Simulation of Molecules in LAMMPS

David Keffer
Department of Materials Science & Engineering
University of Tennessee, Knoxville

date begun: February 22, 2016
date last updated: February 23, 2016

Table of Contents

I. Purpose of Document

The purpose of this document is to provide a practical introduction to the simulation of molecules in LAMMPS. This is particularly challenging since there is significant work that must be done to create the initial configuration, including positions and connectivity, as well as the potential file. This must be done outside LAMMPS. There are a variety of tools available to achieve this end and this example demonstrates the use of one such tool.

II. Intramolecular and Intermolecular Potentials

The simulation of a "molecule" is different from the simulation of a Lennard-Jones fluid or a metal because a molecule contains many different kinds of interaction potentials. One must remember that a classical simulation uses empirical potentials to describe the interaction between atoms or more generally particles. (A particle could also be a "united atom", such as a CH_2 group in an alkane, or a molecule, such as CH_4.)

In general, when describing molecular systems, the interaction potential is divided into intermolecular (between molecules) and intramolecular (within a molecule) components.

$$U_{\text{total}} = U_{\text{inter}} + U_{\text{intra}} \tag{1}$$

II.A. Intermolecular (Non-bonded) Potentials

The proper terminology for an intermolecular potential is a non-bonded potential because one can imagine in a large macromolecule, such as a polymer or a protein, that atoms separated by many bonds, are unaware that they may be part of the same molecules. In its equilibrium state, a polymer may be coiled up. Segments of the polymer separated by long distances along the backbone may come into close proximity with each other. Atoms in these segments will interact through the same non-bonded potential as if they were part of different molecules.

Often the intermolecular or non-bonded potential is separated into a short-range interaction, such as Lennard-Jones, and a long-range interaction, usually Coulombic.

$$U_{\text{inter}} = U_{\text{inter,SR}} + U_{\text{inter,LR}} \tag{2}$$

II.A.1. Short-range (Dispersive) Potentials

The Lennard-Jones potential is an example of "intermolecular potential" because it describes interactions between particles in different molecules. The particles cannot overlap, necessitating a repulsive potential, and they attract due to an dispersive (induced-dipole-induced-dipole) interaction. Explicitly, the total short-range interaction is often expressed as the sum of pairwise interactions between atoms located on different molecules.

$$
\begin{aligned}
U_{\text{inter,SR}} &= \frac{1}{2}\sum_{im=1}^{N_m}\sum_{jm=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja=1}^{N_{a,t(jm)}} U_{\text{inter,SR},t(ia)t(ja)} \\
&= \frac{1}{2}\sum_{im=1}^{N_m}\sum_{jm=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja=1}^{N_{a,t(jm)}} 4\varepsilon_{t(ia)t(ja)}\left[\left(\frac{\sigma_{t(ia)t(ja)}}{r_{ia,im,ja,jm}}\right)^{12}-\left(\frac{\sigma_{t(ia)t(ja)}}{r_{ia,im,ja,jm}}\right)^{6}\right]
\end{aligned}
\tag{3}
$$

In this expression, N_m indicates the number of molecules in the system, which may be of different types. $N_{a,t(im)}$ indicates the number of atoms in a molecule of type *im*. The Lennard-Jones parameters, $\varepsilon_{t(ia)t(ja)}$ and $\sigma_{t(ia)t(ja)}$, are functions of the types of atoms *ia* and *ja*., irrespective of the molecules in which reside. The separation between particles, $r_{ia,im,ja,jm}$, is the distance between atom *ia* on molecule *im* and atom *ja* on molecule *jm*.

II.A.2. Long-range (Coulombic) Potentials

Of course, as we have seen before, this short-range interaction is usually truncated at some cut-off distance and a long tail correction is added. This can still be done. However, the long-range potential to which we refer in this context is the Coulombic potential.

$$
U_{\text{inter,LR}} = \frac{1}{2}\sum_{im=1}^{N_m}\sum_{jm=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja=1}^{N_{a,t(jm)}} U_{\text{inter,LR},t(ia)t(ja)} = \frac{1}{2}\sum_{im=1}^{N_m}\sum_{jm=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja=1}^{N_{a,t(jm)}} \frac{1}{4\pi\varepsilon_o}\frac{q_{ia,im}q_{ja,jm}}{r_{ia,im,ja,jm}}
\tag{4}
$$

In this expression, $q_{ia,im}$ is the charge on atom *ia* of molecule *im* and ε_o is the permittivity of free space. In static charge simulations, the charge on atom *ia* of a given type of molecule is the same for all molecules of the type. In dynamic charge simulations, the charges evolve in time and will be different for every atom in the system, even for those who occupy analogous locations on a molecule.

II.B. Intramolecular (Bonded) Potentials

Intramolecular potential include bonded interactions for atoms whose relative interactions are dominated by the nature of local chemical bonds and non-bonded interactions for atoms, which are separated by many bonds.

$$
U_{\text{intra}} = U_{\text{intra,bonded}} + U_{\text{intra,non-bonded}}
\tag{5}
$$

In this section we focus on the bonded interactions, because the functional form of the intramolecular non-bonded interactions are the same as those for the intermolecular interactions described above.

The intramolecular bonded potential energy is typically composed of four components.

$$U_{\text{intra,bonded}} = \sum_{im=1}^{N_m} \left(U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} \right) \tag{6}$$

We shall describe each of these four components below. It is important to understand that the intramolecular components are empirical in nature and are defined in order to reproduce known geometries of molecules.

II.B.1. Bond (Stretching) Term

The first term in the intramolecular, bonded interaction is the "bond" or "stretching" term. This potential describes the interaction between two atoms connected by a chemical bond. Often, this interaction is described by a Hookean spring,

$$U_{\text{bond}} = \sum_{bonds} \frac{k_{r,t(ia),t(ja)}}{2} \left(r_{ia,ja} - r_{eq,t(ia),t(ja)} \right)^2 \tag{7}$$

where the force constant and the equilibrium bond distance, $k_{r,t(ia),t(ja)}$ and $r_{eq,t(ia),t(ja)}$, are defined for a bond connecting atoms of type *ia* and type *ja*. In other words the force constant and equilibrium separation for a carbon-carbon bond are different than that of a carbon-hydrogen bond. Besides type, the only thing that matters in this bond is the distance between atoms, $r_{ia,ja}$. Therefore, this potential describes bond stretching between adjacent atoms.

Note that some potentials may absorb the factor of two into the force constant. Therefore, one must always read the original papers carefully to make sure that one understands the functional form into which the parameters are being inserted.

It is known that the use of a parabola to describe the energy of a chemical bond is an approximation. Compression is typically more difficult than expansion. Consequently there is an asymmetry in the energy that cannot be captured with a second order polynomial. If you are investigating a phenomenon in which this asymmetry is thought to play an important role, then you must use a potential that includes this asymmetry. There are a few examples of third-order polynomials used to describe bond stretching. One can also use tabulated data from first principles calculations.

II.B.2. Angle (Bending) Term

The second term in the intramolecular, bonded interaction is the "angle" or "bending" term. This potential describes the interaction between three atoms connected in sequence by two chemical bonds. Often, this interaction is described by a Hookean spring applied to the angle defined by the three atoms, $\theta_{ia,ja,ka}$, rather than distance between them.

$$U_{\text{angle}} = \sum_{angles} \frac{k_{\theta,t(ia),t(ja),t(ka)}}{2} \left(\theta_{ia,ja,ka} - \theta_{eq,t(ia),t(ja),t(ka)}\right)^2 \tag{8}$$

where the force constant and the equilibrium bond distance, $k_{\theta,t(ia),t(ja),t(ka)}$ and $\theta_{eq,t(ia),t(ja),t(ka)}$, are defined for an angle of a series of atoms of type *ia*, type *ja* and type *ka*. In other words the force constant and equilibrium separation for a hydrogen-carbon-hydrogen bond are different in methane ($\theta_{eq,HCH} \approx 109.47^\circ$) than that of an oxygen-carbon-oxygen bond in carbon dioxide, ($\theta_{eq,OCO} = 180^\circ$). Besides type, the only thing that matters in this bond is the angle, $\theta_{ia,ja,ka}$, defined by the three atomic positions,

$$\cos\left(\theta_{ia,ja,ka}\right) = \frac{\left(r_{ia} - r_{ja}\right)\cdot\left(r_{ka} - r_{ja}\right)}{\left|r_{ia} - r_{ja}\right|\left|r_{ka} - r_{ja}\right|} \tag{9}$$

Therefore, when one derives the force for this potential, one simply invokes the chain rule for differentiation,

$$F_{ia,\alpha} \equiv -\frac{\partial U}{\partial r_{ia,\alpha}} = -\frac{\partial U}{\partial \theta_{ia,ja,ka}} \frac{\partial \theta_{ia,ja,ka}}{\partial r_{ia,\alpha}} \tag{10}$$

There are many variations in functional form for the angle term. Again, the factor of one half may be absorbed in the force constant. The spring may be a function not of $\theta_{ia,ja,ka}$ but of $\cos\left(\theta_{ia,ja,ka}\right)$. A clear understanding of the potential from the literature is required.

The functional form of a Hookean spring is an approximation. There are examples in the literature of potentials that include a "stretch-bend" terms, in which corrections to the angle energy are made on the basis of the lengths of the two bonds involved in defining the angle. Such terms may be required for the explicit reproduction of spectroscopic signatures.

II.B.3. Dihedral (Torsion) Term

The third term in the intramolecular, bonded interaction is the "dihedral" or "torsion" term. This potential describes the interaction between four atoms connected in sequence by three chemical bonds. This interaction potential is needed to capture the difference between responsible for the preference of the gauche conformation over the eclipsed conformation in, for example, butane. Figure 1 illustrates the dihedral angle.

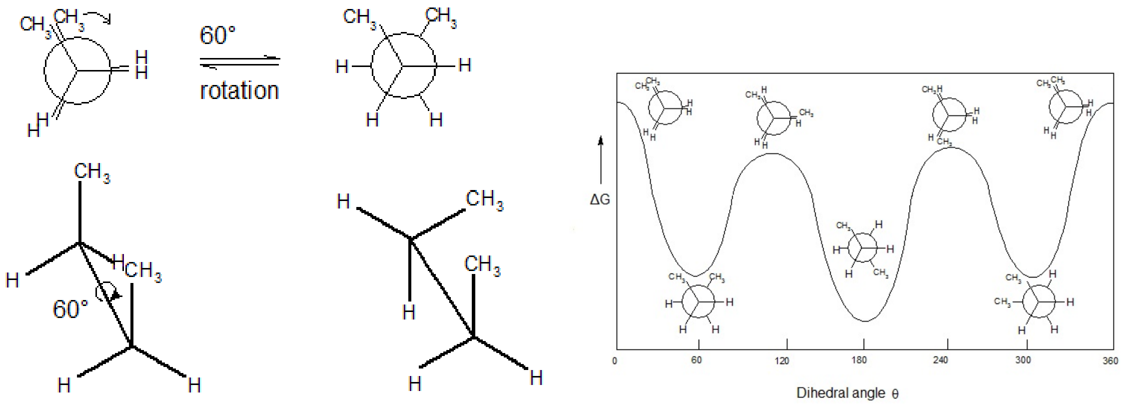

Figure 1. Dihedral torsion in butane. Plots from https://en.wikipedia.org/wiki/Conformational_isomerism .

Often, this interaction is described by a combination of trigonometric functions, based on the torsion angle, $\phi_{ia,ja,ka,la}$, defined by the positions of atoms, ia, ja, ka and la. One such example is

$$U_{dihedral} = \sum_{dihedrals} \begin{pmatrix} \frac{V_1}{2}\left[1+\cos(\phi-\phi_1)\right]+\frac{V_2}{2}\left[1-\cos(2\phi-\phi_2)\right] \\ +\frac{V_3}{2}\left[1+\cos(3\phi-\phi_3)\right]+\frac{V_4}{2}\left[1-\cos(4\phi-\phi_4)\right] \end{pmatrix} \tag{11}$$

Note that we dropped the subscripts on ϕ and on the eight parameters, V_1, V_2, V_3, V_4 and $\phi_1, \phi_2, \phi_3, \phi_4$. Note also that depending on the symmetry, many of these parameters may be zero.

In the figure above, we used an example of a carbon with sp^3 bond hybridization, resulting in a nominally tetrahedral geometry. Of course, dihedral torsion also applies to other systems. For example, in ethylene, H_2CCH_2, there is a double bond between the two carbon atoms, also known as sp^2 bond hybridization. In this case, the equilibrium torsion angle between opposite hydrogen is 0° or 180° degrees, resulting in a planar molecule. A torsion angle for a system with a triple bond, such as acetylene, HCCH, with sp bond hybridization, is undefined.

Aside from pathological cases, the dihedral angle, $\phi_{ia,ja,ka,la}$, is defined by the positions of the atoms defining the angle and is compute by calculating the angle of the $\left(r_{ia}-r_{ja}\right)$ and $\left(r_{la}-r_{ka}\right)$ vectors projected onto a plane with a normal vector defined by $\left(r_{ja}-r_{ka}\right)$.

Again, there are many variations in functional form for the dihedral term. A clear understanding of the potential from the literature is required.

II.B.4. Improper (Torsion) Term

The fourth term in the intramolecular, bonded interaction is the "improper" or "improper torsion" term. This potential describes the interaction between four atoms, in which atoms three of the four atoms are connected to the fourth three chemical bonds. This interaction potential is typically used to induce planarity in molecules such as formaldehyde, CH_2O, in which there is sp^2 bond hybridization of the central carbon. One might note that specifying the angles correctly (to 180°) all combinations of HCH, OCH_1 and OCH_2, should be sufficient to ensure planarity and this is true at 0 K. At finite temperature, one finds that the force constants for the angle potential developed for sp^3 carbon don't apply well to angles around sp^2 carbon, so an extra term is needed to drive the system to planarity. The improper torsion angle is defined as the angle between two planes. The first plane is defined by the i, j and k atoms and the second plane is defined by the j, k and l atoms. One example of a corresponding functional form for the improper torsion is

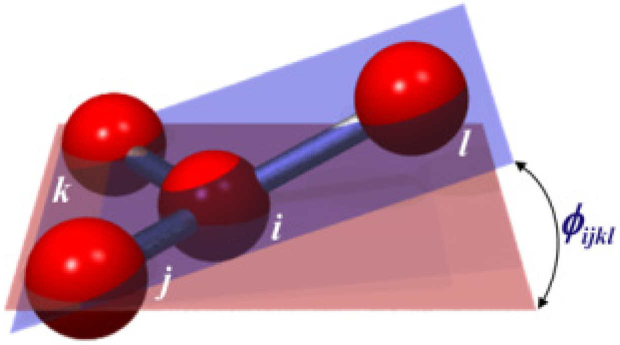

Figure 2. Improper torsion angle. Figure borrowed from http://cbio.bmt.tue.nl/pumma/index.php/Theory/Potentials .

$$U_{improper} = \sum_{improper} \left(K\left[1 + d\cos(n\phi)\right] \right) \tag{12}$$

This form looks like one term of the dihedral potential in which the equilibrium angle is set to zero.

II.B.5. Intramolecular Non-bonded Term

As noted above, the intramolecular non-bonded term has the same functional form as the intermolecular non-bonded term.

$$U_{\text{intra,non-bond}} = U_{\text{intra,non-bond,SR}} + U_{\text{intra,non-bond,LR}} \tag{13}$$

$$U_{\text{intra,non-bond,SR}} = \frac{1}{2}\sum_{im=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja^*}^{N_{a,t(im)}} 4\varepsilon_{t(ia)t(ja)}\left[\left(\frac{\sigma_{t(ia)t(ja)}}{r_{ia,im,ja,im}}\right)^{12} - \left(\frac{\sigma_{t(ia)t(ja)}}{r_{ia,im,ja,im}}\right)^{6}\right] \tag{14}$$

$$U_{\text{intra,non-bond,LR}} = \frac{1}{2}\sum_{im=1}^{N_m}\sum_{ia=1}^{N_{a,t(im)}}\sum_{ja^*}^{N_{a,t(im)}} \frac{1}{4\pi\varepsilon_o}\frac{q_{ia,im}q_{ja,im}}{r_{ia,im,ja,im}} \tag{15}$$

The only remaining uncertainty is to resolve how many bonds must separate atoms within a molecule before they interact via the non-bonded potential. To my knowledge there are three options. First, atoms that interact via bond, angle or torsion interactions, do not interact via non-bonded interactions; this means atoms separated by four or more bonds interact via non-bonded interactions. Second, atoms that interact via bond or angle interactions, do not interact via non-bonded interactions; this means atoms separated by three or more bonds interact via non-bonded interactions. This also means that the outermost atoms in a dihedral angle interact via both the dihedral potential and the non-bonded potential. The third common option is the same as the second except that atoms that are separated by three bonds, and therefore interact via the dihedral potential, interact via the non-bonded potential with a weakened factor, usually 0.5. LAMMPS can allow any of these rules.

II.C. Transferability

The dream of interaction potentials is broad transferability. In other words, the potential parameters are developed by optimization for a given parameter set based on a set of compounds and conditions. The hope is that this potential can also be used in other conditions and for other compounds. There are thus (at least) two kinds of transferability. The first corresponds to thermodynamic state. A transferable potential can be used to describe the system in a given range of temperature, pressure and composition. Certainly, potentials developed to describe molecular geometries at ambient temperatures will not function well at temperatures where the molecule decomposes. Thus there is a range where these potentials function. In terms of composition, transferability implies that the interaction potential between molecules of type A and A will not change when mixed with molecules of type B. It is true that one will now need a B-B potential and an A-B potential, but, for a transferable potential, the A-A potential doesn't change.

A transferrable potential should also work for similar compounds. In other words, a potential developed to describe bonds, angles torsions and non-bonded interactions in short alkanes should work for long alkanes as well. While this discussion is brought up in the section on intramolecular potentials, it also applies to atomic potentials. A potential parameterized to describe crystalline silicon, Si, likely does not work well as the Si-Si component in a potential describing silica, SiO_2.

II.D. Examples

Over the years, many intramolecular potentials have been introduced. Some of the commonly used ones include,

- OPLS-aa (all atom)

  Title: Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids
  Author: William L. Jorgensen , David S. Maxwell , and Julian Tirado-Rives
  Journal: J. Am. Chem. Soc., 1996, 118 (45), pp 11225–11236
  DOI: 10.1021/ja9621760

- OPLS-ua (united atom)

  Title: The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin
  Author: William L. Jorgensen, Julian. Tirado-Rives
  Journal: J. Am. Chem. Soc., 1988, 110 (6), pp 1657–1666
  DOI: 10.1021/ja00214a001

- TraPPE (united atom)

  Title: Transferable potentials for phase equilibria. 1. United-atom description of n-alkanes

Author: M.G. Martin, and J.I. Siepmann
Journal: J. Phys. Chem. B, 102, 2569-2577 (1998).

- TraPPE (all atom)

  Title: Transferable potentials for phase equilibria. 3. Explicit-hydrogen description of n-alkanes
  Author: B. Chen, and J.I. Siepmann
  Journal: J. Phys. Chem. B, 103, 5370-5379 (1999).

- CHARMM

  Title: Combined Ab initio/Empirical Approach for the Optimization of Lennard-Jones Parameters
  Authors: Yin, D. and MacKerell, Jr., A.D.
  Journal: Journal of Computational Chemistry 19: 334-348, 1998.

- UFF: Universal Force Field

  Title: UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations
  Author: A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. GoddardIII, W. M. Skiff
  Journal: J. Am. Chem. Soc., 1992, 114 (25), pp 10024–10035
  DOI: 10.1021/ja00051a040

III. Procedure

There is a set of tasks that must be accomplished before running a simulation of a molecular system. In this section, we perform this sequence of tasks for a simple example, ethane. With the exception of the first step in the sequence, the tools that we use in this example for ethane, will work for arbitrarily complex molecules.

There are numerous tools provided by a variety of individuals and groups to accomplish the tasks described below. In this example, I use the tools that I am most familiar with. Certainly, there are other tools by which this same task could be achieved.

The sequence of steps is as follows.

Step 1. Find or build a reasonable model of the isolated molecule.

Step 2. Build a complete input configuration for the isolated molecule.

Step 3. Build a LAMMPS input configuration for the isolated molecule.

Step 4. Build a LAMMPS input file with the appropriate potential.

Step 5. Build a LAMMPS input file with a system of molecules.

Step 6. Run LAMMPS simulation.

III.A. Step 1. Find or build a reasonable model of the isolated molecule.

To begin the process, you must find or build a model of your isolated molecule in a conformation that is close to its equilibrium state. For common molecules, the recommended path is to find a model. The NIST Chemistry webbook maintains structures for tens of thousands of common compounds. The relevant URL is http://webbook.nist.gov/chemistry/ .

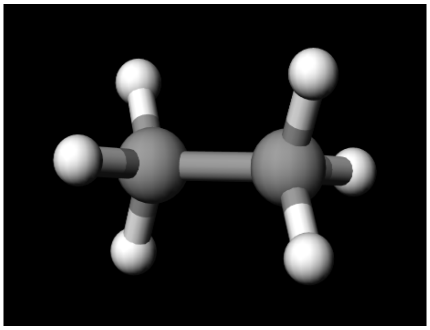

Figure 3. Ethane. Screen capture from http://webbook.nist.gov/cgi/cbook.cgi?Str3View=C74840&Type=JSmol .

The NIST chemistry webbook is set up with JSMol visualizations of the molecules. For ethane, a screen capture of this structure is shown in Figure 3. From NIST, you can also download the molecular structure in `sdf` format. The contents of this file is shown in Appendix A. This file contains both the coordinates of all of the atoms in ethane as well as all of the connectivity (bonding) information. This is the essential information necessary to begin the process.

If you have a very uncommon molecule, it may not appear in the NIST data base. In that case, you have to build the molecule yourself, using some molecule building software. This very practical activity is, due to time limitations, beyond the scope of these notes.

III.B. Step 2. Build a complete input configuration for the isolated molecule.

As noted above there are numerous computational tools available on the internet to build a LAMMPS input configuration. In this example, I use one of my own creation, which requires as a starting point a configuration file in `pdb` (Protein Database) format with both atomic coordinates and connectivity information. There is no special value to the `pdb` format other than the fact that it is common and that it includes connectivity information as the `xyz` file does not.

Software for converting between file formats is also relatively common. In this example, I follow a relatively convoluted path in order to make a point. First, I am going to assume that we don't have the file in the `sdf` format provided by NIST. The reason for this is because the `sdf` file already contains connectivity information, but sometimes we may only have `xyz` coordinates and we may need to build the connectivity ourselves. So, for the purposes of this example, I converted the `sdf` file to an `xyz` file, using an excel spreadsheet, posted on the website. The resulting `xyz` file is provided in Appendix B.

I converted the `xyz` file to a `pdb` file using the program `make_pdb_v02.f`, which is provided on the website. This code created a standard `pdb` file located in Appendix C. Note that this file includes a column with the number of bonds per atom and the charge of each atom. This same code also created a modified `pdb` file located in Appendix D, which contains an additional column with an "1" if the atom is the center of an improper torsion mode. Otherwise it is "0".

The program `make_pdb_v02.f`, requires, in addition to the `xyz` file, three additional input files, (1) `make_pdb_input.txt`, (2) `charge_input.txt` and (3) `improper_input.txt`, which are provided in Appendices E, F and G respectively. In the ethane example, the charges for C and H are set to zero and neither atom is involved in an improper torsion.

The modified `pdb` file located in Appendix D is the input file for the next step.

III.C. Step 3. Build a LAMMPS input configuration for the isolated molecule.

The program, `lammps_input_readpdb.f`, available on the course website converts the modified pdb file into a LAMMPS input configuration file. In addition to the `pdb` file, one additional input files, `intra_input.txt`, is required and is provided in Appendix H.

This file generates two important output files. The first is a configuration file, `config_ethane1.txt`, which is shown in Appendix I. This configuration file contains all of the necessary information including all intramolecular modes for the molecule to be simulated.

The second output file, `intra_output_001.txt`, which is shown in Appendix J, provides the essential information needed to build the potential. Importantly, this file identifies the number and type of each mode. There are two types of atoms (C &H). There are three types of non-bonded interactions (C-C, C-H & H-H). There are two types of stretching (bond) interactions (C-C & C-H). There are two types of bending (angles) interactions (C-C-H and H-C-H). There is one type of dihedral torsion interaction (H-C-C-H). Potential parameters for every one of these modes must be provided in the LAMMPS input file. Therefore this file is crucial to the next step of the procedure.

III.D. Step 4. Build a LAMMPS input file with the appropriate potential.

The next step is to build a LAMMPS input file. This input file, `input_ethane1.txt`, is shown in Appendix K. Important to this example is to observe how the potential is described in the input file. In this example, we are using the OPLS-aa potential. The functional forms of the OPLS-aa potential are mapped onto existing functional forms in LAMMPS as follows:

```
# interaction styles
pair_style lj/cut/coul/cut 12.0
bond_style harmonic
angle_style harmonic
dihedral_style opls
pair_modify mix geometric tail yes

# OPLS considers 1-4 interactions with 50%.
special_bonds lj/coul 0.0 0.0 0.5
```

Numerical values of the parameters for the OPLS-aa potential are inserted into the input file as follows. Here CT is tetrahedral (sp^3) carbon and HC is hydrogen bound to carbon. Note that every type of mode that is identified in `intra_output_001.txt` of Step 3 is explicitly represented here, except the non-bonded CT-H parameter, which uses mixing rules.

```
# force field parameters
# missing nonbonded parameters are inferred from mixing.
pair_coeff   1 1  0.066 3.5      # CT = tetrahedral (sp3) carbon
pair_coeff   2 2  0.03 2.5      # HC = H bound to carbon
bond_coeff   1   268.0 1.529    # CT-CT
bond_coeff   2   340.0 1.09     # CT-HC
angle_coeff  1    37.5 110.7    # CT-CT-HC
angle_coeff  2    33.0 107.8    # HC-CT-HC
```

```
dihedral_coeff 1 0.000 0.000 0.318 0.000 # HC-CT-CT-HC
```

It is worth pausing a moment in the procedure to note that at this point we could run a LAMMPS simulation of an isolated ethane molecule using the LAMMPS input file, `input_ethane1.txt`, of Step 4 (Appendix K) and the initial configuration file, `config_ethane1.txt`, of Step 3 (Appendix I). We will use the LAMMPS input file without any changes for a simulation of a system of many molecules, but the initial configuration file must be modified to include the positions and connectivity of the other molecules.

III.E. Step 5. Build a LAMMPS input file with a system of molecules.

The program, `make_lammps_system_config_v01.f`, available on the website generates a configuration for a system of molecules using only the configuration file for a single molecule. It does so by generating a simple cubic grid and placing the center of mass of each ethane on that grid point. The connectivity information for each molecule is the same, but the atom indices must be shifted. The input file, `input_make_lammps_system_config.txt`, is given in Appendix L. This specifies the number of molecules and the size (thus density) of the system.

The output of this program, `config_ethane_sys.txt`, is shown in part in Appendix M.

III.F. Step 6. Run LAMMPS simulation.

We can now run a LAMMPS simulation of a system of ethane molecule using the LAMMPS input file, `input_ethane1.txt`, of Step 4 (Appendix K) and the initial configuration file, `config_ethane_sys.txt`, of Step 5 (Appendix M).

If the system is dilute, (a gas) and the molecules are spaced far enough apart that we have no overlap between atoms, then we are done. However, most of the time, we are interested in condensed phases. Therefore, the molecules are not space far enough apart to avoid overlap and if the simulation is started using this initial configuration, it will immediately blow up.

Generating stable initial configurations of condensed liquid or disordered solid (e.g. polymer) phases is a subject unto itself. There are a variety of choses for generating a dense configuration. One simple suggestion is to start with this dilute system and use a barostat to drive the system to the desired density. The practical problem with this approach is that it usually takes too long for the barostat to work.

Therefore we turn to computationally less expensive approaches. One way is to rely on the "energy minimization feature of LAMMPS to push the molecules away from their overlapping configurations". This sometimes works. It certainly can lower the energy of the system, but its ability to arrive at a state that doesn't blow up in the first few MD steps is not, in general, reliable. This can be accomplished by inserting something like

```
minimize 1.0e-8 1.0e-8 1000 100000
```

before the simulation runs.

Another way is to recognize the source of the instability, which is the repulsive component of the Lennard-Jones interaction. If one shrinks the size of the sigma, then the instability will vanish. Then one simply artificially “grows” the size of the atoms in a gradual way until they are the appropriate size. At that point, you have a stable state at the proper density.

A LAMMPS input that accomplishes this task is given in Appendix N.

Appendix A. Ethane: `sdf` file.

```
Ethane
  NIST    05100211103D 1   1.00000     0.00000
Copyright by the U.S. Sec. Commerce on behalf of U.S.A. All rights reserved.
  8  7  0     0  0              1 v2000
    0.5480    1.7661    0.0000 H   0  0  0  0  0  0           0  0  0
    0.7286    0.8978    0.6468 C   0  0  0  0  0  0           0  0  0
    0.4974    0.0000    0.0588 H   0  0  0  0  0  0           0  0  0
    0.0000    0.9459    1.4666 H   0  0  0  0  0  0           0  0  0
    2.1421    0.8746    1.1615 C   0  0  0  0  0  0           0  0  0
    2.3239    0.0050    1.8065 H   0  0  0  0  0  0           0  0  0
    2.8705    0.8295    0.3416 H   0  0  0  0  0  0           0  0  0
    2.3722    1.7711    1.7518 H   0  0  0  0  0  0           0  0  0
  1  2  1  0     0  0
  2  3  1  0     0  0
  2  4  1  0     0  0
  2  5  1  0     0  0
  5  6  1  0     0  0
  5  7  1  0     0  0
  5  8  1  0     0  0
M  END
>  <DATE>
2002-05-10

>  <COPYRIGHT>
Collection (C) 2012 copyright by the U.S. Secretary of Commerce on behalf of the United States of America. All rights reserved.

>  <NAME>
Ethane

>  <CAS.NUMBER>
74-84-0

>  <METHOD>
PM3

>  <CONTRIBUTOR>
Karl Irikura

>  <REVIEWER>
Karl Irikura and Joel Liebman

$$$$
```

Appendix B. Ethane: `xyz` file.

```
8
ethane from NIST Chemistry Webbook
H       0.548     1.7661    0
C       0.7286    0.8978    0.6468
H       0.4974    0         0.0588
H       0         0.9459    1.4666
C       2.1421    0.8746    1.1615
H       2.3239    0.005     1.8065
H       2.8705    0.8295    0.3416
H       2.3722    1.7711    1.7518
```

Appendix C. Ethane: standard `pdb` file.

```
REMARK natm =        8
REMARK                STANDARD PDB FORMAT
REMARK        INCLUDES # of BONDS in COL 7
REMARK           INCLUDES CHARGES in COL 8
HETATM    1  H    UNK     1    0.548   1.766   0.000      1  0.00
HETATM    2  C    UNK     1    0.729   0.898   0.647      4  0.00
HETATM    3  H    UNK     1    0.497   0.000   0.059      1  0.00
HETATM    4  H    UNK     1    0.000   0.946   1.467      1  0.00
HETATM    5  C    UNK     1    2.142   0.875   1.161      4  0.00
HETATM    6  H    UNK     1    2.324   0.005   1.806      1  0.00
HETATM    7  H    UNK     1    2.870   0.830   0.342      1  0.00
HETATM    8  H    UNK     1    2.372   1.771   1.752      1  0.00
CONECT       1       2
CONECT       2       1       3       4       5
CONECT       3       2
CONECT       4       2
CONECT       5       2       6       7       8
CONECT       6       5
CONECT       7       5
CONECT       8       5
```

Appendix D. Ethane: modified `pdb` file.

```
REMARK natm =        8
REMARK        INCLUDES # of BONDS in COL 7
REMARK           INCLUDES CHARGES in COL 8
REMARK         INCLUDES  IMPROPER in COL 9
HETATM      1  H    UNK     1    0.548   1.766   0.000     1  0.00000   0
HETATM      2  C    UNK     1    0.729   0.898   0.647     4  0.00000   0
HETATM      3  H    UNK     1    0.497   0.000   0.059     1  0.00000   0
HETATM      4  H    UNK     1    0.000   0.946   1.467     1  0.00000   0
HETATM      5  C    UNK     1    2.142   0.875   1.161     4  0.00000   0
HETATM      6  H    UNK     1    2.324   0.005   1.806     1  0.00000   0
HETATM      7  H    UNK     1    2.870   0.830   0.342     1  0.00000   0
HETATM      8  H    UNK     1    2.372   1.771   1.752     1  0.00000   0
CONECT        1       2
CONECT        2       1       3       4       5
CONECT        3       2
CONECT        4       2
CONECT        5       2       6       7       8
CONECT        6       5
CONECT        7       5
CONECT        8       5
```

Appendix E. Ethane: `make_pdb_input.txt`

```
cintype:  type of input file (xyz or pdb)
xyz
cpos:  name of input file (xyz or pdb)
ethane_0001.xyz
natm:  number of atoms
8
lconnect:  true if computing bonds, false if reading them from pdb file
.true.
ucv1:  unit cell lattice vector 1
100.0 0.0 0.0
ucv2:  unit cell lattice vector 2
0.0 100.0 0.0
ucv3:  unit cell lattice vector 3
0.0 0.0  100.0
lpdb:  apply periodic boundary conditions
.true.
limproper:  include calculation of improper modes
.true.
nimproper: number of entries in improper database
2
cimproper:  name of improper database input file
improper_input.txt
lcharge:  include calculation of charges
.true.
ncharge: number of entries in charge database
2
ccharge:  name of charge database input file
charge_input.txt
ldoublecount:  list i-j and j-i neighbors
.true.
```

Appendix F. Ethane: `charge.txt`

```
C  0.00
H  0.00
```

Appendix G. Ethane: `improper.txt`

```
C  0
H  0
```

Appendix H. Ethane: `intra_input.txt`

```
crunid:  run id
001
csystem:  system name
ethane1
natom:  number of atoms
8
nbond:  number of bonds
7
cdata:  name of data file with composition and connectivity information
make_pdb_output_mod.pdb
lprint:  printer variable
.true.
limproper:  logical variable for computing improper modes
.false.
l567:  logical variable for computing 5-, 6- and 7- atom modes
.false.
xlo:  lower bound of simulation box in x-dimension
0.0d0
xhi:  upper bound of simulation box in x-dimension
100.0d0
ylo:  lower bound of simulation box in y-dimension
0.0d0
yhi:  upper bound of simulation box in y-dimension
100.0d0
zlo:  lower bound of simulation box in z-dimension
0.0d0
zhi:  upper bound of simulation box in z-dimension
100.0d0
```

Appendix H. Ethane: `config_ethane1.txt`

```
LAMMPS INPUT FILE FOR ethane1
       8 atoms
       7 bonds
      12 angles
       9 dihedrals
       0 impropers
       2 atom types
       2 bond types
       2 angle types
       1 dihedral types
       0 improper types
  0.000000000000000E+00   0.100000000000000E+03 xlo xhi
  0.000000000000000E+00   0.100000000000000E+03 ylo yhi
  0.000000000000000E+00   0.100000000000000E+03 zlo zhi

 Masses

       1   0.120107000000000E+02
       2   0.100800000000000E+01

 Atoms

       1        1        2   0.000000000000000E+00   0.548000000000000E+00   0.176600000000000E+01   0.000000000000000E+00  # H
       2        1        1   0.000000000000000E+00   0.729000000000000E+00   0.898000000000000E+00   0.647000000000000E+00  # C
       3        1        2   0.000000000000000E+00   0.497000000000000E+00   0.000000000000000E+00   0.590000000000000E-01  # H
       4        1        2   0.000000000000000E+00   0.000000000000000E+00   0.946000000000000E+00   0.146700000000000E+01  # H
       5        1        1   0.000000000000000E+00   0.214200000000000E+01   0.875000000000000E+00   0.116100000000000E+01  # C
       6        1        2   0.000000000000000E+00   0.232400000000000E+01   0.500000000000000E-02   0.180600000000000E+01  # H
       7        1        2   0.000000000000000E+00   0.287000000000000E+01   0.830000000000000E+00   0.342000000000000E+00  # H
       8        1        2   0.000000000000000E+00   0.237200000000000E+01   0.177100000000000E+01   0.175200000000000E+01  # H

 Bonds

       1        2        1        2
       2        2        2        3
       3        2        2        4
       4        1        2        5
       5        2        5        6
       6        2        5        7
       7        2        5        8

 Angles

       1        2        1        2        3
       2        2        1        2        4
```

```
      3        1        1        2        5
      4        2        3        2        4
      5        1        3        2        5
      6        1        4        2        5
      7        1        2        5        6
      8        1        2        5        7
      9        1        2        5        8
     10        2        6        5        7
     11        2        6        5        8
     12        2        7        5        8

Dihedrals

      1        1        1        2        5        6
      2        1        1        2        5        7
      3        1        1        2        5        8
      4        1        3        2        5        6
      5        1        3        2        5        7
      6        1        3        2        5        8
      7        1        4        2        5        6
      8        1        4        2        5        7
      9        1        4        2        5        8
```

Appendix J. Ethane: `intra_output_001.txt`

```
 nc:  number of components
          2
 np:  number of potentials
          4
 nmn:  number of non-bonded modes
          3
 nms:  number of stretching modes
          2
 nmb:  number of bending modes
          2
 nmt:  number of torsion modes
          1
 nm5:  number of 5-atom modes
          0
 nm6:  number of 6-atom modes
          0
 nm7:  number of 7-atom modes
          0
 list of modes
a C       2
a H       6
N C  C       1
N C  H       1
N H  H       1
S C  C       1
S C  H       6
B C  C  H       6
B H  C  H       6
T H  C  C  H       9
```

Appendix K. Ethane: `intra_ethane1.txt`

```
units real
boundary p p p
atom_style full

read_data config_ethane1.txt

# interaction styles
pair_style lj/cut/coul/cut 12.0
bond_style harmonic
angle_style harmonic
dihedral_style opls
pair_modify mix geometric tail yes

# OPLS considers 1-4 interactions with 50%.
special_bonds lj/coul 0.0 0.0 0.5

# force field parameters
# missing nonbonded parameters are inferred from mixing.
pair_coeff   1 1  0.066 3.5      # CT = tetrahedral (sp3) carbon
pair_coeff   2 2  0.03 2.5      # HC = H bound to carbon
bond_coeff   1   268.0 1.529    # CT-CT
bond_coeff   2   340.0 1.09     # CT-HC
angle_coeff  1    37.5 110.7    # CT-CT-HC
angle_coeff  2    33.0 107.8    # HC-CT-HC
dihedral_coeff 1 0.000 0.000 0.318 0.000 # HC-CT-CT-HC

velocity all create 100.0 53244 dist gaussian mom yes rot yes

timestep 0.25
neigh_modify every 10 delay 20 check yes
thermo 400
thermo_style multi

fix 1 all nve

dump 1 all atom 400 dump.lammpstrj

run 20000
```

Appendix L. Ethane: `input_make_lammps_system_config.txt`

```
crunid:  3 digit run id
001
cconfig1:  input configuration file name
config_ethane1.txt
cconfigsys:  output configuration file name
config_ethane_sys.txt
nmpd:  number of molecules per dimension
5
side:  length of cubic box per dimension (A)
100.0
```

Appendix M. Ethane: `config_ethane_sys.txt`

```
LAMMPS INPUT FILE FOR ethane1
        1000  atoms
         875  bonds
        1500  angles
        1125  dihedrals
           0  impropers
           2  atom types
       2 bond types
       2 angle types
       1 dihedral types
       0 improper types
  0.0000000E+00   100.000000000000       xlo xhi
  0.0000000E+00   100.000000000000       ylo yhi
  0.0000000E+00   100.000000000000       zlo zhi

 Masses

           1   12.0107000000000
           2   1.00800000000000

 Atoms

         1          1          2   0.0000000E+00  -0.8874330E+00   0.8795335E+00  -0.9040670E+00
         2          1          1   0.0000000E+00  -0.7064330E+00   0.1153352E-01  -0.2570670E+00
         3          1          2   0.0000000E+00  -0.9384330E+00  -0.8864665E+00  -0.8450670E+00
         4          1          2   0.0000000E+00  -0.1435433E+01   0.5953352E-01   0.5629330E+00
         5          1          1   0.0000000E+00   0.7065670E+00  -0.1146648E-01   0.2569330E+00
         6          1          2   0.0000000E+00   0.8885670E+00  -0.8814665E+00   0.9019330E+00
         7          1          2   0.0000000E+00   0.1434567E+01  -0.5646648E-01  -0.5620670E+00
         8          1          2   0.0000000E+00   0.9365670E+00   0.8845335E+00   0.8479330E+00
         9          2          2   0.0000000E+00  -0.8874330E+00   0.8795335E+00   0.1909593E+02
        10          2          1   0.0000000E+00  -0.7064330E+00   0.1153352E-01   0.1974293E+02
        11          2          2   0.0000000E+00  -0.9384330E+00  -0.8864665E+00   0.1915493E+02
        12          2          2   0.0000000E+00  -0.1435433E+01   0.5953352E-01   0.2056293E+02
        13          2          1   0.0000000E+00   0.7065670E+00  -0.1146648E-01   0.2025693E+02
```

```
        14          2          2   0.0000000E+00   0.8885670E+00  -0.8814665E+00   0.2090193E+02
        15          2          2   0.0000000E+00   0.1434567E+01  -0.5646648E-01   0.1943793E+02
        16          2          2   0.0000000E+00   0.9365670E+00   0.8845335E+00   0.2084793E+02

many lines omitted

       992        124          2   0.0000000E+00   0.8093657E+02   0.8088453E+02   0.6084793E+02
       993        125          2   0.0000000E+00   0.7911257E+02   0.8087953E+02   0.7909593E+02
       994        125          1   0.0000000E+00   0.7929357E+02   0.8001153E+02   0.7974293E+02
       995        125          2   0.0000000E+00   0.7906157E+02   0.7911353E+02   0.7915493E+02
       996        125          2   0.0000000E+00   0.7856457E+02   0.8005953E+02   0.8056293E+02
       997        125          1   0.0000000E+00   0.8070657E+02   0.7998853E+02   0.8025693E+02
       998        125          2   0.0000000E+00   0.8088857E+02   0.7911853E+02   0.8090193E+02
       999        125          2   0.0000000E+00   0.8143457E+02   0.7994353E+02   0.7943793E+02
      1000        125          2   0.0000000E+00   0.8093657E+02   0.8088453E+02   0.8084793E+02

 Bonds

           1          2          1          2
           2          2          2          3
           3          2          2          4
           4          1          2          5
           5          2          5          6
           6          2          5          7
           7          2          5          8
           8          2          9         10
           9          2         10         11
          10          2         10         12
          11          1         10         13
          12          2         13         14
          13          2         13         15
          14          2         13         16

many lines omitted

         868          2        989        992
         869          2        993        994
         870          2        994        995
         871          2        994        996
         872          1        994        997
         873          2        997        998
```

```
       874          2        997        999
       875          2        997       1000

 Angles

         1          2          1          2          3
         2          2          1          2          4
         3          1          1          2          5
         4          2          3          2          4
         5          1          3          2          5
         6          1          4          2          5
         7          1          2          5          6
         8          1          2          5          7
         9          1          2          5          8
        10          2          6          5          7
        11          2          6          5          8
        12          2          7          5          8
        13          2          9         10         11
        14          2          9         10         12
        15          1          9         10         13
        16          2         11         10         12
        17          1         11         10         13
        18          1         12         10         13
        19          1         10         13         14
        20          1         10         13         15
        21          1         10         13         16
        22          2         14         13         15
        23          2         14         13         16
        24          2         15         13         16
        25          2         17         18         19

many lines omitted

      1488          2        991        989        992
      1489          2        993        994        995
      1490          2        993        994        996
      1491          1        993        994        997
      1492          2        995        994        996
      1493          1        995        994        997
      1494          1        996        994        997
      1495          1        994        997        998
```

```
       1496          1        994        997        999
       1497          1        994        997       1000
       1498          2        998        997        999
       1499          2        998        997       1000
       1500          2        999        997       1000

 Dihedrals

          1          1          1          2          5          6
          2          1          1          2          5          7
          3          1          1          2          5          8
          4          1          3          2          5          6
          5          1          3          2          5          7
          6          1          3          2          5          8
          7          1          4          2          5          6
          8          1          4          2          5          7
          9          1          4          2          5          8
         10          1          9         10         13         14
         11          1          9         10         13         15
         12          1          9         10         13         16
         13          1         11         10         13         14
         14          1         11         10         13         15
         15          1         11         10         13         16
         16          1         12         10         13         14
         17          1         12         10         13         15
         18          1         12         10         13         16

many lines omitted

       1116          1        988        986        989        992
       1117          1        993        994        997        998
       1118          1        993        994        997        999
       1119          1        993        994        997       1000
       1120          1        995        994        997        998
       1121          1        995        994        997        999
       1122          1        995        994        997       1000
       1123          1        996        994        997        998
       1124          1        996        994        997        999
       1125          1        996        994        997       1000
```

Appendix N. Ethane: `intra_ethane_grow.txt`

```
#
units real
boundary p p p
atom_style full

read_data config_ethane_sys.txt

# interaction styles
# you must turn off coulombic interactions for this equilibration process
#pair_style lj/cut/coul/cut 6.0
pair_style lj/cut 10.0
bond_style harmonic
angle_style harmonic
dihedral_style opls
pair_modify mix geometric tail yes

# OPLS considers 1-4 interactions with 50%.
special_bonds lj/coul 0.0 0.0 0.5

#
#  make LJ sigma variables so that you can grow a dense system
#
print "beginning variable declarations;"

variable sigma_11   equal 3.5
variable sigma_22   equal 2.5
print "here at A;"
variable sigma_12   equal 0.5*(${sigma_11}+${sigma_22})

print "sigma_11 (A) = ${sigma_11};"
print "sigma_12(A) = ${sigma_12};"
print "sigma_22 (A) = ${sigma_22};"

#
#  sigma growth is linear y = mx + b
#  define intercept as minimum size at time zero
#

variable facmin equal 0.01

variable sigma_11_min equal ${facmin}*${sigma_11}
variable sigma_12_min equal ${facmin}*${sigma_12}
variable sigma_22_min equal ${facmin}*${sigma_22}

print "facmin  = ${facmin};"
print "sigma_11_min (A) = ${sigma_11_min};"
print "sigma_12_min (A) = ${sigma_12_min};"
print "sigma_22_min (A) = ${sigma_22_min};"

#
#  define slope  by rise over run
#
```

```
variable step_grow equal 1000

variable slope_11 equal (${sigma_11}-${sigma_11_min})/${step_grow}
variable slope_12 equal (${sigma_12}-${sigma_12_min})/${step_grow}
variable slope_22 equal (${sigma_22}-${sigma_22_min})/${step_grow}

print "step_grow  = ${step_grow};"
print "slope_11 (A/step) = ${slope_11};"
print "slope_12 (A/step) = ${slope_12};"
print "slope_22 (A/step) = ${slope_22};"

#
# define linear function
#

variable sigma_11_funk equal ${sigma_11_min}+${slope_11}*step
variable sigma_12_funk equal ${sigma_12_min}+${slope_12}*step
variable sigma_22_funk equal ${sigma_22_min}+${slope_22}*step

# force field parameters
# missing nonbonded parameters are inferred from mixing.
pair_coeff   1 1  0.066 ${sigma_11}      # CT = tetrahedral (sp3) carbon
pair_coeff   2 2  0.03 ${sigma_22}      # HC = H bound to carbon
bond_coeff   1   268.0 1.529    # CT-CT
bond_coeff   2   340.0 1.09     # CT-HC
angle_coeff  1    37.5 110.7    # CT-CT-HC
angle_coeff  2    33.0 107.8    # HC-CT-HC
dihedral_coeff 1 0.000 0.000 0.318 0.000 # HC-CT-CT-HC

# initialize random velocities and run a few steps
# of MD to break symmetries.
velocity all create 180.0 53244 dist gaussian mom yes rot yes
#fix 1 all nve
#run 100
# and minimize for a bit to get a near 0K structure
#minimize 1.0e-12 1.0e-12 10000 1000000

# grow LJ sigma during simulation
fix 1 all adapt 1 pair lj/cut sigma 1 1 v_sigma_11_funk scale no reset no
fix 2 all adapt 1 pair lj/cut sigma 1 2 v_sigma_12_funk scale no reset no
fix 3 all adapt 1 pair lj/cut sigma 2 2 v_sigma_22_funk scale no reset no

timestep 0.25
neigh_modify every 10 delay 20 check yes
thermo 100
#thermo_style multi
thermo_style custom step v_sigma_11_funk v_sigma_12_funk v_sigma_22_funk

dump 2 all atom 400 dump.lammpstrj

fix 4 all nve

run ${step_grow}
unfix 1
unfix 2
unfix 3
```

```
thermo_style multi
#run_style respa 2 10 bond 1 angle 1 dihedral 1 improper 1 pair 2
#run_style respa 4 2 2 2 bond 1 angle 2 dihedral 3 improper 3 pair 4
run 10000
```