

MedeA Flowcharts: Design and Automate Simulation Workflows

Contents

- *[Flowcharts Overview](#page-0-0)*
- *[Initialization and Control](#page-2-0)*
- *[Tables and Graphs](#page-3-0)*
- *[General Properties](#page-5-0)*
- *[Methods](#page-5-1)*
- *[Building and Editing](#page-5-2)*
- *[Structures Lists](#page-7-0)*
- *[Forcefield](#page-7-1)*
- *[Analysis](#page-7-2)*
- *[Available Variables](#page-7-3)*

The *MedeA* environment provides graphical flowcharts to support the efficient construction of complex computational protocols. These flowcharts can be easily created to describe and control the flow of calculations, allowing simple and straightforward access to the various computation engines of *MedeA*, such as *LAMMPS*, *GIBBS*, *VASP*, *MOPAC* and *Gaussian*. These different engines may be combined within a single flowchart, so that a VASP optimization of a unit cell, for example, may be employed as a prelude to a larger scale LAMMPS simulation using an embedded atom method (EAM) forcefield. In addition, the structure can be modified within the flowchart using the provided building and editing capabilities.

The backbone of the flowchart infrastructure is the Tcl language (see [https://www.tcl.tk/\)](https://www.tcl.tk/) which can be directly accessed within custom-scripting stages. Once created the flowcharts can be saved as ASCII files allowing them to be reused and distributed.

Although flowcharts are customizable and flexible their use is straightforward. In this section the basic concepts are summarized, allowing for the construction of both simple and complex flowcharts. Additional details for specific tools are provided elsewhere within the *MedeA* User's Guide.

1 Flowcharts Overview

Flowcharts are accessed via the Jobs $>>$ New Job... menu item. This command yields the main flowchart user interface, where flowcharts are constructed, read from disk or previous calculations or saved to the disk.

Flowcharts provide a variety of control capabilities, such as a While Loop, For Loop, Foreach Structure Loop, If and Foreach Loop. This allows for considerable flexibility in the automation of simulations. For example, a scan of possible system temperatures and pressures for molecular dynamics calculations may be achieved with nested Foreach loops.

Additionally, custom Tcl scripting may be included in a flowchart. Key results from simulation stages are available as variables, which can be operated on, tabulated, and also used to control subsequent stages,

allowing the convenient construction of highly automated procedures for selected applications. If more complex manipulation of the results are required, a custom Tcl stage can be used to process the outputs of the current and any previous stages.

Materials Design flowcharts begin with the Start command which is automatically placed on the lefthand panel of the flowchart dialog. Typically after the Start stage, several variables will be set using a Set Variables stage. The stage is added to the flowchart by clicking the Set Variables button on the righthand panel.

In general, stages are added to a flowchart by clicking the appropriate element on the right-hand pane of the flowchart dialog or user interface. Once added to a flowchart, most stages may be edited by either right-clicking that element and selecting Edit , or double-clicking on that stage.

A summary of the functionality and use of specific flowchart elements is provided below.

2 Initialization and Control

This section of the flowchart interface provides the following stages: Subchart, Set Variables, Print Variables, Custom Tcl Script, and loop related commands: For Loop, Foreach Loop, While Loop, Foreach Structure Loop , Catch and If .

As discussed above, every flowchart begins with a Start stage, this is employed as the starting point of execution for the flowchart and is automatically included in the flowchart. The Subchart stage is simply a container that can hold other flowcharts. This provides a convenient way to break up a large more complicated flowchart into simpler parts and provides a simple way to include a previous flowchart as a building block into a more complicated protocol.

The Set Variables stage allows the user to set the value and units of specific variables that will be employed in later stages in the flowchart. For example, the Set Variables stage can be used to set:

These variables will then be used throughout subsequent stages in LAMMPS calculations, which employ T, tstep, and P as default variables for temperature, timestep, and pressure, respectively. This centralization of the variable setting is convenient in exploring the effect of the system variables on simulated properties. Once declared variables can be accessed, following the Tcl paradigm, by adding a $\frac{1}{2}$ sign in front of the variable name, for example, the variable tstep can be accessed with \$tstep.

The Print Variables reports (calculated) variables matching a specific name, including variables created and made accessible by any previous stage. These variables can be printed to tables or added to structure lists.

The Custom Tcl Script allows you to enter, or retrieve from a file Tcl commands (see [https://www.tcl.tk/\)](https://www.tcl.tk/) which may be used to carry out specific actions during the execution of a flowchart. This can be useful in preparing result summaries or in determining whether specific simulation conditions have been met.

The For Loop includes a flowchart executed repeatedly as defined by the control variables and test conditions.

The Foreach Loop and While Loop stages allow for the introduction of control structures within a flowchart. In each case, allowing for the introduction of a complete flowchart to be executed repeatedly until the foreach vector (like $\{T, P\}$) of variables is exhausted, or the while condition is no longer true.

The Foreach Structure Loop takes structures contained in a Structures List or a previous Trajectory and

performs on each of these structures the same calculation protocol as defined in the Foreach Structure Loop flowchart. This stage is part of the *MedeA HT-Launchpad* module.

Stages within the loops are executed either sequentially, meaning that each iteration is completed before the next is started, or simultaneously, where multiple iterations run in parallel. The **simultaneous** operation mode can be switched on by checking Run the different loop iterations simultaneously , specifying the de-

gree of parallel operation by the parameter Maximum number of jobs to submit simultaneously . All looping structures have an option to Catch and ignore errors in the iterations .

The Catch stage allows for more elaborate error handling, that is in case the flowchart runs into an error, it is caught and another flowchart is executed.

The If stage allows execution of a flowchart if a condition is true (then), or an alternative flowchart (else).

3 Tables and Graphs

This section allows defining a Table , insert results as soon as they occur with Add Row , and finally Print the entire table. The flowchart below is just an illustration, leaving out all looping and calculation.

With Table you define a Table by giving it a name and adding columns. Each column has a

- Title: Name of the column
- Units: Results are converted into units chosen. If no unit is specified, the default unit is used.
- Justification: Left, right, or center
- Width: How much "space" on the left and right
- Format: Can be provided in the printf format. For example: $\frac{1}{6}$ #d or $\frac{1}{6}$ #i for integer values, $\frac{1}{6}$ # \pm for float/double values, %#c for characters, %#s for strings, and %#x for hexadecimal values.

Hint: The maximum number of characters/digits that can be used to print a given variable can be defined by an integer value # between the % sign and the character. In the case of a float value two numbers, separated by a comma, can be used. In this case, the second number defines the number of digits after the comma used to print the value. For example, $88.2f$ displays a floating-point variable with up to 8 characters in total and two digits after the comma.

Table: Add Row adds a line of results to a specified table. Variables can be added, following the Tcl paradigm, by adding a $\frac{1}{2}$ sign in front of the variables name, e.g., the variable Etotal calc can be printed with \$Etotal calc into a table. See *[Available Variables](#page-7-4)* for a list of available variables that can be added.

Table: Print the selected Table. You can also save or append the result to a file as formatted text, comma separated values (csv), tab-delimited, or delimited by a different Separator like ; .

4 General Properties

Hill-Walpole bounds : applies Hill-Walpole statistics on top of Mechanical Properties (see the section on Hill-Walpole bounds for amorphous systems).

Mechanical Properties determines the mechanical properties with LAMMPS or VASP based on given Strains . For more information refer to the chapter on the Theory of elasticity.

Deformation performs deformation simulations beyond the elastic regime with LAMMPS or VASP. For more information refer to the MedeA Deformation section.

Effective Mass : uses VASP to determine the effective mass for a specified k-point. For a more detailed description of the use of this stage see the section on Accurate Effective Mass calculations.

5 Methods

The Methods section of the flowchart interface provides access to a computational engine such as GIBBS, LAMMPS, MOPAC, GAUSSIAN or VASP (VASP flowcharts or a stage accessing the VASP GUI otherwise accessible from the Tools menu). In each case, the interface provided allows you to start a new flowchart and execute relevant specific commands within this new flowchart.

In addition, there are methods available making use of the computational engines above to calculate vibrational properties, phonon dispersion, and phonon density of states (Phonon), to optimize cluster expansions for alloys and other disordered systems and perform Monte Carlo simulations for larger ensembles (UNCLE), to predict properties of polymers using correlations (P3C) or group contributions (QSPR), and to optimize force field parameters by fitting to *ab initio* data (Forcefield Optimizer and MLP Generator).

Additional information on each of the Methods available in this section of the flowchart interface is available in the section of the **User's Guide** dealing with each method.

6 Building and Editing

Flowchart building commands allow for the construction and adjustment of atomic models.

6.1 Set Cell

The Set Cell stage permits the adjustment of unit cell dimensions for periodic systems. Here a variety of options are supported, the density or volume may be specified, or an expansion factor applied to the current system, or specific unit cell dimensions set. This stage may be combined with an appropriate Foreach Loop to explore the volumetric or density-specific behavior of a given system property.

6.2 Change periodicity

A stage to turn on or remove periodic boundary conditions. In this stage the specify a gap to leave between the periodic building blocks. This stage may be used to combine a non-periodic calculation, for example with MOPAC, with a periodic VASP calculation.

6.3 Supercell

The Supercell command may be applied to increase the size of a given model. As this building process is executed on the JobServer this is an efficient mechanism to create large systems.

6.4 Amorphous builder

The Amorphous Builder stage can be used to create amorphous models. *System composition source* can be set to the current system , where the current structure is split into molecules and recombined, to composition files saved on the local machine, and to a composition used in a previous job with jobserver . The *system geometry* of the amorphous material can be either bulk cell or layer and is build according to the set Temperature and cell dimensions. The latter is set with *specify cell* and can be a mixture of the specific cell dimensions a, b, and c in combination with density. Accordingly, additional options become available. The number of mols of the defined composition is set with *Nmols*. As this is a stochastic process the *number of configurations* that is to be created can be set. Multiple amorphous structures can be used to determine the average property of a given model. See the section on Amorphous Materials Builder for more information.

6.5 Thermoset Builder

The Thermoset Builder creates a single configuration of a densely crosslinked thermoset structure from an equilibrated bulk amorphous system. See the section on Thermoset Builder for more information.

6.6 Translate Atoms

The Translate Atoms command can be used to adjust the position of atoms within the current system. Thereby it is possible to explicitly select atoms of the active structure or to select a subset . Atoms can be translated by a translation vector or to a point. The displacement can be in fractional or cartesian coordinates.

6.7 Docking

The Docking stage combines two structures. The host structure is the 'stream' structure upon which the Flowchart is operating. The Guest from Job is specified as the final.sci of the specified job on the current Job Server, the Host is specified as the input structure submitted with the flowchart. Number of guests defines how many guest molecules are put into the host, for each guest this process is repeated for Maximum iterations . The Maximum displacement (Ang) should be larger than the diameter of any ring to avoid the interlocking of molecules. Scale rotations by and Temperature (K), as well as the underlying theory, is explained in the section on Docking.

6.8 Randomly Substitute Atoms

Replaces a defined number of atoms, accounting for the symmetry of the structure, of *element A* with either *vacancies* or atoms of *element B*. For more information refer to the section on Random Substitution.

6.9 Simple Dynamics & Minimization

This stage uses the same simple dynamics and minimization algorithm that is also accessible from the MedeA GUI. The Number of dynamics steps and the Number of minimization steps can be set separately. Use this stage to pre-optimize a structure before applying a computationally more demanding algorithm to it.

6.10 Subset Manager

Manipulate existing subsets in a structure.

7 Structures Lists

These are a collection of commands to handle Structures Lists:

New List creates a new structure list that is accessible by the current flowchart. The List name and the File name of the structure lists can be defined. If no subfolder is specified the list will be placed in the job folder. The list can be initialized from an initial list that is saved locally on the machine running the MedeA GUI. Note, that multiple structure lists can be active at the same time.

Save to List adds the current structure to any of the active structure lists. In addition, the structure can be saved with multiple properties.

Extract from List will extract a structure, specified by Structure index and Configuration index , from any of the active lists.

Sort List will sort any of the active structure lists according to a specific property.

Remove Duplicates from List remove all duplicates from a given list. Duplicates are identified by making use of symmetry.

Compute Descriptors on a list define and compute descriptors on a list.

Apply QSAR model on a List by loading the equation from an XML QSAR model file.

For more details, please see the MedeA High-Throughput section.

Documentation HT

8 Forcefield

Set Forcefield allows you to specify a forcefield.

Assign Atom Types and Charges allows you to readjust or set atomic forcefield parameters. This requires a forcefield with auto-typing.

9 Analysis

Orientation determines the orientation of a subset to a given reference vector: $(x, y, or z)$.

10 Available Variables

Properties listed below are made available in the flowchart after a specified stage as **variables**. To access the value of any of these variables use a $\frac{1}{2}$ sign in front of the property name. For example, the value of the variable Etotal_calc can be accessed with \$Etotal_calc.

10.1 Structure-Specific Properties

Different properties of structures (systems) are available through variables without performing any calculation. The name of the active system can be accessed with \$system_name.

Table1: System-specific variables

Values of the properties listed in the below table can change after calculations, i.e. can be different prior and after calculations are performed. The syntax to access the value of any of the properties is [\$system get property], whereby [and] is the Tcl syntax to declare a string in command (see [https://www.tcl.tk/\)](https://www.tcl.tk/), and Property should be replaced by one of the properties in the below table. For example, the Tcl command [\$system get density] returns the value 1.2345 kg/m^o3, i.e. a number together with a unit. The Tcl command [\$system get cell.a] returns only a single value without a unit but the value is in the unit of \AA .

10 of [28](#page-27-0)

forcefield to which the beads of the | tures (periodic and non-periodic)
V. 3.8 Materials Design Copyright @ 2024 Materials Design, Inc. All rights reserved.
Materials Design Sara Media Media (Fig. Reference of Materials D

mesoscale structure belong

10.2 Structure Lists

Values saved with structures in a structure list can be accessed for each entry in that list using the For Each Structure stage. Within that loop variables are accessed for the active structure with \$system_properties(NAME) where NAME defines the name of the desired and available property. So, if the variable T has been saved with the structure then its value can be fetched with \$system properties(T).

10.3 *MedeA* **LAMMPS**

Table3: Output from Single Point Energy and Single Point Forces stage

When performing a molecular dynamics simulation the mean value of a property is returned with {property} calc and its uncertainty with {property} uncertainty calc. The value of the uncertainty is given in the same unit as the calculated property. In addition, {property} converged calc is returned with the value of 1 if the calculated property has converged during the simulation otherwise it is returned with the value of 0.

Table 4 – continued from previous page

Additional properties available when using *LAMMPS* property modules.

Table5: Cohesive Energy Density

Table6: Thermal Conductivity

Table7: Viscosity

Table8: Diffusion

Table9: Surface Tension

Table10: MD Phonon

10.4 *MedeA VASP*

Different properties are calculated depending on the type of VASP calculation configured within the VASP stage. Properties listed in the first table are always available after a successful VASP calculation.

Table11: All calculation types (last structure)

Table 11 – continued from previous page

Table12: Single Point and Structure Optimization (optimized structure)

Table13: Ab initio Molecular Dynamics

Table14: (Total, valence) charge density, Bader analysis

Property	Unit	Description
BaderCharges_calc		electron charge. Bader charge per atom for each site
BaderChargeTransfers_calc		electron charge. Charge transfer per atom for each
		site
BaderVolumes calc	Å ³	Bader volume per atom for each site
BaderDistances calc	А	Bader distance for each site
BaderVacuumCharge_calc		electron charge. Bader charge of the vacuum region
BaderVacuumVolume calc	\AA^3	Bader volume of the vacuum region
BaderTotalVolume_calc	\AA^3	Total Bader volume

Table15: Density of states and Optical spectra

Table16: Optical spectra

Table 16 – continued from previous page

Table17: Zone center phonons

Table18: Response tensors

Table 18 – continued from previous page

Table19: NMR chemical shifts

Table20: NMR chemical shifts

Table21: Electric field gradients

Table23: Work function (surfaces only)

Table24: Work function (surfaces only) - additional properties resulting from dipole correction

Table25: Energy of formation

10.5 *MedeA* **Mechanical Properties / MT - Elastic Constants.**

This applies to the Mechanical Properties stage and the MT - Elastic constants type of calculation in the VASP 5.4 and VASP 6 stage

Table 26 – continued from previous page

Table 26 – continued from previous page

10.6 *MedeA* **Phonon (Flowchart)**

This applies to the Phonon stage, but not to the Phonon GUI (from the Tools menu), which cannot be integrated into a Flowchart

10.7 *MedeA* **GIBBS**

Table27: All calculation types

Table28: NVT & NPT with test insertions, GEMC, GCMC

Table29: GEMC constant V, pure compounds

10.8 *MedeA* **MOPAC**

Property Unit Description \AA^2 Molecule area DipoleMoment_calc Dipole moment Ef calc kJ/mol Standard heat of formation, at 298 K Ehomo_calc eV HOMO energy Elumo_calc eV LUMO energy Emopac calc eV Total electronic energy EmpiricalFormula_calc Chemical empirical formula IP_calc eV Ionic potential PointGroup_calc Symmetry point group $Vcosmo_calc$ A^3 Molecule volume

Table30: Structure and Energy, available after any MOPAC calculation

Table31: Property accessible after a Thermodynamics calculation

10.9 *MedeA* **Gaussian**

Table32: Single point energy and geometry optimization

If the calculation uses a post-Hartree-Fock method then an appropriate subset of the following total energies

and corrections will also be available, where a correction indicates the contribution to the total energy from the specified level of theory only. All energies have units of Hartree.

Table33: Energies

If the calculation uses either Hartree-Fock or density functional theory then the dipole and quadrupole moments are also available. The dipole moment has units of e*a0 and the quadrupole moment has units of e^* a0².

If the calculation method has analytic derivatives, which may not be the case for higher-order Moller-Plesset or coupled-cluster calculations, a frequencies calculation also produces a dipole and quadrupole moment. The dipole moment has units of e^* a0 and the quadrupole moment has units of e^* a0².

If the calculation method has analytic second derivatives (and hence analytic frequencies - HF, DFT, and MP2), which includes the HF, DFT and MP2 methods, a frequencies calculation also yields the polarizability

and hyperpolarizability (both in atomic units) and APT charges.

Table37: Polarizabilities

Polarizability

Variables available after a frequencies calculation include all of those from a commensurate single point energy calculation and the dipole moment and polarizability. The dipole moment has units of e*a0 and the polarizability is in atomic units.

Table38: Dipole moments

If the calculation method is either Hartree-Fock or density functional theory then a polarizability calculation also produces the hyperpolarizability in atomic units.

Table39: Polarizabilities

10.10 *Machine-Learned Potential Generator*

Different values are made available depending on the type of machine learning potential (MLP) set in the **MLP Generator** stage. This can be either **SNAP** or **NNP**. Properties listed in the first table are always available after the **MLP Generator** stage. The parameters and properties in the second table are available after a **SNAP** fit and those in the third table after an **NNP** fit.

Table40: Properties made available independent of the MLP type.

Table41: Additional parameters made available when using SNAP.

Table42: Additional properties and parameters made available when using NNP.

10.11 Additional Notes on Flowcharts

- The final structure obtained in the previous stage is passed to each subsequent stage. Hence, the endpoint of an NPT calculation, for example, is passed to the next stage in a simulation.
- The principal computed results are placed in local variables accessible by subsequent stages. The names of these variables are formed from the name of the property, with \lrcorner calc, \lrcorner uncertainty calc and converged calc appended in the case of averages.
- The input and output files associated with each stage in a particular Job are stored on the JobServer. The directory hierarchy employed beneath each Job reflects the structure of the flowchart, which was employed in that calculation.