theory of intermolecular potential functions of such systems and their relationship with their structural characteristics need to be formulated.

For atoms and rather simple molecules, quantum mechanical *ab initio* calculation methods [42–44] have been successful in producing accurate intermolecular potential functions. For complex molecules and macromolecules, the computation is prohibitively difficult to produce accurate potential data.

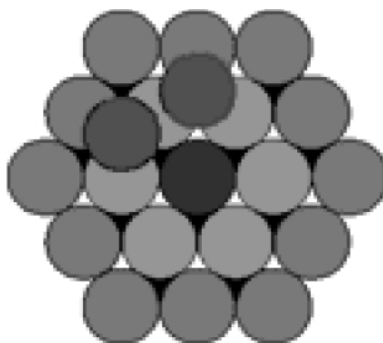


Figure 3. Three-dimensional close-packing options resulting from differences in intermolecular potential energies.

Classical potentials are not always very accurate, and they usually have a simple analytical form, so that numerical computations can be made fast. Thus, when doing modeling or simulation, one should be perfectly aware of the power and the limitations of the potential models being used. One should know in what range and for what purposes the potential model was fitted and made. Classical, and even tight-binding, potential models always suffer from the problem of transferability. For example, if the potential is fitted to reproduce the mechanical properties of the FCC phase of a crystal, one usually cannot expect it to work in situations in which the environment of atoms is different, as in a simple cubic structure with six nearest neighbors instead of 12 in face-centered cubic. Or one can not expect to obtain correct defect energies if the potential model is only fitted to the elastic properties of the bulk phase [45–48]. The analytic forms of those modern interparticle potentials are usually not so simple. Although their two-body interaction forms are complicated, as are discussed below, they could have more than pair interactions in them. Three-body or four-body terms may be added to the two-body interaction. In some other instances, the potential has a two-body form with parameters that now have become dependent on the number and distances of the neighboring atoms.

Now that we have emphasized the importance and limitations of the interatomic potentials in theoretical modeling and simulation, we also discuss the general forms adopted for these potentials and the way they are determined. Interparticle forces and potential energies are known to be, in general, orientation dependent and can be written as a sum of repulsive, London dispersion, hydrogen bonding, and electrostatic energies; that is,

$$\phi(r, \theta_i, \theta_j, \varphi) = \phi_{\text{rep}} + \phi_{\text{disp}} + \phi_{\text{dipole}} + \phi_{\text{quad.}} + \phi_{\text{ind.dipole}} + \dots \quad (28)$$

where r is the separation of the two bodies, θ_i and θ_j are the angles between the molecule axes and the bonds between molecules, and φ is the azimuthal angle. For neutral and spherically symmetric molecules, when the separation r is very small, an exponential repulsive term

$$\phi_{\text{rep}} = \alpha \cdot \exp(\beta/r) \quad (29)$$

dominates, and the potential is strongly positive. Hence, ϕ_{rep} describes the short-range repulsive potential resulting from the distortion of the electron clouds at small separations. For neutral and spherically symmetric molecules, when the separation r is large, the London dispersion forces dominate. From the equation above, the interaction force is

$$F = -[\partial\phi(r)/\partial r]_{\theta_i, \theta_j, \varphi} = F_{\text{rep}} + F_{\text{att}} \quad (30)$$

On the experimental front, the most significant developments were brought about by the invention of the scanning tunneling microscope (STM) in 1982 [49], followed by the atomic force microscope (AFM) [50] in 1986. These are tip-based devices that allow for a nanoscale manipulation of the morphology of the condensed phases and the determination of their electronic structures. These probe-based techniques have been extended further and are now collectively referred to as scanning probe microscopy (SPM). The SPM-based techniques have been improved considerably, providing new tools in research in such fields of nanotechnology as nanomechanics, nanoelectronics, nanomagnetism, and nanooptics [51].

The use of AFM and its modification to optical detection [52] has opened new perspectives for direct determination of interatomic and intermolecular forces. For atoms and molecules consisting of up to 10 atoms, quantum mechanical *ab initio* computations are successful in producing rather exact force–distance results for interparticle potential energy [43, 53–56]. For complex molecules and macromolecules, one may produce the needed intermolecular potential energy functions directly through the use of AFM. For example, AFM data are often used to develop accurate potential models to describe the intermolecular interactions in the condensed phases of such molecules as C_{60} [57], colloids [58], biological macromolecules [59], and so forth. The noncontact AFM can be used for attractive interactions force measurement. Contact AFM can be used for repulsive force measurement. Intermittent-contact AFM is more effective than noncontact AFM for imaging larger scan sizes. Because of the possibility of using AFM in liquid environments [60, 61] it has become

possible to image organic micelles, colloids, biological surfaces such as cells and protein layers, and generally organic nanostructures [4] at nanoscale resolution under physiological conditions.

In addition, making microelectrophoretic measurements of zeta potential allows us to calculate the total interparticle energies indirectly [62]. From the combined AFM and microelectrophoretic measurements, accurate force–distance data could be obtained [63]. From the relation between the force and distance, an interparticle force versus distance curve can be created. Then with the use of the phenomenological potential functions [36], the produced data can be readily fitted to a potential energy function for application in various nanotechnology and nanoscience computational schemes.

Theoretically, there are tremendous challenges because there is no simple way to extract potential energy surfaces from force measurements. A given potential energy surface is a function of the many coordinates of the nanoparticles. Particularly for large molecules, there might be many different potential surfaces corresponding to the same force measurement. Nevertheless, one may be able to tackle this problem with a three-pronged approach:

First, based on modeling experimental data, one may proceed with a top-down strategy. Namely, starting from the experimental data, one can construct a model for different molecules and clusters while drawing on empirical methods to help extract the essential features and build an appropriate potential model for a given compound.

Second, one may also explore a bottom-up approach, by using first-principles methods (density functional theory [64] and quantum chemical techniques [65]) to calculate the potential surfaces for simple molecules. Here we are made virtually free of any input parameters by constructing the model potentials and testing them on much larger molecules. This step can be iterated many times by comparing the quality of the potential with *ab initio* calculations. Once a reliable potential is obtained, it may be used to compute the properties associated with different structures for real materials. At this stage, almost no *ab initio* calculations can be done practically because they would exhaust the present available computational power. This readily leads us to the third step.

Third, one may directly compare theoretical predictions with the experimental models obtained from Eq. (1). This step is central to a project because from the comparison, it will become possible to update the experimental models and initiate new developments with *ab initio* calculations. Once good consistency is found, it should be possible to predict and design new materials. Figure 4 shows how these steps are linked together, and details are presented following the figure.

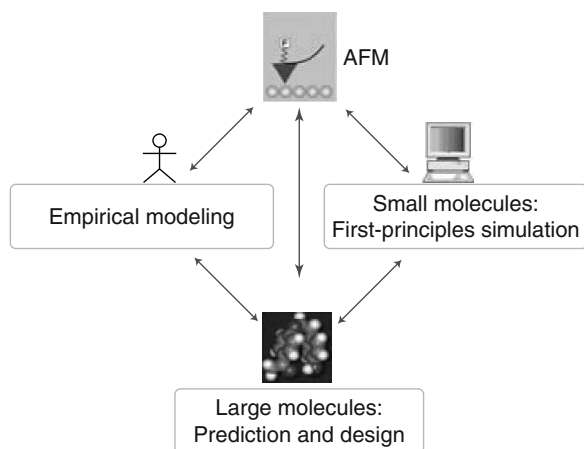


Figure 4. Flowchart for the experimental plus theoretical interparticle force study. Starting from the atomic force microscopy results, an empirical model could be constructed. Both first-principles calculations and experimental results could then be used to predict the properties of materials and, ultimately, to design new materials.

3.1. Step 1: Atomic Force Microscopy Measurement and Empirical Modeling

AFM can provide a highly accurate measure of particle–particle forces by scanning a very sharp tip across a sample. This force depends on the surface morphology and local type of binding—covalent, ionic, metallic, or van der Waals. The change of force versus distance reflects the nature of the potential between the molecules on the surface and the molecules on the tip. There are two tasks at this step: first, as the electronic, Coulomb, and exchange interactions are functions of the inverse distance between electrons, normally the variation of potential versus distance is nonlinear. The information from such experimental relation between force and distance will be used to construct a model potential,

$$\phi(r_{i,j}) = \sum_{i,j} c_{i,j} f(r_{i,j}) \quad (31)$$

where $c_{i,j}$ is the fitting parameters and $r_{i,j}$ represents the coordinates of the atoms. In general, f should be a function of both distance and angle. If we assume that f is a function of distance, there are several commonly used functional forms for f in the literature, as shown in Table 1.

The potential function also includes higher-order interactions (three-body, four-body, etc.). Because normally the potential is also a function of angle–bend, stretch–bend, torsion and rotation, based on the first approach given above, one may perform angular-dependent force measurements. This would provide detailed information about the interactions between different molecules at different angles. After finishing these two tasks, it should be possible to build sophisticated potential surfaces that are functions of both distance and angles. These potentials could then be used to construct intramolecular and intermolecular potentials. Both of them are of critical importance for understanding and designing biochemical molecules.

The obtained potentials could be used to calculate the images of the nanostructures under investigation in different AFM scanning regimes (analogous to those shown in Fig. 5). Such theoretical images could be verified with the highly accurate AFM measurements.

By using potentials that have several levels of accuracy in a computer simulation calculation or a molecular-based theoretical model, one must know what kind of result to expect and always keep in mind for what kind of calculations these potentials can be used. For instance, if a potential is fitted to reproduce elastic properties of the diamond phase of carbon, then one will use this potential to perform simulation or calculation of the diamond phase and cannot expect, in general, to get accurate answers for the elastic or other properties of another phase of carbon with a different coordination number.

3.2. Step 2: Theoretical Modeling

One may use the full-potential linearized augmented plane wave (FLAPW) method within the density functional theory (WIEN package [66] and an appropriate pseudopotential code [67]), along with the available quantum chemical techniques (Gaussian98 [42] and MOLPRO [43]), to compute the potential surfaces for a sample molecule. In the following text, we briefly review the FLAPW method, which is among the most accurate density-functional-based methods, whereas a full description can be found in Ref. [68].

Table 1. Some pair potential function.

Name	Form	Applicable Range
Harmonic oscillator	$(1/2)K(r_1 - r_2)^2$	Intermediate (chemically bonded)
Morse	$D[e^{-2a(r_1-r_2)} - 2e^{-b(r_1-r_2)}]$	Intermediate (chemically bonded)
Mie ^a	$\left[\frac{m}{(m-n)}\right] \left(\frac{m}{n}\right)^{\frac{n}{(m-n)}} \varepsilon \left[\left(\frac{\sigma}{r}\right)^m - \left(\frac{\sigma}{r}\right)^n\right]$	Long-range (nonbonded)
Coulomb	$\frac{q_1 q_2}{r}$	Ionic bonding (partial or full electron transfer)

^aLennard-Jones is the special case when $m = 12$, $n = 6$.

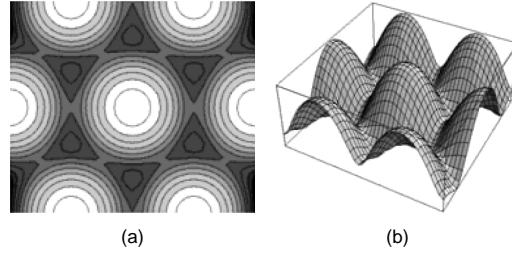


Figure 5. Theoretical simulation of the scanning process at a constant height in the atomic force microscopy (AFM) repulsive mode above the closely packed surface. Both (a) two-dimensional, and (b) three-dimensional images are shown for the vertical force component acting on the AFM tip, depending on its position above a surface. In (a) the darker tone denotes weaker force and the lighter tone the stronger force. Analogous but more complicated figures can be obtained for any model potential describing the intermolecular interaction. Comparison of such kinds, of theoretical figures with the experimental ones will help to make a choice of the most adequate model of the potential.

The LAPW method is among the most accurate methods for performing electronic structure calculations for crystals. It is based on the density-functional theory for the treatment of exchange and correlation, and utilizes, for example, the local spin density approximation (LSDA). Several forms of LSDA potentials exist in the literature, but recent improvements using the generalized gradient approximation are available as well. For valence states, relativistic effects can be included either in a scalar relativistic treatment or with the second variational method, including spin-orbit coupling. Core states are treated fully relativistically. This adaptation is achieved by dividing the unit cell into nonoverlapping atomic spheres (centered at the atomic sites) and an interstitial region (Fig. 6). In the two types of regions, different basis sets are used.

Inside an atomic sphere of radius R_{MT} (where MT refers to “muffin tin” sphere), a linear combination of radial functions times spherical harmonics $Y_{lm}(\hat{r})$ is used,

$$\phi_{k_n} = \sum_{lm} [A_{lm} u_l(r, E_l) + B_{lm} \dot{u}_l(r, E_l)] Y_{lm}(\hat{r}) \quad (32)$$

where $u_l(r, E_l)$ is (at the origin) the regular solution of the radial Schrödinger equation for the energy E_l (chosen normally at the center of the corresponding band with l -like character), and the spherical part of the potential inside the sphere is the energy derivative of u_l (i.e., \dot{u}) taken at the same energy E_l . A linear combination of these two functions constitutes the linearization of the radial function. The coefficients A_{lm} and B_{lm} are functions of k_n (see below), determined by requiring that this basis function matches (in value and slope) the corresponding basis function of the interstitial region; these are obtained by numerical integration of the radial Schrödinger equation on a radial mesh inside the sphere.

In the interstitial region, a plane wave expansion is used,

$$\phi_{k_n} = \frac{1}{\sqrt{\Omega}} e^{ik_n r}, \quad (33)$$

where $k_n = k + K_n$, K_n are the reciprocal lattice vectors and k is the wave vector inside the first Brillouin zone. Each plane wave is augmented by an atomic-like function in every

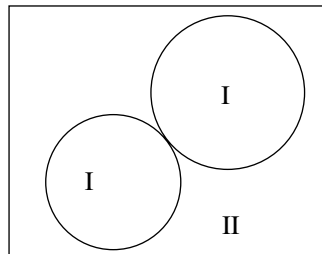


Figure 6. Partitioning of the unit cell into atomic spheres (I) and an interstitial region (II).

atomic sphere. The solutions to the Kohn–Sham equations are expanded in this combined basis set of LAPWs according to the linear variation method,

$$\psi_k = \sum_n c_n \phi_{k_n} \quad (34)$$

and the coefficients c_n are determined by the Rayleigh–Ritz variational principle. The convergence of this basis set is controlled by a cutoff parameter $R_{MT}K_{\max} = 6 - 9$, where R_{MT} is the smallest atomic sphere radius in the unit cell, and K_{\max} is the magnitude of the largest K vector.

In its most general form, the LAPW method expands the potential in the form

$$\phi(R) = \sum_{lm} \phi_{lm}(r) Y_{lm}(\hat{r})$$

for inside the sphere and

$$\phi(R) = \sum_K \phi_K e^{iKr} \quad (35)$$

for outside the sphere, and the charge densities analogously. Thus, no shape approximations are made. This procedure is frequently called the “full-potential” LAPW method (i.e., FLAPW). The forces at the atoms are calculated according to Yu et al. [69]. For implementation of this formalism with the WIEN package, see Kohler et al. [70]. An alternative formulation by Soler and Williams [71] has also been tested and found to be equivalent [72].

The advantage of these theories is that they are normally parameter-free. This is usually an independent check of experimental results and the model potentials. Moreover, with more than 50 years of development, these theories have become quite sophisticated and have strong reliability and predictability. They essentially can treat all kinds of atoms and molecules. For instance, *ab initio* theory was recently used to simulate the reaction path for a retinal molecule in rhodopsin [53–56] (see Fig. 7). It is shown that not only the ground state (Fig. 8a) but also the excited states (Fig. 8b) can be calculated precisely. It is well known that the potential resulting from electrons is much more difficult to treat than that from atoms; this is because the electron density is extremely sensitive to the structure change and external perturbations. Experimentally, the electronic contribution is often very difficult to probe as an external intervention will affect the intrinsic electron distribution substantially. Theoretically, one has the flexibility to calculate these intrinsic properties. One may start from small molecules where no difficulty is expected. This way the potential should be computed easily and is the microscopic potential between molecules and atoms. Building on this effort, one may then calculate larger molecules and clusters. This is a very important leap, where two levels of calculations could be done: from the above microscopic potential, it is then possible to construct a model potential for larger clusters and nanoparticles, and massive parallel computations of realistic potentials for larger nanoparticles consisting of many atoms could be carried out. One may exploit the advantage of parallel computing to find the potential. In particular, for nanoparticles, one can use the chemist’s view to treat different atoms by different computers or nodes. Interactions between these nodes could be communicated through fast Internet connections. With the increase of computational power, one is able to compute even larger clusters, where the potential is “exact.” The model potential resulting from the microscopic potential can be directly compared with this exact potential by improving the quality of the model potential and predictability. This step needs

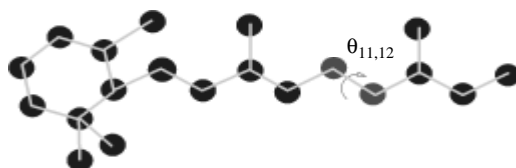


Figure 7. Rhodopsin segment. Isomerization occurs around the bond 11–12 with angle $\theta_{11,12}$.

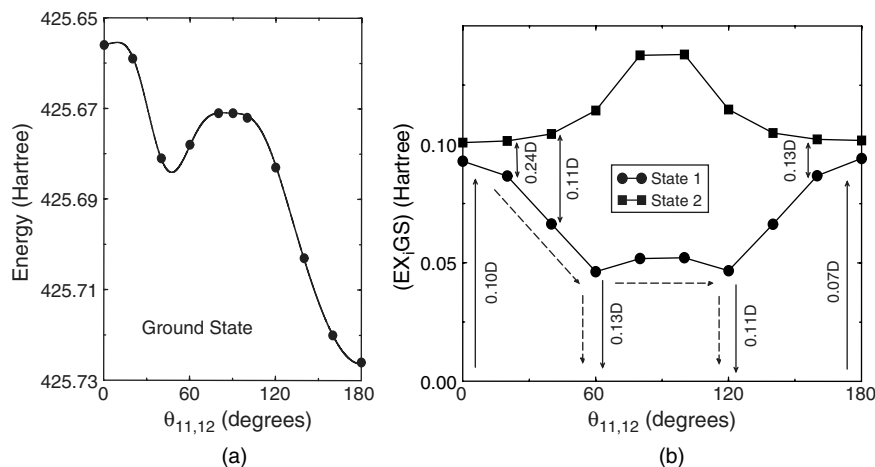


Figure 8. (a) Ground-state potential versus angle $\theta_{11,12}$ in a segment of rhodopsin. (b) Potential for the excited state versus the angle around the bond 11–12. Importantly, the transition matrix elements are also accurately obtained.

much computational power and is usually iterated many times until one could find a reliable potential. This is a very critical step before proceeding to Step 3 below.

3.3. Step 3: Development of Nanoparticle Potentials

When a cluster or nanoparticle becomes larger, a first-principles calculation becomes very expensive and prohibitive. Provided one already has reliable potentials from small and medium-size nanoparticles, it will be possible to construct independent potentials for large cluster or nanoparticle. The resultant potential can be directly compared with empirical models (Step 1). From the comparison, one can refine the experimental models. Such comparison could be repeated until the model potential is optimized. Another direct consequence from this comparison is that it helps one to identify some areas for further theoretical developments. Experimentally, features such as impurities, structural disorders, grain boundaries, and oxidation may be present in real samples. Theoretically, depending on the level of sophistication—self-consistent field, configuration interaction, multireference self-consistent field and multireference configuration interaction—the present first-principles calculations may give different potential surfaces. This is especially true when the compounds have transition metal elements, as often is the case in biological molecules, where the electron correlation is strong. Here, perturbation theory breaks down, although one may be able to develop a very accurate density-matrix renormalization group technique for model systems [53–56]. The initial results are quite promising, where one sees a clear improvement over traditional quantum chemistry methods. The main idea of this method is to retain the most relevant information while properly treating the boundary condition. This method is probably one of the most accurate methods for treating large systems. With the improved quality of theoretical calculations and model simulations, one should be ready to predict and design new materials with desirable features.

4. NANOSCALE SYSTEMS COMPUTER-BASED SIMULATIONS

4.1. Classification of the Methods

The computer-based simulation methods, being developed for nanosystems, generally consist of a computational procedure performed on a limited number of atoms, molecules, or macromolecules confined to a limited, but small, geometrical space. The most important input to such computation is the interparticle energy/force function for interaction between the entities composing the nanosystem. Accurately administered computer simulations can help in three different ways: First, they can be used to compare and evaluate various molecular-based theoretical models. For this purpose, we can use the same intermolecular potential

function in the simulation as well as in the theoretical calculation. Second, for nanosystems, computer simulations can help evaluate and direct an experimental procedure that otherwise may not be possible. Third, an ultimate use of computer simulation is its potential to replace an experiment, provided accurate intermolecular potentials are available to be used in their development. Methods used for simulation of several properties of nanoscale systems differ in their level of accuracy and in the computer time necessary to perform such calculations. Accordingly, the timescales that each of these methods can handle can be from a single total energy for the most accurate calculations, to picoseconds for *ab initio* molecular dynamics simulations, and up to microseconds or even more for classical molecular dynamics. Much longer timescale phenomena such as growth is treated with even less accuracy and very simplified potentials by using the Monte Carlo method. There are extensive references on the subject; however, in what follows, we will mainly give references to books that include a more general treatment of the subject and in which more detailed references can be found.

1. Methods with the highest degree of accuracy (very CPU intensive):

- Input: atomic species and their coordinates and symmetry of the structure; eventually interaction parameters (for model calculations)
- Output: total energy, charge and spin densities, forces on atoms
- Purpose: investigation of both electronic and atomic ground state, optical and magnetic properties
- Examples: *Ab initio* methods for electronic structure calculation of correlated systems [73–75]; Quantum Monte Carlo: variational, fixed-node, Green’s function, path-integral [74, 76]; Quantum chemistry: configuration–interaction [77, 78]; Many-body: GW [79], coupled-cluster [80].

2. Methods with the second-highest degree of accuracy

- Input: atomic species and their coordinates and symmetry of the structure; eventually pseudopotentials or Hamiltonian parameters for the species considered
- Output: total energy, charge and spin densities, forces on atoms, electron energy eigenvalues, capability of doing molecular dynamics (Car–Parinello molecular dynamics of short timescale phenomena <100 ps), phonon spectrum and vibrational modes
- Purpose: accurate calculation of ground state structure by local optimization; calculation of mechanical, magnetic, and optical properties of small clusters and perfect crystals; estimation of reaction barriers and paths
- Examples: *Ab initio* methods for normal Fermi liquids systems (one electron theories) based on either Hartree–Fock (chemistry) [75, 77, 78, 81] or Density functional theories (Physics) [82–86].

3. Semi empirical methods

- Input: atomic species and their coordinates; parameters of the interparticle potential, temperature, and parameters of the thermostat or other thermodynamical variables
- Output of tight-binding (still quantum mechanical): total energy, charge and spin densities, forces on atoms, particle trajectories (capability of doing molecular dynamics: timescales of up to 1 ns or more), phonon calculation; and mechanical, magnetic, and optical properties (approximate) of clusters and crystals
- Output of classical potentials (faster than TB, easily parallelizable): total energy, forces on atoms, particle trajectories (capability of doing molecular dynamics: timescales of up to a microsecond), phonon calculation; mechanical properties (approximate) of large clusters and imperfect crystals (still using periodic boundary conditions)
- Purpose: search for ground-state structure by genetic algorithms (GA), simulated annealing (SA), or local optimization if a good guess for the structure is known; simulation of growth or some reaction mechanisms; calculation of response functions (mostly mechanical or thermal properties) from correlation functions; thermodynamic properties

- Examples: Semiempirical methods for large systems or long timescales [87, 88]; tight-binding or LCAO (quantum mechanical) [88, 89]; molecular dynamics based on classical potentials or force fields [87, 90, 91].

4. Stochastic methods

- Input: parameters of the interparticle potential, temperature, and parameters of the thermostat or other thermodynamical variables
- Output: statistics of several quantities such as energy, magnetization, atomic displacements, and so forth
- Purpose: investigation of long timescale nonequilibrium phenomena such as transport, growth, diffusion, annealing, and reaction mechanisms and also calculation of equilibrium quantities and thermodynamic properties (all with approximate and simple interparticle potentials)
- Examples: Monte Carlo (walk toward equilibrium) [76, 92–94]; kinetic or dynamic Monte Carlo (growth and other nonequilibrium phenomena) [95–97].

In what follows, we will discuss the two nonquantum methods used to deal with the dynamics of atoms in nanoscale systems. The first part of this report introduces the very popular Monte Carlo method, which uses random numbers to perform calculations. Monte Carlo is used in many areas, from the estimation of large-dimensional integrals to generating thermodynamic ensembles, to compute thermal averages of physical quantities of interest, to simulation of non-equilibrium phenomena such as growth and computation of distribution functions out of equilibrium (kinetic Monte Carlo).

The second part concerns the molecular dynamics method, which deals simply with predicting the trajectories of atoms subject to their mutual interactions and eventually an external potential. Some physical quantities of interest such as response functions (viscosity, elastic moduli, thermal conductivity), thermodynamic quantities (such as the total energy and heat capacity), or dynamical properties such as phonon spectra can be deduced from molecular dynamics simulations. Molecular dynamics deals with atomic phenomena at any timescale from femto- to pico-, nano-, or even microseconds. This section ends with the introduction of optimization (local and global) techniques. A comparison between the two methods of Monte Carlo and molecular dynamics in ground state optimization of clusters can be found in [97].

4.2. Monte Carlo Method and Random Numbers

4.2.1. Generating Random Numbers [98]

The basis of Monte Carlo is random numbers. One typically needs to generate random numbers distributed according to some given distribution function. Let us first see how uniformly distributed random numbers can be generated. In general, however, we recommend the reader use the compiler's own random, or rather "pseudorandom," numbers, or subroutines written by professionals. For the sake of completeness, we will briefly mention how one can generate a set of pseudorandom numbers. The prefix "pseudo" is used because the generated numbers are not really random, and they can be regenerated exactly if the same initial seed is chosen, and more importantly, the generated set is periodic, but we try to make the period as large as possible to avoid any correlations among the numbers.

As long as there is no stochastic process occurring in the computer chip, it cannot generate random numbers. All one can produce are pseudorandom numbers generated according to some very specific algorithm. The most widely used algorithm is called the linear congruential random number generator. The pseudorandom numbers are generated according to the following rule [98]:

$$s_{k+1} = \text{mod}[gs_k + c, p] \quad \text{It follows that for all } k \quad 0 \leq s_k < p \quad r_k = s_k/p \quad (36)$$

The numbers r_k thus defined are pseudorandomly distributed in $[0, 1]$. The starting point s_0 is called the seed, so that the same set of numbers can be reproduced if one uses the same seed. The numbers g , c , and p are the generator, increment, and modulus, respectively.

The sequence produced in this way is periodic of period p , but one can make the period very large by choosing appropriate values for the above parameters. The choice of $p = 2^N$ simplifies the mod operation: In the binary representation of a given number, one only needs to keep the p right most digits. N should be chosen as large as possible: $N > 35$ would be a good choice to make the period as large as possible. Usually, N is taken as the number of bits in the largest positive integer. A large value for g will reduce serial correlations. The number c must be prime to p ; $g - 1$ should be a multiple of every prime factor of p . In this case, the period of the sequence would be 2^N . One can also chose $c = 0$; $\text{mod}[g, 8] = 3$ or 5 ; and $s_0 =$ a large odd number. The period in this case would be 2^{N-2} .

Generating random numbers in $[a, b]$ according to the a given distribution function $P(x)$: Importance Sampling. There are a few ways to tackle this problem. We will mention three such methods.

For the sake of simplicity, we can mention one simple algorithm that can easily be implemented and used in one dimension. For a given interval $[a, b]$ and a desired number of random points N , one will first divide the interval into \sqrt{N} pieces (this number is somehow arbitrary) and then generate a uniform set of points in each piece, the number of generated points being

$$NP(x_i) / \sum_i P(x_i)$$

where x_i is a point in the middle of each piece. Clearly, the generated points are distributed according to P , and they are almost N , in number. It must be noted, however, that the generated points are strongly correlated in position and need a reshuffling before being used.

It is also possible to analytically map the uniform random numbers in $[0, 1]$ to $[a, b]$ if the distribution function $P(x)$ is simple enough. If r is uniform in $[0, 1]$, then we have

$$p(r) dr = dr = P(x) dx \Rightarrow r = \int_a^x P(t) dt = g(x, a) \quad g(b, a) = 1 \quad (37)$$

If possible, this relation needs to be inverted to obtain x as a function of r . If r is uniform in $[0, 1]$, then x will be distributed in $[a, b]$ according to P .

Yet, for an arbitrarily complicated function P , a third way is to use the well-known Metropolis algorithm [99]. In this algorithm, one can start from an arbitrary point in $[a, b]$, say $x_0 = (a + b)/2$, and add a random number to it, $x_1 = x_0 + d(2r - 1)$, where r is the random number in $[0, 1]$ and d is the magnitude of the step. If $P(x_1)/P(x_0) > 1$, then the move is accepted; otherwise the ratio is compared to a random number in $[0, 1]$. If the ratio is greater, then the move is accepted; otherwise it is rejected and the old point x_0 is kept, and one starts another trial from this point. There is also the possibility that the generated points go out of $[a, b]$, in which case, they will be rejected. Compared with the first method, this method is by far less efficient because there will be many rejections. In the following paragraphs, we will see why this algorithm generates the correct distribution function. In one dimension, one can see that the efficiency of this method depends strongly on the choice of step size d . With a small d , most trials will be accepted, but a good sampling will require many trials. If d is large, however, one can sample the $[a, b]$ interval quicker, but there might be many rejections. In this simple one-dimensional example, $d = (b - a)/10$ would be a reasonable choice.

A more efficient method from Von Neumann, in 1951, consists of considering a function f [for x in $[a, b]$; $f(x) > P(x)$] for which one can use method the second above and generate points distributed according to f (see Fig. 9). Once such a point x is generated in $[a, b]$, one compares a random number r in $[0, 1]$ to the ratio $P(x)/f(x)$. If r is less than the ratio, the number x is accepted; otherwise, it is rejected and one repeats the process with another point x distributed according to f . In particular, the function f can be taken as a constant larger than the maximum of P .

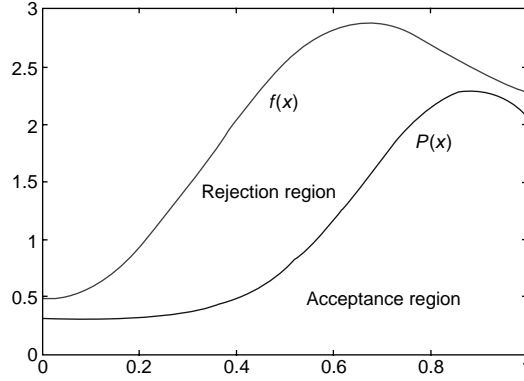


Figure 9. Von Neumann algorithm involves selecting two random numbers in $[0, 1]$, the first one, distributed according to f , is the x -coordinate; and the second one, uniform, the y -coordinate between 0 and $f(x)$. If it falls below $P(x)$, the point is accepted.

4.2.2. Monte Carlo Integration Method

Standard methods of numerical integration of a function are well known (see, e.g., [100]). The simplest method is the trapezoidal method, which adds up the function values at equally separated points and multiplies the result by the width of the intervals. More accurate methods include Simpson's, which instead of approximating the function by linear segments on each interval, replaces it by a quadratic function, and an even more sophisticated method is Gaussian quadrature, where the integral is the weighted sum of the function evaluated at a certain special points. These methods, although quite accurate in one dimension, soon become very prohibitive in higher dimensions, as the number of function evaluations grows as the exponent of the dimension. In this situation too, random numbers can become very handy.

One can approximate the integral by a sum in which the function is evaluated at random points:

$$I = \int_a^b f(x) dx \approx S = \frac{b-a}{N} \sum_{i=1}^N f(x_i) \quad (38)$$

The points x_i are randomly and uniformly distributed on the $[a, b]$ interval. As the number of points N grows, the central limit theorem (CLT) guarantees the convergence of the sum to the real value of the integral; with the error being on the order of $1/\sqrt{N}$. More precisely, the CLT states that if we repeat this process of generating N random numbers and computing the above sum S , we generate a set of random variables S . These variables have a Gaussian distribution of mean I (the real value of the integral) and variance $\sigma = \bar{v}/\sqrt{N}$, where \bar{v} is the mean deviation of $(b-a)f(x)$ from the mean value I , but we do not know what its value is. The term \bar{v} can be estimated by considering the deviation from the mean in a given sample: the usual standard deviation. Of course, v defined as above is sample-dependent; one, therefore, can take its average over all the taken samples. The CLT implies that as the number of samples is increased, the mean obtained from the samples converges to the real mean I like $1/\sqrt{m}$, where m is the number of samples, and the error bar or uncertainty on the accuracy of the result is reduced. The error may be estimated from the CLT:

$$I = \int_a^b f(x) dx \approx S = \frac{b-a}{N} \sum_{i=1}^N f(x_i) \pm (b-a) \sqrt{\frac{\langle f^2 \rangle - \langle f \rangle^2}{N}} \quad (39)$$

where the numerator under the square root may be estimated from the taken sample.

The above method, although convergent, might not be very efficient. If the function has large variations in the interval, for instance, $f(x) = e^{10x}$ in the interval $[0, 1]$, clearly a large number of points and function evaluations will be "wasted" near 0, where the function is much smaller (by more than four orders of magnitude). It would be better to sample the interval in a nonuniform fashion so as to have more points near 1 and fewer points near 0.

This is what is meant by importance sampling. We choose a nonuniform distribution function $P(x)$ according to which points x_i will be sampled. The integral can also be written in the following form:

$$I = \int_a^b f(x) dx = \int_a^b P(x)[f(x)/P(x)] dx \approx \frac{b-a}{N} \sum_{i \in P}^N f(x_i)/P(x_i) \quad (40)$$

Notice that now the function f/P is integrated on points distributed according to P . The function P is of course analytically known and normalized to 1. There would be fewer fluctuations in the integrand f/P if the distribution function P , which we can choose at will, looks like the original function f . Notice that if we take $f = P$ the problem is solved not because the integrand is 1, but because we already know the integral of P .

At any rate, the central idea in importance sampling is to choose P so that it looks like f and therefore f/P does not fluctuate much and the variance of the result is small, or in other words, the estimate of the integral is accurate, even with a relatively small number of points. Table 2 shows a comparison of the estimation of the integral using uniform sampling and importance sampling of the function $f(x) = e^{10x}$ with a quartic distribution function in $[0, 1]$. Note that we preferably want the distribution function P to be nonzero in the regions in which the function f is nonzero; otherwise, the ratio f/P leads to infinity if a random point is generated in that region. This is the reason why a small number was added to $P(x)$ to avoid divergence near zero.

The error can be estimated in the following way: the mean deviations of the function from its mean value here is about 5000 (just look at the plot of e^{10x} in $[0, 1]$, but a better estimate for it can be computed from the deviations from the numerically calculated mean or integral), so $\bar{v} \approx 5000$ and therefore the error (one standard deviation) is of the order of $5000/\sqrt{N}$ for uniform sampling (from the CLT). As for importance sampling, one must evaluate the mean standard deviations of the function f/P in $[0, 1]$ and then multiply it by $1/\sqrt{N}$. We can conclude from the above numerical data that the agreement of the last line to within six digits is not expected and therefore, is fortuitous. To put the statement of CLT in mathematical terms:

$$I = \int_a^b P(x)f(x)/P(x) dx \approx S = \frac{b-a}{N} \sum_{i \in P}^N \frac{f(x_i)}{P(x_i)} \pm (b-a) \sqrt{\frac{\langle (f/P)^2 \rangle - \langle f/P \rangle^2}{N}} \quad (41)$$

From this formula, it is clear that the error is much reduced if the function f/P has weaker fluctuations than f itself. Note that for functions f that change sign, as P has to be always positive, the ratio f/P will always have sign fluctuations, and therefore its variance will never go to zero even as N goes to infinity. This problem can be overcome only if the roots of f are known and the integral can be split to sums of integrals over regions in which the sign of f does not change.

The integration method by Monte Carlo is used to evaluate multidimensional integrals that appear in the evaluation of the ground-state energy of many-particle systems such as few electrons in an atom, a molecule, or a quantum dot [74, 76]:

$$E_0 = \langle \psi | H | \psi \rangle / \langle \psi | \psi \rangle \quad |\psi\rangle = \Psi(r_1, r_2, \dots, r_n) \quad (42)$$

Table 2. Results of the integral $I = \int_0^1 e^{10x} dx = \frac{e^{10}-1}{10} \approx 2202.5466$ using uniform and importance sampling with $P(x) = 5x^4 + 0.0001$.

N	Uniform	Importance	Exact
100	2243.5107	2181.7468	2202.5466
1000	2295.6793	2208.6427	2202.5466
10000	2163.6349	2207.1890	2202.5466
100000	2175.3975	2203.7892	2202.5466
1000000	2211.7200	2202.8075	2202.5466
10000000	2203.8063	2202.5431	2202.5466

This method is called variational Monte Carlo. The trial wavefunction contains a few variational parameters. The term $|\psi|^2$ is taken as the distribution function, and the integral is computed by sampling $H\psi/\psi$ for several values of the variational parameter. The energy is then minimized as a function of the latter. This, as mentioned in the introduction, is the basis for the most accurate calculations of the ground-state properties of a many-particle quantum system.

It can also be used to compute the partition function and average the thermodynamic quantities of a few interacting particles in a finite box [76]:

$$E = \frac{3}{2}NkT + \langle \phi \rangle = \frac{3}{2}NkT + \int dr_1 \dots dr_N \phi(r_1, \dots, r_N) e^{-\beta V(r_1, \dots, r_N)} / \int dr_1 \dots dr_N e^{-\beta \phi(r_1, \dots, r_N)} \quad (43)$$

The typical way importance sampling is done in such calculations is to generate a series of random points distributed according to $|\Psi^2|$ in the quantum case and to $e^{-\beta \phi(r_1, \dots, r_N)}$ in the classical case. In the former, one then computes an average of the “local energy” $E(R) = \hat{H}\Psi(R)/\Psi(R)$; $R = (r_1, r_2, \dots, r_N)$, and in the latter, just the average of the potential energy $\phi(R)$.

Notice that this is just a choice, and one could sample the points R in the three-dimensional space according to any other distribution function. As said before, the idea is to sum a function [$E(R)$, or $\phi(R)$ in our case], that has as little fluctuation as possible. So if the above functions still have large fluctuations, one could use another importance sampling provided that the distribution function is analytically integrable so that the weight is properly normalized. One way such averages may be done is, for instance, to assume a simple function $W(R)$ that “looks like” $\phi(R)$, at least near its minima, and the exponential of which is integrable [a quadratic function, the harmonic approximation of $\phi(R)$, comes naturally to mind]; use sampling points distributed according to $e^{-\beta W(R)}$; and sample the function $\phi(R)e^{-\beta[\phi(R)-W(R)]}$. However, this same function will sample the denominator simultaneously,

$$Z = \int dR e^{-\beta \phi(R)} = \int dR e^{-\beta[\phi(R)-W(R)]} e^{-\beta W(R)} \quad (44)$$

so that both numerator and denominator are computed in one shot. The advantage of this method is that it has also computed the partition function, which might be of use in some other calculations. Note that instead of Boltzmann distribution, used for macroscopic systems, one can choose any other distribution function, such as Tsallis’s, introduced in the first section, to sample the phase space.

The real difficulty in computing integrals accurately, other than the appropriate choice of the weighting function P , is actually generating random numbers that are distributed according to P and that are preferably uncorrelated. This would be the subject of the following paragraphs, in which we will make use of the famous Metropolis algorithm, already mentioned for importance sampling, and the Markov process.

4.2.3. Equilibrium Statistical Mechanics and Monte Carlo

The most widely use of Monte Carlo method has been in solving problems in equilibrium statistical mechanics (SM) calculations [74, 76, 92–94]. Naturally, both fields Monte Carlo and SM deal with statistics and averaging, such as that mentioned in the paragraph on numerical integration. In equilibrium statistical mechanics, one is interested in generating an ensemble of states at the thermodynamic equilibrium so that averages of physical properties can be numerically computed.

The walk toward equilibrium from an arbitrary initial state is done by using a Markov Process (MP). The same process is then used to generate the ensemble of equilibrium states. To avoid correlations between the generated states in the ensemble, one must only take every few of the generated states in order to perform averages, because in a “continuous” walk, points (states) following one another are correlated.

What we mean by “walk” in this Markovian dynamics is going from one state of the system to another by some stochastic process. If this walk is done correctly, meaning any

state is within reach (Ergodicity) provided the walk is continued long enough, then after reaching some relaxation time, one can perform the statistics. The relaxation time is the time necessary for the system to reach thermodynamic equilibrium. It should be mentioned that there is no simple way of finding the relaxation time in advance. The easiest way is to let the system evolve once and then from the time series of different physical quantities estimate the relaxation time.

4.2.3.1. Markov Process The distribution function of the system out of equilibrium usually satisfies a time evolution equation also known as the master equation. It describes how the system jumps from one state to another:

$$\frac{dP_i}{dt} = \sum_j (P_j T_{j \rightarrow i} - P_i T_{i \rightarrow j}) \Rightarrow P_i(t + \Delta t) = \sum_j P_j(t) T_{j \rightarrow i} \quad (45)$$

Here $T_{i \rightarrow j}$ is the transition rate (or probability, if properly normalized) to go from state i at time t to the state j at time $t + \Delta t$. Assuming the states, which are really infinite in number, to be discrete, one can think of T as a matrix that we also call the transition matrix. The term $T_{i \rightarrow j}$ is the element in line j and column i , and $P_i(t)$ is, of course, the probability of finding the system in the state i at time t . In what follows, in case we need to introduce the temperature, we will be introducing and using the inverse temperature β to avoid any confusion with the transition matrix T .

The transition matrix T follows the sum rule (in units of the time step Δt):

$$\sum_j T_{i \rightarrow j} = 1 \quad (46)$$

This simply means that the sum of the transition probabilities from a given state i is 1. We made use of this rule in writing down the second part of Eq. (45), and this form will be used later.

The matrix T might be known for some systems if the physical processes inducing the transitions are known. For our purpose, however, which is just to find a way to generate an ensemble of states at equilibrium, we will design this matrix to achieve this goal. If T is known, we can simulate the time evolution of the distribution function. This will be the subject of nonequilibrium Monte Carlo simulations that which will be discussed later.

A Markov process is a stochastic process by which one goes from a state i at time t to a state j at time $t + \Delta t$. The corresponding transition matrix must have the following properties: the transition rate should not depend on the history of the system, and at any time the transition probability $T_{i \rightarrow j}$ is always the same regardless of what state the system was in at prior time steps; it only depends on the states i and j . To reach the correct equilibrium state of the system, we require the transition matrix modeling our Markov process to have additionally the following properties: Ergodicity and Detailed Balance.

Ergodicity, as we mentioned before, means that all states can be reached in principle if we wait long enough. This is required to guarantee that we will not miss the ground state or get stock in some local minima and not be able to make the correct statistics at equilibrium. Detailed balance, however, ensures that the final stationary state is the thermodynamic equilibrium state. The stationary state condition is $dP_i/dt = 0$, meaning we can have currents in and out of a state, but they add up to zero and there is no net flow:

$$\frac{dP_i}{dt} = 0 \Rightarrow \sum_j P_j^{eq} T_{j \rightarrow i} = \sum_j P_i^{eq} T_{i \rightarrow j} = P_i^{eq} \quad (47)$$

If we think of the distribution probabilities as an array (or a vector) from the above equation, we conclude that the equilibrium distribution is the right eigenvector of the transition matrix with eigenvalue 1. To get rid of the factor Δt , we have chosen the units for time such that $\Delta t = 1$. The time evolution Eq. (1), in its second form, implies that to reach the stationary state, we need to keep multiplying P by T . The sum rule [Eq. (46)] says that the sum of the elements in each column of the transition matrix is always equal to 1. Raising this matrix to “infinite” power will yield a matrix that will have identical columns (we are just stating

this property of the transition matrix without proving it). As a consequence, multiplied by any initial state, we will obtain a state proportional to the columns of T^∞ , which is in fact the equilibrium state. Thus, to obtain the equilibrium probabilities, all one needs to do is to raise T to high powers, until this process converges. The equilibrium probabilities can then be read on any of the columns of T^∞ .

It is also possible to obtain the relaxation time from the matrix T (now in units of Δt). The highest eigenvalue, as we said, will be 1, associated with the eigenvector which contains the equilibrium probabilities. The second-highest eigenvalue is related to the relaxation time. This can be proven if we work in the basis of the right eigenstates of T . Formally, one may write the solution to Eq. (45) as: $P(t) = e^{(T-1)t}P(0)$. If $P(t)$ is expanded on the right eigenstates of T , its “equilibrium component,” corresponding to the eigenvalue 1, has no time evolution. The other components, however, will decay to 0 as time goes by, because all other eigenvalues are real and less than 1. The slowest component is the one corresponding to the second-largest eigenvalue. The timescale over which this decays to 0 would be $-1/(\omega_\lambda - 1)$, where ω_λ is the second-highest eigenvalue of the transition matrix (in units of $1/\Delta t$). Had we integrated the second form of Eq. (45), we would have found that the relaxation time is $-1/\text{Log}\omega_\lambda$, which is very close to the above result, as ω_λ is smaller but close to 1.

What is said above is not much of practical use for doing MC simulations. It just illustrates some of the properties of the transition matrix. If the latter can be estimated or approximated in some way, then it would be possible to obtain the equilibrium distribution by raising T to an infinite power, or by just diagonalizing it. In the latter case, one can also get an estimate of the relaxation time, which contains information on the dynamics of the reactions occurring in the system. For instance, in an adaptive simulated annealing run (see section on optimization), where the cooling rate is dictated by a simple equation as

$$dT/dt = -\left(vT/\varepsilon\sqrt{C}\right) \quad (48)$$

the relaxation time ε , which is $-1/\text{Log}\omega_\lambda$, appears in the denominator. The variable C is the heat capacity of the system, which can be computed either from the knowledge of the transition matrix or from the fluctuations in the energy.

Now let us go back to the detailed balance and the form of the transition matrix we would like to adopt. To make sure detailed balance is satisfied and no cyclic motion appears in the “equilibrium” state, we can impose the following (sufficient but not necessary) microscopic reversibility relation on T :

$$P_i T_{i \rightarrow j} = P_j T_{j \rightarrow i} \quad (49)$$

Equation (48) is the celebrated detailed balance condition. It simply says the currents from i to j cancel the currents from j to i , which is a sufficient condition for being in steady state.

Usually, we know the equilibrium distribution we want to reach. In most instances, we are dealing with the Boltzmann distribution, where $P_i^{eq} = e^{-\beta E_i}/Z$ and Z is the partition function $Z = \sum_i e^{-\beta E_i}$. One can therefore deduce from the detailed balance condition the ratio $T_{i \rightarrow j}/T_{j \rightarrow i} = P_j^{eq}/P_i^{eq} = e^{-\beta(E_j - E_i)}$.

The last point about sampling using a Markov process is the correlation in the time series. When performing ensemble averages, data separated by a time larger than the correlation time should be used. The latter being the time when the autocorrelation function of a given property A decays to 0.

4.2.3.2. Choice of the Transition Function There are many possible choices for the transition matrix with the given constraints of detailed balance and ergodicity and the sum rule of Eq. (46). If one thinks of T as a matrix, if the upper triangular part is known, detailed balance yields the lower triangular part, and the sum rule makes sure the sum of the elements in a column is equal to 1, and thus defines the diagonal elements; ergodicity enforces that not all elements (other than the diagonal terms) in a row or column are 0; otherwise, one will never transit to the state corresponding to that row or column. Therefore, once the upper triangular part is carefully defined, one can obtain all other transition probabilities.

The most widely used transition matrix is the one proposed by Metropolis–Rosenbluth–Rosenbluth–Teller [99]. In this work, the authors used for the first time, in 1953, the MC method to numerically compute the equation of state of a hard-sphere system in two dimensions. In the Metropolis algorithm, the transition probabilities are given by

$$T_{i \rightarrow j} = \text{Min}(1, P_j^{eq}/P_i^{eq}) \quad (50)$$

This means the move from state i to state j is accepted if the latter is more probable; otherwise, it will be accepted with the probability P_j/P_i . In practice, the way this is done is to pick a random number r in $[0, 1]$; the move is rejected if $r > P_j/P_i$, and otherwise it will be accepted. In case of rejection, one starts from state i and tries another move, and when doing the statistics, state i is counted again as many times as the moves were rejected.

Although Metropolis sampling of the phase space works for almost any problem, the fact that there is rejection in the moves makes the algorithm slower, and for any specific problem, one can usually devise better and more adapted algorithms. Some of these more advanced methods will be discussed later. The Monte Carlo method combined with the Metropolis algorithm was used for the first time in 1964 by McMillan [101] to treat a quantum system by the variational method. Below, we will illustrate the results obtained by this method.

For example, one application of the Monte Carlo method is to, as we said, sample the phase space and produce an ensemble of equilibrium states at a given temperature T . As an illustration, we show in Fig. 10 the graphs of the energy fluctuations and distribution function obtained for a 13-atom cluster of argon atoms interacting via the Lennard–Jones potential. After some relaxation time, the ground-state configuration, which is an icosahedron, is obtained and averages can be performed after the ground-state structure is reached (after 20,000 steps in the graph below) to obtain the average energy $E(T)$ and the heat capacity, which can be deduced either from the derivative of E or, even better, from its fluctuations:

$$C = \frac{\partial \langle E \rangle}{\partial T} = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_B T^2}$$

using the definition

$$\langle E \rangle = -\frac{\partial \text{Log } Z}{\partial \beta}$$

In Figure 11, three configurations of the cluster at the initial, middle, and final steps are shown.

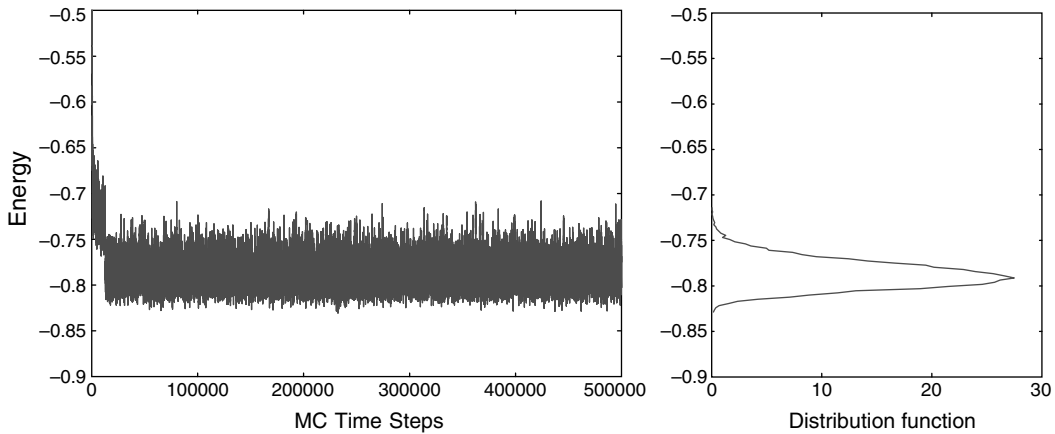


Figure 10. The graphs of the energy fluctuations and distribution function obtained for a 13-atom cluster of argon atoms interacting via the Lennard–Jones potential.

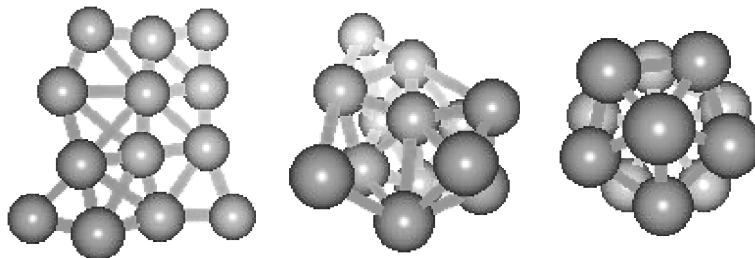


Figure 11. Three configurations of the cluster at the initial, middle, and final steps respectively.

4.2.3.3. Acceptance Ratios and Choice of the Moves So far, we have not discussed how to make the moves. If dealing with discrete degrees of freedom such as in the Ising model, a move could consist of a single spin flip at a time, but this could turn out to be a very slow process if one is near the Curie temperature where fluctuations are large, and all such moves, will almost be accepted. Flipping many spins at the same time could also result in many rejections, especially if the temperature is low. So, depending on the situation (conditions of temperature or other parameters), different choices need to be made about making moves in the phase space. A useful concept in this case is the Acceptance Ratio [74, 96], which says what percentage of tried moves were actually accepted during a given number of trials. A good rule of thumb for the Metropolis algorithm is that moves should be chosen such that the acceptance ratio is kept between 20% and 50%.

As we said, we have a choice of how to make the move, irrespective of its acceptance probability. This freedom in choice can allow one to make moves with high acceptance probability, and thus have an algorithm that can be more efficient than the simple “blind” Metropolis. For instance, if one is sampling the canonical ensemble, moving to points with high energy have a good chance of being rejected. One could instead move to points with “not so high” energy, with a larger chance of being accepted. This presumes a knowledge of the local energy landscape. If one has such information, it must be used to have an efficient sampling. Therefore, although Metropolis works almost for any problem, one can always find a more suitable moving algorithm for the specific problem at hand, which is more efficient in terms of sampling the largest part of the phase space in the shortest time and having the smallest variance.

One such example is the Swendsen–Wang [96, 102] algorithm, which is used near the Curie temperature. Indeed, at the transition temperature, fluctuations in the magnetization are large and the correlation length diverges. Therefore, a single spin flip is considered a small move and is almost always accepted. The idea is to use multiple spin flips, as in a cluster flip, but one can not consider solid clusters of same spin because they would always keep their shape. We need, therefore, to connect parallel spins with some probability p . Choosing the appropriate $0 < p < 1$ is the key to this fast algorithm. It can be shown, though we will not do it here, that the optimal p satisfies $1 - p = e^{-2\beta J}$. Clusters constructed with this value of p will always be flipped (large moves with acceptance ratio = 1), and one obtains a fast and good sampling of the canonical ensemble even near the Curie temperature.

To conclude, the best choice of algorithm is really problem dependent, and one can not design one that works well for all problems. Simulation algorithms are still an active field of research.

4.2.3.4. Other Tricks to Improve the Speed, Faster than the Clock Algorithm In the case of simulation of statistical properties of discrete systems such as an Ising lattice, the walk toward equilibrium is usually done with single spin flips. In this case, it is not necessary to recompute the total energy at each step, but just the change resulting from the single flip:

$$\Delta E(i) = -2 \sum_j J_{ij} S_j = -2 \sum_{j \text{ near } i} J_{ij} S_j \quad (51)$$

This is a finite sum because the exchange integral J is usually short-ranged. All possible neighboring configurations can be considered, and all possible energy changes and transition

probabilities $e^{-\beta\Delta E}$ can be computed and tabulated once and for all before the Monte Carlo run, so that in case of need they will only be looked up and not computed every time.

Another way to speed up the Monte Carlo runs is to use the heat bath algorithm [92]. For a spin chosen at random, one can compute the local probability distribution directly, meaning the energies of all possible states of the spin, and therefore the Boltzmann probabilities of being in these states are computed. By generating a single random number r , one can decide in which “local equilibrium state” the spin will be. If there are m possible states $1 \leq j \leq m$, and

$$\sum_{l=1}^{j-1} P_l \leq r < \sum_{l=1}^j P_l \quad P_l = e^{-\beta E_l} / \sum_{i=1}^m e^{-\beta E_i} \quad (52)$$

Then the state j is chosen as the equilibrium state. This will be done for all spins many times until the system relaxes to equilibrium, after which, relevant statistics may be done. The advantage of this method is that the local equilibrium is accessible in one step. The disadvantage is that updates are still local, and the probability distribution may be difficult or time-consuming to compute.

There are also what is called faster than the clock algorithms. Instead of trying flips at every time step and computing transition probabilities (Boltzmann factors), one can compute how many steps it takes before a flip occurs. Imagine the simple case in which the transition rate is a constant: The probability of transition between time t and $t + \Delta t$ is $\Delta t/\tau$. Therefore, the probability of no transition from 0 to t , and transition between t and $t + \Delta t$, is $p(N\Delta t) = (1 - \Delta t/\tau)^N \Delta t/\tau$, so that the total probability is

$$\sum_{N=0}^{\infty} p(N\Delta t) = \sum_{N=0}^{\infty} (1 - \Delta t/\tau)^N \Delta t/\tau = 1 \quad (53)$$

To find out when the transition occurs, one needs to generate a random number r and find N , for which $p(N\Delta t) > r \geq p[(N + 1)\Delta t]$. This implies

$$N \approx \text{Log}(r\tau/\Delta t)/\text{Log}(1 - \Delta t/\tau) \Rightarrow t/\tau = \text{Log}(\Delta t/\tau) - \text{Log } r \quad (54)$$

This is called a “faster than the clock” algorithm because you can find out the time of transition and flip the particle before the “natural time” it needs. Another advantage of this algorithm is that it has no rejection. Furthermore, its output (flipped moves and their times) can not be distinguished from a Metropolis algorithm. These kinds of algorithms are also called dynamic Monte Carlo methods [96, 103].

4.2.4. Application of Monte Carlo to Nonequilibrium Problems

The Monte Carlo method can be used for simulating the time evolution of a stochastic system in which the transition rates are known. Here, in contrast to the previous section, which dealt with equilibrium, where we could choose the transition rates at will, provided that detailed balance and ergodicity were preserved, the rates are modeling a real physical phenomenon and are given in advance. They dictate the dynamics of the system. Typical problems that are solved using this method are problems of general transport: carrier transport in electronic devices, for instance, dealing with solving the Boltzmann transport equation [95]; mass transport; and physical reactions such as diffusion, desorption, and adsorption, growth [97, 104], and so forth.

To the master equation, which represents the time evolution of a distribution function, one can associate an equation of motion for particles subject to external forces, as well as a stochastic term (in case of contact with a heat bath). This is the Langevin equation. If the particles follow this equation, then their distribution function obeys the master equation. Therefore, instead of solving the master equation to find the time evolution of the distribution function, one can consider a set of particles and move them according to the corresponding Langevin equation and then record the quantities of interest such as total energy, density, magnetization, current, and so forth as a function of time for statistics purposes, as ultimately, one wants to compute the average of these physical observables.

For simplicity, we will consider noninteracting particles. In this case, the many-particle distribution function can be factored into the product of single-particle distribution functions. One can then consider the Langevin equation for a single particle, simulate the time evolution of the particle, and record averages. For interacting systems, it is also possible to use mean-field theory and do the same factorization. Many particles are moved in the external plus random field, to which one should also add the field caused by other particles.

A simple example is the study of transport using the Boltzmann equation. This equation can be solved in some limits or within some approximations, such as the relaxation time approximation. The obtained solutions, though not very accurate, can give insight into the physics of the problem. More accurate solutions can be obtained using the dynamic (also called kinetic) Monte Carlo method to numerically integrate the master equation (again by considering a population of particles) as the transition rates involved are not usually so simple and an analytic treatment is often impossible. We will give below the general algorithm for finding the time evolution of the distribution function in a very general and abstract way, though in practice the way the algorithm is implemented is quite different from the one stated here. However, it gives the idea of how this should be done in general. For now, we will not worry about the efficiency of such an algorithm.

Consider the master equation [Eq. (1)] with known transition rates:

$$\frac{dP_i}{dt} = \sum_j (P_j T_{j \rightarrow i} - P_i T_{i \rightarrow j}) \Rightarrow P_i(t + dt) = \sum_j P_j T_{j \rightarrow i} \quad (55)$$

$i = 1, \dots, s$ labels the s different (here discretized) states of the system. For a noninteracting system, s is just the number of one-particle states. If one is dealing with a particle in a box, the latter can be discretized into a lattice and s will be the number of lattice sites times the number of possible momenta, which would depend on a chosen momentum cutoff. You can see that soon this number becomes very large. For an N -particle system, the number of possible states is really $S = s^N$, and the state index i is really going from 1 to S . The state of the system can be thought of as one filled box among the S available ones. The initial box i is chosen at random; then the particle is moved from this box to another one according to the transition rules defined by $T_{i \rightarrow j}$. If the transition matrix ($S \times S$) is known and calculated at the beginning only once (which is not true for an interacting system in which the transition rates might depend on the present state density) of the system, in which case this matrix must be updated at every time step), then the state at the next time step is simply obtained by choosing a random number $0 < r < 1$ and seeing to which transition $i \rightarrow j$ it corresponds. The state j is chosen as the next step if

$$\sum_{l=1}^{j-1} T_{i \rightarrow l} \leq r < \sum_{l=1}^j T_{i \rightarrow l} \quad \text{since} \quad \sum_{l=1}^s T_{i \rightarrow l} = 1 \quad (56)$$

This process can be iterated on the j th column and so on until the equilibrium is reached and statistics can be performed. It has no rejection: Each moving step does not correspond to any physical time step, as some of the chosen rates are large and some others small. Furthermore, physical parameters of the system such as temperature are included in the transition rates. Now, in general, for a real system, many of the transitions are very improbable or simply impossible, and therefore the T matrix has a majority of zeros, not considering the fact that its size is huge.

This algorithm, although very general and straightforward, is impractical in most cases because of the large number of states the system can have. It can be refined, however, in many cases. In most practical cases, a set of particles is assumed and moved according to the Langevin equation corresponding to the master equation. From their trajectories after the relaxation time has passed, the steady-state distribution function and other averages can be computed. The short time-equilibration process may also be observed and may turn out to be useful in such simulations. This method, also called kinetic Monte Carlo, is also used to investigate diffusion and growth processes that take too long to be simulated by molecular dynamics methods.

Consider a population of N interacting particles subject to external fields and undergoing scattering and diffusion processes of known rates. External and mean fields come in the left side of the master equation, where a time derivative is present, such as in the Boltzmann equation. Scattering and diffusion processes come in the right side of the master equation, where transition rates are present. Assume that there are $i = 1, 2, \dots, p$ types of different processes, each with a given rate w_i .

For a noninteracting system, one can only consider one particle and decide by using a random number which process it will undergo: if $0 < r < 1$ is random, and

$$\frac{1}{z} \sum_{i=0}^{j-1} w_i \leq r < \frac{1}{z} \sum_{i=0}^j w_i$$

where

$$z = \sum_{i=1}^p w_i + w_0$$

then the process w_j is chosen. The rate w_0 is a fictitious rate called the “self-scattering” rate to allow the computation of “free flight” or “waiting” time of the particle. For more details, see the book by Singh and references therein [95], which discusses the Monte Carlo method in transport in semiconductors. (If the process w_0 is selected, then there is no scattering at all.)

The particle is then scattered according to this process into some other state, which might also require a random number to be selected. Next, one has to find out how long it takes before the particle undergoes the next scattering. This time depends clearly on the present scattering rates. As explained in a simple example before, the waiting time is calculated according to

$$T = -\ln r \bigg/ \sum_{i=0}^p w_i$$

where again r is random in $[0, 1]$. This amounts to advancing the “physical clock” by T , whereas the “simulation clock” is only incremented by 1. Again, a new scattering process must be chosen at random. This needs to be done for a long time on a single particle, and then ensemble averages are replaced by time averages. If large amounts of memory and a small number of states are available, one can compute and tabulate all the scattering rates in advance, store them, and just look them up as needed.

For interacting particles, one can map to a similar problem as that just mentioned. Instead of p processes, we now have Np processes. One of these processes must be picked at random. This is equivalent to picking a particle and a process at random. Let it scatter and update the state of the system (potentials and forces). In this case, the scattering rates may need to be computed as one goes on, and they cannot be tabulated because they depend on the state of the system. This method suffers, however, from the approximation in which all particles are not moved at the same time. However, it can still be reasonable as the total rate, the denominator z , is now much larger, and therefore the free flight time becomes much smaller—by a factor of N to be more precise. This calculation soon becomes prohibitive as the number of particles is increased as the needed CPU becomes proportional to N . An alternative in this case is to adopt the idea of molecular dynamics, and fix the time step Δt such that it is at least an order inferior to the smallest timescale (or largest transition rate), and to move all particles at the same time. As the term $w\Delta t$ is generally small, most of the time particles will have a “free flight” according to the external and mean field potentials. The way scatterings are decided at each time step is to pick a random number in $[0, 1]$, and make the following comparison:

$$\sum_{i=1}^{j-1} \Delta t w_i \leq r < \sum_{i=1}^j \Delta t w_i$$

Note that the index I starts from 1: One only considers physical processes. Because $\Delta t w_i$ is chosen to be small, scattering events do not occur very often. There might be other methods to handle the interacting problem of which the authors are not aware.

4.3. Molecular Dynamics Method

Atoms in a solid or a cluster, except for hydrogen and helium, which have a relatively small mass, are considered classical objects. Indeed, their thermal wavelength at temperature T is $\Lambda = [h^2/2\pi m k_B T]^{1/2} = 1/\sqrt{M} \text{Å}$, where M is expressed in units of proton mass. For hydrogen, this is 1 Å, and for carbon less than 0.3 Å at 300 K (at zero temperature, one must consider the energy of the particle instead of $k_B T$). One needs to compare this length to the size of the atom to decide whether or not to treat it classically. Therefore, for small atoms, the zero-point quantum motion is important, but for larger ones it is often neglected, and we use the classical Newtonian equations to describe their motion. Molecular dynamics is the method to numerically integrate the classical equations of motion and to deduce atomic trajectories over some finite timescale on the order of nano- or microseconds at the most. From the trajectories, one obtains information about the dynamics of the atoms and can visualize a reaction or compute the mechanical and thermodynamic properties of a given system.

4.3.1. Boundary Conditions

To treat solids or bulk materials, to avoid surface effects resulting from a finite size of the system one is simulating, the particles are put in a box with periodic boundary conditions (for free clusters, however, one uses the free-, or no-boundary conditions). That means if they move out of the box from one side, we put them back in from the other side of the box. The way this is usually done to avoid using IF statements is as follows (this example is in FORTRAN language):

$$X(i) = X(i) - \text{box} * \text{anint}(X(i)/\text{box})$$

where the $\text{anint}(x)$ function picks the nearest integer to x . This operation ensures that the particle coordinate remains in $[-\text{box}/2, \text{box}/2]$. In this way, the bulk system is replaced by a periodic crystal with atoms moving exactly in the same manner in all the unit cells. One important issue to consider, therefore, is fully handling the effect of long-range forces coming from particles in the neighboring boxes. Typical examples are systems with ionic bonding such as biological or organic molecules, or with van der Waals interactions. If particles are charged because of the ionic nature of the bonds, then long-range Coulomb interaction is present between the particles. The standard technique to handle this problem using the Ewald sum method, attributed originally to Ewald for summing the $1/R$ potential in three dimensions. However, it can be generalized to $1/R^n$ in any dimension provided that the sums are convergent [84]. Other methods that calculate the multipole moments of the potential for atoms situated far away are also used.

van der Waals interactions, which are usually of the Lennard–Jones type, have algebraic decay, but in practice they are truncated by using a cutoff distance and a shift:

$$\begin{aligned} W(r) &= \phi(r) - \phi(R_{\text{cut}}) & r < R_{\text{cut}} \\ W(r) &= 0 & r > R_{\text{cut}} \end{aligned} \quad (57)$$

4.3.2. Computational Load and Neighbor Lists

The number of operations for calculating the force on a particle is in principle proportional to the number of particles. Therefore, in general, the tasks of computing all the forces at every time step is on the order of N^2 . In practice, because of screening, each particle only sees the force caused by its neighbors, and the task of computing all the forces becomes $O(N)$. Exceptions are long-range forces such as coulomb forces. In practice, many efforts, however complicated, are made to have an order N algorithm, in which case, it is always useful and necessary to have a neighbor list. Making a neighbor list is itself unavoidably $O(N^2)$, but it needs to be done only at the beginning of the run. However, every few time

steps, it needs to be updated if one is simulating a liquid, though this task itself is $O(N)$ because one only needs to check which particles in the “second neighbor shells” entered the neighbor list of the considered atom. For this reason, to create a neighbor list, one needs to use a larger cutoff than that of the forces; typically, one can take as the cutoff distance of the neighbor list $R = R_{cut} + n\Delta t v_{\max}$, where n is the number of steps after which the neighbor list is updated.

4.3.3. Long-Range Forces and Particle–Mesh Methods

For Coulomb or gravitational forces that are long ranged, the task of computing forces is $O(N^2)$. To make this task lighter, one can use what is called a particle–mesh method: The space is discretized and the potential energy is calculated and represented on a grid. If M is the number of grid points, the computation of the potential is $O(NM)$ because it can be first calculated in the Fourier space and then Fourier transformed (FFTd) back in the real space as follows: for each $q = 1, \dots, M$,

$$\phi(q) = \sum_{i=1}^N \frac{4\pi}{q^2} e^{iqR_i} \quad \vec{F}(q) = \sum_{i=1}^N \frac{4\pi}{q^2} i\vec{q} e^{iqR_i} \quad (58)$$

where R_i is the position of the particle i and we have used the Fourier transform of the Coulomb potential. The computation of $\phi(q)$ for all q is an $O(MN)$ task, and its FFT is $O(M \log M)$. Once the forces are known on all grid points, an interpolation can give their value at the actual particle positions. This treatment assumes that the potential resulting from Coulomb interactions, or whatever long-range interaction we have, is a smooth function; in effect, a field.

It is also possible to assign the charges on each grid point and to avoid altogether the sum over particles $i = 1, N$, and then solve a Poisson equation. Charge assignment to a grid may be done by adding up the charges in the volume around that grid point. With free boundary conditions such as a gravitational problem, one can directly solve Poisson equation real space; discretized on a mesh of size M , it is equivalent to solving a linear system problem, which is $O(M^2)$. With periodic boundary conditions, one may use the FFT to solve Poisson. Two FFTs that are $O(M \log M)$ are needed to obtain the potential in the real space. The general algorithm may be summarized as follows:

1. Discretize the space
2. To each grid point assign a charge
3. Solve Poisson to get potential and forces
4. Interpolate forces at the particle positions
5. Integrate the equations of motion for particles
6. The new particle positions are found; GOTO 2

The interpolation scheme could be tricky: If charges are calculated so that they change continuously as particles move from one cell to another, the computational load would increase, but calculations would be more accurate. To compute forces more accurately, it is possible to subtract explicitly the contribution of neighboring cells and to add exactly the contribution of the particles in those cells (additional short-range forces are, of course, calculated exactly). For more details, the reader can refer to the book *Computer Simulations Using Particles* by Hockney and Eastwood [105]. A good reference book, though not very modern, on the subject of molecular dynamics is *Computer Simulation of Liquids* by Allen and Tildesley [87].

4.3.4. Update Algorithms

The particles, assumed to be like a point, are moved according to the Newton’s equation of motion:

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i \quad (59)$$

where the force F_i is coming from any external field plus the interparticle interactions.

This simple form of the equations is for an isolated system that has total energy and total momentum conserving. In statistical mechanical terms, we are performing a microcanonical simulation. We will see later how to include the effect of a thermostat (heat bath) for canonical simulations in the equations of motion.

4.3.5. Verlet Method

To numerically integrate the equations of motion, one needs to discretize the time and, by using a Taylor expansion, write the position at a later time step as a function of the present position, velocity, acceleration, and eventually, higher derivatives.

$$\begin{aligned} r(t + \Delta t) &= r(t) + \Delta t v(t) + \frac{(\Delta t)^2}{2} a(t) + \frac{(\Delta t)^3}{6} b(t) + O(\Delta t^4) \\ v(t + \Delta t) &= v(t) + \Delta t a(t) + \frac{(\Delta t)^2}{2} b(t) + O(\Delta t^3) \end{aligned} \quad (60)$$

Because the equations of motion are second order, we need two initial conditions to start the integration. Usually the positions and velocities are given. It is then possible to also compute the forces acting on atoms at time $t = 0$ and to use the first three terms of the above equation to update the position, and the first two terms to update the velocity. The acceleration, being proportional to the force, can be computed exactly at every time step. The third-order term $b(t)$ or higher can not be computed (other than from finite differences of accelerations), and we have no use for them in the simulation. From the above equations, we see then that the error in v is second order in Δt , and the error in r is third-order in Δt , and if Δt is not chosen carefully, the trajectories might diverge very soon from their true path. Thus, we need algorithms to reduce the error in the time integration while keeping as large a Δt as possible. The simplest way is to expand the position at $t - \Delta t$ as a power series in position, velocity, and acceleration at time t , and to add the two equations to obtain the position at $t + \Delta t$ and subtract them to obtain the velocity at t :

$$\begin{aligned} r(t + \Delta t) &= 2r(t) - r(t - \Delta t) + (\Delta t)^2 a(t) + O(\Delta t^4) \\ v(t) &= [r(t + \Delta t) - r(t - \Delta t)]/2\Delta t + O(\Delta t^2) \end{aligned} \quad (61)$$

This method, originally attributed to Verlet, works better, but it can also become inaccurate because of round-off errors coming from subtraction of two large numbers and addition of a small number of the order of Δt^2 . A better method is the Leap-Frog method from Hockney.

In this method, the velocities are calculated at $n\Delta t + \Delta t/2$; that is, at half steps between the positions, whereas positions and accelerations are computed at multiples of Δt . This could also be a problem at the start of the run, but assuming we are given $v(t - \Delta t/2)$ and $r(t)$, we can first compute $a(t)$, and then the velocity is computed Δt seconds later, and next the position at $t + \Delta t$:

$$\begin{aligned} v\left(t + \frac{\Delta t}{2}\right) &= v\left(t - \frac{\Delta t}{2}\right) + \Delta t a(t) \\ r(t + \Delta t) &= r(t) + \Delta t v\left(t + \frac{\Delta t}{2}\right) \end{aligned} \quad (62)$$

Note that if we have periodic boundary conditions, and at $t + \Delta t$ the particle moves to the other side of the box, the simple Verlet algorithm will not give the correct velocity, but the Leap-Frog method does. Errors in Leap-Frog are on the order of Δt^2 , but the algorithm is more stable than Verlet.

Yet a better and more widely used method is called the velocity-Verlet (VV) algorithm. Here, all positions, velocities, and accelerations are computed at multiples of Δt . First the position is updated according to the usual equations of motion, then the acceleration is

computed at that position, and then the velocity is computed from the average of the two accelerations (particles moving out of the box cause no problem with this algorithm):

$$\begin{aligned} r(t + \Delta t) &= r(t) + \Delta t v(t) + \frac{(\Delta t)^2}{2} a(t) \\ v(t + \Delta t) &= v(t) + \frac{\Delta t}{2} [a(t + \Delta t) + a(t)] \end{aligned} \quad (63)$$

All a , v , r need to be stored in this method, but in terms of stability, this algorithm is superior to the two others because it is based on an implicit method: the velocity at $t + \Delta t$ is computed from the acceleration at t and at $t + \Delta t$. Stability means it allows us to take a larger Δt to update positions. That means that with fewer steps, we can run longer simulations. As it is, after the update of the accelerations, the old accelerations are also needed. That makes $4N$ dimensional arrays to be stored instead of $3N$ (N being the number of degrees of freedom, or three times the number of particles in three dimensions). The way VV is implemented to save memory is to update first the velocities $\Delta t/2$ step later:

$$v(t + \Delta t/2) = v(t) + a(t)\Delta t/2 \quad (64)$$

then positions are updated:

$$r(t + \Delta t) = r(t) + \Delta t v(t + \Delta t/2) \quad (65)$$

and finally, from the new positions, new accelerations $a(t + \Delta t)$ are calculated and velocities are updated according to:

$$v(t + \Delta t) = v(t + \Delta t/2) + a(t + \Delta t)\Delta t/2 \quad (66)$$

If memory is no objection, then this two-step update of velocities is unnecessary, and one can just use Eq. (63).

4.3.6. The Predictor-Corrector Method [106]

To obtain more accurate trajectories with the same time step, or to be able to use larger time steps, one can use a higher order method called predictor-corrector (PC). Assuming that the trajectories are smooth functions of time, in this method, the information on the positions two or three previous time steps ago is also used. First the positions, velocities, and accelerations are predicted at $t + \Delta t$ from the finite difference formula up to order n in principle. Below we show the third-order formula:

$$\begin{aligned} r^P(t + \Delta t) &= r(t) + \Delta t v(t) + \frac{(\Delta t)^2}{2} a(t) + \frac{(\Delta t)^3}{6} b(t) + O(\Delta t^4) \\ v^P(t + \Delta t) &= v(t) + \Delta t a(t) + \frac{(\Delta t)^2}{2} b(t) + O(\Delta t^3) \\ a^P(t + \Delta t) &= a(t) + \Delta t b(t) + O(\Delta t^2) \end{aligned} \quad (67)$$

After the position is predicted at $t + \Delta t$, however, we can compute the actual correct acceleration at that time a^C . The error we have made in predicting the acceleration from the above formula is $\Delta a = a^C - a^P$. Using a perturbative idea, one can argue that the error on the predicted positions and velocities is also proportional to Δa , and hence, we can correct for that error:

$$\begin{aligned} r^C(t + \Delta t) &= r^P(t + \Delta t) + c_0 \Delta a \\ v^C(t + \Delta t) &= v^P(t + \Delta t) + c_1 \Delta a \\ a^C(t + \Delta t) &= a^P(t + \Delta t) + c_2 \Delta a \end{aligned} \quad (68)$$

where $c_2 = 1$, and the other coefficients may be determined, as Gear has done it [106], to achieve highest accuracy. In a third-order method, $c_0 = 0$, $c_1 = 1$, and $c_2 = 1$, and in a fourth-order method, $c_0 = 1/6$, $c_1 = 5/6$, $c_2 = 1$, and $c_3 = 1/3$. Higher-order methods can also be used, but they require more memory and can consume more CPU power without considerably improving the accuracy [87].

4.3.7. How Should One Determine Δt ?

The important parameter in a MD simulation is Δt . In a microcanonical simulation, the total energy of the system must be conserved. If Δt is too large, steps might become too large, and the particle may enter the classically forbidden region in which the potential energy is an increasing function. This will occur when two particles collide or when a particle hits the “wall” imposed by the external potential. Entering the classically forbidden region means that the new potential energy has become higher than the maximal value it is allowed to have. In this case, the total energy has increased, and this phenomenon keeps occurring for large step sizes until the total energy diverges. So, depending on the available total energy, Δt should be chosen small enough so that the total energy remains constant at all times, but not too small that it takes a large number of steps to perform the simulation. The optimal value of Δt is usually found by trial and error. 1 femto second is a good trial guess, but the optimal value really depends on the initial energy and the kind of potential one is dealing with.

4.3.8. Systems in Contact with a Heat Bath: Thermostats

In canonical simulations, the total energy is not a conserved quantity, rather, the temperature T is a constant, and particles exchange energy with a heat bath external to the system, so that the kinetic energy per particle remains $3k_B T/2$ on average. Modeling this energy exchange mechanism can be done in many ways:

4.3.8.1. Velocity Scaling Because the average kinetic energy is $3k_B T/2$ per particle, one can simply force the total kinetic energy to be $3Nk_B T/2$ at all times, and therefore multiply all velocities by a scale factor to make the kinetic energy equal to this number. This is the simplest way that contact with a heat bath at T can be modeled, but one may argue that it is not very physical because the kinetic energy is a fluctuating quantity, and only its average is $3Nk_B T/2$.

If one is simulating rigid molecule with more than three degrees of freedom per particle that is formed of two or several atoms, the equipartition theorem must be applied carefully. One has $3k_B T/2$ for the translation of the center of mass and $3k_B T/2$ for the rotation around the center of mass ($k_B T$ only for diatomic molecules). The kinetic energy of $3k_B T$ (or $5k_B T/2$ for diatomic molecules) must be shared by all the atoms in the molecule. Even if the assumption of rigidity is absent, velocities must be scaled such that the center of mass has kinetic energy of $3k_B T/2 + 3k_B T/2$ for rotations and translations, and $3(N - 2)k_B T/2$ for the vibrational (internal) degrees of freedom. It means that velocities of the particles need to be decomposed to center of mass + relative velocity. Then each component should be separately scaled and then added back to obtain the absolute speed of the particles.

4.3.8.2. Nose–Hoover Thermostat [107, 108] There is a single degree of freedom ξ for the Nose–Hoover thermostat, which is interacting with the particles. The particle Lagrangian is modified to also include the thermostat. One can then deduce the coupled equations of motions for the particles and thermostat. This time the Helmholtz free energy of the whole system of particles + thermostat is a conserved quantity. The equations of motion in this scheme are:

$$\begin{aligned} \frac{d^2 \vec{r}_i}{dt^2} &= \frac{\vec{F}_i}{m_i} - \dot{\xi} \frac{d\vec{r}_i}{dt} \\ \frac{d\dot{\xi}}{dt} &= \frac{1}{\tau^2} \left[\sum_{i=1, N} (m_i v_i^2 / f k_B T) - 1 \right] \end{aligned} \quad (69)$$

where f is the total number of degrees of freedom, and τ is a free parameter modeling the strength of the coupling of the system to the thermostat. A small value of τ yields a large derivative for $\dot{\xi}$ and hence a large additional friction force in the motion of the particles. Small τ then means strong coupling to thermostat. The term τ has the dimensions of a time and can thus be called a “relaxation time.” It can be taken to be about 1 ps, but again,

its value should be adjusted depending on how and what damping process one desires to model. The Hamiltonian from which these equations of motion can be derived is

$$H = \sum_i \frac{1}{2} m_i v_i^2 + \sum_{i<j} v(r_{ij}) + \left(\frac{1}{2} \tau^2 \dot{\xi}^2 + \xi \right) f k_B T$$

One can adjust Δt so that the above Hamiltonian is a conserved quantity. Note that if the total kinetic energy per particle is just $3k_B T/2$, then $\dot{\xi}$ is a constant that we can set initially to be 0. If the term in parentheses is positive, meaning that there is too much kinetic energy, then $\dot{\xi}$ increases to become positive and there is a friction force that slows down the particles, and if that term is negative, $\dot{\xi}$ decreases to a negative value and pushes the particles further along their momentum so as to increase their kinetic energy. One can see that this coupling tries to keep the kinetic energy of the particles around $3Nk_B T/2$.

This thermostat can be used for a NVT canonical simulation. If one wishes to perform an NPT simulation with a variable box volume $V(t)$ but constant pressure P_0 , it is also possible to add an additional ‘‘barostat’’ degree of freedom coupled to the system. The resulting equations are

$$\begin{aligned} \frac{dr}{dt} &= v(t) + \eta(t) [r(t) - R_0] \\ \frac{dv}{dt} &= \frac{F}{m} - [\eta(t) + \chi(t)] v(t) \\ \frac{d\chi}{dt} &= \frac{1}{\tau_p^2} \left[\sum_{i=1, N} (m_i v_i^2 / f k_B T - 1) \right] \\ \frac{d\eta}{dt} &= \frac{(P - P_0) V(t)}{\tau_p^2 N k_B T} \\ \frac{dV}{dt} &= 3\eta(t) V(t) \end{aligned} \quad (70)$$

where τ_p is the pressure relaxation time, P_0 is the target pressure, R_0 the center of mass coordinate, and V the system volume. In this case, the Gibbs free energy G is conserved, where $G_{NPT} = U_{NVT} + P_0 V(t) + f k_B T \tau_p^2 \eta^2(t)/2$, as proposed by Melchionna and coworkers [109].

4.3.8.3. Langevin Thermostat Yet a third method of exchanging energy with the particles is using the Langevin thermostat. Again the particles trajectories are modified by two additional forces: a friction force and a random force of time average 0, with white noise of magnitude $2mk_B T/\tau \Delta t$ per degree of freedom:

$$m \frac{d^2 x}{dt^2} = F_x - \frac{m}{\tau} \frac{dx}{dt} + f \quad \langle f \rangle = 0 \quad \langle f(t) f(t') \rangle = \delta(t - t') 2mk_B T/\tau \quad (71)$$

Note that when implementing the algorithm, the Dirac delta function is replaced in the magnitude calculation by $1/\Delta t$. So, in practice, one generates random numbers with Gaussian distribution of mean 0 and variance $2mk_B T/\tau \Delta t$ for each direction of the force. One simple way of generating random numbers of Gaussian distribution is the following [87]: generate two random numbers a and b in $[0, 1]$; calculate $A = M + \sigma \sqrt{-2 \ln a} \cos 2\pi b$ and $B = M + \sigma \sqrt{-2 \ln a} \sin 2\pi b$. The numbers A and B have a Gaussian distribution of mean M and variance σ . In this thermostat, τ is the relaxation time and represents the mean collision time between the real particles of the system and some fictitious particles that collide at random with them. This thermostat could be interesting because it physically models some stochastic collisions with virtual particles. It can be shown that particles following this dynamic will have an average kinetic energy of $3k_B T/2$ per particle in three dimensions. The collision rate $1/\tau$ models the strength of the interaction with the thermostat. Large τ means weak friction and weak random force; that is, weak interaction with the bath as particles do not collide so often with the bath particles. A reasonable value for τ would be about 1000 time steps Δt ; it must be adjusted based on the kind of thermal contact one wants to model. Again, some trial and error is necessary to adjust this parameter. In Figs. 12 and 13, we can

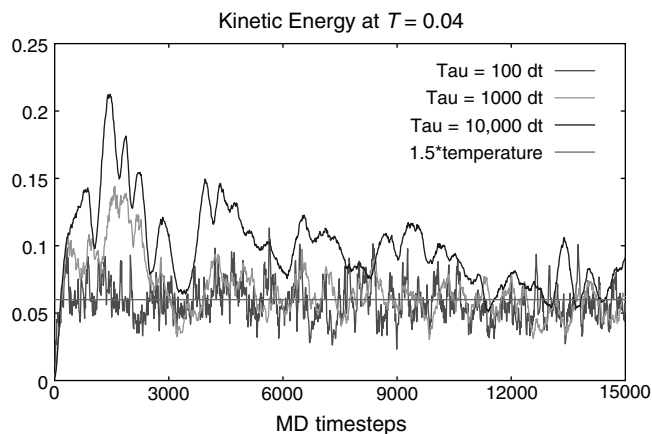


Figure 12. Time evolution of the kinetic energy of a 13-atom argon cluster for three different relaxation times $\tau = 100 \Delta t$, $1000 \Delta t$, and $10,000 \Delta t$. It can clearly be seen that the blue curve corresponding to $10,000 \Delta t$ relaxes to the value 0.06 after about 10,000 or more steps, and the green curve needs about 3000 steps to relax.

see the evolution of the kinetic energy to $3k_B T/2$ as a function of time for several values of the relaxation time. The initial configuration is a deformed 13-atom cluster of icosahedral shape. Three values were chosen for this illustration.

The Langevin thermostat with a small relaxation time would also be a useful tool for performing structural relaxations. Indeed, the particles move along the steepest descent direction (i.e., along the forces) but are also given some random momentum to escape eventual potential barriers. A small value of τ would be better in relaxation calculations, and it can eventually be increased toward the end of the run. It is also possible to lower the temperature towards the end of the run, just as is done in simulated annealing (see later sections regarding optimization techniques).

4.3.9. Response Functions

From the trajectory of the particles, it is possible to obtain the autocorrelation function of the positions, velocities, or energy flow (particle energy times its velocity). They respectively yield the structure factor, which can be compared directly to experiments, the diffusion coefficient in the case of fluids (and the elastic constants or phonon frequencies in the case of solids), and the thermal conductivity. This is one powerful application of the MD method. It is not just about making an animation and seeing how things move, one can also compute and predict physical properties quantitatively. For a dynamical observable $\vec{S}_i(t)$, associated

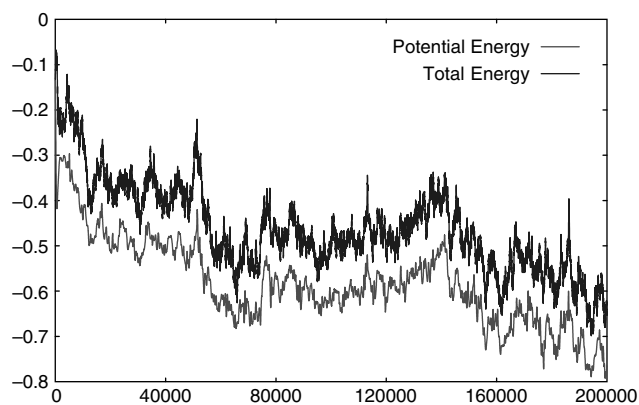


Figure 13. Evolution of the total and potential energy of a 13-atom argon cluster at $T = 0.08$. After about 200,000 steps.

with every particle i , the autocorrelation function and its FFT are defined as

$$\begin{aligned}
 F(\tau) &= \frac{1}{N} \sum_{i=1}^N \left\langle (\vec{S}_i(\tau) - \langle \vec{S}_i \rangle) \cdot (\vec{S}_i(0) - \langle \vec{S}_i \rangle) \right\rangle \\
 &= \frac{1}{T} \int_0^T \frac{1}{N} \sum_i (\vec{S}_i(t + \tau) - \langle \vec{S}_i \rangle) \cdot (\vec{S}_i(t) - \langle \vec{S}_i \rangle) dt \\
 \tilde{F}(\omega) &= \int F(\tau) e^{i\omega\tau} d\tau \quad \left(= \sum_{\lambda} \delta(\omega - \omega_{\lambda}) |c_{\lambda}|^2 \right)
 \end{aligned} \tag{72}$$

where the ensemble average in the second equation is replaced by the time average which needs to be performed on a long timescale T . The third line is only valid in the case in which the quantity $\vec{S}_i(t)$ oscillates about its equilibrium or average value. One will then observe peaks in the FTT of the autocorrelation function corresponding to the eigenfrequencies of the system. Examples are phonon frequencies if $\vec{S}_i(t)$ is the displacement vector of atom i , or magnon frequencies if $\vec{S}_i(t)$ is the magnetization vector for each atom. The function F can also be written as a function of the FTT of the function S for nonzero frequencies: $\tilde{F}(\omega) = \tilde{S}(\omega)\tilde{S}(-\omega) = |\tilde{S}(\omega)|^2$. It can, therefore, more simply be obtained from the square of the FTT of S . The advantage is that the result will always be positive, and only a single FFT is required to find the answer.

4.3.10. Search for the Ground State of a System: Energy Minimization

The other important application of the molecular dynamics method is in calculating the ground-state structure of molecules and clusters. Nowadays, many such calculations are performed on very large organic molecules such as proteins and DNA to understand their folding and unfolding mechanisms and some of the reactions that take place on small timescales. In fact, some of these reactions may occur in milliseconds up to seconds, and with today's computers it is impossible to run such long timescales (remember that a typical molecular dynamics time step is on the order of a femtosecond). The structure of the grains in granular materials, or clusters adsorbed on surfaces, or simply in the gas phase is also of importance for obvious technological reasons.

4.3.10.1. Local Optimization Methods The simplest of the optimization methods, which may date to two centuries back or even more, as they carry Newton's name, deal with local optimization. A particle is inside a not-necessarily-harmonic well in an N -dimensional space (N being the number of degrees of freedom typically equal to three times the number of particles), and we would like to go to the minimum of that well by using some dynamics, not necessarily Newtonian, as fast as possible. The landscape is defined by the "cost function" $E(x)$, which, in our case of interest, is the potential energy of the system.

Steepest Descent The simplest way that comes to mind to achieve this minimum is just to move along the force applied on the particle. This is called the steepest descent (SD) method because the force is minus the gradient of the potential energy that we want to minimize. The simple algorithm is as follows:

$$x(t + \Delta t) = x(t) + \lambda f(x(t)) \tag{73}$$

where λ is a positive constant coefficient, and $f(x) = -\nabla E(x)$ is the force acting on the particle at x . Here one has to think of x and f as N -dimensional arrays, and the time is really discretized and represents the minimization step. It is better to write this as mathematicians do: $x(i + 1) = x(i) + \lambda f(x(i))$ (i is the iteration number). This method surely works because at each step the potential energy is lowered, but its convergence is often too slow. One may also adopt damped Newtonian dynamics, which also fall in the category of the SD and is not any faster than SD. These methods could be very slow, as the well is very anisotropic. The particle will keep oscillating around the valley that leads down to the minimum, and it can take many steps before reaching it.

Conjugate Gradients A more intelligent method is called the conjugate gradients (CG). This method will find the minimum of the anisotropic valley much faster than SD. In principle, it finds the minimum of a quadratic function in N dimensions in exactly N line minimization steps. Provided that one is near the minimum, the convergence to the minimum is almost as fast because the function is “almost quadratic.”

For the sake of completeness and brevity we will just mention the algorithm to implement and will bypass its explanation. Interested readers can consult specialized books on the subject [100].

1. Perform a line minimization along the steepest descent direction; find λ such that $x(1) = x(0) + \lambda f(0)$ minimizes the energy [Assuming $c(0) = f(0)$].
2. Calculate the next SD direction $f(i + 1) = -\nabla E[x(i + 1)]$
3. Calculate the conjugate direction $c(i + 1) = f(i + 1) + b_i c(i)$ where $b_i = \frac{\|f(i+1)\|^2}{\|f(i)\|^2}$
4. Update x by doing a line minimization: $x(i + 1) = x(i) + \lambda_i c(i)$ (this defines λ_i)
5. Go to line 2 until the calculation converges.

Quasi-Newton Method Newton’s method is an algorithm to find the zero of a function f in one dimension. The iterative process is as follows:

$$x(i + 1) = x(i) - f(x(i))/f'(x(i)) \quad (74)$$

In our minimization problem, we are also looking for the zero of the force vector. That is where the minimum is. This method can easily be generalized to N dimensions. The only difference is that the derivative of the vector f , the Hessian, becomes an $N \times N$ matrix, and furthermore, it needs to be inverted and then multiplied by the force vector. This computation is very difficult and computationally demanding. The computation of the matrix elements of the Hessian is very time consuming and complicated, and its inversion is also a very CPU intensive task. One resorts, therefore, to quasi-Newton methods that try to approximate the Hessian as the iterations proceed. These methods only need the forces (or gradients of the potential energy) as input, start with the steepest descent that approximates the Hessian by the identity matrix times a constant, and at each step, using the information on all previous forces and positions, improve their guess of the inverse Hessian matrix. This is also known as Broyden’s method because he was the first to propose such a method.

Again, we will just mention the algorithm to implement [110, 111]:

$$\text{Step 1: } x(2) = x(1) - \lambda f(1)$$

$$\text{Step } i: x(i + 1) = x(i) - \lambda f(i) - \sum_{j=2}^i u(j) \langle v(j) | f(i) \rangle$$

where

$$u(j) = -\lambda (f(j) - f(j - 1)) + x(j) - x(j - 1) - \sum_{l=2}^{j-1} u(l) \langle v(l) | f(j) - f(j - 1) \rangle$$

and

$$v(j) = \frac{f(j) - f(j - 1)}{\langle f(j) - f(j - 1) | f(j) - f(j - 1) \rangle} \quad (75)$$

all these quantities except for λ , which is just the magnitude of the first step along the SD direction, are to be interpreted as vectors, and the bracket $\langle f | g \rangle$ notation is just the dot product of these vectors. The index i, j , or l in parentheses is the iteration number and not the component of these vectors.

This method has been proven to be very useful in relaxing systems provided the starting point is close enough to the local minimum. There is some freedom with the step size λ , which needs to be adjusted by trial and error. In addition to optimizations, Broyden’s method can of course be used to solve a set of nonlinear equations $\vec{f}(\vec{x}) = 0$.

4.3.10.2. Global Optimization Methods A much more challenging task is to find the global minimum of a multivalley energy landscape. Below, we will mention briefly two of the most widely used methods in the field of structure optimization of clusters.

Simulated Annealing Simulated annealing (SA) is how metallurgists experimentally reach the ground state of their sample: by annealing it. The sample is put in an oven at high temperature for a long time so that energy barriers can be overcome and the system will hopefully find the lowest-energy structure. It can be shown that if the cooling rate of the sample is low enough, the ground state is always reached at the end. The implementation of this method was first proposed by Kirkpatrick et al. [112].

In simulations, again, it does not matter what dynamics is followed to cross the barriers—one could use Newtonian dynamics along with the Langevin thermostat for example—but this is not the only choice. One can also make moves at random and perform a Monte Carlo walk according to the Metropolis algorithm, for instance. The cooling schedule is quite important in this problem. Either the user finds it by trial and error, or one can use an adaptive algorithm as proposed by Andersen and Gordon [113]. Simple cooling schedules can be linear, stepwise, or exponential. In the adaptive simulated annealing (ASA) method, the temperature evolves according to:

$$dT/dt = -vT/\varepsilon\sqrt{C} \quad (76)$$

In this equation T is the temperature, t is time, v is the velocity of annealing (just a constant coefficient controlling the overall cooling rate), C is the heat capacity, and ε is the relaxation time coming from the second largest eigenvalue of the transition matrix (see the corresponding chapter on Monte Carlo methods) $\varepsilon = -\text{Log}(\omega_2) \approx 1 - \omega_2$. To construct the transition matrix, one can use a simple model and classify states according to their energy. First some energy range is defined within which the total energy of the system is fluctuating. This energy interval is then divided into M equal parts (M could be on the order of 100 or even more). Each interval corresponds to one state, meaning that if the energy of the system is between E_i and E_{i+1} , we say the system is in the state i ($i = 1, M$).

Now we let the system evolve at the temperature T according to some dynamics, be they Langevin dynamics or a Markov process defined by the Metropolis algorithm, or even Newtonian dynamics. During some time steps P , say of the order of $P = 1000$ (this really depends on the size of the system and the complexity of the energy landscape; larger or more complex systems require a larger number of steps), we record the energy of the system and increment the matrix element (i, j) of the transition matrix by one unit every time the system goes from state i to state j . If the energy remains in the same interval j , the (j, j) element must of course be incremented by 1. At the end, the elements of each column are rescaled so that their sum is equal to 1. This is the sum rule we mentioned in the chapter on Monte Carlo. One now needs to diagonalize this matrix. Its highest eigenvalue must be 1, and the corresponding eigenvector contains the equilibrium probabilities for the system to be in each of the M states. From the second eigenvalue, one obtains the relaxation time ε .

The heat capacity may be found from the fluctuations in the energy during these P steps run: $C = \langle E^2 \rangle - \langle E \rangle^2 / k_B T^2$.

Now that all the elements of the cooling equation are known, one can update the temperature according to $T = T - vT/\varepsilon\sqrt{C}$, rerun the dynamics for another P steps, collect statistics to calculate C and the new transition matrix, and then recalculate the temperature and so on, until it reaches a desired (low) value.

Genetic Algorithms The idea in all global optimization problems is to sample the phase space as quickly and as completely as possible. One can do it by molecular dynamics, but it will obviously take a huge number of steps, and even then many barriers might never be crossed. It is also possible to use a Markov process driven by Metropolis (sampling the equilibrium configurations at temperature T). Yet another way is to use genetic algorithms (GA) [114]. They are based on concepts coming from genetics and the survival of the fittest idea; being fit meaning having a lower energy. Genes represent the identity of the individuals. In our case, individuals are clusters, and their genes are their coordinates. In GA, individuals are mated among themselves and their genes may also mutate, and this way, one obtains a new generation of individuals selected based on their fitness (potential energy).

The mating, also called the crossover operation, consists of taking two individuals at random and combining half the genes of one with half the genes of the other to obtain two new

individuals. Note that there are many ways to perform this operation, and one may even produce, from a single pair, a next generation of many more individuals that are formed from taking many combinations of the genes of their parents. A priori, if we start with N individuals in generation P , one may end up after mating, with say, $20N$ individuals in the next generation. One can eventually perform mutations on them. Mutations consist in randomly modifying part of their genes. In our context, mutation means displacing one or more atoms in a cluster at random.

From the $20N$ obtained individuals, one can first perform a quick structure optimization and then sort their energies and pick the N lowest-energy individuals as survivors for the next generation. One must, however, be careful to choose candidates of different geometrical forms, so as to keep diversity. It would be completely useless to keep N clusters of the same shape for the next generation, even though their energies were the lowest. Now, one can iterate this process and perform mating, mutation, relaxation, and final selection based on shape diversity to obtain a new generation and so on. One is hopeful that after few hundred or thousand generations (again this depends on the cluster size and complexity of the problem), the fittest individuals have been found. There is, however, no guarantee of the convergence of the process.

In SA or GA, one is just sure to have found a good minimum if not the absolute one. The choice of genes, mating and, mutations in GA and the dynamics in SA depend on the problem and can be tailored to the specific problem at hand, just like sampling in Monte Carlo is not unique to Metropolis.

5. CONCLUSIONS

The single most important barrier to the development of nanotechnology is the lack of an exact predictable description of the nature of matter in nanoscale systems. In this review we have introduced the state-of-the-art theories, intermolecular forces and potentials, and computer simulation methods needed for statistical mechanical modeling and prediction of the behavior of matter in nanoscale. With a thorough knowledge on this subject, we will be able to predict the possible arrangement of atoms and molecules in all the ways permitted by physical law, which is the main aim of nanotechnology. To be able to build a wide range of molecular structures and develop the technology and know-how for self-assembly of molecules and self-replication, their statistical mechanical prediction is necessary. An accurate theoretical prediction capability will also provide us with deep, penetrating insight for building precise molecular tools, molecular machine components, and products. This will influence the research and development climate and redirect research funding and targets.

Through their intrinsic nature, the theories and simulation methods outlined above are adapted to treating systems formed of a finite number of particles. The other advantage of these methods is that, compared to experiments, they are relatively inexpensive and very safe in determining properties of a given nanosystem. Theoretical calculations and simulations can even be performed on systems under very unrealistic conditions of very low or very high temperatures and pressures. These would normally be either very expensive or unsafe, if not impossible, with an experimental setup. Nowadays, the stability of proposed nanosystems can be estimated rather reliably before they are made experimentally.

GLOSSARY

AFM Atomic force microscope. A tip-based picoscanning equipment of nano- and micro-size surfaces in three dimensions.

FLAPW Full-Potential Linearized Augmented Plane Wave.

Intermolecular forces Forces resulting from attraction and repulsion between molecules.

Intermolecular potential Potential energy resulting from attraction and repulsion between molecules.

LAPW Linearized Augmented Plane Wave.

LSDA Local Spin Density Approximation.

Markov process (MP) The random walk toward equilibrium from an arbitrary initial state, with no memory from previous steps.

Metropolis algorithm In this algorithm, one can start from an arbitrary point in $[a, b]$, say $x_0 = (a + b)/2$ and add a random number to it: $x_1 = x_0 + d(2r - 1)$, where r is the random number in $[0, 1]$ and d is the magnitude of the step. If $P(x_1)/P(x_0) > 1$, then the move is accepted; otherwise the ratio is compared to a random number in $[0, 1]$.

Molecular dynamics simulation A computer simulation based on the equations of motion and energy of a finite number of particles and their resulting statistical averaging.

Monte Carlo simulation A computer simulation based on random number generation of the variable to estimate the function statistically.

Nanoscale One billionth of meter scale.

Nanostructure Geometrical structures in nanoscale.

Nanosystem Controlled volume or controlled mass systems defined in nanoscale.

Small system A system defined in nanoscale (nanosystem).

ACKNOWLEDGMENTS

We thank Dr. Guoping Zhang for his input to Section 3.2 “Step 2: Theoretical Modeling” of Intermolecular Forces and Potentials.

REFERENCES

1. K. E. Drexler, “Nanosystems: Molecular Machinery, Manufacturing and Computation.” Wiley, New York, 1992.
2. M. C. Roco, S. Williams, and P. Alivisatos, Eds., “Nanotechnology Research Directions: IWGN Workshop Report—Vision for Nanotechnology R&D in the Next Decade.” WTEC, Loyola College in Maryland, Baltimore, MD, 1999.
3. M. J. Madou, “Fundamentals of Microfabrication: The Science of Miniaturization,” 2nd ed. CRC Press, Boca Raton, FL, 2002.
4. G. A. Mansoori, *Nanotechnol. U.N. Tech. Monitor* 53, (2002).
5. W. Heisenberg, “Physics and Philosophy.” Harper and Row, New York, 1958.
6. J. M. Haile and G. A. Mansoori, Eds., “Molecular-Based Study of Fluids,” Advanced Chemistry Series 204, American Chemical Society, Washington, DC, 1983.
7. E. Matteoli and G. A. Mansoori, Eds., “Fluctuation Theory of Mixtures.” Taylor & Francis, 1990.
8. L. Hill, “Thermodynamics of Small Small Systems” (W. A. Benjamin, Ed.), New York, 1963; *ibid*, Vol. 2, 1964.
9. T. L. Hill, *Nano Lett.* 1, 273 (2001).
10. G. A. Mansoori, “Proceedings of the first Conference on Nanotechnology—The Next Industrial Revolution,” Vol. 2, p. 345. 2002.
11. R. P. Feynman, *Eng. Sci. Mag.* 23, 22 (1960).
12. L. Onsager and S. Machlup, *Phys. Verhandlungen* 3, 84 (1952).
13. I. Prigogine, “Introduction to Thermodynamics of Irreversible Processes.” Wiley, New York, 1967.
14. D. H. E. Gross, “Microcanonical Thermodynamics,” Vol. 65. World Scientific Lecture Notes in Physics, 2001.
15. G. M. Wang et al., *Phys. Rev. Lett.* 89, 050601 (2002).
16. J. Wilks, “The Third Law of Thermodynamics.” Oxford University Press, London, 1961.
17. G. R. Vakili-Nezhaad and G. A. Mansoori, *J. Comp. Nanosci. Nanotechnol.* (2004) (unpublished).
18. O. Penrose, “Foundations of Statistical Mechanics. A Deductive Treatment, International Series of Monographs in Natural Philosophy,” Vol. 22. Pergamon Press, Oxford, 1970.
19. C. G. Chakrabarti and D. E. Kijal, *J. Math. Math. Sci.* 23, 243 (2000).
20. J. Kestin, “A Course in Statistical Thermodynamics.” Academic Press, 1971.
21. D. A. McQuarrie, “Statistical Thermodynamics.” University Science Books, 1985.
22. D. A. McQuarrie, “Statistical Mechanics.” University Science Books, 2000.
23. R. S. Berry, J. Ross, S. A. Rice, and S. R. Berry, “Matter in Equilibrium: Statistical Mechanics and Thermodynamics,” 2nd ed. Oxford University Press, 2001.
24. C. Tsallis, *J. Stat. Phys.* 52, 479 (1988).
25. A. K. Aringazin and M. I. Mazhitov, *Physica A* 325, 409 (2003).
26. Q. A. Wang, *Phys. Lett. A* 300, 169 (2002); Q. A. Wang and A. Lé Méhauté, *J. Math. Phys.* 43, 5079 (2002); Q. A. Wang, *Euro. Phys. J. B* 26, 357 (2002); Q. A. Wang, L. Nivanen, A. Lé Méhauté, and M. Pezeril, *J. Phys. A* 35, 7003 (2002); Q. A. Wang and A. Le Méhauté, *Chaos, Solitons Fractals* 15, 537 (2003).
27. C. Tsallis, *Chaos, Solitons Fractals* 6, 539 (1995).
28. R. Laing, *J. Theoret. Biol.* 54, 63 (1975).
29. J. Rebek, Jr., *Sci. Am.* 271, 48 (1994).

30. Y. Gogotsi, J. A. Libera, A. G. Yazicioglu, and C. M. Megaridis, *Appl. Phys. Lett.* 79, 1021 (2001); C. M. Megaridis, A. G. Yazicioglu, and J. A. Libera, and Y. Gogotsi, *Phys. Fluids* 14, L5 (2002).
31. W. Gans and J. C. A. Boeyens, Eds., "Intermolecular Interactions." Plenum, 1998.
32. P. L. Huyskens, W. A. P. Luck, and T. Zeegers-Huyskens, Eds., "Intermolecular Forces: An Introduction to Modern Methods and Results." Springer, New York, 1991.
33. A. Stone, "The Theory of Intermolecular Forces," International Series of Monographs on Chemistry, No. 32. Oxford University Press, Oxford, 1997.
34. K. Terakura and H. Akai, "Interatomic Potential and Structural Stability," Proceedings of the 15th Taniguchi Symposium. Springer, Kashikojima, 1993.
35. M. Edalat, F. Pang, S. S. Lan, and G. A. Mansoori, *Int. J. Thermophys.* 1, 177 (1980).
36. H. Rafii-Tabar and G. A. Mansoori, "Encyclopedia Nanoscience and Nanotechnology." American Scientific Publishers, 2004 (unpublished).
37. A. R. Massih and G. A. Mansoori, *Fluid Phase Equilibria* 10, 57 (1983).
38. E. H. Benmekki and G. A. Mansoori, *Fluid Phase Equilibria* 41, 43 (1988).
39. G. A. Mansoori, *Fluid Phase Equilibria* 4, 61 (1980).
40. F. Firouzi, H. Modarress, and G. A. Mansoori, *Eur. Polymer J.* 34, 1489 (1998).
41. E. Z. Hamad and G. A. Mansoori, *J. Phys. Chem.* 94, 3148 (1990).
42. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, Jr., R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, Petersson, G. A. Ayala, P. Y. Q. Cui, K. Morokuma, P. Salvador, J. J. Dannenberg, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, A. G. Baboul, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, J. L. Andres, C. Gonzalez, M. Head-Gordon, E. S. Replogle, and J. A. Pople, Gaussian, Inc., Pittsburgh, PA, 2001.
43. H.-J. Werner and P. J. Knowles, MOLPRO.
44. A. K. Sum, S. I. Sandler, R. Bukowski, and K. Szalewicz, *J. Chem. Phys.* 116, 7627 (2002).
45. W. R. Busing, Ed., "Intermolecular Forces and Packing in Crystals," Transactions of the American Crystallographic Association Series: Vol. 6. Polycrystal Book Service, Huntsville, AL, 1970.
46. J. G. Kirkwood, "Dielectrics Intermolecular Forces Optical Rotation." Gordon & Breach Science, 1965.
47. H. Margenau, "Theory of Intermolecular Forces," 2nd ed. Pergamon Press, 1971.
48. J. Israelachvili, "Intermolecular and Surface Forces: With Applications to Colloidal and Biological Systems," 2nd ed. Academic Press, 1992.
49. G. Binnig and H. Rohrer, *Helv. Phys. Acta* 55, 726 (1982).
50. G. Binnig, C. F. Quate, and C. Gerber, *Phys. Rev. Lett.* 56, 933 (1986).
51. D. Sarid, "Scanning Force Microscopy with Applications to Electric Magnetic and Atomic Forces." Oxford University Press, Oxford, 1994.
52. G. Meyer and N. M. *Am. Appl. Phys. Lett.* 53, 1045 (1988).
53. G. P. Zhang, X. F. Zong, and T. F. George, *J. Chem. Phys.* 110, 9765 (1999).
54. G. P. Zhang, *Phys. Rev. B* 60, 11482 (1999).
55. G. P. Zhang and X. F. Zong, *Chem. Phys. Lett.* 308, 289 (1999).
56. G. P. Zhang, T. F. George, and L. N. Pandey, *J. Chem. Phys.* 109, 2562 (1998).
57. Z. Gamba, *J. Chem. Phys.* 97, 553 (1992).
58. W. A. Ducker, T. J. Senden, and R. M. Pashley, *Nature* 353, 239 (1991).
59. G. U. Lee, L. A. Chrisey, and R. J. Colton, *Science* 266, 771 (1994).
60. W. A. Ducker, T. J. Senden, and R. M. Pashley, *Langmuir* 2, 1831 (1992).
61. H. Sekiguchi, H. Arakawa, T. Okajima, and A. Ikai, *Appl. Surface Sci.* 188, 28, 489 (2002).
62. A. T. Andrews, "Electrophoresis." Oxford University Press, 1992; J. Hunter, Ed., "Zeta Potential in Colloid Science." Academic Press, New York, 1981.
63. G. A. Mansoori, L. Assoufid, T. F. George, and G. Zhang, "Proceedings of Conference on Nanodevices and Systems, Nanotech 2003," pp. 23–27. San Francisco, CA, 2003.
64. P. Honenberg and W. Kohn, *Phys. Rev.* 136, 864B (1964).
65. A. Szabo and N. S. Ostlund, "Modern Quantum Chemistry." Dover, Mineola, NY, 1996.
66. P. Blaha, K. Schwarz, P. Sorantin, and S. B. Trickey, *Comput. Phys. Commun.* 59, 399 (1990).
67. G. P. Zhang et al., *Phys. Rev. B* 65, 165107 (2002); G. P. Zhang et al., *Phys. Rev. Lett.* 88, 077401 (2002); 88, 189902 (Erratum) (2002).
68. J. L. Hutter and J. Bechhoefer, *J. Appl. Phys.* 73, 4123 (1993).
69. R. Ru, D. Singh, and H. Krakauer, *Phys. Rev. B* 43, 6411 (1991).
70. B. Kohler, S. Wilke, M. Scheffler, R. Kouba, and C. Ambrosch-Draxl, *Comp. Phys. Commun.* 94, 31 (1996).
71. J. M. Soler and A. R. Williams, *Phys. Rev. B* 40, 1560 (1989).
72. H. G. Krimmel, J. Ehmman, C. Elsässer, M. Fähnle, and J. M. Soler, *Phys. Rev. B* 50, 8846 (1994).
73. M. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, *Rev. Mod. Phys.* 73, 33 (2001); L. Mitas, *Comp. Phys. Commun.* 96, 107 (1996).
74. D. M. Ceperley, *AIP Conf. Proc.* 690, 378 (2003); D. M. Ceperley, *Rev. Mod. Phys.* 67, 279 (1995).
75. P. Fulde, "Electron Correlations in Molecules and Solids," 3rd ed., Vol. 100. Springer Series in Solid State, 1995.

76. K. Binder and D. W. Heerman, "Monte Carlo Methods in Statistical Physics," 4th ed. Springer, 2002; K. Binder, "The Monte Carlo Method in Condensed Matter Physics," Topics in Applied Physics, Vol. 71. Springer, 1992; K. Binder, "Applications of the Monte Carlo Method in Statistical Physics," Topics in Current Physics, Vol. 36. Springer, 1984.
77. K. D. Sen, "Celebration of the Contribution of R. G. Parr," Rev. Mod. Chem. World Scientific, 2002.
78. S. R. Langhoff, "Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy." Kluwer Academic, 1995.
79. L. Hedin and S. Lundqvist, *Solid State Phys.* 23, 1 (1969).
80. T. J. Lee and G. E. Scuseria, in "Quantum Mechanical Electronic Structure Calculations with Chemical Accuracy" (S. R. Langhoff, Ed.), p. 47. Kluwer Academic, 1995.
81. W. Ekardt, "Metal Clusters," Series in Theoretical Chemistry, Wiley, 1999.
82. R. G. Parr and W. T. Wang, "Density Functional Theory of Atoms and Molecules." Oxford University Press, 1989.
83. J. R. Chelikowski and S. G. Louie, "Quantum Theory of Real Materials." Kluwer Academic, 1996.
84. K. Ohno, K. Esfarjani, and Y. Kawazoe, "Computational Materials Science, From ab initio to Monte Carlo," Springer series in Solid State, 1999.
85. M. C. Payne, M. P. Teter, D. C. Allan, T. A. Arias, and J. D. Joannopoulos, *Rev. Mod. Phys.* 64, 1045 (1993).
86. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
87. M. P. Allen and D. J. Tildesley, "Computer Simulation in Chemical Physics," Nato ASI Series C, Vol. 397. Kluwer Academic, 1992.
88. P. E. A. Turchi, A. Gonis, and L. Colombo, "Tight-Binding Approach to Computational Materials Science" 1998; P. E. A. Turchi and A. Gonis, "Electronic Structure and Alloy Phase Stability," JPCM 13, 8539 (2001).
89. W. A. Harrison, "Elementary Electronic Structure." World Scientific, Singapore, 1999.
90. M. Sprik, in "Computer Simulation in Chemical Physics," Nato ASI Series C, Vol. 397, p. 211. Kluwer Academic, 1992.
91. J. Israelachvili, "Intermolecular and Surface Forces," 2nd ed. Academic Press, 1995.
92. M. E. J. Newman and G. T. Barkema, "Monte Carlo Methods in Statistical Physics." Clarendon Press, Oxford, 1999.
93. D. P. Landau, K. K. Mon, and H. B. Schuttler, "Computer Simulation in Condensed Matter Physics IV." Springer, Berlin, 1992.
94. H. Gould and J. Tobochnik, "An Introduction to Computer Simulation Methods," 2nd ed. Addison-Wesley, 1996.
95. J. Singh, "Electronic and Optoelectronic Properties of Semiconductor Structures." Cambridge University Press, 2003.
96. W. Krauth and O. Pluchery, *J. Phys. A: Math Gen.* 27, L715 (1994); W. Krauth, in "Advances in Computer Simulation" (J. Kertesz and I. Kondor, Eds.), Lecture Notes in Physics. Springer, 1998; W. Krauth, "New Optimization Algorithms in Physics" (A. K. Hartmann and H. Rieger, Eds.), Wiley-VCh (unpublished); J. M. Pomeroy, J. Jacobsen, C. C. Hill, B. H. Cooper, and J. P. Sethna, *Phys. Rev. B* 66, 235412 (2002).
97. K. K. Bhattacharya and J. P. Sethna, *Phys. Rev. E* 57, 2553 (1998).
98. D. E. Knuth, "The Art of Computer Programming," 3rd ed., Vol. 2. Addison-Wesley, Reading, MA, 1998.
99. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, *J. Chem. Phys.* 21, 1087 (1953).
100. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. R. Flannery; "Numerical Recipes, The Art of Parallel Scientific Computing." Cambridge University Press, 1996.
101. M. C. McMillan, *Phys. Rev.* 138, A442 (1964).
102. R. H. Swendsen and J. S. Wang, *Phys. Rev. Lett.* 63, 86 (1987).
103. M. A. Novotny, *Phys. Rev. Lett.* 74, 1 (1995), Erratum: 75, 1424 (1995).
104. A. Zangwill, "Physics at Surfaces." Cambridge University Press, 1988.
105. R. W. Hockney and J. W. Eastwood, "Computer Simulation Using Particles." McGraw-Hill, 1981.
106. C. W. Gear, "Numerical Initial Value Problems in Ordinary Differential Equations." Prentice Hall, 1971.
107. S. Nose, *Mol. Phys.* 52, 255 (1984).
108. W. G. Hoover, *Phys. Rev. A* 31, 1695 (1985).
109. S. Melchionna, G. Ciccotti, and B. L. Holian, *Molec. Phys.* 78, 533 (1993).
110. C. G. Broyden, *Math. Comput.* 19, 577 (1965).
111. D. Singh, *Phys. Rev. B* 40, 4528 (1989).
112. K. S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* 220, 671 (1983).
113. B. Andersen and J. M. Gordon, *Phys. Rev. E* 50, 4346 (1994).
114. D. E. Goldberg, "Genetic Algorithms in Search, Optimization and Machine Learning." Addison-Wesley, 1989.
115. M. Holzmann and W. Krauth, *Phys. Rev. Lett.* 83, 2687 (1999).
116. J. C. Grossman, L. Mitás, and K. Raghavachari, *Phys. Rev. Lett.* 75, 3870 (1995).
117. V. Kumar, K. Esfarjani, and Y. Kawazoe, in "Clusters and Nanomaterials Theory and Experiment," Springer Series in Cluster Physics, 2002, and references therein; D. M. Deaven and K. M. Ho, *Phys. Rev. Lett.* 75, 288 (1995).
118. V. Kumar and Y. Kawazoe, *Phys. Rev. Lett.* 88, 235504 (2002); V. Kumar and Y. Kawazoe, *Phys. Rev. B* 64, 115405 (2001).
119. S. Datta, "Electronic Transport in Mesoscopic Systems." Cambridge University Press, 1995; Y. Meir and N. S. Wingreen, *Phys. Rev. Lett.* 68, 2512 (1992).

120. M. C. Petty and M. Bryce, Eds., "An Introduction to Molecular Electronics." Oxford University Press, New York, 1995; J. Reichert, et al., *Phys. Rev. Lett.* 88, 176804 (2002); J. Heurich, J. C. Cuevas, W. Wenzel, and G. Schön, *Phys. Rev. Lett.* 88, 256803 (2002).
121. C. Joachim, J. K. Gimzewski, and A. Aviram, *Nature (London)* 408, 541 (2000).
122. A. A. Farajian, K. Esfarjani, and Y. Kawazoe, *Phys. Rev. Lett.* 82, 5084 (1999); K. Esfarjani, et al., *Appl. Phys. Lett.* 74, 79 (1999); S. J. Tans, M. Devoret, H. Dai, A. Thess, R. E. Smalley, L. J. Geerligs, and C. Dekker, *Nature (London)* 386, 474 (1997).
123. W. Kohn, *Phys. Rev. Lett.* 76, 3168 (1996); J. M. Millam and G. E. Scuseria, *J. Chem. Phys.* 506, 5569 (1997); R. Baer and M. Head-Gordon, *J. Chem. Phys.* 109, 10159 (1998); S. Goedecker, *Rev. Mod. Phys.* 71, 1085 (1999).
124. L. W. Wang and M. Teter, *Phys. Rev. B* 45, 13196 (1992).
125. R. M. Dreizler and E. K. U. Gross, "Density Functional Theory." Springer, Berlin, 1990.
126. J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria, *Phys. Rev. Lett.* 91, 146401 (2003).