Frequently Asked Questions

Contents

1 System Related

1.1 I cannot use Update from the Web

Please download an installation medium (Windows Installer or ISO image) from our website or contact us.

1.2 Which result files do we need to backup?

Please backup the Jobs directory `/MD/Jobs` and its database `MD/Databases/MDJobs.db`.

2 *MedeA*

2.1 I cannot start up *MedeA*

Please check if you have a valid license file in the `C:/MD` or `~/MD` directory. If you do not have a license file or if *MedeA* fails to read your license file, please contact support@materialsdesign.com.

If *MedeA* stops during startup, there may be a problem connecting to the databases. Try restarting your computer and if that does not help, please contact support@materialsdesign.com.

2.2 *MedeA* shows a message "The databases are not created yet. Please try again in a minute"

On startup *MedeA* needs to access the databases which can take some seconds. While doing so, *MedeA* indicates progress in the lower left message bar.

2.3 The bonds of a structure displayed by *MedeA* look strange

You may have to update the bonds. To do so, right-click on the structure and select Edit bonds . In the menu that appears check that Reset all bonds is ticked, click Apply and close with OK .

2.4 How can I display the charge density computed by VASP?

Use Analysis >> Charge Densities >> Total Valence Charge Density to visualize charge densities.

2.5 Can *MedeA* display VASP Charge Densities including core electrons?

Use Analysis >> Charge Densities >> Total Charge Density to visualize charge densities.

2.6 Can I draw the electronic/phonon dispersion relation/density of states (DOS) using Excel?

Click on Analysis >> Export >> Bandstructure and Density of States to select a band structure or DOS job to be written to an Excel spreadsheet. You have to have a current version of Microsoft Office as the export function connects to an API provided by Office.

2.7 Can I save a Phonon animation as a movie file?

Yes, use the red record button to create snapshots saved in your `C:\Users\[username]\MedeA\Movies` or `~/MedeA/Movies` directory, then use any movie maker program to create a movie from a series of snapshots. To automate this process link your `ffmpeg` executable in *MedeA* preferences (File >> Preferences and switch to the **Programs** tab).

2.8 How do I create a vacancy?

You can edit a structure by right-clicking in the structure window. To delete an atom, first select the atom, then right-click Atom >> Delete selected atom . Note, you would need to lower the symmetry to P1 first by right-click Edit symmetry .

Alternatively, you can use the random substitution tool accessible via Builders >> Random substitution . Configure it to randomly substitute one atom with a vacancy.

A more systematic way of analyzing substitutional sites with respect to symmetry and inserting vacancies or defects is using the Substitutional Search Tool. To use this tool, select the structure you want to edit and, from the *MedeA* Builders menu, select Substitutional search.

3 Jobs

3.1 How do we check the progress of a running job?

To check the progress of a running job, select Jobs >> View and Control Jobs from the *MedeA* menu. Click on a job number then on Tasks to see a list of running tasks. Click on a task number to see the output files. For a VASP calculation, you can check the progress in the `OSZICAR` file.

3.2 What are JobServer and TaskServer?

MedeA uses a (tiered) client-server model to distribute computational tasks throughout a network. You can think of the JobServer as the central place (middle tier) where all computational tasks are configured and where all result files end. A TaskServer (end tier) is a machine that receives a specific task (e.g. a VASP single point run), performs the calculation and returns all result files to the JobServer. The JobServer controls computational jobs, which can consist of many separate tasks. *MedeA* connects to a JobServer machine to start jobs and retrieves results. In most cases you only work with one JobServer and multiple TaskServers.

3.3 I cannot access JobServer with the browser

The JobServer and/or TaskServer machines which *MedeA* tries to access might be down. Check whether the machine starts-up and if there are network problems.

On the machine with the JobServer, check if the JobServer is actually running:

- Windows:

   Click `Ctrl-Alt-Delete` to bring up the Windows Task Manger and in the Processes tab click on the column header `Image Name` to sort the list of processes alphabetically. Look for a process *mdJobServer*. If it is not present you need to start the JobServer, either as a normal program (click `Start` >> `All Programs` >> `Materials Design` >> JobServer) or as a service (click `Start` >> `All Programs` >> `Materials Design` >> `MedeA Maintenance` and then `Manage Services`)

- Linux:

   Log on to the JobServer and/or TaskServer machine via a terminal, and use

   ```
   wget localhost:32000
   ```

   or

   ```
   wget localhost:23000
   ```

   to check the status of JobServer and TaskServer, respectively.

Maybe the network IP address of your JobServer has changed or the port number for the Job Server was changed. Check your settings in `File` >> `Preferences` >> `Job Servers` or the server configuration file, located in *C:\MD\servers.dat* or *˜/MD/servers.dat*. For each JobServer configured in *MedeA* there should be one uncommented line. Make sure the settings for your machine agree with your network configuration.

3.4 The Job Server claims a job is running, but it seems to be finished on that machine

Depending on types of computational tasks, the JobServer may still need some more time to post-process results (plotting phonon dispersion curves, creating trajectory files, etc) even after the computational tasks have finished on the compute node. Please wait patiently for the JobServer to finish its post-processing.

If for extended time the Job status is still running from the JobServer, a communication error from the TaskServer back to the JobServer might have occurred. Ensure the JobServer **Host address** from *http://localhost:32000/ServerAdmin/manager.tml* is an IP address or a hostname that the TaskServer can resolve. Consult your IT department of contact Materials Design Support for further details. If you have used a hostname try replacing it with an IP address.

3.5 Can I remove a job that has failed from the Job Server?

You can remove a Job and all its associated files by ticking the checkbox next to a Job entry on the job list page of the JobServer and then click `Delete Selected` at the bottom of the page.

DOCUMENTATION

4 *MedeA VASP*

4.1 How can I see the results of a VASP calculation?

Job.out summarizes the computational settings and results of a specific calculation. To access the file via the *MedeA* menu, click on Jobs >> View and Control Jobs (brings up your default web browser) >> Jobs >> Job # >> Job.out .

To load a minimized structure, select File >> Open >> From previous calculation from the *MedeA* menu.

The `ListOfResults.sli` database, containing all calculated properties of a VASP calculation, can be added to a structure list of your choice using the structure list editor (File >> Structure List Editor). This can be done by opening an existing or a new structure list and then adding `ListOfResults.sli` databases from various jobs via Add structures >> Structure list from job .

To display graphical results such as band structures, density of states, optical spectra, etc, select the corresponding entry from the Analysis menu.

4.2 How can I preview an INCAR file to be used in a VASP calculation in *MedeA*?

In the VASP interface, select Preview Input and choose INCAR from the **File** menu.

4.3 How do I add extra VASP parameters?

You can do this via the Add to Input tab in the VASP interface window. Please note that only the first occurrence of a keyword in `INCAR` will be used by VASP. To perform a constrained minimization selected the atoms, that should be frozen, right-click and select Selection >> Freeze selected atoms... . To apply a specific magnetic ordering, right-click on an atom and set the magnetic moment via Atom >> Edit , or switch to the molecular spreadsheet and set these parameters in the corresponding columns.

4.4 How can I see the initial parameters of calculated jobs?

In *MedeA* click on Jobs and select View and control jobs . Click on the number of the job you want to check parameters for and next inspect settings used in the calculation in *Job.out*.

4.5 Is there a detailed description of the VASP parameters?

For a detailed description refer to the *MedeA* user guide, section *MedeA VASP 5.4* or *MedeA VASP 6*, or to the VASP manual, https://www.vasp.at/wiki/index.php/The_VASP_Manual. You can get access a short explanation for each parameter by right-clicking on it in the VASP interface and selecting Help .

4.6 When is a VASP calculation converged?

Single Point: A single point calculation is converged when the electronic energy difference between two self-consistent steps changes less than the SCF convergence criterion demands. Default for the SCF convergence is 10^{-5} eV.

Structure Optimization: A structure optimization is converged when the residual forces on the atoms are smaller than the convergence criterion. The default value for structure relaxation is 0.02 eV/Å.

> **Note:** A *MedeA* job may finish although a calculation is not converged!

DOCUMENTATION

For example, if a calculation does not converge within the maximum number of steps defined in the VASP interface, *MedeA* will finish the job anyway.

Defaults for the number of steps for the electronic SCF are `60` steps, and for the structure optimization `100` geometry steps.

4.7 What is the typical accuracy of a VASP calculation?

For a structure calculation, the deviation from experimental lattice parameter is typically within a 2-3 percent. This depends on the functional used. A more accurate functional, such as meta GGA, will yield better results with smaller deviation. However, it will come at a higher computational cost.

4.8 What is different between *Real space* and *Reciprocal space*?

Real space and *Reciprocal space* refer to different numerical techniques of performing integrals. Real space is recommended for "larger" unit cells (a dimension larger than 5 Å) where it is faster than the reciprocal space integration scheme. For smaller cells *reciprocal space* is recommended.

Note: If you would like to perform a very accurate calculation, use *reciprocal space* in any case.

4.9 How do I select the right smearing method?

Please consult the detailed description of the *MedeA VASP* user interface MedeA VASP.

4.10 I feel the values of the partial charges in *Job.out* are somewhat strange

VASP evaluates charges within spheres of a specific radius for each atom. The total charge is the sum of the charges within all (non-overlapping) spheres and the contribution from the interstitial space. As a default, *MedeA* applies covalent radii for all elements. You can set the sphere radii for each atom by specifying the parameter `RWIGS` in the **Add to Input** tab.

4.11 Why are the RWIGS values in INCAR files different from POTCAR ?

MedeA applies covalent radii for all elements, not the `RWIGS` values of the `POTCAR` file. If different radii should be applied, you can specify the parameter `RWIGS` in the **Add to Input** tab. See also the previous question.

4.12 I get an error trying to perform a VASP calculation on a hexagonal system

In the VASP interface SCF tab, try ticking Shift origin to Gamma or Use odd size grids .

4.13 Error when performing a DOS and Band structure calculation with VASP

Check the *Job.out* for the following message:

```
Can't use empty string as operand of "\*" while executing.
```

If this error occurs, divide the band structure calculation into a number of tasks. To do so, edit the **Number of k-points** per task on the **Band Structure** tab of the VASP interface.

DOCUMENTATION

4.14 What are the specific differences of *MedeA VASP* 5.2 and 4.6 versus *VASP* 5.4?

The support for *VASP* 4.6 and *VASP* 5.2 has ended.

Meta-GGAs and van der Waals functionals, as well as properties such as electric field gradients, hyperfine parameters and NMR chemical shifts are implemented in *VASP* 5.4 or newer and require the new potentials which are provided since *MedeA* 5.4.

The *VASP* 4.6 GUI applied *VASP* 4.6.36 executables and the entire range of potentials created in the early days of VASP development. This includes the projector augmented wave (PAW) potentials created with three different approximations to exchange and correlation, LDA (local density approximation), GGA-PW91 (generalized gradient approximation of Perdew and Wang from 1991), and GGA-PBE (GGA of Perdew, Becke and Ernzerhof) potentials. Furthermore, ultrasoft pseudopotentials (US) and for some elements norm conserving pseudopotentials (NC) were available. Spin-orbit splitting and non-collinear magnetism are available, however, non-local exchange, meta-GGAs and methods for including Van der Waals interactions are not implemented in the VASP 4.6.36 code. Property calculations are limited to electronic density of states, band structure, pseudocharge density, electron localization functions, local potential, and work functions for surfaces.

The *VASP GUI* 5.2 applied VASP 5.2.12 executables and an updated set of PAW potentials created with LDA and GGA-PBE approximations to exchange and correlation. The other potentials from the early days of VASP were dismissed because their usage was discouraged and not any more supported by the VASP development group. Features such as the non-local exchange based functionals (hybrid functionals, screened exchange and Hartree-Fock) as well as the coverage of Van der Waals interactions by means of forcefields parameterized by Stefan Grimme are exclusively implemented for PAW potentials. Furthermore, the GUI provided access to additional properties such as zone center phonons from finite differences and response tensors from the linear response technique, as well as accurate quasiparticle energies from the GW approach.

5 *MedeA* Phonon

5.1 How can a Phonon calculation give imaginary frequencies in a system that might be stable?

Phonon calculates the Eigen-frequencies from the dynamical matrix. The dynamical matrix is build from force constant matrix which contains the second derivative of the energy with respect to atomic displacements. So, this approach is based on the harmonic approximation. Imaginary frequencies (negative in a *MedeA* Phonon dispersion plot) can occur even though the system is thermodynamically stable if the harmonic model fails. For example in a perovskite. However, another reason for imaginary modes can be that the forces were not calculated precisely enough. How precisely the forces need to be calculated for a given system depends on the shape of the energy hyper surface. Therefore, a general answer is difficult. However, we recommend checking the following settings if negative modes occur for a system which might be stable:

Geometry Optimization: Make sure the undisplaced system is really in equilibrium regarding the atomic positions. Run the initial atomic relaxation separately with *VASP* before running Phonon. Choose a force convergence of 0.001-0.005 eV/Å and for the electronic (SCF) convergence choose 10ˆ-6, or 10ˆ-7 eV.

k-points sampling: You might need more k-points. Even for large supercells Gamma-point only is not recommended.

Accuracy of forces: Try using the following parameters for a more accurate calculation; :highlightgray:' Projection:' Reciprocal space; :highlightgray:' Precision:' Accurate; Check Extrafine augmentation grid for accurate forces in the **Advanced/Restart tab** (*ADDGRID = .TRUE.*)

Allow for + / - **displacements** if this option has not been set.

5.2 How can one obtain Infrared and/or Raman spectra for a previous Phonon calculation, which either didn't run the spectra part, or where the spectra part failed?

The *Phonon GUI* offers two options using the ⃞Action⃞ variable (located in the *Task* subpanel to the right side of the **Calculation** tab of the *Phonon GUI*):

- ⃞Rerun VASP based properties as selected⃞ for (re)running required VASP calculations for the IR spectra only.

- ⃞Rerun VASP based properties + Raman as selected⃞ for (re)running VASP calculations needed for IR and Raman spectra.

Both of these options are based on a previous Phonon calculation which needs to be set in the *Task* subpanel via ⃞...⃞. The detailed steps to be taken are:

- Load the minimized structure of the previous Phonon job (the initial, if a minimized one isn't available) into *MedeA*

- Open the *Phonon GUI* for this structure and select as ⃞Action⃞, for instance, ⃞Rerun VASP based properties + Raman as selected⃞

- Click the ⃞...⃞ button for *using files from*, select the same previous Phonon job as above

- As properties use the checkboxes to the left to select ⃞Infrared spectra and LO/TO mode split⃞ and ⃞Raman spectra⃞, for instance.

- Enter a proper title and submit the job

For rather large systems, the default VASP options for the spectra calculations might be too demanding, which could cause failure of these VASP calculations. In such cases you may repeat the above procedure, but modifying the VASP options before submitting the job. This is done by clicking the ⃞VASP Settings⃞ buttons for the IR and/or Raman part which becomes available when the corresponding property has been ticked. The main handle for decreasing computational demands is the ⃞K-mesh in Brillouin Zone for SCF⃞ in the **SCF** tab of the *VASP GUI*, where the ⃞Spacing of k-points⃞ could be increased from the default of 0.25 1/Ang, however, avoid a 1x1x1 mesh (Gamma-point only).

6 *MedeA GIBBS*

6.1 Which molecules should I declare as rigid, flexible and semi-flexible?

The molecules that must be declared as flexible or semi-flexible are those having at least one bending or torsion angle varying in the course of the simulation: n-alkanes beyond propane, branched alkanes, cyclic alkanes, olefins beyond propene, alkylbenzenes, thiols... Among these molecules, the selection between "flexible" and "semi-flexible" is made as follows:

- when the molecule comprises at least one fixed bending or torsion angle or if bears electrostatic charges, it must be declared as semi-flexible. This is for instance the case of alkylbenzenes (because of the planar geometry of the cycle) and of polar flexible molecules such as alcohols and thiols... because they bear electrostatic charges.

- when the molecule does not include any electrostatic charge and when all its bending angles and torsions angles vary, it may be declared as flexible. This is the case of n-alkanes, branched alkanes, cyclic alkanes.

When the "flexible" case is selected, the algorithms are slightly more rapid. The other molecules, *i.e.*, those behaving like a rigid body, are normally declared as rigid. The molecules bearing no electrostatic charge with one force center (Ar, Kr, CH_4...) or two force centers (ethane, ethylene...) may be declared as rigid or flexible. Apart from the way of writing the `molecule.dat` file, the only difference between the rigid and flexible option for these molecules is that the bias algorithm is different in the case two centers of force.

6.2 How to declare cyclic molecules?

Flexible cyclic molecules (cyclopentane, cyclohexane. . .) are automatically recognized by the GIBBS code from the information contained in molecule.dat file on the basis of a simple test (the first force center has two neighbors). It is not necessary to label them explicitly.

Similarly, alkyl-substituted cycloalkanes (alkylcyclohexane, alkycyclopentane) must be declared as flexible. The molecules containing aromatic cycle(s) that can be considered as rigid (benzene, toluene, naphthalene. . .) do not need to be declared in a specific way compared to other rigid molecules. Aromatic molecules comprising a flexible alkyl chain (alkylaromatics. . .) must be declared as semi-flexible.

Some among the more complex structures (polycyclic, several alkyl chains) can be handled, but the code is not accepting all possible structures.

6.3 Which molecules should we declare as electrostatic?

Any molecule bearing electrostatic charges according to its molecule.sci file must be declared electrostatic, whether it is fully flexible or partly flexible.

6.4 When should I use Ewald summation? Which parameters?

The Ewald summation is a method for integrating precisely the electrostatic interactions, which is recommended for the strongly polar systems (water, adsorbents bearing electrostatic charges. . .), because a simple spherical cut-off cannot account for long distance electrostatic interactions.

This method decomposes the evaluation of the electrostatic energy into two parts: a summation in the direct space and a summation in the reciprocal space.

The number of vectors in the reciprocal space must be sufficiently high to avoid a significant dependence of the electrostatic energy with this parameter. The total number of vectors corresponds to all k-vector included in a sphere of radius `kmax`. For instance, it is often considered that a kmax value of 7 is adequate for liquid water. The dimensionless parameter n $_{|alpha|}$ used to compute the α parameter of the Ewald method, which characterizes the integration in the direct space: $a = n_a \dfrac{\pi}{L}$

where L is the smaller dimension of the simulation box. The criterion to be respected is that $e^{-\alpha L} \ll 1$

which is generally obtained with n approximately equal to 2.

> **Important:** The variable cutoff must be set to zero when the Ewald summation is used (the cut-off radius in the direct space is then set automatically to L/2).

6.5 How to select the cut-off and cut-in radius?

Ideally, it is recommended to use the same cut-off radius as used by the developers of the intermolecular potential. If it is unknown, 10 Å is a reasonable value that is often used, as it leads generally to a reliable evaluation of dispersion-repulsion interactions. The user must control that the cut-off radius is smaller than half the size of the simulation box at each stage of the simulation.

If the user selects the option that the cut-off radius is half the box size (which is possible by setting the cut-off radius to zero) the above condition is always respected. However, this option involves more calculations for large boxes, so that it may increase significantly computer time.

> **Note:** This option is mandatory if Ewald summation is used.

For the Lennard-Jones potential, the cut-in (threshold distance between force centers or between charges, below which the energy is considered infinite) is comprised between zero and the half of the Lennard-Jones diameter. This parameter is aimed at avoiding useless or meaningless calculations for overlapping molecules, and it has therefore no other influence on final results.

In the case of non polar molecules (without electrostatic charges) the cut-in can be set to rather high values (typically 1.5 Å or 2 Å), in order to minimize useless calculations. However, in some difficult initializations where molecules tend to overlap strongly, it is recommended to use a very low value for this parameter (such as 0.1 Å, or even less). Indeed, a large cut-in like 1.5 or 2 Å does not allow the algorithm to make the difference between a strong and a slight overlap, as both yield an infinite energy, and the simulation may fail to separate the molecules from each other.

In the case of exp-6 potential, the user must take care of that the cut-in is greater than the distance where the potential rises its maximum. If not, it will lead to unphysical interactions at short distances. In the case of electrostatic molecules, the cut-in has also the role of preventing unphysical configurations by maintaining the distance between charges above the cut-in distance. Without this parameter, the superposition of a positive and a negative charge would indeed yield an infinitely negative energy. The charges belonging to two different molecules may get closer than their dispersion-repulsion forces centers, because they are often off-centered. Consequently, the cut-in is selected smaller in such cases.

For molecules like H_2S, H_2O, CO_2 a value of 0.5 Å is adequate for instance.

6.6 Which statistical ensemble is adapted to my problem?

In the case of monophasic systems, the answer is contained in the ensemble's name:

- if properties have to computed for a **fixed molar volume or density**, the NVT ensemble is used and it is possible to determine the resulting average pressure (among other properties) from the simulation.

- if properties have to be computed at **imposed pressure**, the NPT ensemble is used, allowing to determine the corresponding average density, volume... from the simulation results.

- if a calculation at **imposed chemical potential** is desired the μVT ensemble (also named Grand Canonical ensemble) is used, and the average number of molecules can be determined from the simulation. This ensemble is useful to compute the amount adsorbed on a solid for a fixed gas pressure, since the chemical potential of the gas components can be readily determined from the pressure and composition of the gas.

In the case of fluid phase equilibria, the adequate statistical ensemble is the GIBBS ensemble. If p phases are involved, the phase rule specifies the conditions to apply it:

- if the **number of components is (p - 1)**, the GIBBS ensemble at imposed global volume must be used. The equilibrium pressure is an output of the simulation. Phase properties do not depend on the global volume nor on the global composition, as only the proportions of the phases vary when one of these parameters is changed. This is particularly the case with a unique component (p=1): the equilibrium pressure is the saturated vapor pressure at the temperature considered, and the difference of the molar enthalpies of both phases is the molar enthalpy of vaporization.

- if the **number of components is p**, it is possible to impose either the global volume or the pressure in the GIBBS ensemble calculation. The phase properties (composition, density...) do not depend on the global composition, which influences only the proportions of the phases.

- if the **number of components exceeds p**, it is also possible to impose either the global volume or the pressure in the GIBBS ensemble calculation. Opposite to the previous cases, phase properties (composition, density...) do depend on the global composition.

These rules apply only if the p simulation boxes correspond to separate phases. If chemical reactions are involved, those rules apply but the user must consider that the number of components is reduced by one for each chemical reactions simulated.

DOCUMENTATION

6.7 How to determine the chemical potential or fugacity in Grand Canonical simulations?

In the GIBBS code, the chemical potential is divided by the Boltzmann constant k, so that it reported in Kelvin. The reference state (zero chemical potential) corresponds to an ideal gas of unit density, in the units used by the GIBBS code (one molecule per $Å^3$ i.e. P=1.66 10^6 mole/m^3). This reference state is purely fictitious. Instead of the chemical potential, the user may enter the fugacity f, which is defined by:

$$\mu = \mu_0 + RT \log f$$

In the gas phase, at low pressure (say below 1 bar), the fugacity is approximately the partial pressure $F_i = Py_i$ where y_i is the molar fraction of component i in the gas. Entering fugacity is the most convenient way to simulate adsorption isotherms at low pressure. At higher pressure or in liquid phases, significant departures from ideality are observed, and the fugacity coefficient must be introduced: $f_i = \phi_i y_i P$

A way to determine the fugacity or the chemical potential in such cases is to use a classical thermodynamic model (equation of state, activity coefficient). However this way may not be consistent with the intermolecular potential used in the simulation.

Another way to determine the fugacity or the chemical potential is by performing test insertions during a preliminary NPT simulation of the fluid phase at the desired pressure.

6.8 How many runs and iterations should I do?

The number of output files generated by a simulation increases as the number of runs. In order to minimize file handling after the simulation, this number has to be kept minimal. In the case of most monophasic simulations, a single run is sufficient. When initialization is difficult (i.e. when there is a dense phase with large molecules), it may be useful to proceed in two steps:

- a first run without transfer moves, with a low cut-in distance, until the total energy falls down to a reasonable order of magnitude (less than 10^8 K),

- from the final configuration, start a second simulation with transfer moves and with a normal cut-in distance.

6.9 What frequencies for the Monte Carlo moves?

Although it does not influence the statistical averages, the selection of the frequencies of Monte Carlo moves is a delicate point because it controls the efficiency of the simulation. This point is discussed in the textbooks cited in the reference list. If N is the total number of molecules, the following approximate guidelines may be considered:

- **volume changes:** zero if the statistical ensemble involves no volume fluctuations (monophasic NVT, Grand Canonical ensemble); frequency may be set to *1/N* for monophasic NPT simulations, to *(p-1)/N* for the GIBBS ensemble at imposed global volume with p phases, or to *p/N* for the GIBBS ensemble at imposed pressure with p phases.

- **molecule transfers between phases.** This frequency must be zero for all monophasic simulations. In the GIBBS ensemble, the frequency of transfer moves may be set between 30% at high temperature and 80% at low temperature. Ideally it should be adjusted to get a total number of accepted moves at least equivalent to a tenth of the accepted internal moves (translations, rotations. . .) but this is hard to do in dense phases. Note that the efficiency of molecule transfers is not only controlled by their frequency but also by the statistical bias parameters.

- **insertions/deletions.** This frequency is zero in the GIBBS ensemble simulations. In Grand canonical simulations (μVT) insertions/deletions are essential moves, the frequency of which may be selected according to the same rules as transfer moves in the GIBBS ensemble. In the NVT and NPT ensembles, the frequency of the Widom test insertions (which are defined by the same parameter) can be done in a comparable way, i.e. frequencies up to 80% may be necessary for an efficient estimation of the chemical potentials. Here again, the overall efficiency of the insertions/deletions depends also on the statistical bias parameters.

- **Internal moves (translations, rotations, partial regrowth, internal rotation, reptation, pivot).** The frequencies of these moves must complement the frequencies of volume changes, transfers and insertions/deletions so that the sum is 1. The distribution of these internal moves depend on the type of system considered:

\# **monoatomic molecules without electrostatic charges** (CH $_4$, Ar, Kr...): translations exclusively

\# **monoatomic molecules bearing electrostatic charges** (H $_2$O-TIP4P, H $_2$S) and rigid polyatomic molecules (ethane, ethylene, benzene...): translations and rotations in equal parts,

\# **flexible molecules of moderate size** (10 force centers or less): translations, rotations and partial regrowth, with equal frequencies. If linear molecules are present, reptation may be included.

\# **flexible molecules of large size** (more than 10 force centers): translations, rotations, regrowth and internal rotations, with equal frequencies. Include as well reptation if linear molecules are present. Include pivot for branched molecules having two branched atoms (such as 2,4 dimethyl-hexane for instance)

In the case of mixtures, the molecules needing the larger array of Monte Carlo moves must serve as reference for selecting the frequencies of internal moves. When the system relaxation is difficult, a part of the translation and rotation moves may be replaced by displacements.

6.10 What is the difference between translations and displacement moves?

In a translation, the orientation and the internal conformation of the molecule are preserved. In order to reach the desired success ratio, the amplitude of the move adjusts to values that are alike the average intermolecular spacing because these avoids excessive overlaps.

In the displacement move, the molecule is deleted in its initial position and inserted in a new location which is selected at random in the same simulation box. This move makes use of the same statistical bias procedure as the transfer moves in the GIBBS ensemble or the insertion/deletion moves in the Grand Canonical ensemble. If the molecule is flexible, it is reconstructed in a different conformation with the configurational bias. If it is rigid or cyclic, several locations and several orientations are tested through the reservoir bias algorithm. For these reasons, the displacement move requires more computer time than the translations or rotation moves, and its acceptance ratio is lower in dense phases. However, it is a useful move - if not necessary - when using the NVT or NPT monophasic ensembles in some specific cases, as when very high energy barriers make relaxation slow or impossible with other moves.

6.11 Which frequencies should I select to store results and to save configurations?

It is generally useless to store results in the `Aavgi.res` files and other result files more frequently than every 500 steps, because a higher storage frequency produces very large files and may slow down the program due to input/output limitations. Often, a storage frequency of 1000 is convenient. For systems changing slowly, when simulation have to be performed over tens of millions of Monte Carlo steps, a storage every 2500 or 5000 iterations allows you to limit the size of result files. The storage frequency of the chemical potential in the `Cpi.res` files must not be done so frequently (typically every 10^4 to 10^6 iterations), otherwise the chemical potential displays excessive fluctuations (particularly in dense liquid phases). Saving the configuration file every 5000 or 10000 iterations is generally sufficient. As this file is rather large, it is not desirable to save it more often because this would require significant input/output time.

6.12 How many molecules should I consider?

- When investigating fluid phases, the total number of molecules is generally comprised between 100 and 2000. The lower end of this interval corresponds to complex molecules for which simulations take a long computer time. The upper end corresponds to systems where specific difficulties require large systems: proximity of critical conditions (large systems limit the amplitude of density fluctuations), multicomponent systems, phase equilibria with low solubilities, associating fluids.

DOCUMENTATION

- With pure compounds, it is generally recommended that the number of molecules is at least 200 molecules in the liquid phase(s) and 10 molecules in the vapor phases. When the GIBBS ensemble is used for mixtures, a minimum number of 3-4 molecules of every type should be present on average in each phase. If a molecular type does not fulfill this condition, its concentration in the phase considered may be biased. For other ensembles (NVT, NPT) applied to multicomponent systems, the determination of phase properties is still valid when a molecular type is represented by one molecule only.

Nevertheless, representing adequately the composition of a multicomponent fluid requires large systems (for instance 500 molecules allow to represent the molar fractions with an accuracy of 0.2 %) If associating species are present (water, alcohol. . .) care must be taken that the distribution of possible cluster sizes is adequately sampled. This may require larger systems than usual. When adsorption on solids is considered through the Grand Canonical ensemble, the number of molecules is not imposed, but it is controlled by the size of the simulation box.

6.13 How to select the size of simulation boxes?

In the case of adsorption in microporous solids, the box lengths in the x, y and z directions must be multiples of the elementary unit cell. These lengths will be different for orthorhombic solids. Often, a single elementary unit cell is too small for a representative system. It is important to select a number of unit cells in the x, y and z directions sufficiently large that a minimum number of 3-4 molecules of each species is found on average. Otherwise, the adsorbed amount may be biased for the species considered. In the case of fluid systems, the simulation boxes must be cubic.

When fluids are simulated at constant pressure, the initial box length has no influence on the final result, provided the liquid phase(s) converge(s) toward the right density. In order to satisfy this condition, the initial liquid density must be sufficient to make the molecules feel the attractive interactions that are responsible for the liquid cohesion. Typically, a density half of the expected liquid density is often well adapted, as it avoids excessive overlaps between molecules and yet corresponds to a significant attractive energy. The relationship between the fluid density ρ (kg/m^3) and the box length L (m) is:

$$\frac{N M_w}{N_A L^3} = \rho$$

where N_A is Avogadro's number and M_w is the average molecular weight (kg/mol). As a consequence of the possible range for the number of molecules (see above), the typical size of a simulation box is 20 to 40 Å for a liquid phase and 30 to 200 Å for a vapor phase. When phase equilibrium is simulated at imposed global volume, the selection of the initial phase volumes is important because it influences the results (for instance the proportion of each phase).

It is important that the global volume is neither too small (the vapor box would contain few molecules) nor too large (the liquid box would contain few molecules). At low pressure (i.e. a few atmospheres) this leads to configurations where the vapor box size is much larger than the liquid. Also, care must be taken that the initial density of the liquid box is adequate to avoid excessive overlaps, i.e. approximately the half of the expected equilibrium density.

7 MedeA LAMMPS

7.1 LAMMPS in General

Should I use a minimization stage prior to my simulation run? Why?

The presence of large gradients may cause LAMMPS to lose atoms when attempting to run MD.

To avoid this situation, it is generally a good idea to include a short minimization stage at the beginning of the job (say 1000 steps and, if periodic, without relaxing the cell). There are occasional cases in which large forces may be present, so including a short minimization stage in the beginning is always a good precaution and rarely does any harm (the idea is to keep it short to remove overlaps, etc without otherwise changing the structure).

DOCUMENTATION

Should I be careful when using PPPM to compute long-range electrostatic interactions?

PPPM invokes a particle-particle particle-mesh Ewald solver which maps atom charge to a 3d mesh, used 3d FFTs to solve Poisson's equation on the mesh, then interpolates electric fields on the mesh points back to the atoms. The cost of traditional Ewald summation scales as $N^{3/2}$ where N is the number of atoms in the system. The PPPM solver scales as *N log(N)* due to the FFTs, so it is almost always a faster choice.

However, running an initial NPT stage on an unequilibrated system with PPPM long-range electrostatics in effect often leads to failure due to losing track of atoms. Therefore, it is recommended that any equilibration NPT stage always be performed using atom based cutoffs with the default 9.5 Å distance. After the initial density adjustment has occurred, switching to PPPM should be fine for the subsequent production and analysis stages.

Why do we need Set Cell to "rho_calc"?

This is related to the above point: the instantaneous position of a trajectory, which defines the instantaneous temperature, pressure and other properties, varies a great deal from step to step. However, the averages converge with time to fixed values. At the end of a NPT trajectory the pressure, volume and temperature are the last instantaneous values, which fluctuate around the average value. Rather than taking whatever the last value for the density or volume of the cell happens to be, we use Set Cell to set the density to the average values, using the variable `rho_calc`, which is set automatically to the average value from the previous NPT stage. Equivalently, we could set the volume to `V_calc`. For each stage, the properties listed in the table in the output, with an `_calc` appended have the average value. The uncertainty has `_uncertainty_calc` appended. Finally, for properties that are not averaged, such as the time, the last value is in the variable with `_last_calc` appended to the name of the property. For more information refer to the chapter on Flowcharts.

7.2 Green-Kubo methods for thermal conductivity

Why are my results different than the example in the documentation?

A dynamics trajectory is very sensitive to small perturbations. Indeed, if slightly perturbed it diverges exponentially from the original trajectory. However, due to the "ergodic theorem", for a given set of conditions the trajectory sweeps out a certain volume of phase space, so after a time it will be back close to any point it has previously visited. Thus, though the instantaneous behavior may appear different, long time averages will converge to the same value. Since dynamics typically starts the velocities random numbers, and the round-off errors are different on different machines, calculations that appear to be identical may produce different answers. However, the averages should be the same within the statistical error bars. If not, either at least one of the simulations is too short, or something is truly different between them.

Why is *lambda_avg* not the average of *lambda_x, lambda_y, lambda_z*?

That is because the autocorrelation functions are averaged and then fitted. Typically the fit comes out to a different time than each x, y and z fit, so the result is not simply the average of the three directions.

Can the Green-Kubo method be used on systems with non-zero net charges?

Green-Kubo method for the calculation of the thermal conductivity, will work for systems without charges. At the moment, it is not known how to implement the *Green-Kubo* method for systems with charges in a satisfactory way, since it is not obvious how to efficiently calculate the per-atom energies needed using a long-range summation method like *Ewald* sums or *PPPM*. And, in general, using atom-based cutoffs, which is the only cutoff method implemented in *LAMMPS*, will not work well.

DOCUMENTATION

How can we check if the calculated thermal conductivity is correct and has converged?

The computation itself gives many clues to this. In the Green-Kubo method, we calculate the thermal conductivity in the x, y and z directions and then average these to get the best answer. If we examine the output, we can look at the x, y and z components of the thermal conductivity. If the system is isotropic, the thermal conductivity in the x, y and z directions should be the same. If they are not the same, the run might not be long enough. A longer run will slowly clean this up, giving more confidence in the result. Empirically, however, we have found that while the individual components vary with the run length, the average is not very sensitive, so the final results is probably better than we might think.

The code also provides another check. The thermal conductivity is obtained form the integral of the autocorrelation function of the heat flux. The job plots the autocorrelation of the heat flux, and the integral, in files that are in the last task directory under the job itself. Clicking on the appropriate link in the main page for the job let's you see the files in the subdirectory corresponding to the task. The plots are in `JJt_fit_average.gif`, `JJt_fit_x.gif`, etc. Clicking on the as plain text link on the right of the file name will bring the picture up in your browser (clicking on the file name itself brings up the data of the picture, which is probably **not** what you want!). The black line is the autocorrelation function of the heat flux (left axis), the blue is the numerical integral of the heat flux, alcreated converted to thermal conductivity (right axes), and the red line is the fit that the program made to the integral. We can see from this graph the extent of the convergence of the curves to a final value at large times and the quality of the fit.

Finally, where possible check the results with experimental results for similar systems. Remember that there are two sources of error in the calculation: statistical and convergence errors, which we checked for above, and systematic errors, which are harder to check for. The principle source of systematic error is the forcefield itself. After all, the forcefield is an approximation of, and fit to, physical reality, and is by no means perfect. Thus, a well-converged lengthy simulation with small error bars is only as good as the underlying forcefield. Since forcefields tend to have similar errors for classes of materials, if possible benchmark against experiment for similar compounds.

7.3 Müller-Plathe method for thermal conductivity

What kind of simulation box should be used?

An orthorhombic cell should be used, with the length in one direction (the direction on which the temperature profile is being created) being approximately three times larger than that in the other two directions. This is the most efficient shape for both the dynamics and for the extraction of the thermal conductivity from the heat flux. A cubic cell could also be used, but the cell size should be large enough to allow the creation of a temperature profile in the direction where kinetic energies are being exchanged, which would mean a significant slowing down of the calculation.

Can the Muller-Plathe method be used on systems with non-zero net charges?

The method works for systems with and without electrostatic terms.

Why don't I need to use a thermostat?

Swaps conserve both momentum and kinetic energy, even if the masses of the swapped atoms are not equal. Thus you should not need to thermostat the system. Indeed, thermostatting the system defeats the purpose, since it will alter the temperatures, affecting the temperature profile that we are carefully setting up. It will also affect the amount of heat transferred in an unmeasured way, so thermostatting the system will actually produce invalid results for the thermal conductivity.

How important is the initial system used for the thermal conductivity calculation run?

If the thermal conductivity stage (NVE run) follows a NPT run, then the initial density and temperature of the NVE stage are the ones of the final configuration of the previous NPT run, which differ from the average

DOCUMENTATION

values of density and temperature of that run. The density can be corrected (attain the average value) by using Set Cell and using `rho_calc`. The initial temperature however will be the final temperature of the previous run, which will deviate from the average value. If this deviation is significant, the subsequent NVE run will be at a different temperature than the target. A way to obtain the best possible initial configuration (density and temperature very close to the average density and temperature of the NPT run) is to add an NVT run in between the NPT and the NVE runs. The density can be set (using Set Cell and *rho_calc*) to its average value and a Nose thermostat can be used choosing a high Drag value (i.e. 4). This will dampen the temperature fluctuations so that we end with a temperature close to what we want.

DOCUMENTATION