

Statistical Thermodynamics

Contents

- *Statistical Ensembles and Partition Functions*
- *NPT ensemble*
- *Determination of Average Properties*
- *Determination of Derivative Properties from Fluctuations*
- *Possible Ways of Simulating Ensembles*

This chapter's purpose, largely inspired by a textbook by Ungerer, Tavitian & Boutin [1], is to help *MedeA GIBBS* and *MedeA LAMMPS* users understand the link between forcefield-based molecular simulation methods (Monte Carlo, molecular dynamics) and classical thermodynamics.

In contrast to quantum chemistry, the detailed electronic structure of each molecule is not considered in molecular simulation with classical forcefields. Indeed, such complex treatment is generally unnecessary for thermodynamic properties, as the related molecular interactions do not involve significant electronic structure modifications. In compensation, molecular simulation can consider larger systems which often makes it possible to derive macroscopic properties (i.e., that can be compared with measured quantities).

For systems at equilibrium, this connection is provided by the well-established framework of statistical thermodynamics [2], which we recall briefly in the section *Statistical ensembles and partition functions*. As the distribution of energy among molecules plays a central role in statistical mechanics, we will then review how potential energy is modeled in forcefields, see section Forcefields for Materials Simulations and section Monte Carlo Methods. Monte Carlo Methods introduces the various types of Monte Carlo algorithms implemented in *MedeA GIBBS*, while more practical aspects of simulation and numerical methods are addressed in Morphology.

1 Statistical Ensembles and Partition Functions

The link between microscopic and macroscopic properties is not intuitive because fluctuations of properties are significant in microscopic systems. It is therefore required to collect multiple snapshots of a microscopic system for computing a meaningful *average* property. In statistical thermodynamics, this collection of snapshots is called a *statistical ensemble*.

A statistical ensemble is a collection of various states of the system, which differ in positions and velocities of the component particles. The space of all possible system states, of dimension $6N$ for N particles, is called the *phase space*.

There are several different types of statistical ensemble, depending on the type of physical system we are investigating and the conditions in which it is placed.

[1] Philippe Ungerer, B Tavitian, and A. Boutin, Applications of Molecular Simulation in the Oil and Gas Industry: Monte Carlo Methods, (Editions Technip, 2005).

[2] D.A. McQuarrie, *Statistical Mechanics*, (Harper & Row, 1975); M P Allen and D J Tildesley, *Computer Simulation of Liquids, Book*, (Clarendon Press, 1987); Erwin Schrödinger, *Statistical Thermodynamics*, (Dover Publications, 1989); Daan Frenkel and Berend Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, 1st ed., (Academic Press, 1996).

1.1 Canonical or NVT Ensemble

A system at imposed volume V and temperature T is represented by a statistical ensemble called the canonical or NVT ensemble. The number of molecules N and the global volume V are identical for all system states belonging to the ensemble, but they differ in total energy E , which is a fluctuating variable in this ensemble. Each state j of the canonical ensemble occurs with a probability proportional to $\exp(-E_j/k_B T)$ where k_B is the Boltzmann constant ($k_B = R/N_A$, $R = 8.314 \text{ J/mol/K}$ the gas constant and $N_A = 6.0222 \cdot 10^{23}$ the Avogadro number) and E_j is the total energy (kinetic + potential) of the system in state j . Here we use the symbols E for total energy (kinetic and potential) and U for the potential energy.

The Boltzmann factor $\exp(-E_j/k_B T)$ expresses that low-energy states are favored compared with high-energy states. Increasing temperature broadens the energy distribution in the ensemble with the consequence that the average energy is increased. A general notation used in statistical mechanics is:

$$\beta = \frac{1}{k_B T} \quad (1)$$

Thus, the probability of a given state j existing in the phase space is given by:

$$P_j = \frac{\exp(\beta E_j)}{Q_{NVT}} \quad (2)$$

Where

$$Q_{NVT} = \frac{1}{N!} \sum_{r_i} \sum_{p_i} \exp(-\beta E(r_i, p_i)) \quad (3)$$

The expression Q_{NVT} , known as the *partition function*, is the sum of the Boltzmann factors for all possible different states in the phase space. The factor $1/N!$ originates from the possible combinations of N identical particles corresponding to the same state. Consistently with quantum mechanics, the finite summation may be transformed in a continuous integral:

$$Q_{NVT} = \frac{1}{h^{3N} N!} \int_{r_i} \int_{p_i} \exp(-\beta E(r_i, p_i)) dr_i dp_i \quad (4)$$

where the factor $1/h^{3N}$ may be understood as the volume of an individual quantum state in the phase space (involving the Planck constant $h = 6.626 \cdot 10^{-34}$ Js).

The partition function provides the link with key thermodynamic functions. In the case of the canonical ensemble, the Helmholtz free energy A is expressed as:

$$A = -k_B T \ln Q_{NVT} \quad (5)$$

Using this relationship, we can derive numerous properties. [3] If we have a finite collection of system states, representative of the NVT statistical ensemble, average properties can be obtained by simple arithmetic averaging over the n states composing the collection without explicitly computing the partition function. For instance, the average energy is:

$$\langle E \rangle = \frac{1}{n} \sum_{i=1}^n E_i \quad (6)$$

Other statistical ensembles can be defined in the same way as the canonical ensemble, as summarized below:

- If an intensive variable is fixed, the associated extensive variable fluctuates.
- If the temperature is fixed, energy fluctuates.
- If the pressure is fixed, volume fluctuates,
- If the chemical potential is fixed, the corresponding number of molecules fluctuates.

The NVT and NPT ensembles are available in both *MedeA LAMMPS* and *MedeA GIBBS*, but the Grand Canonical Ensemble, the Osmotic Ensemble, the Semi-grand Canonical Ensemble, and the Gibbs Ensemble are available in *MedeA GIBBS* only).

[3] D.A. McQuarrie, Statistical Mechanics, (Harper & Row, 1975).

| Statistical ensemble | Imposed variables | Associated thermodynamic potential | Applications |
|--|---|---|---|
| Canonical ensemble | N, V, T | $A = E - TS$ (Helmholtz free energy) | Phase properties (P, H, C_V, μ, \dots) |
| Grand Canonical Ensemble | μ_i, V, T | PV (i.e. $E - TS - \sum \mu_i N_i$) | Adsorption in microporous solids |
| Isothermal - isobaric ensemble | N, P, T | $G = H - TS$ (Gibbs free energy) | Phase properties $H, C_p, \rho, \mu_v, \dots$ |
| Gibbs Ensemble at imposed global volume (m phases) | $N = N_1 + \dots + N_m$ $V = V_1 + \dots + V_m$ T | $A = E - TS$ (Helmholtz free energy) | Phase equilibrium of pure components and mixtures |
| Gibbs Ensemble at imposed pressure (m phases) | $N = N_1 + \dots + N_m, P, T$ | $G = H - TS$ (Gibbs free energy) | Phase equilibrium of mixtures |

2 NPT ensemble

If we want to simulate a system at imposed pressure and temperature, we use the *isothermal-isobaric* ensemble or NPT ensemble, where volume and energy are fluctuating variables. The probability of any given state-with total energy E_j and volume V_j - is proportional to:

$$\exp(-\beta E_j - \beta P V_j) \quad (7)$$

This ensemble may be used to compute the average energy using Eq. (6) and the average volume using:

$$\langle V \rangle = \frac{1}{n} \sum_{j=1}^n V_j \quad (8)$$

2.1 Grand Canonical or μVT Ensemble

If we want to simulate adsorption isotherms, either for pure compounds or multicomponent systems, we must be able to impose temperature and partial pressures to simulate a given point on the isotherm. For this purpose, the most convenient statistical ensemble is the *grand canonical ensemble*, also termed the μVT ensemble. At low pressure, the chemical potential μ_i is linked with the partial pressures $P_i = P y_i$ where P is total gas pressure y_i is the molar fraction of component i . A more general expression involves the fugacities f_i in the fluid phase in equilibrium with the adsorbent, which are equivalent to partial pressures in the low-pressure limit:

$$\mu_i = \mu_{i_0} + RT \ln \frac{f_i}{P_0} \approx \mu_{i_0} + RT \ln \frac{y_i P}{P_0} \quad (9)$$

where P_0 is a reference pressure. The chemical potential of each species appears explicitly in the probability of each system state in the grand canonical ensemble, which is proportional to:

$$\exp(-\beta E_j + \beta N_1 \mu_1 + \beta N_2 \mu_2 + \dots) \quad (10)$$

A grand canonical simulation provides the average number of molecules $\langle N_i \rangle$ in the system for every species i . In the case of microporous adsorbents, this is the average number of adsorbed molecules. The heat of adsorption may also be computed.

2.2 Microcanonical or NVE Ensemble

The microcanonical or *NVE* ensemble, in which the number of molecules, global volume, and internal energy are fixed, is obtained by solving the equations of motion in molecular dynamics without a thermostat or a barostat. In this ensemble, every system state has the same probability. The temperature fluctuates, and the average value can be obtained by considering the average kinetic energy per degree of freedom of the system:

$$T = \frac{2\langle K \rangle}{k_B N_f} \quad (11)$$

where K is the kinetic energy of the system, expressed as a function of the masses m_i and the velocities v_i of its N constituent particles:

$$K = \sum_{i=1}^N \frac{m_i v_i^2}{2} \quad (12)$$

and $N_f = 3N - N_c$ is the total number of degrees of freedom, N_c being the total number of independent constraints such as imposed bond lengths or imposed bond angles.

2.3 Gibbs Ensemble

The explicit simulation of liquid-vapor interfaces, using either the *NVT* or *NPT* statistical ensemble, requires large systems, and it is mainly used to investigate the properties of the interface itself. A more efficient way of computing phase equilibria by molecular simulation uses the Gibbs Ensemble [4], where a separate simulation box is used for each phase without any explicit interface. In this ensemble, both temperature and the total number of molecules are fixed, and we can impose either global volume (i.e., the sum of phase volumes) or pressure. As a result of transfer moves from one phase to the other, the number of molecules displays opposite fluctuations in each phase, and chemical potentials are equal in both phases. The sketch shows a two-phase Gibbs ensemble for a system comprising three molecular types (cyclohexane, n-butane, and methane). The arrows illustrate the transfer of molecules from one box to the other, which is specific to this statistical ensemble.

When pressure is imposed in the Gibbs ensemble, phase volumes fluctuate independently. Because the phase rule limits the number of imposed intensive parameters, this option can be used only when more than one component is considered.

When global volume is imposed in the Gibbs ensemble, phase volumes fluctuate in opposite directions. This option is applicable either to single or multicomponent systems. In the case of a pure component, it allows the determination of its vapor-liquid coexistence properties.

The Gibbs ensemble is a particular case of the *NVT* or *NPT* statistical ensemble in which interface energy is not accounted for.

Phase volumes and energies may be derived by applying the usual averaging procedure defined by Eq. (6). Pressure may be computed for each phase from the virial equation (see next section). The computation of pressure may be used to check that mechanical equilibrium is reached because, at equilibrium, pressure must be equal in both phases (within the statistical uncertainties).

[4] Athanassios Panagiotopoulos, "Direct Determination of Phase Coexistence Properties of Fluids by Monte Carlo Simulation in a New Ensemble," *Molecular Physics* 61, no. 4 (July 1, 1987): 813-826.

3 Determination of Average Properties

The Virial equation provides the average pressure:

$$\langle P \rangle = \frac{Nk_B T}{V} + \langle W_j \rangle \quad (13)$$

where W_j is a term called the Virial, which accounts for intermolecular interactions through the forces acting on molecule k in the j^{th} configuration of the ensemble:

$$\langle W_j \rangle = \frac{1}{3V} \sum r_k \cdot F_k \quad (14)$$

where r_k is the position of the molecular center of mass in configuration j . If we recall that $k_B = R/N_A$, where N_A is Avogadro's number, the expression reduces to the perfect gas law ($PV = RT$ for one mole) at very low densities when intermolecular forces tend to zero.

The average density (in molecules per unit volume) is derived using

$$\rho = \frac{N}{\langle V \rangle} \quad (15)$$

This is different from the arithmetic average of densities, which has no physical sense.

In statistical ensembles where the temperature is imposed, deriving the average energy of the system through equation (6) is straightforward. The reference state for energy corresponds to zero kinetic energy and zero potential energy, i.e., the conditions of a dilute gas at 0 K. Once energy is known, enthalpy can be derived through $\langle H \rangle = \langle E \rangle + \langle P \rangle V$ in the canonical ensemble or $\langle H \rangle = \langle E \rangle + P \langle V \rangle$ in the NPT ensemble.

The chemical potential is defined as a derivative of the Helmholtz energy:

$$\mu_i \left(\frac{\partial A}{\partial N_i} \right)_{V, T, N_{j, j \neq i}} \quad (16)$$

However, it cannot be evaluated simply as an ensemble average.

For extensive properties, comparison of simulation results with macroscopic measurements is more convenient when properties are expressed on a molar basis. For instance, the molar volume v may be computed from an NPT simulation using:

$$v = \frac{N_a}{N} \langle V \rangle \quad (17)$$

The same conversion may be applied to other extensive variables like energy and enthalpy. Intensive variables such as pressure or chemical potential do not need any correction factor to be compared with measurements, provided a consistent unit system is used.

4 Determination of Derivative Properties from Fluctuations

The amplitude of the fluctuations around the average makes it possible to determine partial derivatives of the property considered. For instance, the standard deviation of energy in the canonical ensemble is linked with the partial derivative of energy versus temperature, i.e., the heat capacity C_V of the system:

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = \frac{1}{k_B T^2} (\langle E^2 \rangle - \langle E \rangle^2) \quad (18)$$

In this expression, the right-hand side represents energy fluctuations around the mean value, which can be computed by molecular simulation. Similarly, fluctuations in volume in the NPT ensemble may be used to obtain the compressibility coefficient:

$$\beta_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_T = \frac{1}{k_B T^2} (\langle V^2 \rangle - \langle V \rangle^2) \quad (19)$$

Comparable fluctuation formulae are available for other derivative properties, including the isosteric heat of adsorption.

4.1 Isosteric Heat of adsorption

The differential molar enthalpy of adsorption (ΔH^S), also called the isosteric enthalpy of adsorption (Q_{st}), is defined as:

$$\Delta H = H_{\text{sorbed}} - H_{\text{gas}} \quad (20)$$

where H_{sorbed} and H_{gas} are the partial molar enthalpies in the sorbed and gas phases, respectively. The enthalpy is the sum of the internal energy and the product of pressure and volume, PV . For the vapor phase, PV is assumed to be equal to RT , and the adsorbed phase's molecular volume is neglected. Assuming molar kinetic energy is equal in the gas and the adsorbed state, we can express the heat of adsorption as a function of the total molar potential energy in the gas phase and the adsorbed phase:

$$-\Delta H^0 = RT - \bar{E}_{\text{sorbed}}^{\text{tot}} + \bar{E}_{\text{gas}}^{\text{tot}} \quad (21)$$

In GCMC simulations, it is equivalent to calculating $-\Delta H^0$ using the partial derivative of the average total energy with respect to the average number of adsorbed molecules in both gas and adsorbed phases:

$$\Delta H^0 = RT - \frac{\partial \langle E_{\text{sorbed}}^{\text{tot}} \rangle}{\partial \langle N_{\text{sorbed}} \rangle} + \frac{\partial \langle E_{\text{gas}}^{\text{tot}} \rangle}{\partial \langle N_{\text{gas}} \rangle} \quad (22)$$

where $\langle N_{\text{sorbed}} \rangle$ and $\langle N_{\text{gas}} \rangle$ are the average number of molecules in the adsorbed and gas phases, respectively. The heat of adsorption Q_{st} can thus be calculated by the fluctuations method:

$$Q_{st} = RT - \frac{\langle E_{\text{sorbed}}^{\text{tot}} N \rangle - \langle E_{\text{sorbed}}^{\text{tot}} \rangle \langle N \rangle}{\langle N_{\text{sorbed}}^2 \rangle - \langle N_{\text{sorbed}} \rangle^2} + \frac{\langle E_{\text{gas}}^{\text{tot}} N \rangle - \langle E_{\text{gas}}^{\text{tot}} \rangle \langle N \rangle}{\langle N_{\text{gas}}^2 \rangle - \langle N_{\text{gas}} \rangle^2} \quad (23)$$

The heat of adsorption calculation requires a Monte Carlo simulation of the vapor phase in the Grand Canonical ensemble to determine the last part of the above equation ((23)). These simulations are usually performed so that the number of molecules fluctuates reasonably around five or six, which often implies simulation box dimensions of about 1,000 Å. However, usually the calculation of Q_{st} is achieved without this second simulation of the vapor phase.

If the vapor phase is assumed to be ideal, the third part of the equation is equivalent to the molar intramolecular energy of molecules in the vapor phase. Moreover, if internal degrees of freedom are considered not to be affected by adsorption, the molar intramolecular energies of the vapor phase and the adsorbed phase are equal. These approximations lead to the following expression for the heat of adsorption, proposed by Nicholson and Parsonage:

$$Q_{st} = RT - \frac{\langle E_{\text{sorbed}}^{\text{ext}} N \rangle - \langle E_{\text{sorbed}}^{\text{ext}} \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \quad (24)$$

where $\langle E_{\text{sorbed}}^{\text{ext}} \rangle$ represents intermolecular interactions in the adsorbed phase.

4.2 Grid Isosteric Heat of Adsorption

Eq. (24), commonly used to determine the isosteric heat of adsorption, is based on severe approximations: ideal gas for the fluid phase and the assumption that intramolecular energies in the vapor phase and the adsorbed are equal. This last hypothesis is only sometimes fulfilled. Nevertheless, at low coverage, when sorbate/sorbate interactions are feeble, the isosteric heat Q_{st} can be determined through equation (24) by considering only the adsorbate/zeolite interactions in the intermolecular potential energy (see (25)):

$$Q_{st,0} = RT - \frac{\langle E_{\text{sorbed}}^{\text{grid}} N \rangle - \langle E_{\text{sorbed}}^{\text{grid}} \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \quad (25)$$

where $E_{\text{sorbed}}^{\text{grid}}$ is the interaction energy of adsorbate and adsorbent (i.e., neglecting adsorbate-adsorbate interactions).

5 Possible Ways of Simulating Ensembles

The simplest way to build a statistical ensemble is to solve the equations of motion for every atom in every molecule, accounting for intramolecular and intermolecular interactions inside the system: this way is known as *molecular dynamics*. An alternate way, known as Monte Carlo simulation, consists of using statistical methods to generate the collection of configurations with the desired probability distribution without solving the equations. Both methods yield identical ensemble averages (this is known as the ergodicity theorem).

Newton's equations of motion, which have to be integrated in molecular dynamics, are expressed in the form:

$$F_i = m_i \frac{d^2 r_i}{dt^2} \quad (26)$$

where r_i is the position of an atom, i.e., a three-dimensional vector (x_i, y_i, z_i)

F_i is the force acting on the atom from the other atoms of the system, which can be computed by deriving the potential energy U versus coordinates:

$$F_i = \nabla U(r_i) \quad (27)$$

The solution of the equations of motion is such that total energy (i.e., the sum of kinetic and potential energy) is constant. Therefore, molecular dynamics simulates the microcanonical or NVE ensemble unless specific steps are taken to force energy or volume to fluctuate.

To perform a molecular dynamics calculation at imposed temperature, the kinetic energy is changed by altering the center of mass velocities at regular intervals: this is the purpose of thermostat techniques, which allow simulation in the NVT ensemble [5].

Molecular dynamics at imposed pressure is possible through barostat or barostat methods that allow for volume variations [6] to obtain the correct probability distribution equation.

Simulating open ensembles (Grand Canonical and Gibbs ensemble) with molecular dynamics is difficult because molecule insertions or deletions cause discontinuities in integrating Newton's equations. Monte Carlo is more often used for such ensembles.

Molecular dynamics is most useful for simulating the dynamic behavior of a system, particularly its transport properties. By transport properties, we mean here the coefficients governing the rate of mass transfer (diffusion coefficients), heat transfer (thermal conductivity), or momentum transfer (viscosity). For instance, the self-diffusion coefficient of a component in a space of dimension d can be expressed from the mean squared displacement, according to Einstein's expression:

$$D = \frac{\langle (r(t) - r(t_0))^2 \rangle}{2d(t - t_0)} \quad (28)$$

Shear viscosity η and thermal conductivity λ may also be determined using molecular dynamics through Green Kubo expressions (see Ungerer, Nietro-Draghi, Rousseau, Ahunbay & Lachet [7] and references therein):

$$\eta = \frac{V}{k_B T} \int_0^{\infty} \langle \sigma_{xy}(t) \sigma_{xy}(0) \rangle dt \quad (29)$$

$$\lambda = \frac{V}{3k_B T} \int_0^{\infty} \langle J_q(t) J_q(0) \rangle dt \quad (30)$$

where $\sigma_{xy}(t)$ and $J_q(t)$ are pressure tensor components and heat flux. It is also possible to use non-equilibrium molecular dynamics, in which perturbations or specific boundary conditions are imposed, to get transport coefficients [8].

[5] Shuichi Nosé, "A Unified Formulation of the Constant Temperature Molecular Dynamics Methods," *Journal of Chemical Physics* 81, no. 1 (1984): 511. William Hoover, "Canonical Dynamics: Equilibrium Phase-Space Distributions," *Physical Review A* 31, no. 3 (March 1985): 1695-1697.

[6] D Evans and G Morriss, "Isothermal-Isobaric Molecular Dynamics," *Chemical Physics* 77, no. 1 (May 15, 1983): 63-66.

[7] Philippe Ungerer, Carlos Nieto-Draghi, Bernard Rousseau, Göktug Ahunbay, and Véronique Lachet, "Molecular Simulation of the Thermophysical Properties of Fluids: From Understanding Toward Quantitative Predictions," *Journal of Molecular Liquids* 134, no. 1 (May 15, 2007): 71-89.

[8] B Y Wang and P T Cummings, "Non-Equilibrium Molecular Dynamics Calculation of the Shear Viscosity of Carbon Dioxide/Ethane

Mixtures," Molecular Simulation 10, no. 1 (April 1, 1993): 1-11; RL Rowley, Statistical Mechanics for Thermophysical Property Prediction, Prentice Hall, 1994; F Müller-Plathe and D Reith, "Cause and Effect Reversed in Non-Equilibrium Molecular Dynamics: an Easy Route to Transport Coefficients," Computational and Theoretical Polymer Science 9, no. 3 (1999): 203-209.