Tutorial 6: Vasp Calculations for Ab Initio Molecular Dynamics

Deyu Lu and Neerav Kharche

Workshop on “Theory and Computation for Interface Science and Catalysis: Fundamentals, Research and Hands on Engagement using VASP”
Nov. 3 – 7, 2014
Outline

• Basic of molecular dynamics

• Ab initio molecular dynamics

• AIMD run for 16 H₂O cell

• Data analysis of precomputed 32 H₂O cell
Molecular dynamics

"for the development of multiscale models for complex chemical systems".

- protein folding,
- catalysis,
- electron transfer,
- drug design
- ...

Winners of Nobel Prize in Chemistry 2013

Ergodicity

**Ensemble average**

\[ \langle A \rangle_{\text{ens}} = \int \int dp^N \, dr^N \rho(r^N, p^N) A(r^N, p^N) \]

Average over all possible states of the system in the phase space

**Time average**

\[ \langle A \rangle_{\text{time}} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^\tau dt \, A(r^N(t), p^N(t)) \]

Average over a sufficiently long time
Ergodicity

If one allows the system to evolve in time indefinitely, that system will eventually pass through all possible states.

The ergodic hypothesis states

$$\langle A \rangle_{ens} = \langle A \rangle_{time}$$

Ensemble average = Time average
Integration of equations of motion

Newton’s equation of motion:  \( F = m \ddot{r} \)

- **Verlet algorithm**
  - the error in new position is \( O(\Delta t^4) \)
  - does not use the velocity to compute the new position
  - the velocity can be derived with an error of \( O(\Delta t^2) \)

- **Leap frog algorithm**
  - evaluates the velocities at half-integer time steps
  - Uses velocities to compute new positions

- **Velocity-corrected Verlet algorithm**
  - the error in both the positions and velocities is \( O(\Delta t^4) \)
  - requires positions and forces at \( t+\Delta t \) to update velocity

- **Higher-order schemes**
### Verlet Algorithm

**Position at step n-1:**
\[
r_{n-1} = r_n - v_n \Delta t + \frac{1}{2} \left( \frac{F_n}{m} \right) \Delta t^2 - O(\Delta t^3)
\]

**Position at step n+1:**
\[
r_{n+1} = r_n + v_n \Delta t + \frac{1}{2} \left( \frac{F_n}{m} \right) \Delta t^2 + O(\Delta t^3)
\]

**Sum of the two term:**

Propagate position
\[
r_{n+1} = 2r_n - r_{n-1} + \left( \frac{F_n}{m} \right) \Delta t^2 + O(\Delta t^4)
\]

**Do a subtraction**

\(v_n\) is one step behind
\[
v_n = \frac{r_{n+1} - r_{n-1}}{2\Delta t} + O(\Delta t^2)
\]
Common thermal dynamic ensembles

• **Microcanonical ensemble (NVE)**
  – Isolated
  – Total energy $E$ is fixed
  – Every accessible microstate has equal probability

• **Canonical ensemble (NVT)**
  – The system can exchange energy with a heat bath
  – $T$ is constant
  – Probability of finding the system at state $i$
    $$p_i = \frac{e^{-E_i/k_BT}}{\sum_i e^{-E_i/k_BT}}$$

• **Isobaric-isothermal ensemble (NPT)**
  – Both $P$ and $T$ are constant

• **Grand canonical ensemble ($\mu$VT)**
Microcanonical ensemble

- Initialize $r_0$ and $v_0$
- Calculate force
- Integrate the equation of motion
- Update $r$ and $v$
Canonical ensemble (NVT)

Maxwell-Boltzmann distribution: \[ P(p) = \left( \frac{\beta}{2\pi m} \right)^{3/2} e^{-\left(\frac{\beta p^2}{2m}\right)} \]

Temperature \leftrightarrow \text{kinetic energy}: \[ k_B T = m \left\langle v_{\alpha}^2 \right\rangle \]

\[ \left\langle E_K \right\rangle = \frac{3}{2} Nk_B T \]

- Berendsen thermostat: Velocity rescaling
- Anderson thermostat: Stochastic coupling
- Nosé-Hoover thermostat: Extended Lagrangian
Temperature fluctuation

\[ P(p) = \left( \frac{\beta}{2 \pi m} \right)^{3/2} e^{-\left(\frac{\beta p^2}{2m}\right)} \]

Relative variance of the kinetic energy:

\[ \frac{\sigma_{p^2}^2}{\langle p^2 \rangle^2} \equiv \frac{\langle p^4 \rangle - \langle p^2 \rangle^2}{\langle p^2 \rangle^2} = \frac{2}{3} \]

Relative variance of temperature:

\[ \frac{\sigma_T^2}{\langle T_K \rangle_{NVT}^2} \equiv \frac{\langle T_K^2 \rangle_{NVT} - \langle T_K \rangle_{NVT}^2}{\langle T_K \rangle_{NVT}^2} \]

\[ = \frac{N\langle p^4 \rangle + N(N-1)\langle p^2 \rangle^2 - N^2 \langle p^2 \rangle^2}{N^2 \langle p^2 \rangle^2} \]

\[ = \frac{1}{N} \frac{\langle p^4 \rangle - \langle p^2 \rangle^2}{\langle p^2 \rangle^2} = \frac{2}{3N} \]
Berendsen thermostat

\[ \lambda = \sqrt{\frac{T_{bath}}{T(t)}} \]

\[ \Delta T = \frac{1}{2} \sum_{i=1}^{N} \frac{2m_i (\lambda v_i)^2}{3Nk_B} - \frac{1}{2} \sum_{i=1}^{N} \frac{2m_i v_i^2}{3Nk_B} \]

\[ = (\lambda^2 - 1) T(t) \]

\[ \frac{dT}{dt} = \frac{T_{bath} - T}{\tau} \]

\[ \lambda^2 = 1 + \frac{\Delta t}{\tau} \left( \frac{T_{bath}}{T} - 1 \right) \]

- Not real canonical ensemble, although close
- No direct proof of Maxwell-Boltzmann distribution
Andersen thermostat

• Start with \( \{ r_0^N, p_0^N \} \) and integrate the equations of motion for \( \Delta t \).
• A number of particles are selected to undergo a collision with the heat bath, if \( p > \nu \Delta t \).
• The new velocity will be drawn from a Maxwell-Boltzmann distribution at \( T_{\text{bath}} \).

✔ Andersen thermostat guarantees the canonical distribution.
✗ The stochastic collisions destroy the correlation of particle velocities, which disturbs dynamic properties.
Nosé-Hoover thermostat

\[ H = \sum_{i=1}^{N} \frac{p_i^2}{2m_is^2} + U(r^N) + \frac{p_s^2}{2Q} + L \frac{Lns}{\beta} \]

- An extended Lagrangian method.
- Deterministic molecular dynamics.
- It produces a canonical due to heat exchange between fictitious degree of freedom and real system.
- \( s \) is a scaling factor of the time step, so the time step fluctuates.
Ab initio molecular dynamics

Hamiltonian containing both nuclear and electronic degrees of freedom

\[
H = \sum_{l=1}^{N} \frac{p_l^2}{2M_l} + \sum_{i=1}^{N_e} \frac{p_i^2}{2m_i} + \sum_{i>j} \frac{e^2}{|r_i - r_j|} + \sum_{I>J} \frac{Z_I Z_J e^2}{|R_I - R_J|} - \sum_{i,I} \frac{Z_I e^2}{|R_I - r_i|}
\]

\[
\equiv T_N + T_e + V_{ee}(r) + V_{NN}(R) + V_{eN}(r, R)
\]

In general, we need to solve

\[
[T_N + T_e + V_{ee}(r) + V_{NN}(R) + V_{eN}(r, R)]\Psi(x, R) = E\Psi(x, R)
\]

- Electronic degree of freedom
- Nuclear degree of freedom
Born-Oppenheimer molecular dynamics

the adiabatic approximation

separation of variables: \( \Psi(x, R) = \phi(x, R) \chi(R) \)

\[
\begin{align*}
[T_e + V_{ee}(r) + V_{eN}(r, R)]\phi_0(x, R) &= \varepsilon_0(R)\phi_0(x, R) \\
[T_N + \varepsilon_0(R) + V_{NN}(R)]X(R, t) &= i\hbar \frac{\partial}{\partial t}X(R, t).
\end{align*}
\]

- Electrons stay in the adiabatic ground state at any instant of time.
- Nuclei move on the ground state Born-Oppenheimer potential energy surface.
- It a good approximation if the energy difference between the electronic ground state and first excited state is large compared to \( k_B T \).
- Minimization is required at each step of the MD simulation and the forces are computed using the orbitals thus obtained.
Car–Parrinello molecular dynamics

• The coupling between nuclear time evolution and electronic minimization is treated efficiently via an implicit adiabatic dynamics approach.

• A fictitious dynamics for the electronic orbitals is invented which, given orbitals initially at the minimum for an initial nuclear configuration, allows them to follow the nuclear motion adiabatically.

• Electronic orbitals are automatically at the approximately minimized configuration at each step of the MD evolution.
Car–Parrinello molecular dynamics

Lagrangian of an extended dynamical system:

\[ L = \mu \sum_i \langle \psi_i | \psi_i \rangle + \frac{1}{2} \sum_{i=1}^N M_i \dot{R}_i^2 - E[\{\psi\}, R] + \sum_{i,j} [\Lambda_{ij} (\langle \psi_i | \psi_j \rangle - \delta_{ij})] \]

a fictitious mass parameter

Car-Parrinello equations of motion:

\[ M_i \ddot{R}_i = -\nabla_i E[\{\psi\}, R] \]

\[ \mu \dot{\psi}_i(r) = -\frac{\delta}{\delta \psi_i^*(r)} E[\{\psi\}, R] + \sum_j \Lambda_{ij} \psi_j(r) \]

\[ -H^{KS}_\psi \psi_i \]

By properly choosing the fictitious mass and time step, the electronic and nuclear motions can be decoupled, so that the electronic subsystem stays cold.

Outline

• Basic of molecular dynamics
• *Ab initio* molecular dynamics
• AIMD run for 16 H$_2$O cell
  – Input parameters
  – Temperature and energy profiles
  – Visualization using VMD
• Data analysis of precomputed 32 H$_2$O cell
  – RDF introduction
  – RDF using VMD
Tutorials: File System – 16H2O MD Run

/software/Workshop14/Tutorials/Tutorial6/16H2O

- MD_run
- MD_run.ref
- VMD_scripts
- README (file)

- VASP Files: INCAR, POSCAR, POTCAR, KPOINTS, vpbs.com
- Perform your calculations in this directory

- TCL scripts
  - vmd_viz_16H2O.tcl
  - connect_broken_OH_bonds_bulk_water.tcl
  - To be used from VMD command line
Sample MD Run: 16 H$_2$O

**Initial atomic structure**
- Density = 1 g/cm$^3$
- 16 H2O in 7.82 Å cubic box
- Initial equilibration
  - Software: GROMACS
  - Classical MD at room temperature (300 K)

**Key simulation parameters**
- Functional: PBE
- Pseudopotential: PAW
- Γ-point sampling
- Elevated simulation temperature 400 K
  - To avoid overstructuring
  - For correct diffusion coefficients
- Time step: 0.5 fs
  - To sample O-H bond fluctuations
- Deuterium mass for Hydrogen
  - Allows for longer time step
- (Today) short MD trajectory: 50 fs i.e. 100 ionic steps
- For statistically meaningful results
  - Trajectories on the order of 5 ps
**MD Input**

**INCAR**

- PREC = Normal
- ENCUT = 400
- ALGO = Fast
- LREAL = Auto
- ISMEAR = 0  ! Gaussian smearing
- SIGMA = 0.05
- ISYM = 0  ! Symmetry off

! MD

- IBRION = 0  ! MD
- POTIM = 0.5  ! Time step = 0.5 fs
- NSW = 100   ! Number of ionic steps
- TEBEG = 400 ! Start temperature
- TEEND = 400 ! Final temperature
- SMASS = 0   ! Canonical (Nose-Hoover) thermostat
- POMASS = 16.0 2.0 ! Deuterium mass for Hydrogen

! Don’t write WAVECAR or CHGCAR

- LWAVE = F
- LCHARG = F

**KPOINTS (Γ-only)**

- 0
- Gamma
- 1 1 1
- 0 0 0

---

\[
\begin{array}{c}
\text{KPOINTS (Γ-only)} \\
\hline
0 \\
\text{Gamma} \\
1 1 1 \\
0 0 0 \\
\end{array}
\]
Temperature and Energy Profiles

- Relevant OSZICAR output

<table>
<thead>
<tr>
<th>RMM: 12</th>
<th>-0.233662671242E+03</th>
<th>0.91672E-04</th>
<th>-0.21448E-04</th>
<th>130</th>
<th>0.439E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>T= 405.</td>
<td>E= -23120407E+03</td>
<td>F= -23366267E+03</td>
<td>E0= -23366267E+03</td>
<td>EK= 0.24586E+01</td>
<td>SP= 0.00E+00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RMM: 4</th>
<th>-0.233736351315E+03</th>
<th>-0.48227E-05</th>
<th>-0.40425E-04</th>
<th>173</th>
<th>0.590E-02</th>
</tr>
</thead>
<tbody>
<tr>
<td>T= 416.</td>
<td>E= -23120583E+03</td>
<td>F= -23373635E+03</td>
<td>E0= -23373635E+03</td>
<td>EK= 0.25298E+01</td>
<td>SP= 0.70E-03</td>
</tr>
</tbody>
</table>

- Extract temperature and energy profiles from OSZICAR
  - > grep "T= " OSZICAR | awk '{ print $1 " " $3 " " $5 }' > T_E.txt

![Graph showing temperature and energy profiles](image)

- Target Temperature = 400 K
- Simulation output:
  - Average = 397.32 K
  - σ(T) = 52.25 K
Visualization using VMD

Load vasprun.xml in VMD

• Start VMD
• Open TCL Console

• Run (source) TCL script vmd_viz_16H2O.tcl

VMD Display
Tutorials: File System – 32H2O MD Data Analysis

/software/Workshop14/Tutorials/Tutorial6/32H2O

- 32H2O_5ps.xyz: Precomputed 5ps (10,000 steps) MD trajectory
- vmd_viz_32H2O.tcl: VMD-TCL script for visualization
- rdf.tcl: VMD-TCL script to calculate radial distribution functions (RDFs)
- README
- RDF.ref: Precomputed RDFs
MD Trajectory for 32 H₂O Cell

Simulation protocol
• Density = 1 g/cm³
• 32 H₂O in 9.86 Å cubic box
• DFT with vdW
• Functional: optB88-vdW
• Pseudopotential: PAW
• Temperature: 350 K
• Time step: 0.5 fs
• Simulation time: 5 ps equilibration followed by 5 ps production

Radial Distribution Function (RDF)

- Describes how density varies as a function of distance from a reference particle
- Definition

\[ g(r[i]) = \frac{n_{pair}[i]}{v[i]} \cdot \frac{V}{N_{pair}} \]

- \( n_{pair}[i] \): Number of pairs in bin \((r_i, r_{i+1} = r_i + dr)\)
- \( v[i] \): Volume of bin
- \( N_{pair} \): Number of pairs
- \( V \): Volume of simulation cell
- Coordination number: Integral over first peak of \( g(r) \)


O-O RDF


RDFs using VMD

- Load 32H₂O_5ps.xyz
  - source vmd_viz_32H₂O.tcl
- Compute RDFs
  - source rdf.tcl
  - RDFs will be written to files rdf_OO.dat, rdf_OH.dat, and rdf_HH.dat
  - Data format

<table>
<thead>
<tr>
<th>r</th>
<th>g(r)</th>
<th>Integrated g(r)</th>
</tr>
</thead>
</table>

- Integral over first peak
  - Theory: 4.5
  - Expt. 4.7 (PNAS 103, 7973 (2006))
RDFs using VMD

Covalent bonds
Hydrogen bonds

O-H RDF

Experimental data: A. K. Soper

H-H RDF