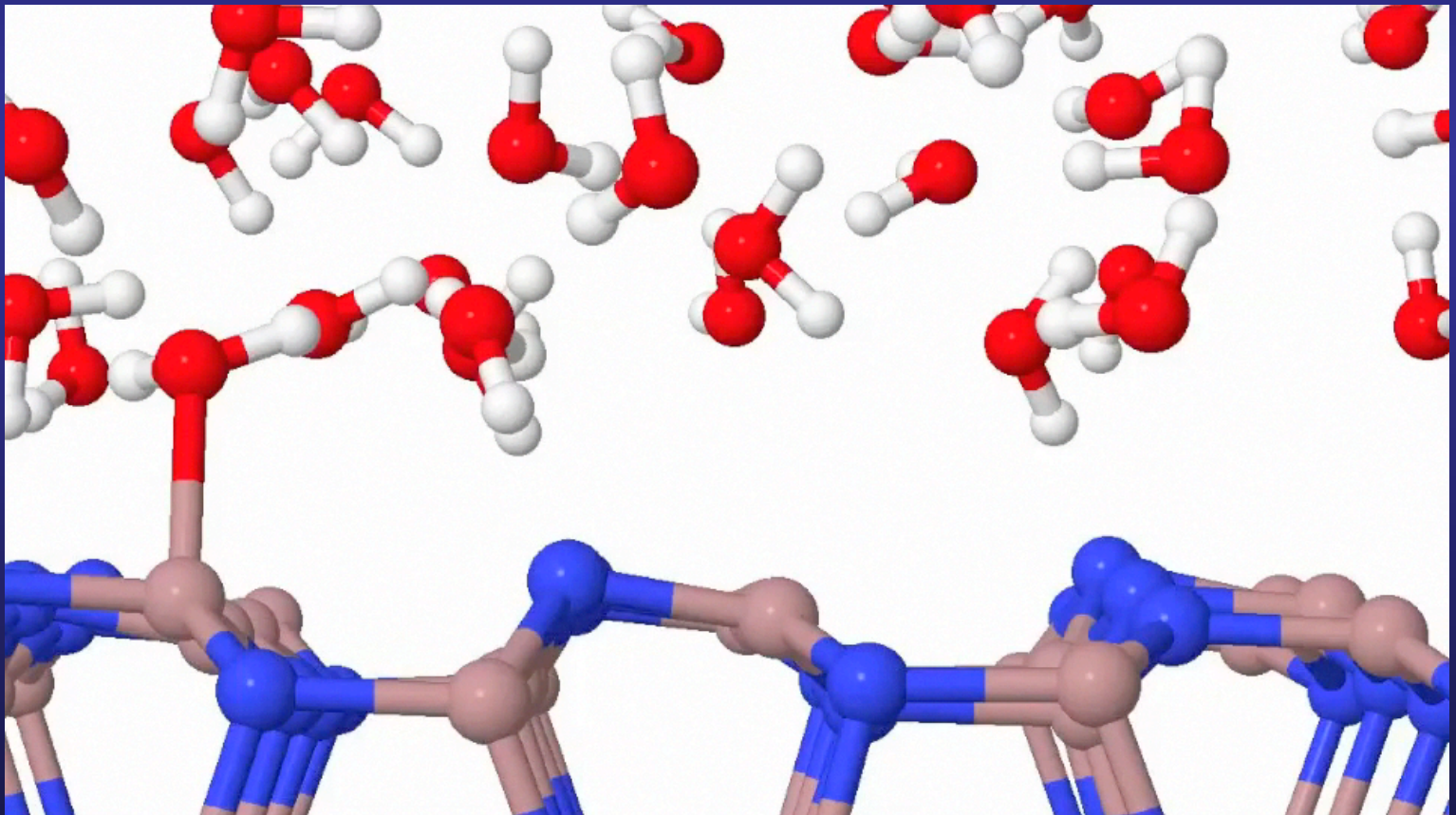


# Molecular dynamics in Siesta

Marivi Fernandez -Serra



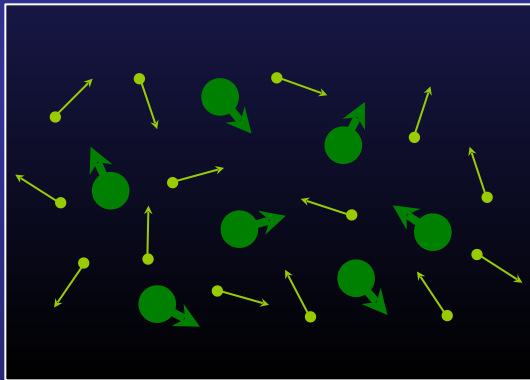
Stony Brook University



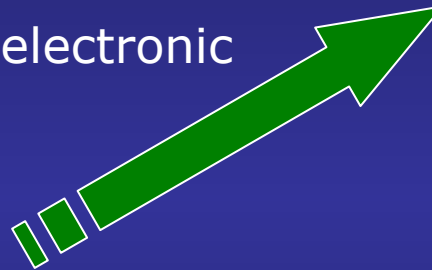
# Born-Oppenheimer dynamics

Nuclei are much slower than electrons

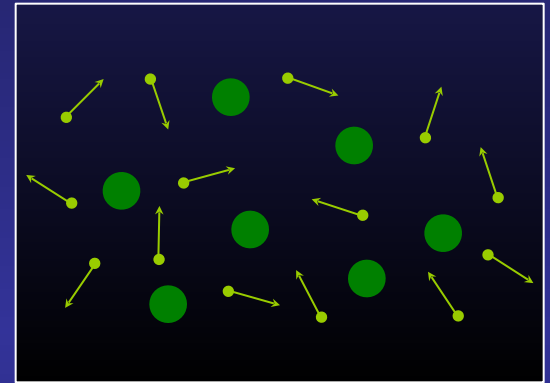
$$m_n \gg m_e$$



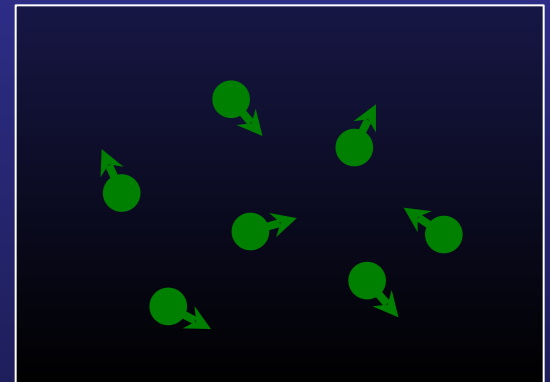
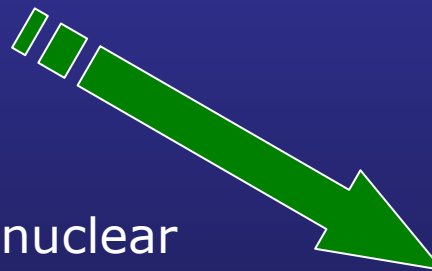
electronic



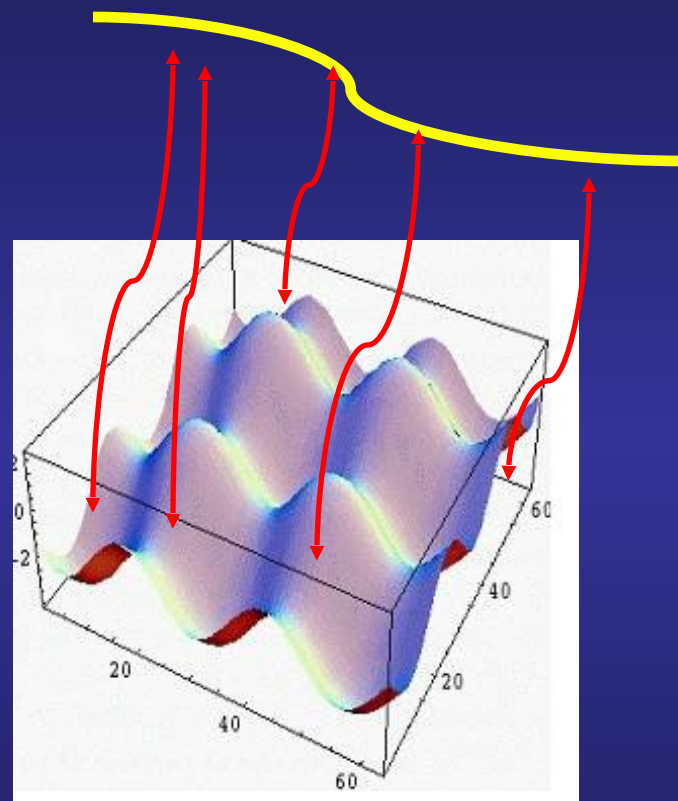
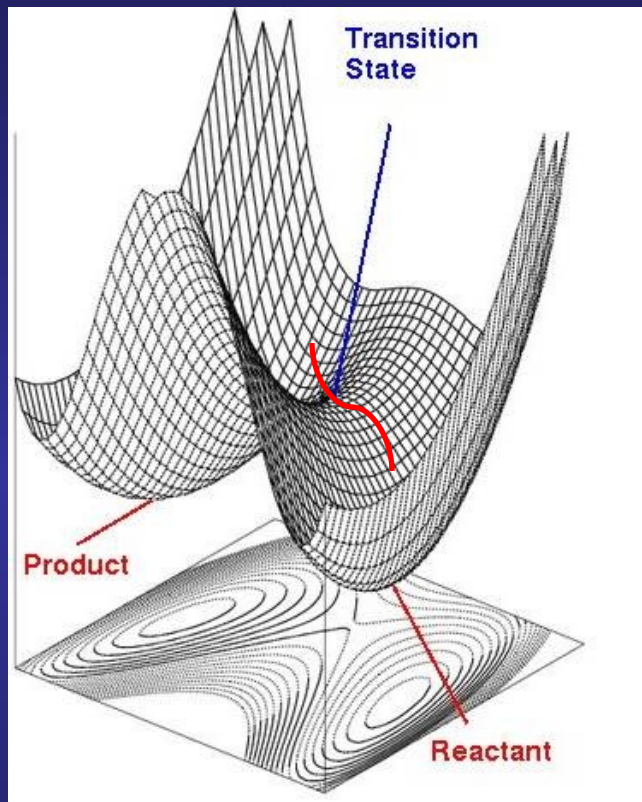
decoupling



nuclear



# Extracting information from the Potential Energy Surface (PES)



## -Optimizations and Phonons:

- We move **on** the PES
- Local vs global minima
- PES is harmonic close to minima

## -MD

- We move **over** the PES
- Good Sampling is required!!

# Molecular Dynamics

- Follows the time evolution of a system
- Solve Newton's equations of motion:

$$\vec{F}(t) = -\vec{\nabla} E = m\vec{a}(t) = m \frac{d^2 r(t)}{dt^2}$$

- Treats electrons quantum mechanically
- Treats nuclei classically
  - Hydrogen may raise issues:
    - tunneling (overestimating Energy barriers)
- Allows study of dynamic processes
- Annealing of complex materials
- Examines the influence of temperature
- Time averages Vs Statistical averages

# Ergodicity

- In MD we want to replace a full sampling on the appropriate statistical ensemble by a SINGLE very long trajectory.
- This is OK only if system is **ergodic**.
- **Ergodic Hypothesis: a phase point for any isolated system passes in succession through every point compatible with the energy of the system before finally returning to its original position in phase space. This journey takes a Poincare cycle.**
- **In other words, Ergodic hypothesis: each state consistent with our knowledge is equally “likely”.**
  - Implies the average value does not depend on initial conditions.
  - $\langle A \rangle_{\text{time}} = \langle A \rangle_{\text{ensemble}}$  , so  $\langle A_{\text{time}} \rangle = (1/N_{\text{MD}}) = \sum_{t=1, N} A_t$  is good estimator.
- Are systems in nature really ergodic? **Not always!**
  - Non-ergodic examples are glasses, folding proteins (in practice) and harmonic crystals (in principle).

# Different aspects of ergodicity

- The system relaxes on a “reasonable” time scale towards a unique equilibrium state (microcanonical state)
- Trajectories wander irregularly through the energy surface eventually sampling all of accessible phase space.
- Trajectories initially close together separate rapidly.(sensitivity to initial conditions). Lyapunov exponent.

**Ergodic behavior makes possible the use of statistical methods on MD of small system.** Small round-off errors and other mathematical approximations may not matter.

# Molecular Dynamics

- Timestep must be small enough to accurately sample highest frequency motion
- Typical timestep is 1 fs ( $1 \times 10^{-15}$  s)
- Typical simulation length = Depends on the system of study!!  
(the more complex the PES the longer the simulation time)
- *Is this timescale relevant to your process?*
- Simulation has two parts:
  - equilibration (redistribute energy)  
**System is equilibrated if averages of dynamical and structural quantities do not change with time.**
  - production (record data)
- Results:
  - diffusion coefficients
  - Structural information (RDF's,)
  - free energies / phase transformations (very hard!)
- *Is your result statistically significant?*

# Choosing the integrator

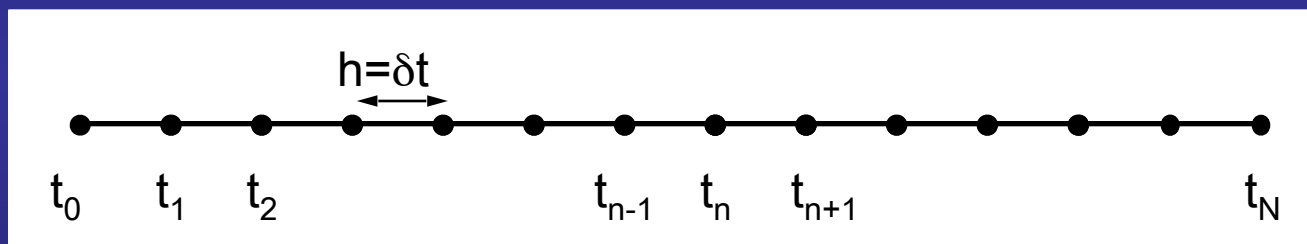
- The interatomic potentials are highly non-linear, often with discontinuous high derivatives, or are evaluated with limited precision.
- Small errors (precision) or minimal differences in the initial conditions lead to completely different trajectories (Ergodicity!). Statistical averages are the relevant quantities; they do not depend on the details of the trajectories (IF the simulation is long enough!!!!).
- Because of this, and since potentials are not perfect (all potential models are approximations to the real ones), one does not need too much accuracy in the integration of the equations of motion (as long as errors are not too large, and they do not affect fundamental properties such as conserved quantities).
- Conservation of energy IS important!!. We can allow errors in the total energy conservation of the order of 0.01 kT.
- CPU time is completely dominated by the calculation of the forces. Therefore, it is preferable to choose algorithms that require few evaluations of the forces, and do not need higher derivatives of the potential.



# Standard method to solve ordinary differential equations: the finite difference approach

Given molecular positions, velocities, and other dynamic information at a time  $t$

We attempt to obtain the position, velocities, etc. at a later time  $t + \delta t$ , to a sufficient degree of accuracy



## Notes:

The equations are solved on a step by step basis

The choice of the time interval  $\delta t$  will depend on the method of solution, but  $\delta t$  will be significantly smaller than the typical time taken for a molecule to travel its own length

# General step of a stepwise Molecular Dynamics simulation

Predict the positions, velocities, accelerations, etc. at a time  $t + \delta t$ , using the current values of these quantities

Evaluate the forces, and hence the accelerations  $\vec{a}_i = \vec{F}_i/m_i$  from the new positions

Correct the predicted positions, velocities, accelerations, etc. using the new accelerations

Calculate any variable of interest, such as the energy, virial, order parameters, ready for the accumulation of time averages, before returning to the first point for the next step



# Desirable qualities for a successful simulation algorithm

It should be fast and require little memory

Since the most time consuming part is the evaluation of the force, the raw speed of the integration algorithm is not so important

It should permit the use of long time step  $\delta t$

Far more important to employ a long time step. In this way, a given period of simulation time can be covered in a modest number of steps

It should duplicate the classical trajectory as closely as possible

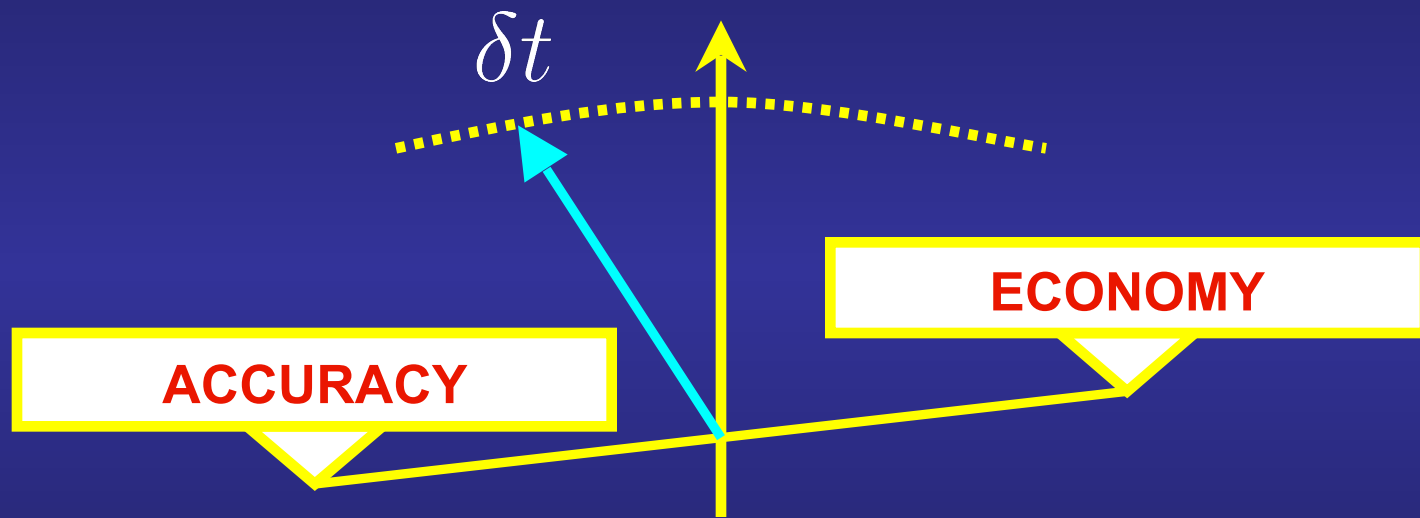
It should satisfy the known conservation laws for energy and momentum, and be time reversible

It should be simple in form and easy to program

Involve the storage of only a few coordinates, velocities,...

# Energy conservation is degraded as time step is increased

All simulations involve a trade-off between



A good algorithm permits a large time step to be used while preserving acceptable energy conservation

# Parameters that determine the size of $\delta t$

- Shape of the potential energy curves
- Typical particle velocities

**Shorter time steps are used at high-temperatures, for light molecules, and for rapidly varying potential functions**

# The Verlet algorithm method of integrating the equations of motion: description of the algorithm

Direct solution of the second-order equations

Method based on:

- the positions  $\vec{r}(t)$
- the accelerations  $\vec{a}(t)$
- the positions from the previous step  $\vec{r}(t - \delta t)$

A Taylor expansion of the positions around  $t$

$$\vec{r}(t + \delta t) = \vec{r}(t) + \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2 + \dots$$

$$\vec{r}(t - \delta t) = \vec{r}(t) - \vec{v}(t)\delta t + \frac{1}{2}\vec{a}(t)\delta t^2 - \dots$$

Adding the two equations

$$\vec{r}(t + \delta t) + \vec{r}(t - \delta t) = 2\vec{r}(t) + \vec{a}(t)\delta t^2 + \mathcal{O}(\delta t^4)$$

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)\delta t^2 + \mathcal{O}(\delta t^4)$$

# The Verlet algorithm method of integrating the equations of motion: some remarks

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)\delta t^2 + \mathcal{O}(\delta t^4)$$

## Remark 1

**The velocities are not needed to compute the trajectories**, but they are useful for estimating the kinetic energy (and the total energy).

They can be computed a posteriori using  
[  $\vec{v}(t)$  can only be computed once  $\vec{r}(t + \delta t)$  is known ]

$$\vec{v}(t) = \frac{\vec{r}(t + \delta t) - \vec{r}(t - \delta t)}{2\delta t} + \mathcal{O}(\delta t^2)$$

## Remark 2

Whereas the errors to compute the positions are of the order of  $\delta t^4$

The velocities are subject to errors of the order of  $\delta t^2$

# The Verlet algorithm method of integrating the equations of motion: some remarks

$$\vec{r}(t + \delta t) = 2\vec{r}(t) - \vec{r}(t - \delta t) + \vec{a}(t)\delta t^2 + \mathcal{O}(\delta t^4)$$

## Remark 3

The Verlet algorithm is properly centered:  $\vec{r}(t - \delta t)$  and  $\vec{r}(t + \delta t)$  play symmetrical roles.

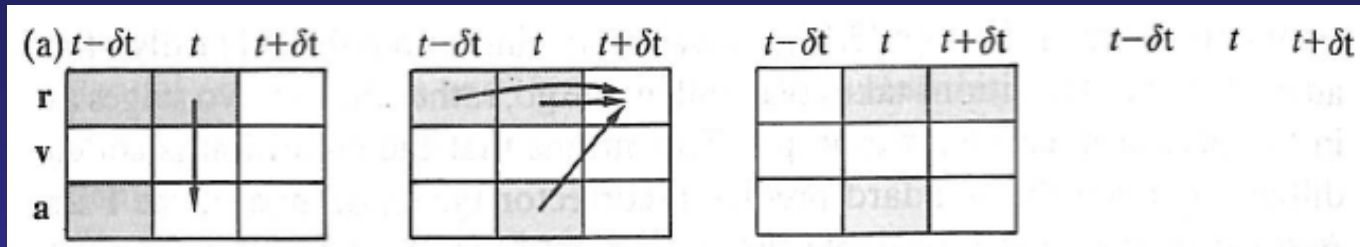
**The Verlet algorithm is time reversible**

## Remark 4

The advancement of positions takes place all in one go, rather than in two stages as in the predictor-corrector algorithm.



# The Verlet algorithm method of integrating the equations of motion: overall scheme



Known the positions at  $t$ , we compute the forces (and therefore the accelerations at  $t$ )

Then, we apply the Verlet algorithm equations to compute the new positions

...and we repeat the process computing the forces (and therefore the accelerations at  $t + \delta t$ )

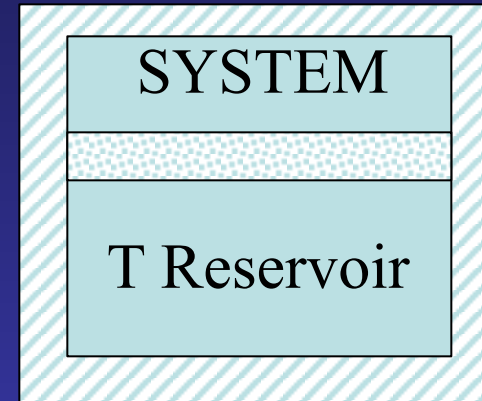
# When do we use MD?

- **Amorphous systems:**
  - Molecular Liquids (H<sub>2</sub>O, CO<sub>2</sub>)
  - Glasses (Si, SiO<sub>2</sub>)
- **Displacive Phase transitions (P and T relevant).**
- **Study of kinetic effects.**
  - Diffusion at surfaces
  - Thermal stability

# Nose-Hoover thermostat

- MD in canonical distribution (TVN)
- Introduce a friction force  $\zeta(t)$

$$\frac{dp}{dt} = F(q, t) - \zeta(t) p(t)$$



Dynamics of friction coefficient to get canonical ensemble.

$$Q \frac{d\zeta}{dt} = \sum \frac{1}{2} m_i v_i^2 - \frac{3N}{2} k_B T$$

Feedback makes  
K.E. =  $3/2 k_B T$

$Q$  = fictitious “heat bath mass”. Large  $Q$  is weak coupling

# Hints

- Nose Mass: Match a vibrational frequency of the system, better high energy frequency

$$Q \approx \frac{KT}{\omega^2}$$

# Which Ensemble should we use?

## • NVE (Verlet): Microcanonical

- Good trajectories.
- Time reversible (up to numerical error)
- Dynamical variables are well defined.
- Initial X and V are relevant: necessity of equilibration.

Same sampling  
In the  
thermodynamic limit

## • NVT (Nose): Canonical

- Good T control
- Equilibrates the system.
- Choice for Structural sampling.
- Sensitive to Nose mass.

## • NPE (Parrinello-Rahman)

- Phase transitions systems under pressure.
- 1 mass parameter (barostat)

## • NPT (Nose-Parrinello-Rahman)

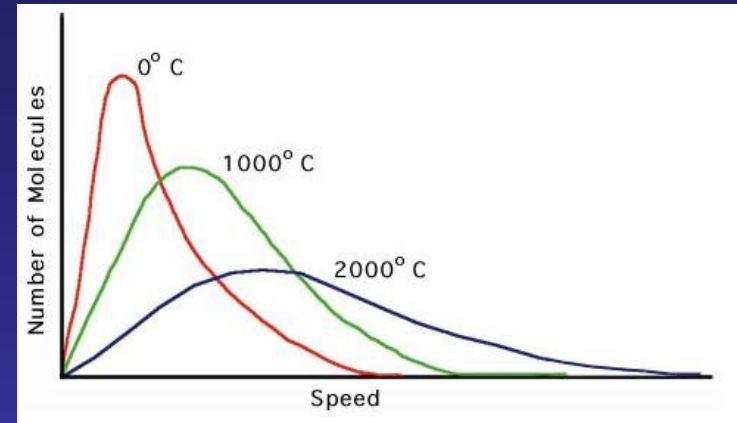
- Phase transitions under P and T
- 2 mass parameters, barostat and thermostat. (Fluctuations!!)

## Molecular Dynamics in SIESTA(1)

- **MD.TypeOfRun Verlet**  
NVE ensemble dynamics
- **MD.TypeOfRun Nose**  
NVT dynamics with Nose thermostat
- **MD.TypeOfRun ParrinelloRahman**  
NPE dynamics with P-R barostat
- **MD.TypeOfRun NoseParrinelloRahman**  
NPT dynamics with thermostat/barostat
- **MD.TypeOfRun Anneal**  
Anneals to specified p and T

## Molecular Dynamics in SIESTA(2)

- Setting the length of the run:  
MD.InitialTimeStep 1  
MD.FinalTimeStep 2000
- Setting the timestep:  
MD.LengthTimeStep 1.0 fs
- Setting the temperature:  
MD.InitialTemperature 298 K  
MD.TargetTemperature 298 K
- Setting the pressure:  
MD.TargetPressure 3.0 Gpa
- Thermostat / barostat parameters:  
MD.NoseMass / MD.ParrinelloRahmanMass



Maxwell-Boltzmann

## Annealing in SIESTA

- MD can be used to optimize structures:  
MD.Quench true  
- zeros velocity when opposite to force
- MD annealing:  
MD.AnnealOption Pressure  
MD.AnnealOption Temperature  
MD.AnnealOption TemperatureAndPressure
- Timescale for achieving target  
MD.TauRelax 100.0 f



# Phonons and MD

1. MD simulations (NVE)
2. Fourier transform of Velocity-Velocity autocorrelation function.
  1. Anharmonic effects:  $\omega(T)$
  2. Expensive, but information available for MD simulations.

$$Ph(\omega) = \int_0^{\infty} dt e^{2\pi i \omega t} \frac{\langle \vec{v}(0) \vec{v}(t) \rangle}{\langle \vec{v}(0)^2 \rangle}$$



# Computing the instantaneous temperature, kinetic energy and total energy

$$\sum_{i=1}^N \frac{|\vec{p}_i|^2}{m_i} = 2\mathcal{K} = 3Nk_B\mathcal{T}$$

# Output of a Molecular Dynamic in Siesta: the Verlet algorithm (NVE-microcanonical ensemble)

SystemLabel.MDE

Conserved quantity



# Step	T (K)	E_KS (eV)	E_tot (eV)	Vol (A <sup>3</sup> )	P (kBar)
1	600.00	-2955.25340	-2954.55542	93.166	166.448
2	536.29	-2955.18330	-2954.55944	93.166	171.567
3	639.06	-2955.30406	-2954.56065	93.166	161.570
4	893.32	-2955.60359	-2954.56439	93.166	136.114
5	1251.77	-2956.02573	-2954.56955	93.166	104.263
6	1647.75	-2956.49212	-2954.57530	93.166	68.292
7	2010.57	-2956.91768	-2954.57879	93.166	29.133
8	2284.08	-2957.23821	-2954.58114	93.166	-8.519
9	2431.77	-2957.40992	-2954.58106	93.166	-48.192
10	2438.54	-2957.41700	-2954.58026	93.166	-86.422
11	2316.90	-2957.27302	-2954.57779	93.166	-118.168
12	2095.96	-2957.01291	-2954.57469	93.166	-145.360
13	1810.62	-2956.67849	-2954.57220	93.166	-166.635
14	1493.81	-2956.30810	-2954.57036	93.166	-184.229
15	1173.12	-2955.93339	-2954.56870	93.166	-199.845
16	873.37	-2955.58295	-2954.56697	93.166	-212.241
17	615.48	-2955.28182	-2954.56584	93.166	-221.756
18	414.73	-2955.04759	-2954.56514	93.166	-228.956
19	281.21	-2954.89192	-2954.56480	93.166	-234.651
20	220.45	-2954.82120	-2954.56476	93.166	-238.258
21	233.83	-2954.83693	-2954.56492	93.166	-240.579
22	319.50	-2954.93689	-2954.56522	93.166	-240.647

Example for MgCoO<sub>3</sub> in the rhombohedral structure

# Output of a Molecular Dynamic in Siesta:

<b>SystemLabel.MD</b>	Atomic coordinates and velocities (and lattice vectors and their time derivatives if the dynamics implies variable cell). (unformatted; post-process with iomd.F)
<b>SystemLabel.MDE</b>	shorter description of the run, with energy, temperature, etc. per time step
<b>SystemLabel.ANI</b>	(contains the coordinates of every Molecular Dynamics step in xyz format)

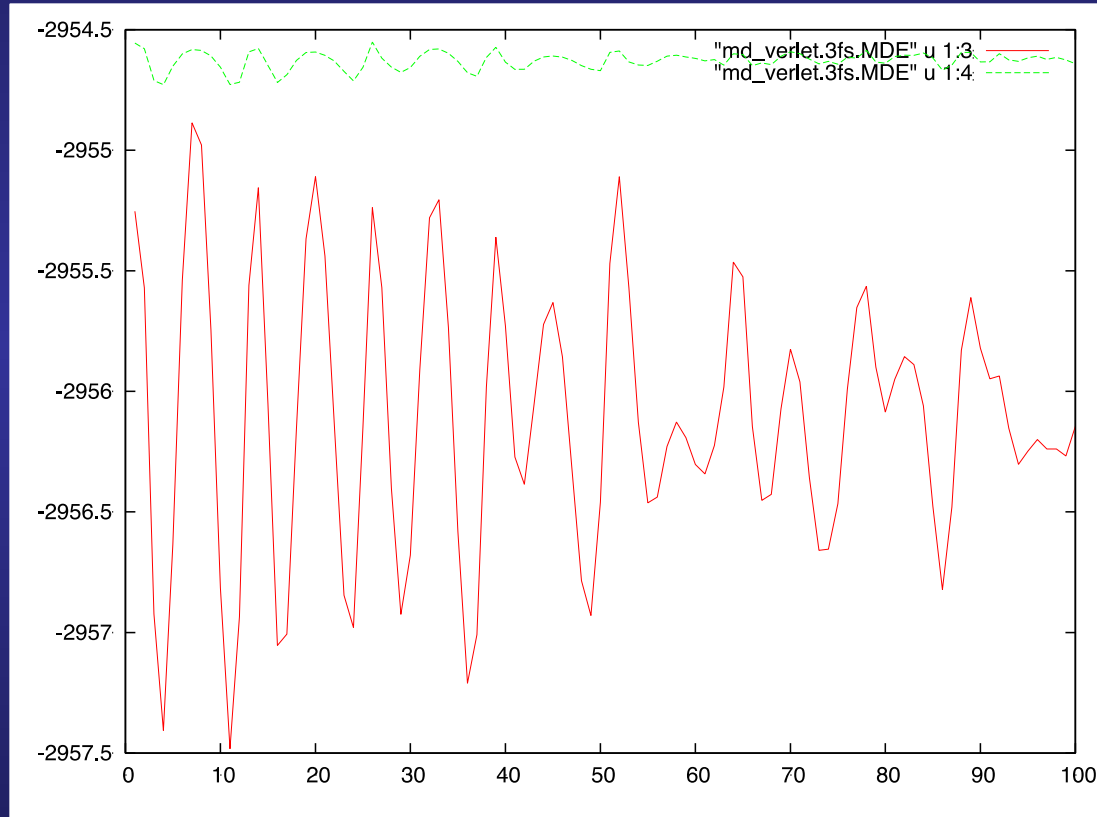
These files are accumulative even for different runs.  
Remember to delete previous ones if you are not interested on them

# Check conservation of energy

Length of time step: 3 fs

```
$ gnuplot
```

```
$ gnuplot> plot "md_verlet.MDE" using 1:3 with lines, "md_verlet.MDE" using 1:4 with lines
```



Compare:  
**Total energy**  
with  
**KS energy**

```
$ gnuplot> set terminal postscript color
```

```
$ gnuplot> set output "energy.ps"
```

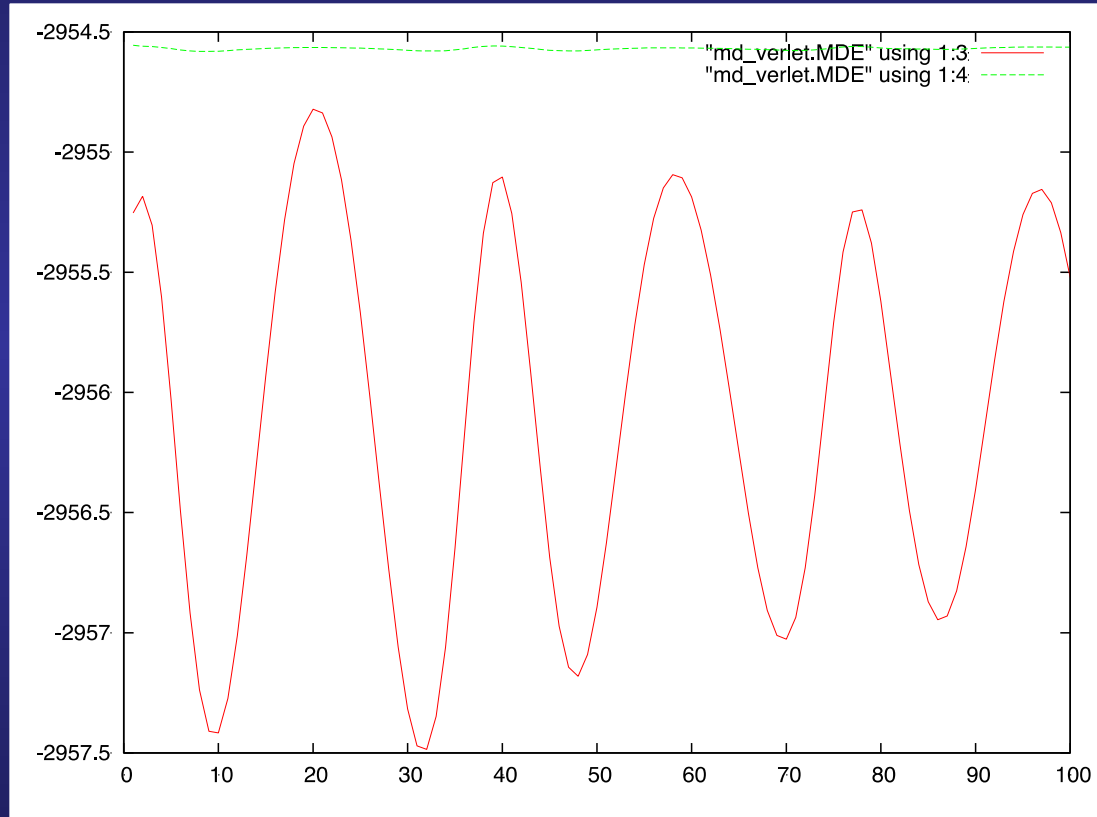
```
$ gnuplot> replot
```

# Check conservation of energy

Length of time step: 1 fs

```
$ gnuplot
```

```
$ gnuplot> plot "md_verlet.MDE" using 1:3 with lines, "md_verlet.MDE" using 1:4 with lines
```



Compare:  
**Total energy**  
with  
**KS energy**

```
$ gnuplot> set terminal postscript color
```

```
$ gnuplot> set output "energy.ps"
```

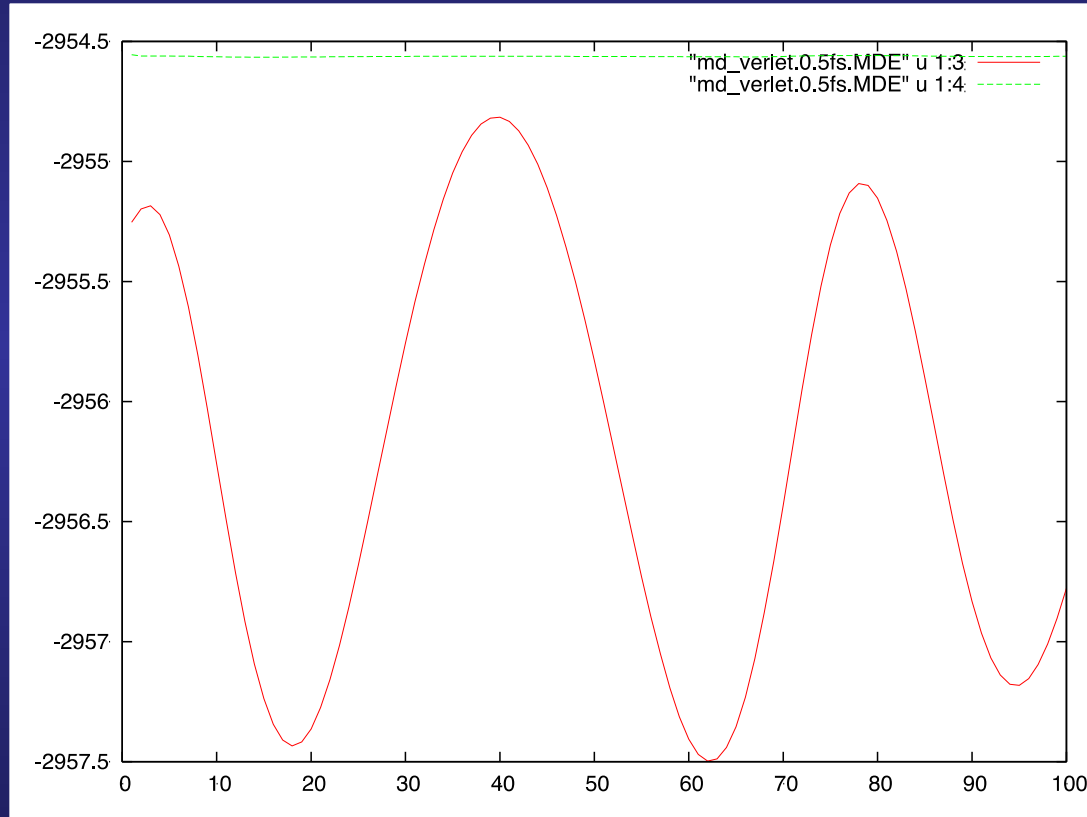
```
$ gnuplot> replot
```

# Check conservation of energy

Length of time step: 0.5 fs

```
$ gnuplot
```

```
$ gnuplot> plot "md_verlet.MDE" using 1:3 with lines, "md_verlet.MDE" using 1:4 with lines
```



Compare:  
**Total energy**  
with  
**KS energy**

```
$ gnuplot> set terminal postscript color
```

```
$ gnuplot> set output "energy.ps"
```

```
$ gnuplot> replot
```

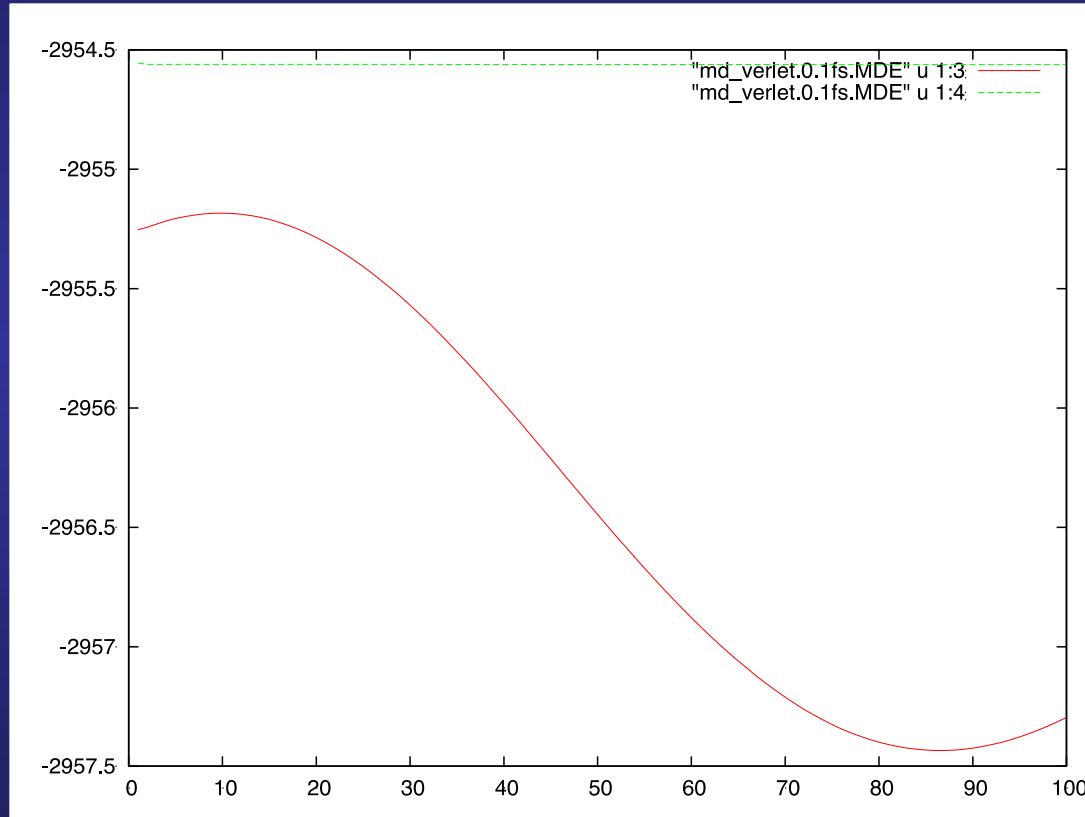


# Check conservation of energy

Length of time step: 0.1 fs

```
$ gnuplot
```

```
$ gnuplot> plot "md_verlet.MDE" using 1:3 with lines, "md_verlet.MDE" using 1:4 with lines
```



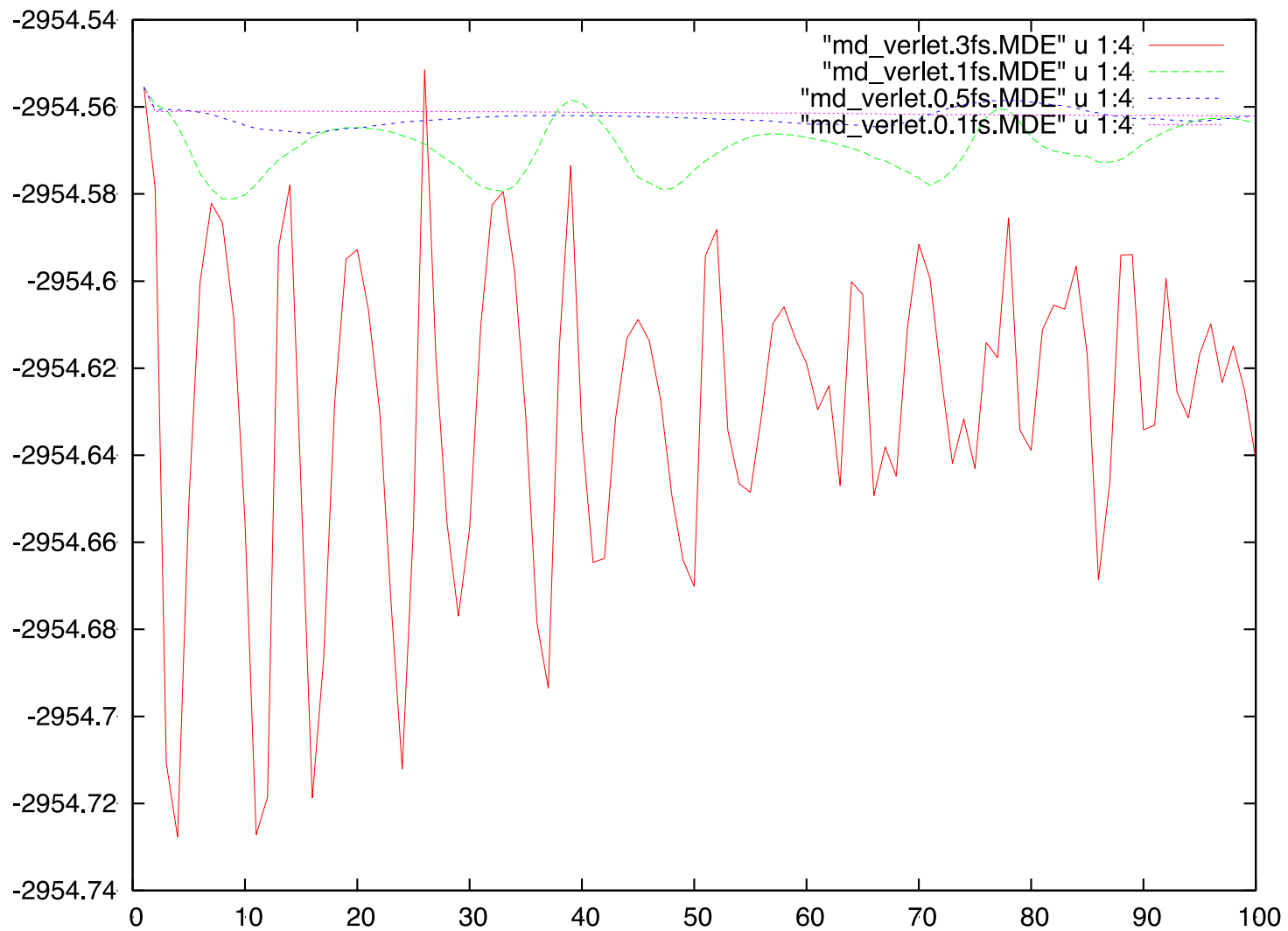
Compare:  
**Total energy**  
with  
**KS energy**

```
$ gnuplot> set terminal postscript color
```

```
$ gnuplot> set output "energy.ps"
```

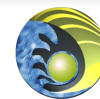
```
$ gnuplot> replot
```

# Check conservation of energy





# Full solvation: Ab Initio Molecular Dynamics



X. Shen, Y.A. Small, J. Wang,  
P. B. Allen, M.V. Fernandez-Serra,  
M. S. Hybertsen, and J.T. Muckerman,  
*J. Phys. Chem. C* 114, 13695 (2010).

J. Wang, L. Pedroza, A. Poissier, and  
M. Fernandez-Serra, *JPCB* (2012).

4x3x5 GaN unit cells

24 GaN surface sites

94 Water molecules

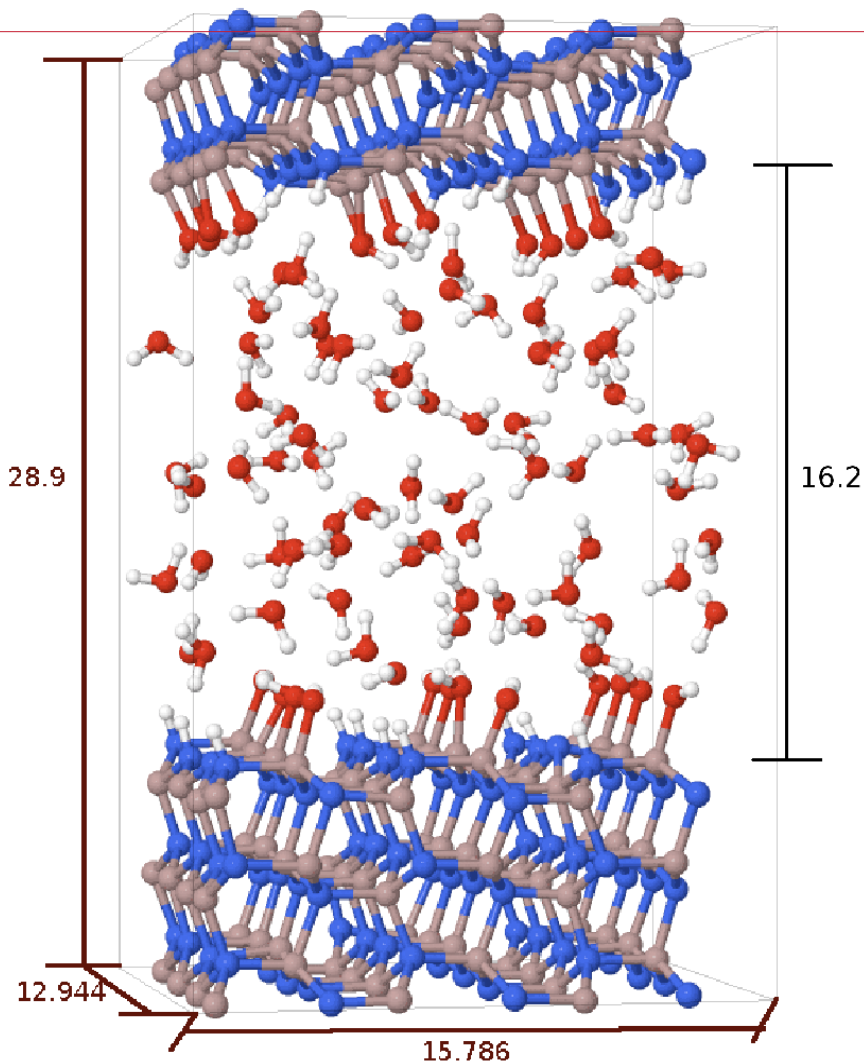
240 GaN atoms

MD Method: DFT (PBE)

Package: SIESTA

AIMD Equilibration: 2ps

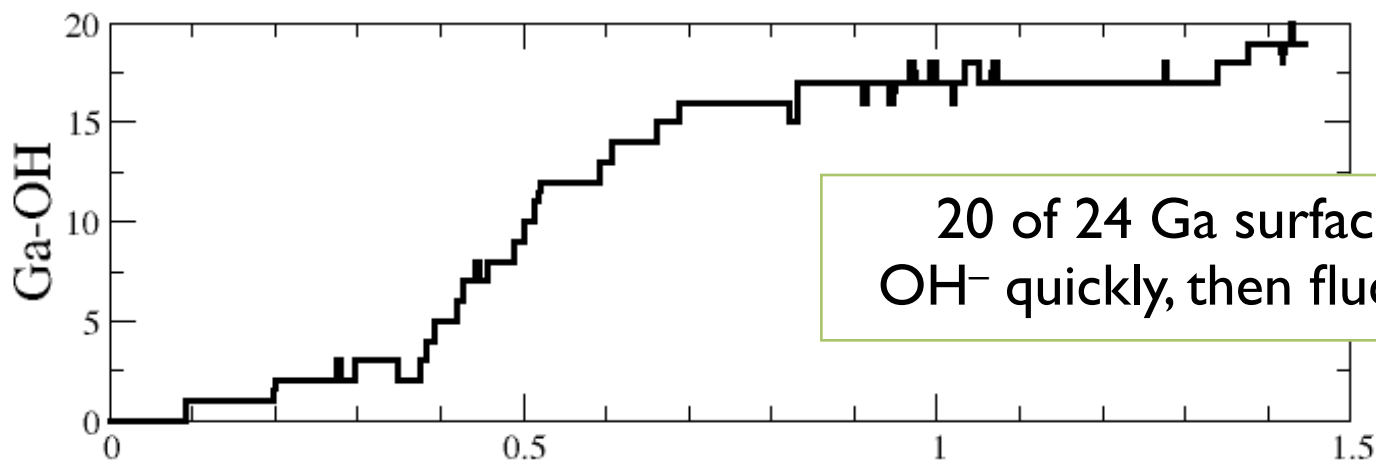
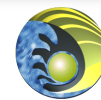
AIMD production: 10ps



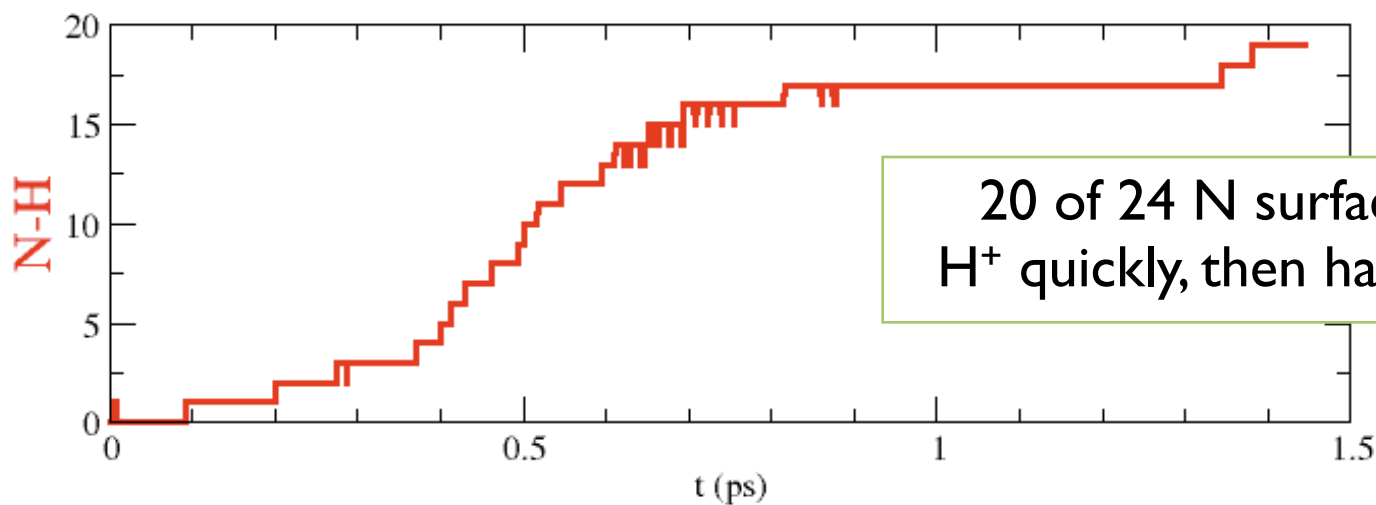
Dissociation occurred in first 1 ps;  
initially undissociated.



# Early history of GaN/liquid water interface equilibration: dissociation of water seen in AIMD run



20 of 24 Ga surface sites bind  $\text{OH}^-$  quickly, then fluctuate slightly.



20 of 24 N surface sites bind  $\text{H}^+$  quickly, then hardly fluctuate.



# GaN/Aqueous Interface: H-bond network

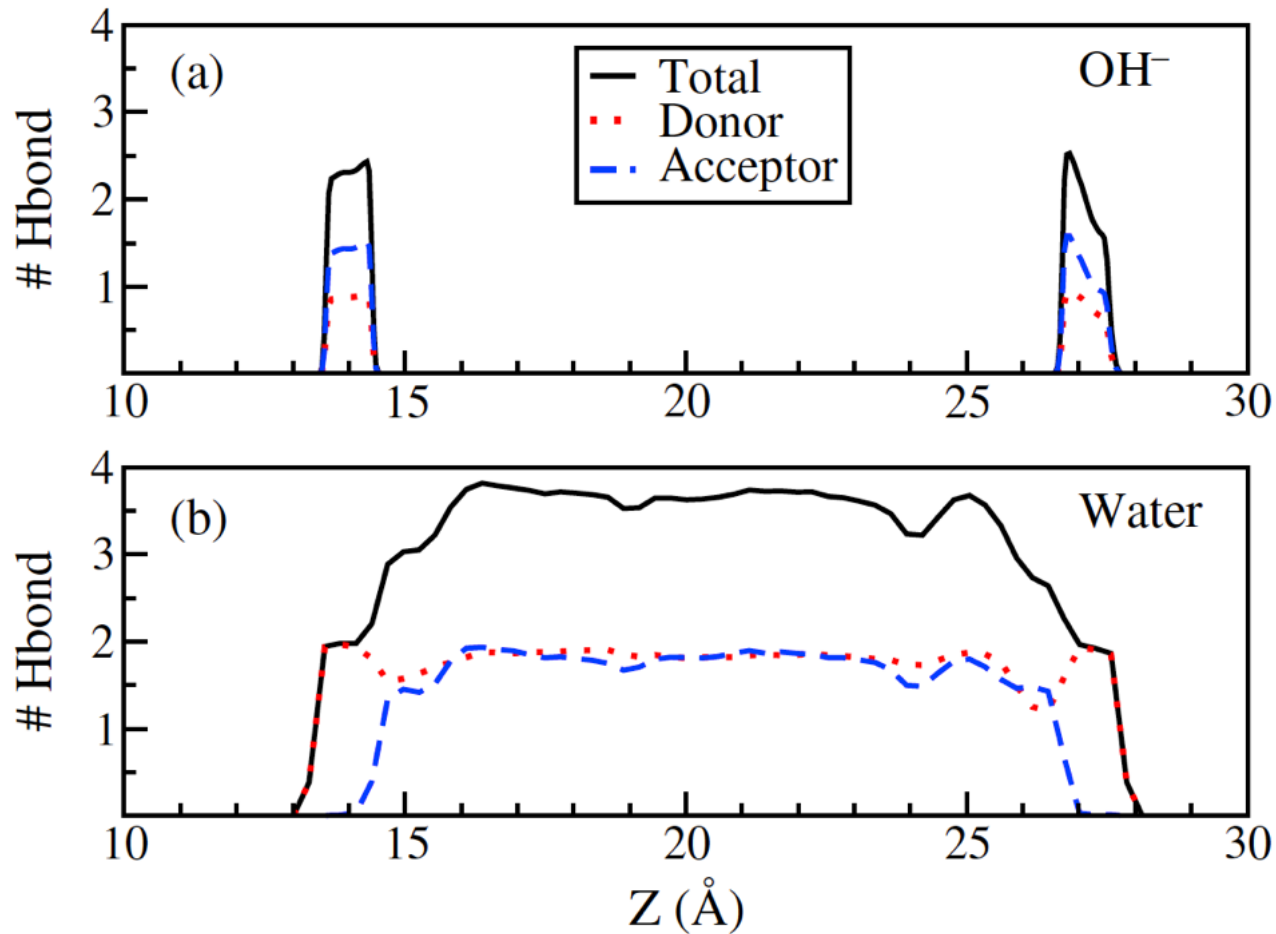


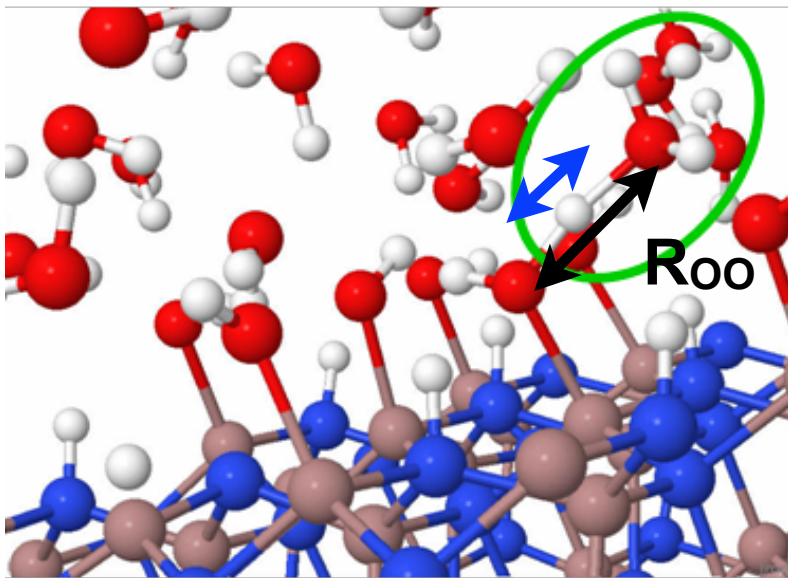
Figure 3: Average number of hydrogen bonds along the direction perpendicular to the GaN surface for (a) the surface OH groups, and (b) all water molecules. Curves display the total number of Hbond (solid), the donor of Hbond (dotted) and the acceptor of Hbond (dashed).



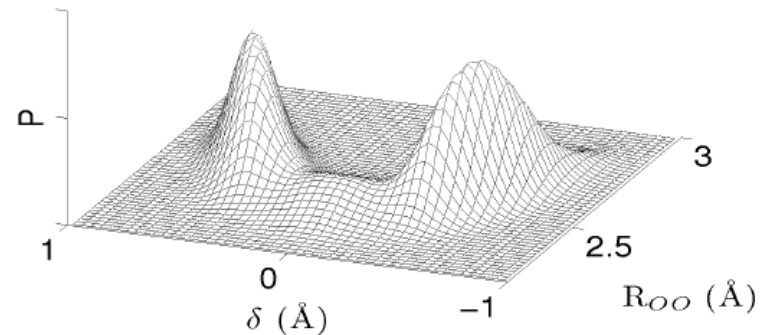
# Proton diffusion



$$\Delta F = -k_b T \ln \left\{ \int dR_{OO} P(R_{OO}, \delta) \right\}$$

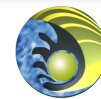


$$\delta = R_{O_a H} - R_{O_b H}$$

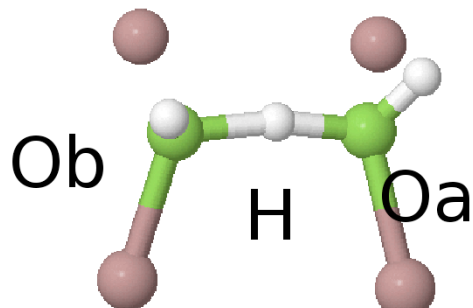




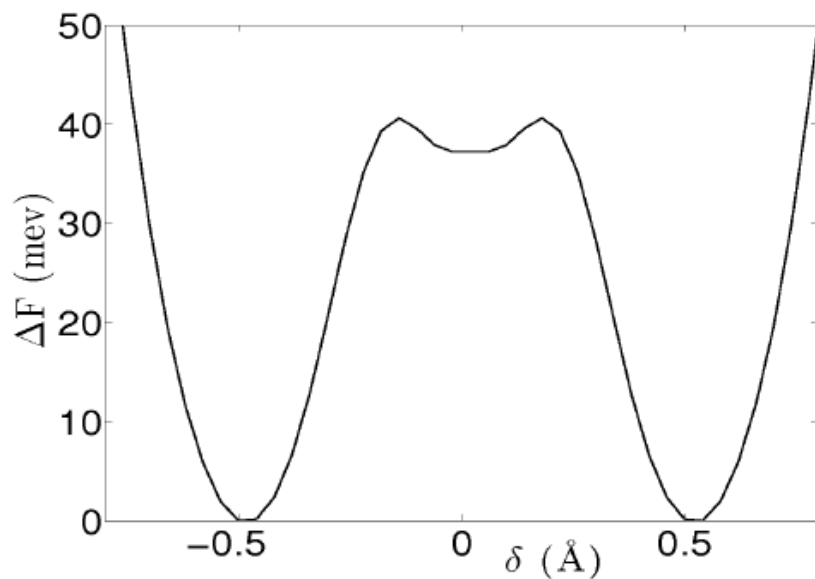
# Surface-Surface diffusion



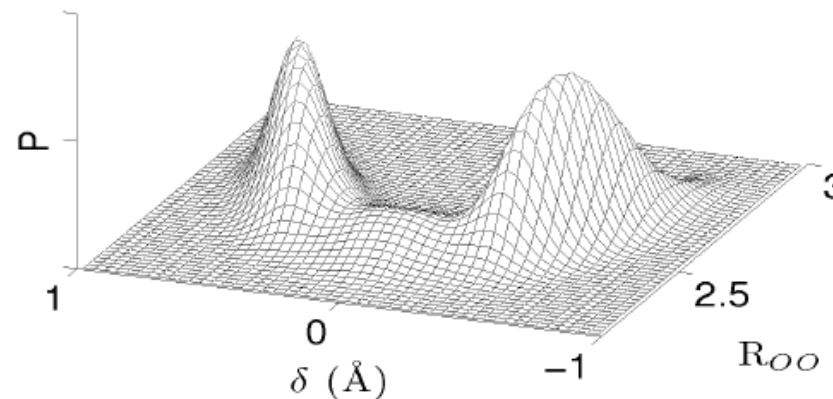
## Diff<sub>S-S</sub>



Symmetric Distribution  
at  $\delta \sim 0$  there is non zero probability:  
centrosymmetric complexes of the Zundel  
( $\text{H}_5\text{O}_2^+$ ) type. Barrier  $\sim 40$  meV

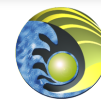


(a)

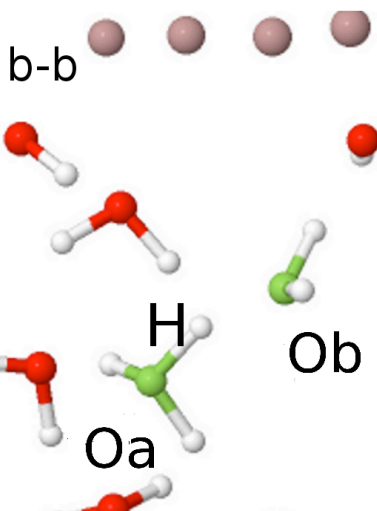




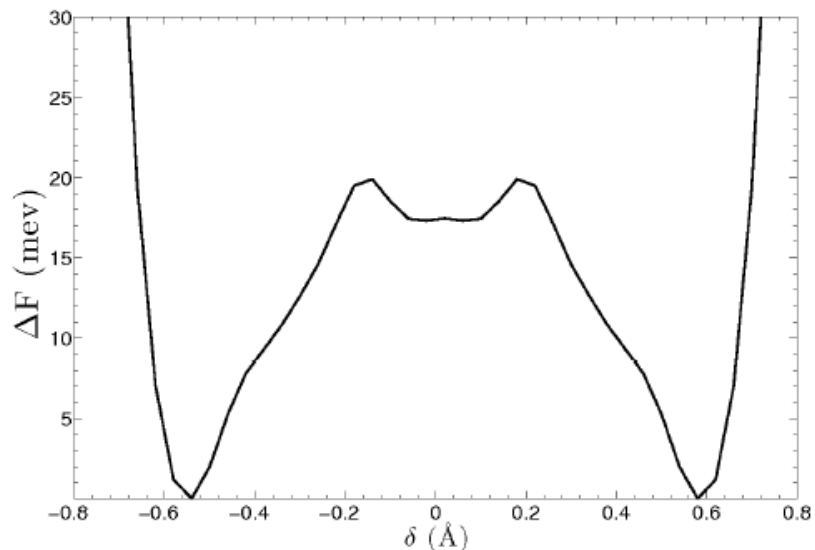
# bulk-bulk diffusion



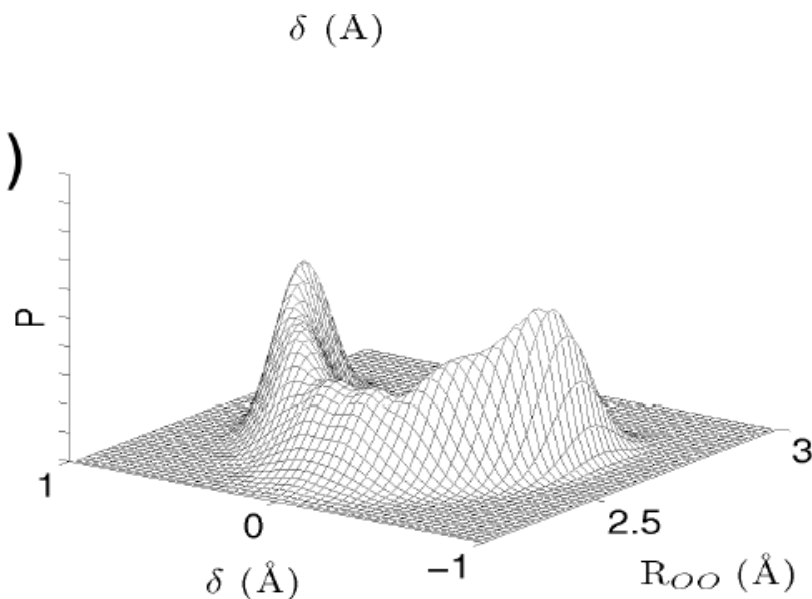
Diff



Symmetric Distribution  
at  $\delta \sim 0$  there is non zero probability:  
centro-symmetric complexes of the Zundel  
( $\text{H}_5\text{O}_2^+$ ) type. Barrier  $\sim 20$  meV



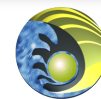
(c)



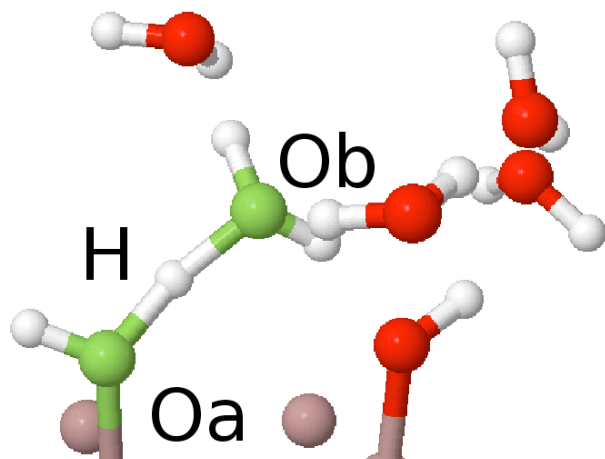




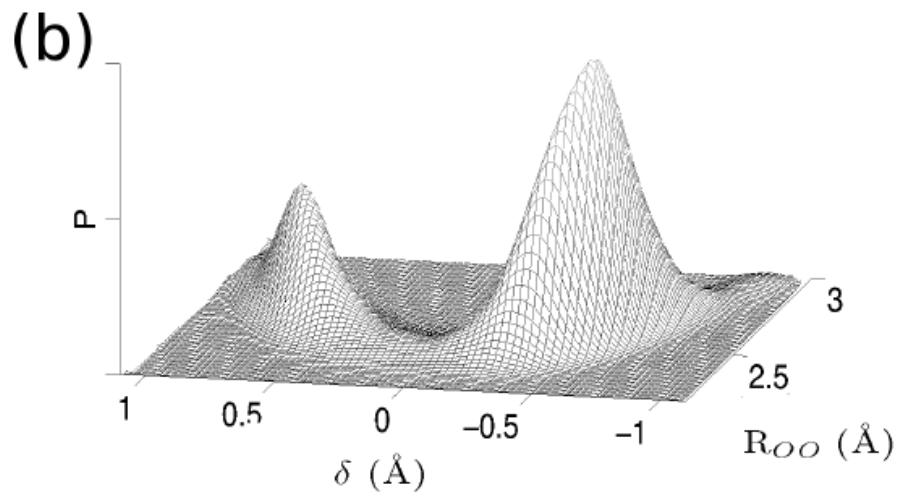
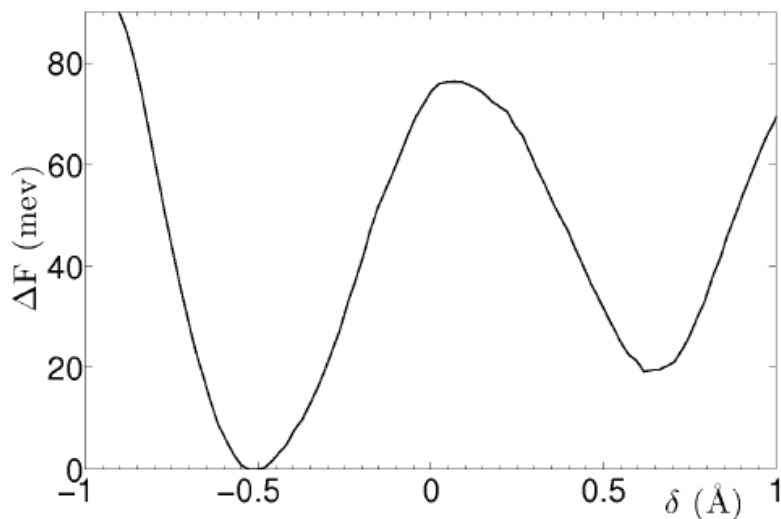
# Surface-bulk diffusion



## Diff<sub>s-b</sub>

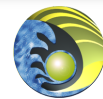


Asymmetric Distribution, Asymmetric Zundel complex  
at  $\delta \sim 0$  there is  $\sim 0$  probability:  
Barrier out  $\sim 75$  meV.  
Barrier in  $\sim 55$  meV





# Deprotonation free energy barrier and pKa



$$\Delta G^{(0)} = -k_B T \ln \left\{ C_0 \int_0^{R_{cut}} dR A(R) \exp[-\beta \Delta F(R)] \right\}$$

$$\text{pK}_a = -\log_{10} \exp(-\beta \Delta G^{(0)})$$

$$\text{pK}_a = 2.9 \pm 0.1$$

Once H<sup>+</sup> leave the surface, the barriers to diffuse within the bulk liquid region are smaller than those to return to the surface. This agrees with the large acidity obtained.



# PHONONS FROM MOLECULAR DYNAMICS

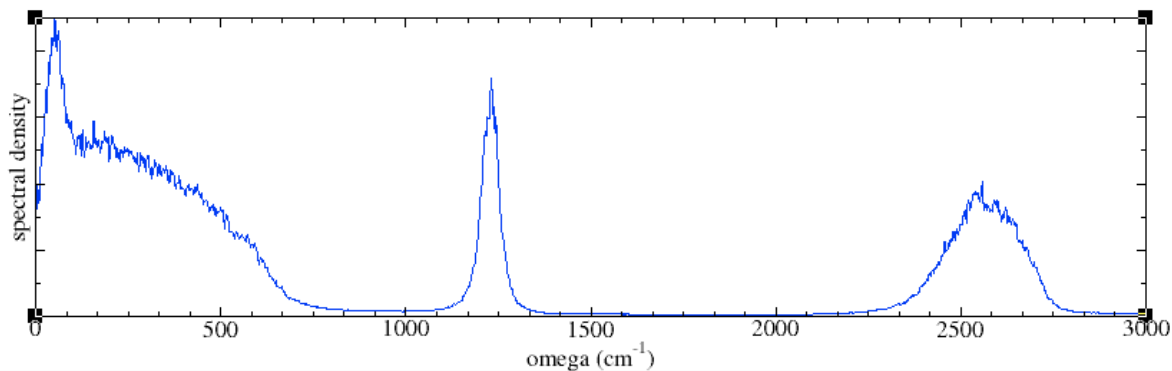


$$v_{\alpha,k}(t) = \left(\frac{\hbar}{NM_k}\right)^{1/2} \sum_{j=1}^3 e_{\alpha,k}^j [-i\omega_j] e^{-i\omega_j t} [2\omega_j]^{-1/2} [a_j + a_j^\dagger]$$

$$\sum_{\alpha,k} M_k \langle v_{\alpha,k}(t) v_{\alpha,k}(0) \rangle = -\frac{1}{V} \sum_{j=1}^3 e^{-i\omega_j t} \langle \hbar\omega_j [n_j + 1/2] \rangle$$

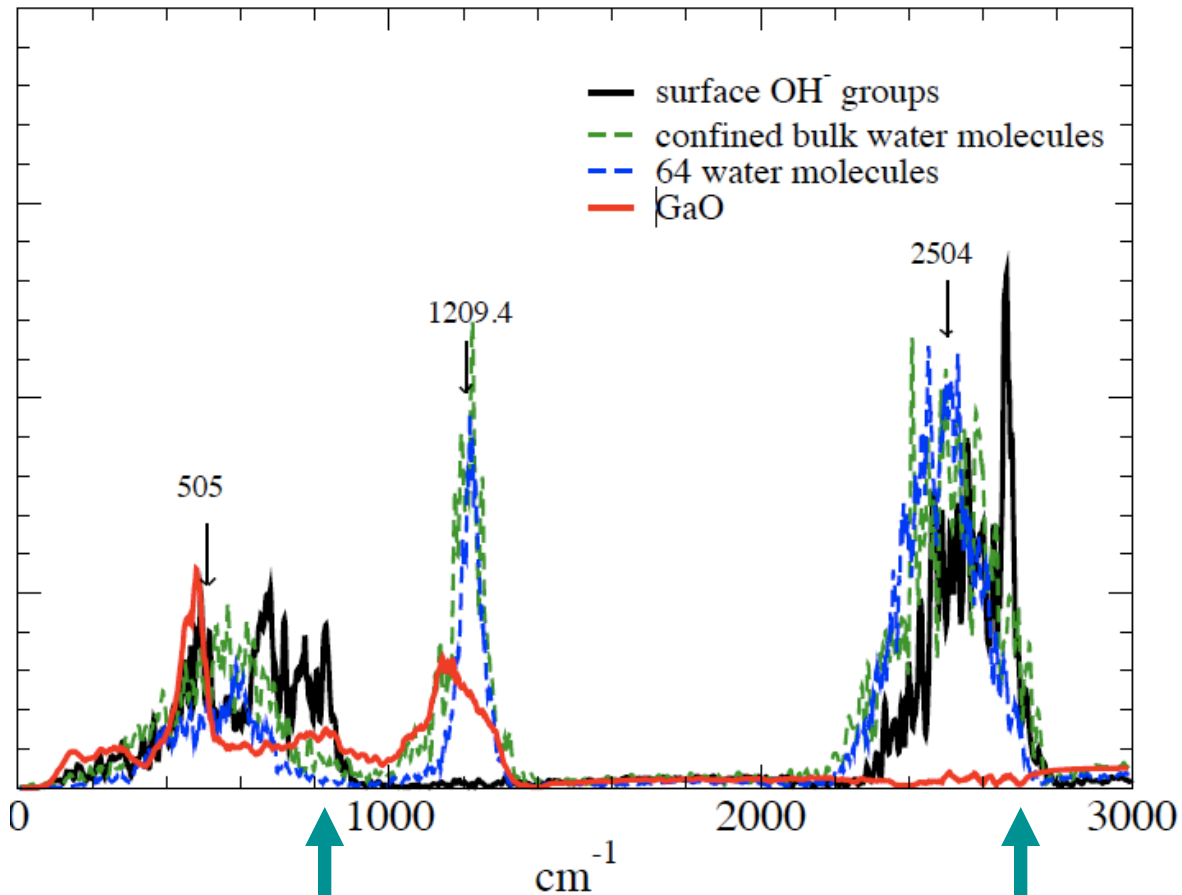
Fourier Transform

$$\sum_{\alpha,k} M_k \langle |v_{\alpha,k}(\omega)|^2 \rangle = -\frac{1}{V} K_b T \sum_{j=1}^3 \delta(\omega - \omega_j)$$





# Vibrational Spectrum



2504  $\text{cm}^{-1}$ -- combination of stretching, anti-stretching and 2\*bending modes

1209.4  $\text{cm}^{-1}$ -- bending mode

505  $\text{cm}^{-1}$ -- librations

New librational modes associated to surface  $\text{OH}^-$ s

Surface  $\text{OH}^-$  stretching mode