# Supplementary Material to
# $ms2$: A Molecular Simulation Tool for Thermodynamic Properties

Stephan Deublein[a], Bernhard Eckl[b], Jürgen Stoll[b], Sergey V. Lishchuk[c], Gabriela Guevara-Carrion[a], Colin W. Glass[d], Thorsten Merker[a], Martin Bernreuther[d], Hans Hasse[a], Jadran Vrabec[e,*]

[a]*Lehrstuhl für Thermodynamik, Universität Kaiserslautern, 67653 Kaiserslautern, Germany*
[b]*Institut für Technische Thermodynamik und Thermische Verfahrenstechnik, Universität Stuttgart, 70550 Stuttgart, Germany*
[c]*Department of Mathematics, University of Leicester, Leicester LE1 7RH, United Kingdom*
[d]*Höchstleistungsrechenzentrum Universität Stuttgart (HLRS), 70550 Stuttgart, Germany*
[e]*Lehrstuhl für Thermodynamik und Energietechnik, Universität Paderborn, 33098 Paderborn, Germany*

**Abstract**

This supplementary material includes detailed descriptions of the equations implemented in $ms2$ as well as detailed information on the input and output files of the program.

## 1. Definitions of the thermodynamic properties accessible in $ms2$

*1.1. Density, pressure, internal enegy and enthalpy*

At constant temperature $T$ and volume $V$, the pressure $p$ is determined by [1]

$$p = \frac{k_{\mathrm{B}}T}{V} + \frac{\langle W \rangle}{V} + \Delta p^L = \frac{k_{\mathrm{B}}T}{V} + \frac{1}{3V} \Big\langle \sum_{i=1}^{N-1} \sum_{\substack{j=i+1 \\ r_{ij} < r_{\mathrm{c}}}}^{N} \boldsymbol{r}_{ij} \boldsymbol{f}_{ij} \Big\rangle + \Delta p^L \,, \tag{1}$$

where $k_{\mathrm{B}}$ is the Boltzmann constant and the brackets $\langle ... \rangle$ indicate the ensemble average. $W$ denotes the virial, which is defined by the force vector $\boldsymbol{f}_{ij}$ acting between two molecules at a separation vector $\boldsymbol{r}_{ij}$ between their centers of mass. The contribution $\Delta p^L$ considers the long-range interactions with molecules beyond the cut-off radius $r_{\mathrm{c}}$.

In the isothermal-isobaric ensemble, the volume is a fluctuating parameter that on average yields the volume corresponding to the specified pair of temperature $T$ and pressure $p_0$. In MD simulations with Andersen's barostat [2], which is implemented in $ms2$, the fluctuations of the volume are damped by a fictive piston mass $Q_{\mathrm{p}}$ according to the equation of motion

$$\ddot{V} = \frac{p - p_0}{Q_{\mathrm{p}}} \,. \tag{2}$$

In MC simulations, the volume is changed randomly and the new volume accepted by applying the Metropolis acceptance criterion

$$P_{acc} = \min\big(1, \exp\big(\frac{\Delta E}{k_{\mathrm{B}}T}\big)\big) \,, \tag{3}$$

---

*Corresponding author: Jadran Vrabec, Warburger Str. 100, 33098 Paderborn, Germany, Tel.: +49-5251/60-2421, Fax: +49-5251/60-3522, Email: jadran.vrabec@upb.de

where $P_{\mathrm{acc}}$ is the probability of accepting the volume change and $\Delta E$ is the difference between the old state with energy $U_{old}$ for the volume $V_{old}$ and the new state with energy $U_{new}$ for the volume $V_{new}$ according to

$$\Delta E = p_0(V_{new} - V_{old}) + (U_{new} - U_{old}) + Nk_{\mathrm{B}}T\ln\frac{V_{old}}{V_{new}} \;. \tag{4}$$

The residual enthalpy $H^{\mathrm{res}}$ is directly linked to the residual internal energy, pressure and volume of the system

$$H^{\mathrm{res}} = U^{\mathrm{res}} + pV - Nk_{\mathrm{B}}T \;, \tag{5}$$

where $U^{\mathrm{res}}$ is the residual potential energy. It is defined as the sum of all pairwise interaction energies $u_{ij}$ and the appropriate long range correction $\Delta U^L$, cf. section 2.4 of this Supplementary Material

$$U^{\mathrm{res}} = \sum_{i=1}^{N-1}\sum_{\substack{j=i+1 \\ r_{ij}<r_{\mathrm{c}}}}^{N} u_{ij} + \Delta U^L \;. \tag{6}$$

### 1.2. Second derivatives

The accessible second derivatives vary with the employed ensemble. In the $NVT$ ensemble, the residual iso-choric heat capacity $c_v^{\mathrm{res}}$ is determined by fluctuations of the residual potential energy $U^{\mathrm{res}}$

$$c_v^{\mathrm{res}} = \frac{1}{N}\Big(\frac{\partial U^{\mathrm{res}}}{\partial T}\Big)_v = \frac{1}{k_{\mathrm{B}}(NT)^2}\big(\langle(U^{\mathrm{res}})^2\rangle - \langle U^{\mathrm{res}}\rangle^2\big) \;. \tag{7}$$

The partial derivative of the potential energy with respect to the volume at constant temperature $\big(\partial U^{\mathrm{res}}/\partial V\big)_T$ is determined by fluctuations of the residual potential energy $U^{\mathrm{res}}$ and the virial $W$

$$\Big(\frac{\partial U^{res}}{\partial V}\Big)_T = \frac{N}{3V}\Big(\frac{1}{k_{\mathrm{B}}T}\big(\langle W\rangle\langle U^{\mathrm{res}}\rangle - \frac{1}{N}\langle WU^{\mathrm{res}}\rangle\big) + \langle W\rangle\Big) \;. \tag{8}$$

In the $NpT$ ensemble, the second derivatives of the Gibbs energy, namely the residual isobaric heat capacity $c_p^{\mathrm{res}}$, the isothermal compressibility $\beta_T$ and the volume expansivity $\alpha_p$ are functions of ensemble fluctuations [1]. The residual isobaric heat capacity $c_p^{\mathrm{res}}$ is related to fluctuations of the residual enthalpy $H^{\mathrm{res}}$

$$c_p^{\mathrm{res}} = \frac{1}{N}\Big(\frac{\partial H^{\mathrm{res}}}{\partial T}\Big)_p = \frac{1}{k_{\mathrm{B}}(NT)^2}\big(\langle(H^{\mathrm{res}})^2\rangle - \langle H^{\mathrm{res}}\rangle^2\big) \;. \tag{9}$$

To obtain the total isobaric heat capacity $c_p$, the solely temperature dependent ideal gas contribution $c_p^{\mathrm{id}}(T)$ has to be added. This ideal property is accessible e.g. by quantum chemical methods and can often be found in data bases [3].

An analogous relationship links the isothermal compressibility $\beta_T$ to volume fluctuations in the $NpT$ ensemble

$$\beta_T = -\frac{1}{V}\Big(\frac{\partial V}{\partial p}\Big)_T = \frac{1}{k_{\mathrm{B}}T\langle V\rangle}\big(\langle V^2\rangle - \langle V\rangle^2\big) \;. \tag{10}$$

The partial derivative of the residual enthalpy with respect to the pressure at constant temperature $\left(\partial H^{\mathrm{res}}/\partial p\right)_T$ is linked to fluctuations of volume and residual internal energy

$$\left(\frac{\partial H^{\mathrm{res}}}{\partial p}\right)_T = V - \frac{N}{T}\left(\langle U^{\mathrm{res}}V\rangle - \langle U^{\mathrm{res}}\rangle\langle V\rangle + p(\langle V^2\rangle - \langle V\rangle^2)\right),\tag{11}$$

and the volume expansivity $\alpha_p$ again to volume and residual enthalpy fluctuations

$$\alpha_p = \frac{1}{V}\left(\frac{\partial V}{\partial T}\right)_p = \frac{N}{T^2\langle V\rangle}\left(\langle H^{\mathrm{res}}V\rangle - \langle H^{\mathrm{res}}\rangle\langle V\rangle\right).\tag{12}$$

Note that Eqs. (7) to (12) are valid for mixtures as well.

### 1.3. Speed of sound

The speed of sound $c$ is defined by the isothermal compressibility $\beta_T$, the volume expansivity $\alpha_p$, the isobaric heat capacity $c_p$ and the temperature $T$ by

$$c = \left(\frac{1}{M(\beta_T\rho - T\alpha_p^2/c_p)}\right)^{0.5},\tag{13}$$

where $M$ is the molar mass. In $ms2$, the speed of sound is calculated both for pure components and mixtures in the $NpT$ ensemble.

### 1.4. Chemical potential - Widom test molecule insertion

For the calculation of the chemical potential of component $i$ according to Widom [4], a so-called "test" molecule $l$ of component $i$ is inserted into the simulation volume at a random position with a random orientation. At constant temperature and pressure its potential energy $\psi_l$, due to its interactions with all other "real" molecules, is related to the chemical potential according to

$$\mu_i - \mu_i^{\mathrm{id}}(T) = -k_{\mathrm{B}}T\ln\frac{\langle V\exp\left(-\psi_l/k_{\mathrm{B}}T\right)\rangle}{\langle N_i\rangle}.\tag{14}$$

The test molecule is removed immediately after the calculation of its potential energy, thus it does not influence the time evolution of the system or the Markov chain, respectively.

The value of $\psi_l$ is highly dependent on the random position of the test molecule. In addition, the density of the system has a significant influence on the accuracy of the calculation. For very dense fluids, test molecules almost always overlap with some real molecules, which leads to a potential energy $\psi_l \to \infty$ and thus to no contribution to Eq. (14) resulting in poor statistics for the chemical potential or even complete failure of the sampling. Within limits, lower statistical uncertainties of the chemical potential can be achieved by inserting a large number of test molecules into the simulation volume, which leads to an increasing computational demand.

### 1.5. Self-diffusion coefficients

The self-diffusion coefficient $D_i$ is related to the mass flux of single molecules within a fluid. Therefore, the relevant Green-Kubo expression is based on the individual molecule velocity autocorrelation function [5]

$$D_i = \frac{1}{3N_i} \int_0^\infty dt \, \langle \mathbf{v}_k(t) \cdot \mathbf{v}_k(0) \rangle, \tag{15}$$

where $\mathbf{v}_k(t)$ is the center of mass velocity vector of molecule $k$ at some time $t$. Eq. (15) is an average over all $N_i$ molecules of component $i$ in the ensemble, since all contribute to the self-diffusion coefficient.

## 2. Details of $ms2$

This section provides a closer insight into $ms2$ and the implemented equations.

### 2.1. Reduced quantities

The simulation program $ms2$ internally uses reduced quantities for its calculations. All quantities are reduced by a reference length $\sigma_\mathrm{R}$, a reference energy $\epsilon_\mathrm{R}$ and a reference mass $m_\mathrm{R}$, respectively.

Table 1: Important physical quantities in their reduced form. Note that $\varepsilon_0$ indicates the permittivity of the vaccum $\varepsilon_0 = 8.854187 \times 10^{-12}$ $\mathrm{A^2 s^4 kg^{-1} m^{-3}}$.

| | |
|---|---|
| length | $l^* = \dfrac{l}{\sigma_\mathrm{R}}$ |
| energy | $u^* = \dfrac{u}{\epsilon_\mathrm{R}}$ |
| mass | $m^* = \dfrac{m}{m_\mathrm{R}}$ |
| time | $t^* = \dfrac{t}{\sigma_\mathrm{R}} \sqrt{\dfrac{\epsilon_\mathrm{R}}{m_\mathrm{R}}}$ |
| piston mass | $Q_\mathrm{p}^* = \dfrac{Q_\mathrm{p} \sigma_\mathrm{R}^4}{m_\mathrm{R}}$ |
| temperature | $T^* = \dfrac{T k_\mathrm{B}}{\epsilon_\mathrm{R}}$ |
| pressure | $p^* = \dfrac{p \sigma_\mathrm{R}^3}{\epsilon_\mathrm{R}}$ |
| density | $\rho^* = \rho \sigma_\mathrm{R}^3 N_\mathrm{A}$ |
| volume | $V^* = \dfrac{V}{\sigma_\mathrm{R}^3}$ |
| chemical potential | $\tilde{\mu} = \dfrac{\mu}{k_\mathrm{B} T}$ |
| point charge | $q^* = \dfrac{1}{\sqrt{4 \pi \varepsilon_0}} \dfrac{q}{\sqrt{\epsilon_\mathrm{R} \sigma_\mathrm{R}}}$ |
| dipole moment | $\mu^* = \dfrac{1}{\sqrt{4 \pi \varepsilon_0}} \dfrac{\mu}{\sqrt{\epsilon_\mathrm{R} \sigma_\mathrm{R}^3}}$ |
| quadrupole moment | $Q^* = \dfrac{1}{\sqrt{4 \pi \varepsilon_0}} \dfrac{Q}{\sqrt{\epsilon_\mathrm{R} \sigma_\mathrm{R}^5}}$ |
| diffusion coefficient | $D^* = \dfrac{D}{\sigma \sqrt{m_\mathrm{R}/\varepsilon_\mathrm{R}}}$ |
| viscosity | $\eta^* = \dfrac{\eta \sigma_\mathrm{R}^2}{\sqrt{m_\mathrm{R} \varepsilon_\mathrm{R}}}$ |
| thermal conductivity | $\lambda^* = \dfrac{\lambda \sigma_\mathrm{R}^2}{k_\mathrm{B} \sqrt{m_\mathrm{R}/\varepsilon_\mathrm{R}}}$ |

## 2.2. Intermolecular interactions

*Point dipole interactions.* The interaction between a point dipole with moment $\mu_i$ and a point charge $q_j$ at a distance $r_{ij}$ is given by [6, 7]

$$u_{ij}^{Dq}(r_{ij}, \theta_i, \mu_i, q_j) \quad = \quad -\frac{1}{4\pi\varepsilon_0}\frac{\mu_i q_j}{r_{ij}^2}\cos\theta_i .$$

(16)

Here, $\theta_i$ is the angle between the distance vector of the point charge and the orientation vector of the point dipole, as illustrated in Figure 2 in the associated paper.

*Point quadrupole interactions.* The interaction potential of a linear point quadrupole $Q_i$ with a point charge $q_j$ at a distance $r_{ij}$ is given by [1, 8]

$$u_{ij}^{Qq}(r_{ij}, \theta_i, Q_i, q_j) = \frac{1}{4\pi\varepsilon_0}\frac{Q_i q_j}{4r_{ij}^3}\big(3\cos^2\theta_i - 1\big),$$

(17)

where $\theta_i$ is the angle between the distance vector of the point charge and the orientation vector of the point quadrupole, as illustrated in Figure 2 in the associated paper.

The interaction between a linear point quadrupole with moment $Q_i$ and a point dipole with moment $\mu_j$ is given by [6, 8]

$$u_{ij}^{Q\mu}(r_{ij}, \theta_i, \theta_j, \phi_{ij}, Q_i, \mu_j) = \frac{1}{4\pi\epsilon_0}\frac{3}{2}\frac{Q_i\mu_j}{r_{ij}^4}\big(\cos\theta_i - \cos\theta_j\big)$$
$$\big(1 + 3\cos\theta_i\cos\theta_j - 2\cos\phi_{ij}\sin\theta_i\sin\theta_j\big),$$

(18)

where the angles $\theta_i$, $\theta_j$ and $\phi_{ij}$ indicate the relative angular orientation of the point dipole $j$ and the point quadrupole $i$, as shown in Figure 2 in the associated paper.

## 2.3. Reaction field method

The truncation of electrostatic interactions of first and second order are corrected for with the reaction field method [9, 10]. Here, all dipoles within the cut-off radius $r_c$ polarize the fluid surrounding the cut-off sphere, which is modeled as a dielectric continuum with a relative permittivity $\varepsilon_s$. The polarization gives rise to a homogeneous electric field within the cut-off radius, called the reaction field $\boldsymbol{E}_i^{\mathrm{RF}}$ of magnitude

$$\boldsymbol{E}_i^{\mathrm{RF}} = \frac{1}{4\pi\varepsilon_0}\frac{2(\varepsilon_s - 1)}{2\varepsilon_s + 1}\frac{1}{r_c^3}\sum_{b=1}^{m}\sum_{\substack{j=1 \\ r_{ij}<r_c}}^{N_b}\boldsymbol{\mu}_{b,j} .$$

(19)

Note that all $m$ dipoles with moment $\boldsymbol{\mu}$ on all $N_b$ molecules within the cut-off sphere have to be summed up. The reaction field acting outside of the cut-off sphere interacts with a dipole $\boldsymbol{\mu}_i$ at the center of the cut-off sphere by [1, 11]

$$u_i^{\mu\mathrm{RF}} = -\frac{1}{2}\boldsymbol{\mu}_i\boldsymbol{E}_i^{\mathrm{RF}} ,$$

(20)

where $u_i^{\mu\mathrm{RF}}$ is its contribution to the potential energy. In Eq. (20), it is assumed that the system is sufficiently large so that tinfoil boundary conditions ($\varepsilon_s \to \infty$) are applicable without a loss of accuracy, e.g. $N \geq 500$ [12, 13, 14, 15].

The reaction field method can also be applied to correct for sets of point charges, as long as they add up to a total charge of zero. Therefore, a point charge distribution is reduced to a dipole vector $\boldsymbol{\mu}^q$ according to

$$\boldsymbol{\mu}^q = \sum_{i=1}^{n} \boldsymbol{r}_i q_i \,, \tag{21}$$

where $n$ denotes the total number of point charges and $\boldsymbol{r}_i$ the position vector of charge $q_i$. The resulting dipole $\boldsymbol{\mu}^q$ is transferred into the correction term according to Eqs. (19) and (20).

### 2.4. Cut-off mode

In the site-site cut-off mode, the LJ contributions of molecules beyond the cut-off radius $r_c$ to the internal energy are estimated by [1]

$$\Delta U^{L*} = \frac{8}{9} N \rho^* \pi \left( \left(\frac{1}{r_c^*}\right)^9 - 3 \left(\frac{1}{r_c^*}\right)^3 \right) \,. \tag{22}$$

The contributions to the pressure are considered by [1]

$$\Delta p^{L*} = \frac{32}{9} (\rho^*)^2 \pi \left( \left(\frac{1}{r_c^*}\right)^9 - \frac{3}{2} \left(\frac{1}{r_c^*}\right)^3 \right) \,. \tag{23}$$

Using the center of mass cut-off mode, the LJ contributions of molecules beyond the cut-off radius $r_c$ are estimated on the basis of the correction terms of Lustig [16]. The correction $\Delta u^{L*}$ for the residual internal energy can be divided into three contributions, the contributions by interactions between two molecule centers (TICC), between one molecule center and one site (TICS), and between two molecule sites (TISS). The first contributions describe the long range contributions between molecular interaction sites that are positioned in the center of the mass of the individual molecules. The TICS terms describe the contributions, where one site is located in the center of mass of its molecule, while the other site is not. The TISS terms correct for contributions between molecular sites that are not positioned in the center of mass of their molecules. Using these formulations, the correction term for the residual internal energy is defined by

$$\Delta u^{L*} = 2\pi\rho \sum_{i=1}^{N_{\mathrm{C}}} \sum_{j=1}^{N_{\mathrm{C}}} \sum_{\alpha=1}^{N_{\mathrm{LJ},i}} \sum_{\beta=1}^{N_{\mathrm{LJ},j}} x_i x_j 4\varepsilon \begin{cases} \left(\mathrm{TICCu}(-6) - \mathrm{TICCu}(-3)\right) \\ \left(\mathrm{TICSu}(-6) - \mathrm{TICSu}(-3)\right) \\ \left(\mathrm{TISSu}(-6) - \mathrm{TISSu}(-3)\right) \,. \end{cases} \tag{24}$$

Here, $N_{\mathrm{C}}$ is the number of components in the system and $N_{\mathrm{LJ},i}$ the number of LJ sites of the molecule of component $i$. The TIXXu functions are calculated by the following equations, where the argument in the brackets defines the exponent $n$

$$\mathrm{TICCu}(n) = \frac{r_c^{2n+3}}{\sigma^{2n}(2n+3)} \,, \tag{25}$$

7

$$\text{TICSu}(n) = \frac{(r_{\rm c} + \tau)^{2n+3} - (r_{\rm c} - \tau)^{2n+3}}{4\sigma^{2n}\tau(n+1)(2n+3)} r_{\rm c} + \frac{(r_{\rm c} + \tau)^{2n+4} - (r_{\rm c} - \tau)^{2n+4}}{4\sigma^{2n}\tau(n+1)(2n+3)(2n+4)} , \tag{26}$$

$$\text{TISSu}(n) = -\frac{(r_{\rm c} + \tau_+)^{2n+4} - (r_{\rm c} + \tau_-)^{2n+4} - (r_{\rm c} - \tau_-)^{2n+4} + (r_{\rm c} - \tau_+)^{2n+4}}{8\sigma^{2n}\tau_1\tau_2(n+1)(2n+3)(2n+4)} r_{\rm c} +$$
$$\frac{(r_{\rm c} + \tau_+)^{2n+5} - (r_{\rm c} + \tau_-)^{2n+5} - (r_{\rm c} - \tau_-)^{2n+5} + (r_{\rm c} - \tau_+)^{2n+5}}{8\sigma^{2n}\tau_1\tau_2(n+1)(2n+3)(2n+4)(2n+5)} . \tag{27}$$

The terms $r_+$ and $r_+$ are defined by

$$\tau_+ = \tau_1 + \tau_2 , \tag{28}$$

$$\tau_+ = \tau_1 - \tau_2 , \tag{29}$$

where $\tau_1$ and $\tau_2$ define the distance between sites 1 and 2, respectively, to the center of mass of the molecules they belong to.

The correction of the pressure for the center of mass cut-off is given by

$$\Delta p^{L*} = \frac{-2}{3}\pi\rho^2 \sum_{i=1}^{N_{\rm C}} \sum_{j=1}^{N_{\rm C}} \sum_{\alpha=1}^{N_{\rm LJ,i}} \sum_{\beta=1}^{N_{\rm LJ,j}} x_i x_j 4\varepsilon \begin{cases} \left(\text{TICCp}(-6) - \text{TICCp}(-3)\right) \\ \left(\text{TICSp}(-6) - \text{TICSp}(-3)\right) \\ \left(\text{TISSp}(-6) - \text{TISSp}(-3)\right) . \end{cases} \tag{30}$$

These expressions follow the same naming scheme as for the internal energy correction. The functions TICCp, TICSp and TISSp are defined by

$$\text{TICCp}(n) = 2n \cdot \text{TICCu}(n) , \tag{31}$$

$$\text{TICSp}(n) = -\frac{(r_{\rm c} + \tau)^{2n+2} - (r_{\rm c} - \tau)^{2n+2}}{4\sigma^{2n}\tau(n+1)} r_{\rm c}^2 - 3 \cdot \text{TICSu}(n) , \tag{32}$$

$$\text{TISSp}(n) = -\frac{(r_{\rm c} + \tau_+)^{2n+3} - (r_{\rm c} + \tau_-)^{2n+3} - (r_{\rm c} - \tau_-)^{2n+3} + (r_{\rm c} - \tau_+)^{2n+3}}{8\sigma^{2n}\tau_1\tau_2(n+1)(2n+3)} r_{\rm c}^2 - 3 \cdot \text{TISSu}(n) . \tag{33}$$

### 3. Modular structure - example

The modular structure of $ms2$ is discussed by highlighting a MD simulation calculation process of a pure fluid modeled by two LJ sites and two point charges in the $NVT$ ensemble. During the initialization process, the modules for memory allocation and definition of the simulation scenario are executed. This process is not unique for any of the hierarchical levels, but takes places on all levels, cf. Figure 4 in the associated paper.

On level one, the simulation is delimited to "molecular dynamics", therefore all modules concerning MC simulation are entirely neglected. On level two, the "initialization" modules define the simulation environment by setting ensemble specific properties, like simulation volume, temperature as well as the thermostat and integration algorithm etc. While the use of the first modules mentioned above is required in each simulation, independently of MD or MC, e.g. the trajectory calculation is specific for MD. However, the modular structure retains the efficiency of the program by initiating only the molecule propagation that is specific for the simulation technique. The same effect of modules can be found in the initiation process on lower levels. Here, the composition of the molecular species in the simulation volume (module component), the description of the molecule (module molecule) and the storage of the positions, velocities and forces of all LJ sites and charges (module sites) are determined. All other modules, e.g. treating additional components for mixtures or containing characteristic properties of dipoles or quadrupoles, are fully omitted by $ms2$ in this case, because they are not needed in this simulation. Once the simulation has been initialized, a set of modules is invoked, which deals with maintaining the simulation run. On level one, this includes modules of global use for any simulation, e.g. limiting the execution time etc. On level two, Newton's equations of motion are numerically integrated, requiring information about the fluid's composition, the positions and orientations of the molecules etc. provided by modules of the respective level. The thermostat is applied and the calculation of thermodynamic properties is performed. The forces, torques and energies are calculated in the second branch on the level component. For a MD simulation, this covers routines for the calculation of energies as well as forces at the same time, whereas for a MC simulation, the modules only calculate energies and virial contributions. A further set of modules deals with the accumulation of data. These modules are designed to store data and evaluate the statistics of the produced data. The actual thermodynamic properties are calculated on the ensemble level. The last set of modules deals with the output, where the results of the molecular simulation are written to file. These modules are called independently on the user settings in all simulation scenarios.

## 4. Input and output

$ms2$ was designed to be an easily applicable simulation program. The structure of the input files as well as the output files is shown in Figure 1.
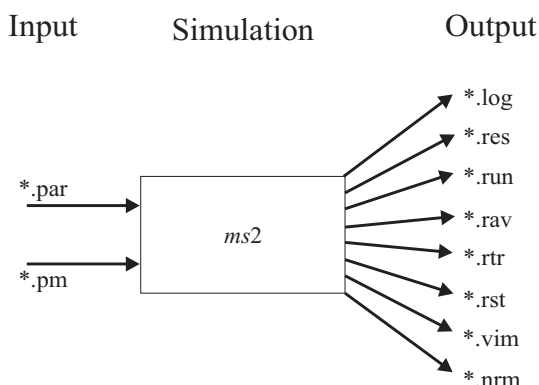
Input     Simulation     Output

*.par

*.pm

*ms2*

*.log
*.res
*.run
*.rav
*.rtr
*.rst
*.vim
*.nrm

Figure 1: File structure needed and generated by $ms2$.

### 4.1. Input file *.par

The simulation program $ms2$ requires one input file (*.par) to specify the simulation parameters and one molecular model file (*.pm) for every molecular species considered. The *.par file contains all input variables for the simulation process, such as simulation type, ensemble, number of equilibration and production steps, time step length etc. Furthermore, the user has to specify temperature, density and fluid composition. Table 2 lists all input parameters and options to be specified in the *.par file. An example for a complete *.par file is given below for a MC simulation of pure ethylene oxide in the $NpT$ ensemble, where the chemical potential is calculated by gradual insertion.

Table 2: Parameters and options specified in the *.par file.

| Parameter | Option | Explanation | Recommended value |
|---|---|---|---|
| Units | SI | Physical properties in the *.par file are given in SI units | SI |
| | reduced | Physical properties in the *.par file are given in reduced units with respect to the reference values of length $\sigma_R$, energy $\epsilon_R$ and mass $m_R$ | |
| LengthUnit | | Reference length $\sigma_R$ in Å | 3.0 |
| EnergyUnit | | Reference energy $\epsilon_R/k_B$ in K | 100.0 |
| MassUnit | | Reference mass $m_R$ in atomic units $u = 1.6605 \times 10^{-27}$ kg | 100.0 |
| Simulation | MD | Molecular dynamics simulation | |
| | MC | Monte-Carlo simulation | |
| Integrator | Gear | Gear predictor-corrector integrator (MD only) | Gear |
| | Leapfrog | Leapfrog integrator (MD only) | |

**Table 2** continued

| Parameter | Option | Explanation | Recommended value |
|---|---|---|---|
| TimeStep | | Time step of one MD step in fs (MD only) | ∼1 |
| Acceptance | | Acceptance rate for MC moves (MC only) | 0.5 |
| Ensemble | NVT | Canonical ensemble | |
| | NVE | Micro-canonical ensemble | |
| | NPT | Isobaric-isothermal ensemble | |
| | GE | Grand equilibrium method (pseudo-$\mu VT$) | |
| MCORSteps | | MC relaxation loops for pre-equilibration | 100 |
| NVTSteps | | Number of equilibration time steps (MD) or loops (MC) in the $NVT$ ensemble | 20000 |
| NPTSteps | | Number of equilibration time steps (MD) or loops (MC) in the $NpT$ ensemble (optional) | 50000 |
| RunSteps | | Number of production time steps (MD) or loops (MC) | 300000 |
| ResultFreq | | Size of block averages in time steps or loops | 100 |
| ErrorFreq | | Frequency of writing the *.res file in time steps or loops | 5000 |
| VisualFreq | | Frequency of saving configurations in the *.vim file for visualization in time steps (MD) or loops (MC) | 5000 |
| CutoffMode | COM | Center of mass cut-off | COM |
| | Site | Site-site cut-off | |
| NEnsemble | | Number of ensembles in the simulation | 1 |
| CorrfunMode | yes | Calculation of autocorrelation functions enabled | |
| | no | Calculation of autocorrelation functions disabled | |
| Temperature | | Specified temperature | |
| Pressure | | Specified pressure | |
| Density | | Specified density | |
| PistonMass | | Piston mass for simulations at constant pressure | |
| NParticles | | Total number of molecules | 864 - 4000 |
| Liqdensity | | Simulation result density | |
| VarLiqDensity | | Statistical uncertainty of density | |
| LiqEnthaly | | Simulation result residual enthalpy | |
| VarLiqEnthaly | | Statistical uncertainty of residual enthalpy | |
| LiqBetaT | | Simulation result isothermal compressibility | |
| VarLiqBetaT | | Statistical uncertainty of isothermal compressibility | |
| LiqdHdP | | Simulation result $(dh^{\mathrm{res}}/dp)_T$ | |
| VardHdP | | Statistical uncertainty of $(dh^{\mathrm{res}}/dp)_T$ | |
| NComponents | | Number of components in the simulation | |

**Table 2** continued

| Parameter | Option | Explanation | Recommended value |
|---|---|---|---|
| Corrlength | | Length of the autocorrelation in time steps | |
| SpanCorrFun | | Time steps separating subsequent autocorrelation functions | |
| ViewCorrFun | | Output frequency of the full autocorrelation functions into the *.rtr file | |
| ResultFreqCF | | Output frequency of transport properties into the *.res file | |
| PotModel | | Potential model *.pm file of a component | |
| MolarFract | | Molar fraction of a component | |
| ChemPotMethod | none | No calculation of the chemical potential for this component | none |
| | Widom | Calculation of the chemical potential for this component using Widoms's test molecule method | |
| | GradIns | Calculation of the chemical potential for this component using gradual insertion | |
| NTest | | Number of test molecules for Widom's test molecule method | 2000 |
| WeightFactors | Guess | For gradual insertion: use user defined initial values for the weight factors with optimization of these factors during simulation | Guess |
| | OptSet | For gradual insertion: user defined values for the weight factors without adjustment during simulation | |
| Cutoff | | Cut-off radius for center of mass cut-off | |
| CutoffLJ | | Cut-off radius for LJ interactions (site-site cut-off) | |
| CutoffDD | | Cut-off radius for dipole-dipole interactions (site-site cut-off) | |
| CutoffDQ | | Cut-off radius for dipole-quadrupole interactions (site-site cut-off) | |
| CutoffQQ | | Cut-off radius for quadrupole-quadrupole interactions (site-site cut-off) | |
| Epsilon | | Dielectric constant | 1.0E+10 |

An example for a *.par file is given in Table 3. The scenario is a MC simulation in the $NpT$ ensemble for ethylene oxide.

Table 3: Parameters and options specified in the *.par file.

| Sim_EOX.par | | |
| --- | --- | --- |
| Units | = | SI |
| LengthUnit | = | 3.0 |
| EnergyUnit | = | 100.0 |
| MassUnit | = | 100.0 |
| | | |
| Simulation | = | MC |
| Acceptance | = | 0.5 |
| Ensemble | = | NVT |
| | | |
| MCORSteps | = | 100 |
| NVTSteps | = | 2000 |
| NPTSteps | = | 10000 |
| RunSteps | = | 50000 |
| ResultFreq | = | 100 |
| ErrorsFreq | = | 2000 |
| VisualFreq | = | 10 |
| | | |
| CutoffMode | = | COM |
| NEnsembles | = | 1 |
| | | |
| Temperature | = | 400.0 |
| Pressure | = | 0.79355 |
| Density | = | 21.09227 |
| PistonMass | = | 0.00001 |
| NParticles | = | 500 |
| NComponents | = | 1 |
| | | |
| PotModel | = | eox.pm |
| MolarFract | = | 1.0 |
| ChemPotMethod | = | GradIns |
| WeightFactors | = | Guess |
| | | 1.00 |
| | | 2.00 |
| | | 3.00 |
| | | 4.00 |
| | | 5.00 |
| | | 10.00 |
| | | 20.00 |
| | | 40.00 |
| | | 60.00 |
| | | |
| Cutoff | = | 5.0 |
| Epsilon | = | 1.E+10 |

## 4.2. Input file *.pm

A *.pm file contains the molecular model for a given substance. It contains the relative positions and the parameters of all sites. The potential model file for methanol is shown in Table 4. Methanol was modeled by two LJ sites and three point charges [17]. All positions and distances in the *.pm file are given in Å, the LJ parameters $\sigma$ and $\varepsilon/k_B$ are given in Å and K, respectively, while the mass is given in atomic units ($u = 1.6605 \times 10^{-27}$ kg). The magnitudes of the charges are specified in electronic charges (e $= 1.602 \times 10^{-19}$ C), while the dipole moments and quadrupole moments are given in Debye (D $= 3.33564 \times 10^{-30}$ Cm) and Buckingham (B $= 3.33564 \times 10^{-40}$ Cm$^2$), respectively. The orientations of the dipole and quadrupole are represented by spherical coordinates, where the azimuthal angle $\phi$ specifies the angle to the positive $x$-axis and the polar angle $\theta$ defines the angle to the positive $z$-axis. Both angles are specified in degrees. Molecular models can be orientated arbitrarily in the *.pm file. All site positions are transformed into a principal axes coordinate system at the beginning of each simulation with $ms2$. The normalized site positions are written to a *.nrm file for each component.

Table 4: Parameters and options specified in the *.pm file for a molecular model of methanol.

| MeOH.pm | | |
|---|---|---|
| NSiteTypes | = | 2 |
| | | |
| SiteType | = | LJ126 |
| NSites | = | 2 |
| | | |
| x | = | 7.660331E-01 |
| y | = | 1.338147E-02 |
| z | = | 0.0 |
| sigma | = | 3.754348 |
| epsilon | = | 120.591759 |
| mass | = | 15.034 |
| | | |
| x | = | -6.564695E-01 |
| y | = | -6.389332E-02 |
| z | = | 0.0 |
| sigma | = | 3.030 |
| epsilon | = | 87.879094 |
| mass | = | 16.00 |
| | | |
| SiteType | = | Charge |
| NSites | = | 3 |
| | | |
| x | = | 7.660331E-01 |
| y | = | 1.338147E-02 |
| z | = | 0.0 |
| charge | = | 0.247461 |
| mass | = | 0.0 |
| shielding | = | 0.1 |
| | | |
| x | = | -6.564695E-01 |
| y | = | -6.389332E-02 |
| z | = | 0.0 |
| charge | = | -0.678742 |
| mass | = | 0.0 |
| shielding | = | 0.1 |
| | | |
| x | = | -1.004989E+00 |
| y | = | 8.145993E-01 |
| z | = | 0.0 |
| charge | = | 0.431281 |
| mass | = | 1.008 |
| shielding | = | 0.05 |
| | | |
| NRotAxes | = | auto |

*4.3. Output files*

$ms2$ yields seven output files:

- *.log file - stores a complete summary of all execution steps taken by $ms2$.

- *.res file - contains the results of the simulation in an aggregated form. The data is written to file in reduced quantities as well as in SI units, along with the statistical uncertainties of the calculated properties. The *.res file is created during simulation and updated every specified number of time steps or loops.

- *.run file - contains the calculated properties of the simulation for a specified time step or loop interval. The file is in tabular form, where the data is given in reduced units. The file is subsequently updated according to the user specification, which is set in the *.par file.

- *.rav file - contains the block averages of the calculated properties. The file is in tabular form, where the data is given in reduced units. The file is subsequently updated according to the user specification, which is set in the *.par file.

- *.rtr file - stores the final values of the autocorrelation functions and their integrals. The number of output lines has to be defined in the *.par file.

- *.rst file - is the restart file of the simulation. It contains all molecular positions, velocities, orientations, forces, torques and block averages for the thermodynamic properties. It is written once at the end of a simulation or immediately after having received a termination signal of the operating system. The *.rst file allows for a stepwise execution of the simulation, necessary e.g. in case of an early interruption of the simulation, time limits on a queuing system or unexpected halts.

- *.vim file - is the trajectory visualization file. It contains the positions and orientations of all molecules in an aggregated ASCII format. The configurations are written to file after a user-specified interval of time steps or loops. The *.vim file is readable by the visualization tool $ms2molecules$, which is also part of the simulation package.

- *.nrm file - stores the normalized coordinates of a potential model after a principal axes transformation.

## References

[1] Allen, M. and Tildesley, D., *Computer Simulation of Liquids*, Clarendon Press, Oxford, 1987.

[2] Andersen, H., The Journal of Chemical Physics **72** (1980) 2384.

[3] Rowley, R. et al., *DIPPR Data Compilation of Pure Compound Properties*, Design Institute for Physical Properties, AIChE, 2003.

[4] Widom, B., The Journal of Chemical Physics **39** (1963) 2808.

[5] Gubbins, K., *Statistical Mechanics*, volume 1, The Chemical Society Burlington House, London, 1972.

[6] Hirschfelder, J., Curtiss, C., and Bird, R., *Molecular Theory of Gases and Liquids*, J. Wiley & Sons, 1954.

[7] Lorrain, P., Corson, D., and Lorrain, F., *Elektromagnetische Felder und Wellen*, Walter de Gruyter, Berlin, 1995.

[8] Gray, C. and Gubbins, K., *Theory of molecular fluids, Volume 1: Fundamentals*, Clarendon Press, Oxford, 1984.

[9] Frenkel, D. and Smith, B., *Understanding Molecular Simulation*, Academic Press, Elsevier, San Diego, 1993.

[10] Nymand, T. and Linse, P., The Journal of Chemical Physics **112** (2000) 6152.

[11] Boettcher, C., van Belle, O., Bordewijk, P., and Rip, A., *Theory of electric polarization, Vol. 1: Dielectrics in static fields*, Elsevier, 1973.

[12] Lisal, M., Budinsky, R., and Vacek, V., Fluid Phase Equilibria **135** (1997) 193.

[13] Garzon, B., Lago, S., Vega, C., and Rull, L., The Journal of Chemical Physics **102** (1995) 7204.

[14] Lisal, M., William, W., and Nezbeda, I., Fluid Phase Equilibria **181** (2001) 127.

[15] Jedlovszky, P. and Mezei, M., The Journal of Chemical Physics **110** (1999) 2991.

[16] Lustig, R., Molecular Physics **65** (1988) 175.

[17] Schnabel, T., Srivastava, A., Vrabec, J., and Hasse, H., The Journal of Physical Chemistry B **111** (2007) 9871.