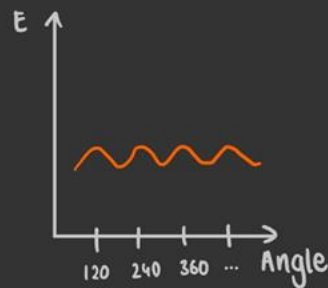
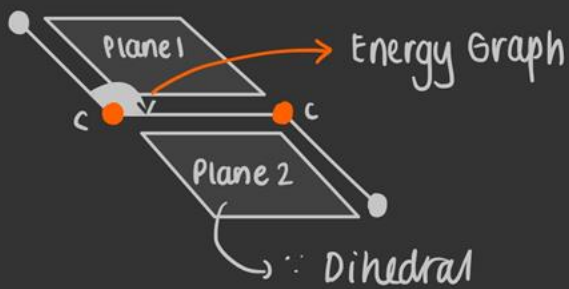


15-C

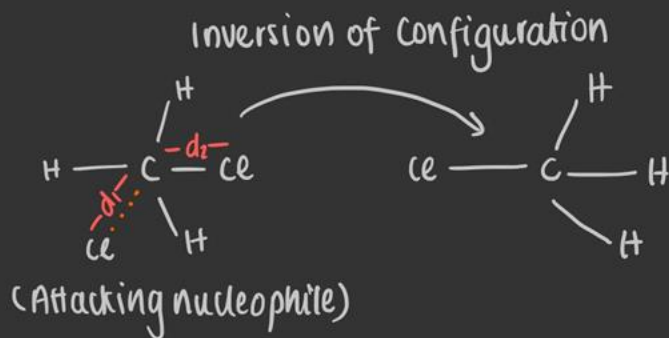
T H E O R Y



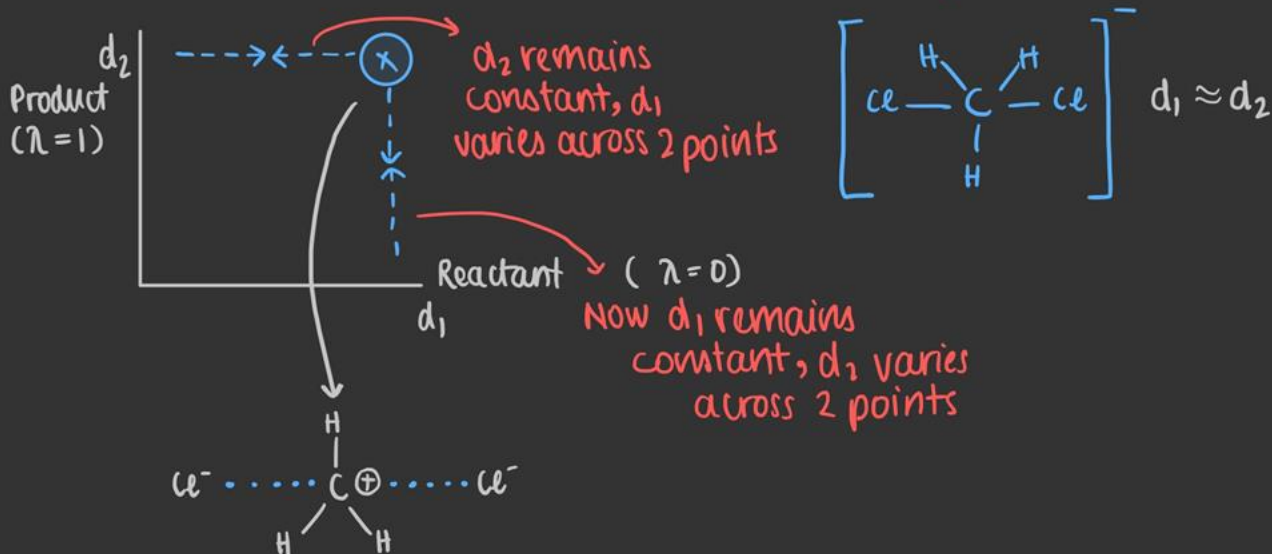
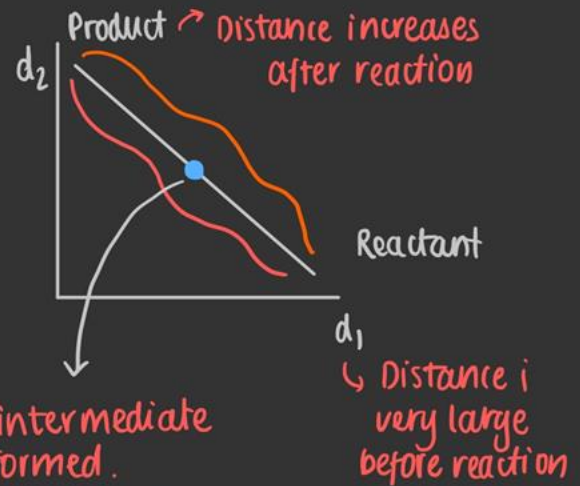
Lecture 1



SN₂ Reaction

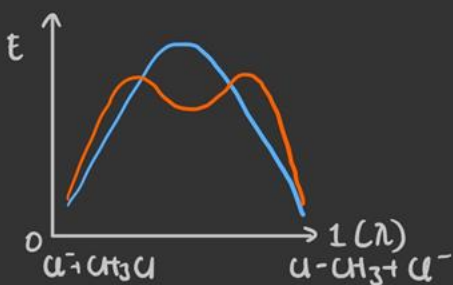


The process can take an alternate path too. (Not SN₂)



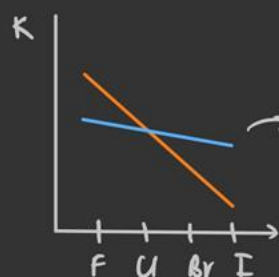
The change of energy state could take one or more forms, but at $\lambda=1$ and $\lambda=0$, the energy level must be the same.

- SN₂
- SN₁



Arrhenius' Law

Rate of Reaction : $k \propto Ae^{-E_a/RT}$



One of the lines represents SN_2 reaction whereas other an SN_1 .

Obtained via quantum calculation

Lecture 2

Subject: CIS1 $\begin{cases} \text{Classical Mechanics} \\ \text{Quantum Mechanism} \end{cases}$

Classical Mechanics MOSTLY suffices for biomolecular simulation.

Newtonian mechanics cares about $\begin{cases} \text{Mass } [m] \\ \text{Position } [r] \end{cases}$

\rightarrow Cartesian coordinates.

Velocity: $\frac{d\vec{r}(t)}{dt} = \frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}$

[Each of these vary with time]



Acceleration: $\frac{d^2\vec{r}(t)}{dt^2} = \frac{d^2x}{dt^2} + \frac{d^2y}{dt^2} + \frac{d^2z}{dt^2}$

$\left[\frac{\Delta x}{\Delta t} \approx \frac{dx}{dt} \text{ if } \Delta t \rightarrow 0 \right]$
Slope \approx Derivative

Momentum: $m\vec{v}$. At a time t , $p_i = m_i v_i$ (useful if mass varies)

Net Force on Object: (F_{ext}) Independent on object of concern, doesn't depend on internal properties.
 $F_{ext} = \frac{dp_i}{dt}$

Interactions Possible w/ object

- (i) Electrical
- (ii) Gravitational (for large enough bodies)
- (iii) Hooke's law: Spring forces.
- \vdots

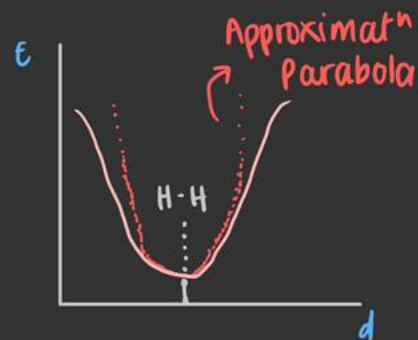
Atomic Structure

Nucleus $\left\{ \begin{array}{l} \text{Protons } (+1) \rightarrow \text{Det. net charge in nucleus.} \\ \text{Neutrons } (+0) \rightarrow \text{Mass, stability of nucleus} \end{array} \right.$

Neutrons DON'T determine chemical properties.

H_1^1, H_1^2, H_1^3 \rightarrow same reactivity and chem prop.

Quantum Chemistry (orbitals and structure)

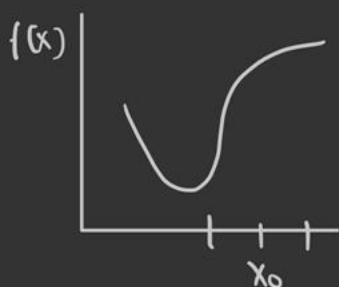


CRUCIAL OBSERVATION! We attempt to model all atoms in a classical mechanics approach using a spring. \rightarrow Most of the times



Assume spring to have constant 'k'. k is obtained by fitting the energy function (as above) into a parabola

Assume function $f(x)$



Taylor expansion: Approximating function about a given point.

Find $f(x_0)$. Then $f'(x_0)$, $f''(x_0)$... until required.

Let $x, x_0 \in [a, b]$, $f(x) = f(x_0)$

Now, given $x \in [a, b]$, $f(x) = ?$
 $x_0 \in [a, b]$, $f(x) \approx f(x_0)$ (0th order approx.)

$f(x) \rightarrow$ Accuracy w/ more terms.

$\Rightarrow f(x_0) + \frac{df}{dx}(x-x_0)$ (1st order approx.)

$\Rightarrow f(x_0) + \frac{df}{dx}(x-x_0) + \frac{1}{2} \left(\frac{d^2f}{dx^2} \right) (x-x_0)^2$ (2nd order approx)

\vdots

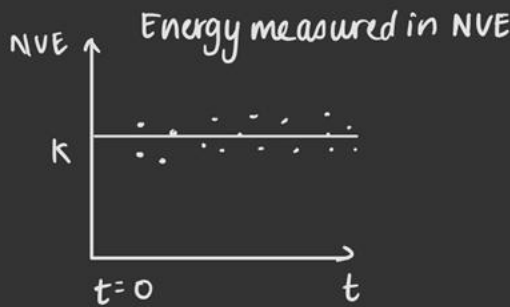
\rightarrow This equation is parabolic.

At minima/maxima, $\left(\frac{df}{dx} \right)_{x_0} = 0$. We can eliminate encircled term. from eq above.

Lecture 5

Constant in Simulation → NVT Number volume Temp
NVE Number volume Energy

- Molecular dynamics → Biomolecular dynamics
- Force field: Setup system / PBC / Temperature minimise



We can measure the same in an NVT simulation also, there E may not be constant

Why? otherwise it would imply non constant temperature.

Trajectory could be of 2 forms: ① At equilibration phase (not constant phase)
② During production phase → determines equilibrium state.

Thermodynamics: Free energy, Average energy, Heat capacity, Entropy

$$E = \frac{1}{M} \sum_i E_i \quad \left| \quad (\Delta E)^2 = (E - \bar{E})^2 = \frac{1}{M} \sum_{i=1}^M (E_i - \bar{E})^2 = \frac{1}{M} \sum_{i=1}^M E_i^2 - \bar{E}^2 \right.$$

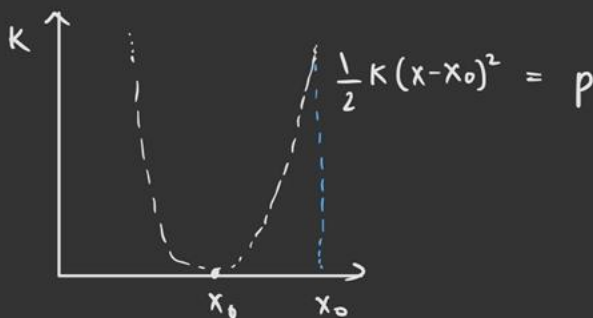
(Proof at the end)

Number of configurations 'Cv'

$$\bar{K} = \frac{1}{M} \sum_i K_i \quad \rightarrow \text{Average KE (Equipartition Theorem)}$$

$$\bar{K} = N \cdot d \cdot \frac{1}{2} K_B T = \frac{3}{2} N K_B T$$

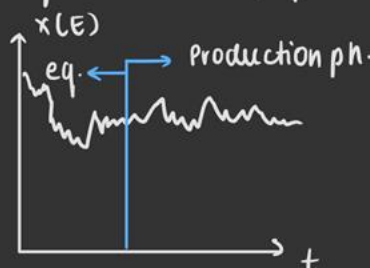
Dimensionality (3 for normies)



simulation at an arbitrary temperature, so, $\vec{p} =$

$$-K(x-x_0) - \gamma \vec{p} + F_R$$

over time, x_0 varies as:



Frictional Const.

$$\dot{x} = \frac{1}{m} p$$

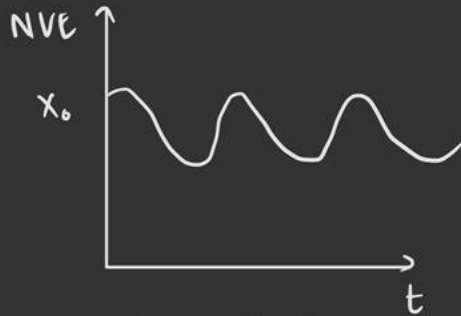
Here, $\epsilon(t)$ is not a constant

(This represents a 1D-STM)

$$\frac{p^2}{2m} + \frac{1}{2} k(x-x_0) \quad [p(t), x(t)]$$

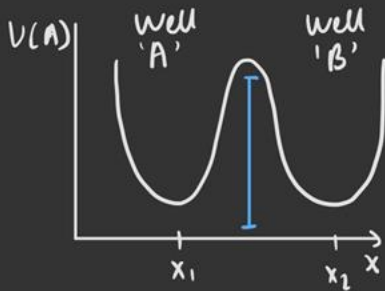
(from the above page.
eqⁿ)

The NVM varies as:

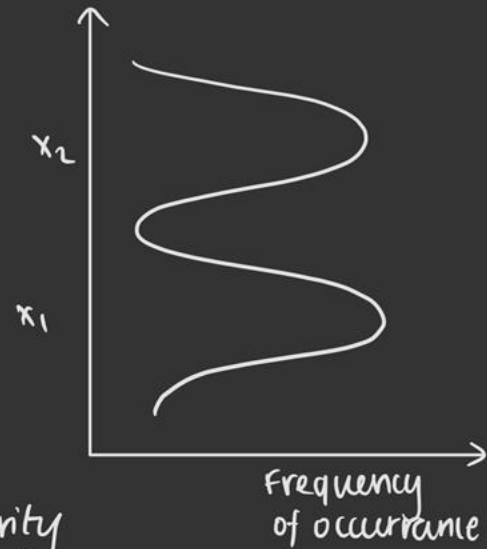
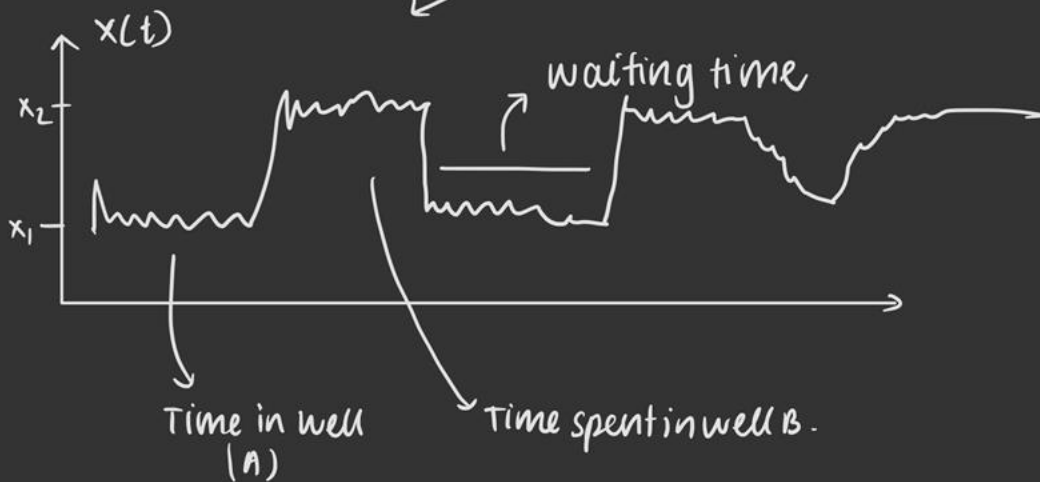


$$x(t) = x_0 + (x(0) - x_0) \cdot \cos \omega t$$

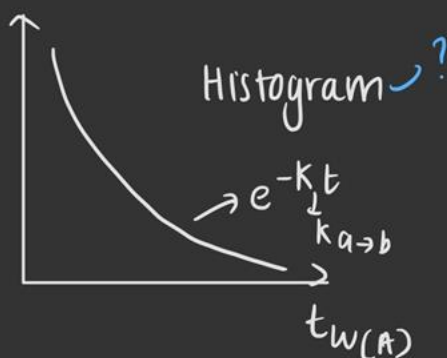
$\Delta U^\ddagger \sim 10kT$ (P.E not at eq)



Such a reaction is like a stochastic process.
some states are preserved for longer periods
of times.



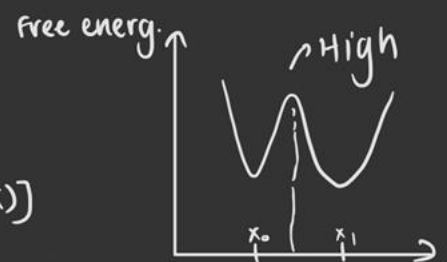
• states that have less free energy. they occur for the majority
of time.



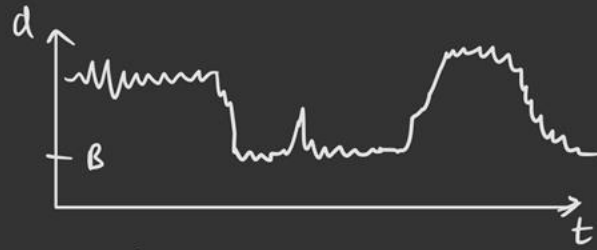
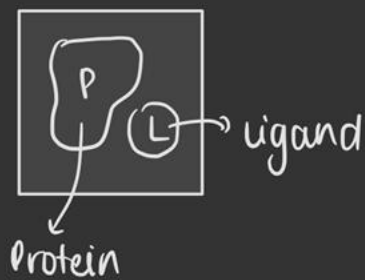
Landau's free energy:

$$F(x) = -kT \ln(P(x))$$

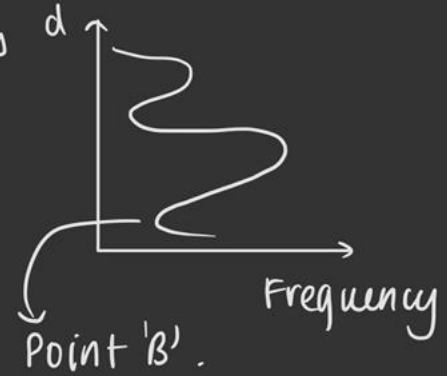
Highest when not at x_1 or x_2 .



Order Parameter (λ)



↓ interpretable as



$$d_i = |\bar{R}_P^{CH}(i) - \bar{R}_L^{CH}(i)|$$

$$\bar{R}_A^{CM} = \frac{1}{M_A} \sum_{i \in A} m_i \bar{r}_i$$

↓ 'A' molecule
 protein ligand

Order parameter: λ

① of a process

$\lambda = 0 \rightarrow$ 'start'

$\lambda = 1 \rightarrow$ 'end'

② Statistics on these \rightarrow properties.

Proof

$$\left(\frac{1}{M} \sum_{i=1}^M (E_i - \bar{E})^2 = \frac{1}{M} \sum_{i=1}^M (E_i^2 - \bar{E}^2) \right)$$

$$\frac{1}{M} \left(\sum_{i=1}^M E_i^2 + \sum_{i=1}^M (\bar{E})^2 - 2 \sum_{i=1}^M E_i \bar{E} \right)$$

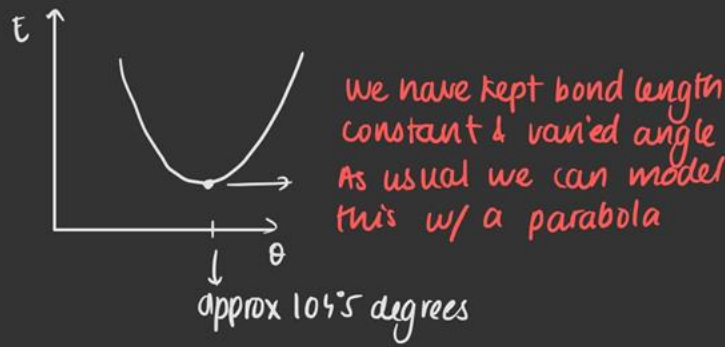
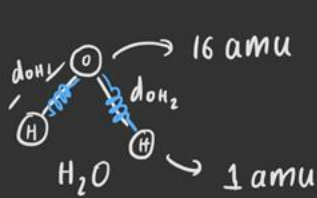
$$\Rightarrow \frac{1}{M} \left(\sum_{i=1}^M E_i^2 + M(\bar{E})^2 - 2\bar{E} \sum_{i=1}^M E_i \right)$$

$$\Rightarrow \frac{1}{M} \left(\sum_{i=1}^M E_i^2 + M(\bar{E})^2 - 2\bar{E}(\bar{E}M) \right)$$

$$\Rightarrow \frac{1}{M} \sum_{i=1}^M E_i^2 + (\bar{E})^2 - 2(\bar{E})^2$$

$$\Rightarrow \frac{1}{M} \sum_{i=1}^M E_i^2 - (\bar{E})^2$$

Lecture 3



$$V_{eq} = \frac{1}{2} k_{\text{OH}} (d_{\text{OH}1} - d_{\text{OH}}^0)^2 + \frac{1}{2} k_{\text{OH}} (d_{\text{OH}2} - d_{\text{OH}}^0)^2 + \frac{1}{2} k_{\text{OH}} (\theta - \theta_0)^2$$

Parameterisation for equilibrium condition.

Here the particles' V cancels out

If config. of atoms is given, position, etc,
we can calculate the force on i^{th} particle as follows:

$$\vec{F}_i(\vec{r}_1, \vec{r}_2, \vec{r}_3) = (-) \left(\frac{\partial V}{\partial \vec{r}_i} \right)_{j \neq i} = - \left(\frac{\partial V}{\partial x_i}, \frac{\partial V}{\partial y_i}, \frac{\partial V}{\partial z_i} \right)$$

we use an approximation to find \vec{F}_i , using which we will be able to calculate the momentum of particles at a time t .

$$\vec{p}_i = m \frac{d\vec{r}_i}{dt} \quad \& \quad \frac{d\vec{p}_i}{dt} = \vec{F}_i \quad \Rightarrow \quad \vec{p}_i(t + \Delta t) \approx \vec{p}_i(t) + \Delta t \vec{F}_i(t)$$

(momentum after an infinitesimal time t)

$$\frac{d\vec{r}_i}{dt} \approx \frac{\Delta \vec{r}_i}{\Delta t} = \frac{[\vec{r}_i(t + \Delta t) - \vec{r}_i(t)]}{\Delta t} = \frac{1}{m} \vec{p}_i(t)$$

using above, we can obtain:

$$\frac{d\vec{r}_i}{dt} = \frac{1}{m} \vec{p}_i(t) \quad \Rightarrow \quad \vec{r}_i(t + \Delta t) = \frac{1}{m} \vec{p}_i(t) \cdot \Delta t + \vec{r}_i(t)$$

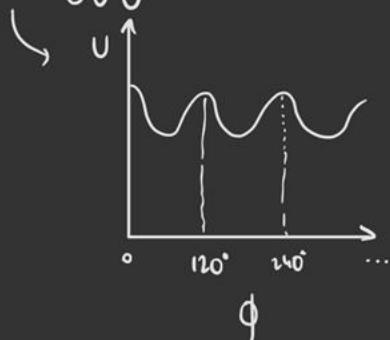
We have not made any assumptions about the particle, and from a classical mechanics standpoint, this suffices for any general particle.

Let's consider an ethane molecule



'H' positions' variation can be evaluated using torsion angle.

Potential energy graph: A 'combination of cosines'



$$\text{Sum} = \sum_i K_i \cos(n\phi - \phi_0)$$

This sum represents torsion

K_i : Some calculated unique/arbitrary const.

ϕ_0 : Different baseline (Eq. librium) angles? eg. H-C-H has one C-H has another } Dihedral angle

n : changes for each molecule

(3 for ethane)

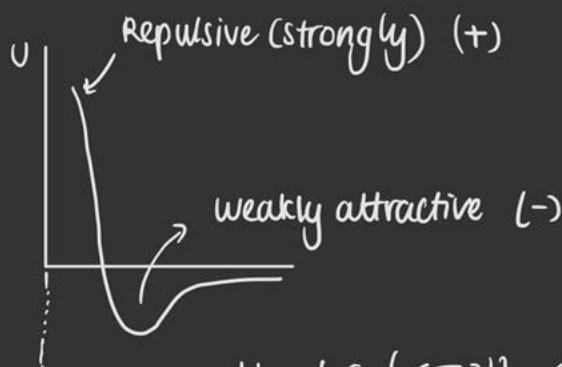
(number of times central angle has to multiply to obtain original config)



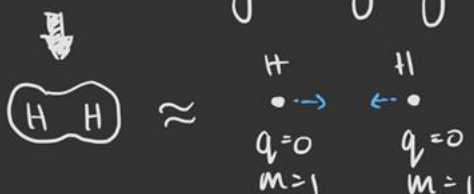
Types of bonds and bond angles in C_2H_6

- | | | |
|---|-------|---------------|
| 1 | C-C | } Bond angles |
| 6 | C-H | |
| 6 | H-C-H | |
| 6 | H-C-C | |

Lenneard Jones' equation potential energies



Partial charges & Hydrogen



$$U = 4\epsilon \left(\left[\frac{\sigma}{r} \right]^{12} - \left[\frac{\sigma}{r} \right]^6 \right)$$

(Lennard Jones')

Here we observe partial charges
The charge on each atom will be at most +1 or -1. (Almost never possible)



Lecture 4

Biomolecular simulation \rightarrow we have motivated the construction of PE as some fn's

Total PE \rightarrow Covalent Bonding (Modelled as a spring)
 \rightarrow Angle bending (same)
 \rightarrow Bond rotation (Quadratic graph)
 \rightarrow Non bonding (Leonard Jones' Vanderwaal's forces)

These are the 4 important interactions.

Potential Energy

$U(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_N) \equiv U(\vec{r}^N) \rightarrow$ This is for the centre of mass
 \downarrow
 Position vector of 1st particle wrt. chosen arbitrary centre. Potential energy of N atoms.

$$\vec{F}_i = - \left(\frac{\partial U}{\partial \vec{r}_i} \right)_{j \neq i} \equiv - \underbrace{\left(\left(\frac{\partial U}{\partial x_i} \right), \left(\frac{\partial U}{\partial y_i}, \frac{\partial U}{\partial z_i} \right) \right)}_{\text{Combination of } 3N \text{ partial derivative}}$$

Equations of motion

- ① $\frac{d\vec{p}_i}{dt} = \vec{F}_i \quad \forall i \in [1, N] \rightarrow 3N \text{ such equations}$
- ② $\frac{d\vec{r}_i}{dt} = \frac{1}{m} \vec{p}_i \quad \forall i \in [1, N] \rightarrow 3N \text{ such equations.}$

We can also obtain integrators:

$$\left\{ \begin{array}{l} \vec{r}_i(t) \\ \vec{p}_i(t) \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \vec{r}_i(t+dt) \\ \vec{p}_i(t+dt) \end{array} \right\}$$

Euler scheme (explaining above)

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \frac{1}{m} \vec{p}_i(t) \cdot \Delta t \rightarrow \text{order of } \Delta t^2 \quad (O(\Delta t^2))$$

$$\vec{p}_i(t+\Delta t) = \vec{p}_i(t) + \vec{F}_i(t) \Delta t \rightarrow O(\Delta t^2)$$

Verlet's integration can be substituted for Euler's

\downarrow
 Because of approximation, we use till 2 terms. Higher order terms are removed.

In a simulation we save configuration every 100 or \times timesteps:

- Saves memory
- Each timestep represents an actual change in configuration.

Newton's second law: Total energy is conserved.

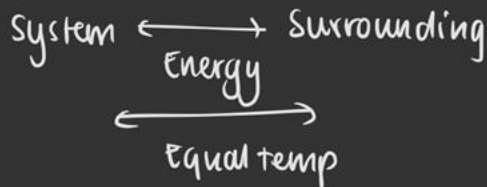
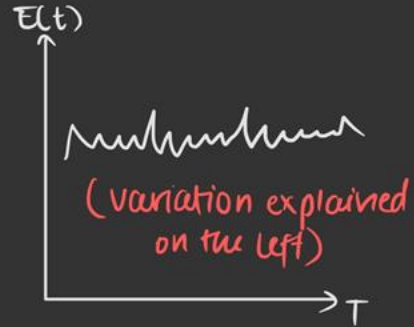
- $\{\bar{r}_i(t), \bar{p}_i(t)\} \rightarrow$ Parameters of i th particle at time t .

$$E = U(\bar{r}^N) + K(\bar{p}^N)$$

$\xrightarrow{\text{Kinetic energy}} \equiv \sum_{i=1}^N \frac{p_i^2}{2m}$

- A lot of collisions $\rightarrow 10^{23}$ ish per second could happen
- Some of them are elastic, but some inelastic.

- Total energy: Varying \because exchanges w/ surrounding.
- KE: Varying \because Inelastic collisions



$$\langle (E(t) - \bar{E})^2 \rangle$$

Related to C_v, T of the system

Langevin

$$\forall i, \frac{d\bar{p}_i}{dt} = \bar{F}_i \rightarrow \text{if the force is against momentum of particle it is frictional, and system loses some of the energy.}$$

$$\bar{F}_i - \gamma \bar{p}_i + \bar{F}_{R,i}$$

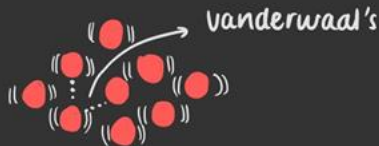
\swarrow \downarrow \searrow
 F of object Frictional force Random force - undetermined

\rightarrow Stochastic Differential eqⁿ (SDE)

- This is theoretical.
- At T , system \longleftrightarrow surrounding. $T(\gamma, F_R)$ (function)

$$\forall i, \frac{d\bar{p}_i}{dt} = \frac{1}{m_i} \bar{p}_i$$

Consider a system of Ar^{18} atoms



- This will have only Lennard Jones' Interactions.
- We observe some kind of scattering.

- Simulation time $\propto (\text{Diameter})^{3/2}$ or $\propto \text{Volume}^2$
- When we take bigger molecules, like protein bulk and edge molecules have diff. properties. We increase volume to homogenise distributions.



Randomness \propto

volume \rightarrow

\propto Simulation time

But this is not good!

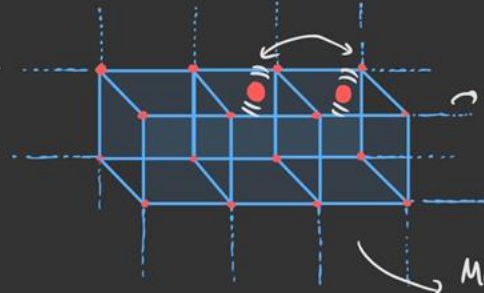
So we choose a moderate value.

Periodic Boundary Condition.

for crystal structures
• Many shapes

• If you insert something
it will still interact w/ infinite
other outside cell.

• This is observed in simulation.



Infinite lattice

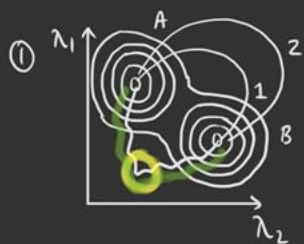
May not be cube, could be



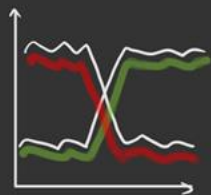
Lecture 6

2D order parameter - λ_1, λ_2

↳ Functions of portions of n-particles

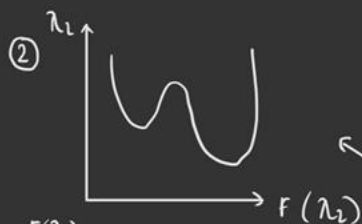


The reaction pathways are:



- Most likely path
- Metastable path
- Least likely path

Why? The reaction MUST follow this. i.e. it must go through the metastate.



Similar graphs.

- Why have the axes been flipped?
- Why λ_1 and λ_2 ?

Reaction mechanism

Consists of metastable states, transition states, local minimums in FEP's states

Minima: Metastable state

Maxima: Transition state

Free energy surfaces

The equation for free energy as a function of λ 's can be given as follows

$$F(\lambda_1, \lambda_2) = -kT \ln P(\lambda_1, \lambda_2)$$

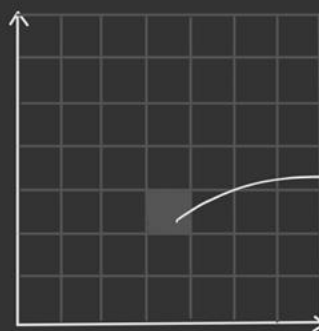
Method 1:

- An extended duration MD simulation
- About 1 or 2 years
- Plot the histogram

Steps:

- ① Grid the graph
- ② Calculate $P(\lambda_1, \lambda_2)$ for each grid
- ③ Use the eqn

$$F(\lambda_1, \lambda_2) = -kT \ln(P(\lambda_1, \lambda_2)) \text{ for each grid (Formula above)}$$

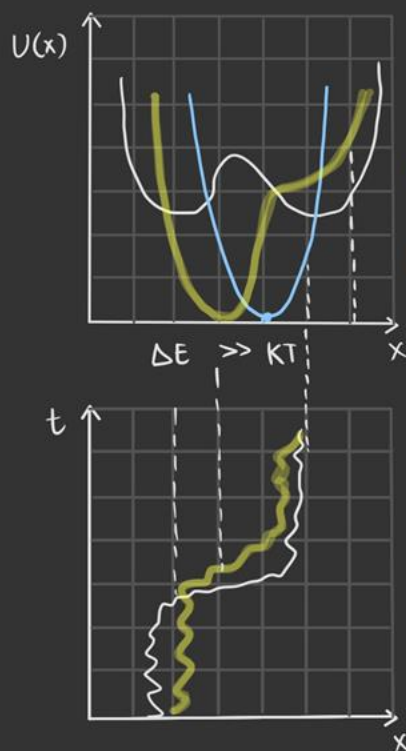


$$P(3,2) = \frac{N(3,2)}{N_{\text{total}}}$$

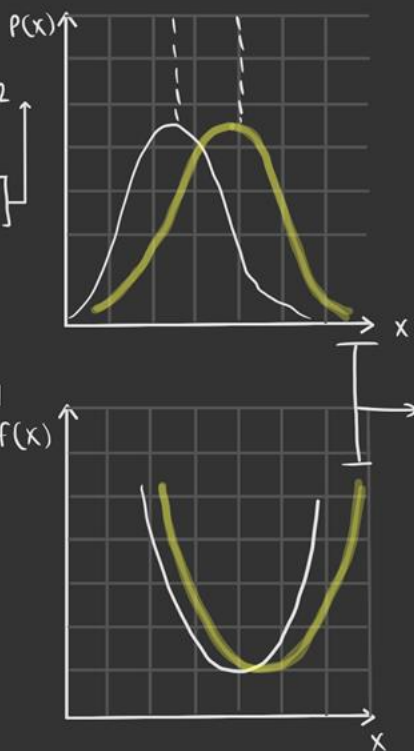
Method 2: (used more often) - developed by Torrie & Valleau

- We call this "Umbrella sampling"
- This is used to improve sampling of a system (or different system) where ERGODICITY

From method 1:



- $U'(x)$
- $U_{bj}(x)$
- $U(x)$



- Biased energy $U_{bj}(x)$
- Histogram
- Weighted analysis method.

Here the sampling is incomplete.
↓
≈ system trapped

- Start simulation from $U'(x)$
- $U'(x) = U(x) + U_{bj}(x)$ → Biased energy

suggested: $\frac{1}{2} k_j (x - x_i)^2$

Why $U'(x)$? Because it is a closer minima.

- Steps:
- ① Decide on an order parameter
 - ② Setup M independent bias simulation such that all of the order parameters are obtained
 - ③ Unbias - (involves CM)
↳ WHAM (Next)

\vec{p} before and after bias:

Before bias:

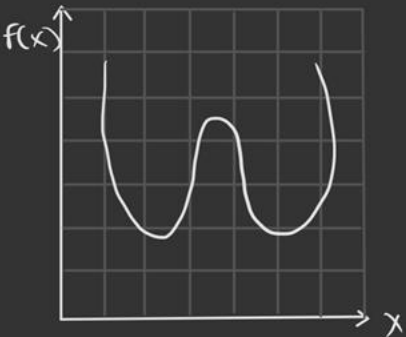
$$\vec{p} = -\frac{\partial U}{\partial x} - r_p + f_r$$

After bias:

$$\vec{p} = -\frac{\partial U}{\partial x} - \frac{\partial U_{bj}}{\partial x} - r_p + f_r$$

WHAM (Weighted Histogram analysis method)

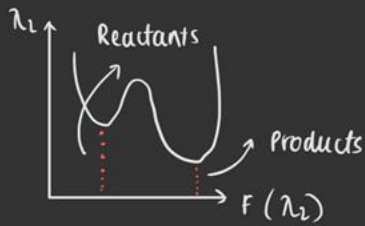
- ① Plot graphs, i.e histograms for all simulations
- ② Take weighted average of the histograms
∵ All processes do not contribute equally to weight.
- ③ Plot the weighted histogram.



- From $F(x)$ v/s x
- we get - $F(R)$ } measured
 - $F(P)$ } thermal props.
 - E_a → Rate (or activation E ?)
 - Identification of TS.
(if order parameter is carefully considered)

Lecture 7

We consider 2 ranges of order parameters.
The system traverses between these values of order parameters.



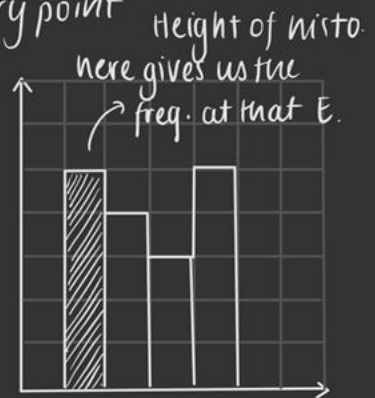
We normally wish to calculate a free energy surfaces' profile.

- The most naive way to calculate using the Landau free energy method.
- Calculate free energy at each and every point of time using the following equation.

$$F(\lambda_1, \lambda_2) = -K T \ln P(\lambda_1, \lambda_2)$$

- We can just make a histogram of free energy
- Using Landau we calc. the probabilities of where a system will be, given the value of the order parameter

Landau's method



- We use the 'frequentist' approach that says: probability = frequency

Here, E: Free energy.

- If the system cannot sample the whole data in the given time, it will resort to sampling a very small region of the sample.
- To avoid this → Use non boltzmann sampling → umbrella sampling.

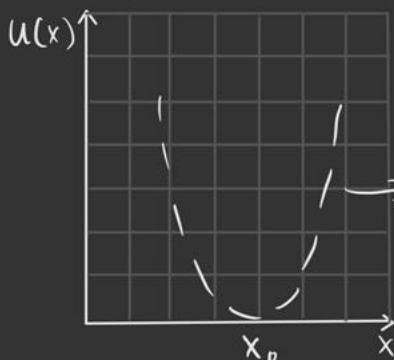
Here we add an additional bias:

This additional bias most commonly is a harmonic potential whose centre will be away from the centre if there wasn't a bias.

Add Bias → Sample (easier to do) → Weighted histogram (remove bias) [WHAM]

Another method: REPLICA EXCHANGE

eg. 1D simple harmonic oscillator



This reaction is done under constant temperature.

$$U(x) = \frac{1}{2} K (x - x_0)^2$$

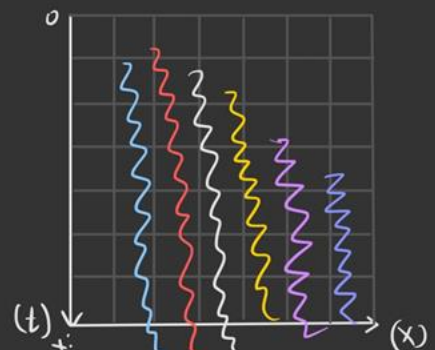
How?

→ Modify Newton's equation:

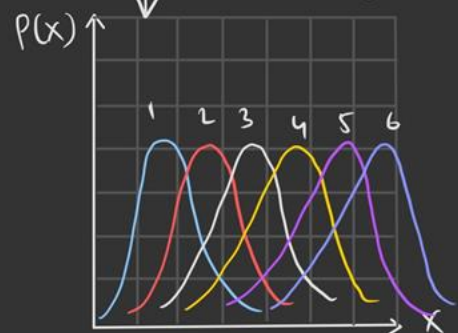
$$\dot{p} = -K(x - x_0) - \gamma p + F_R$$

$$\dot{x} = \frac{p}{m}$$

Too much time

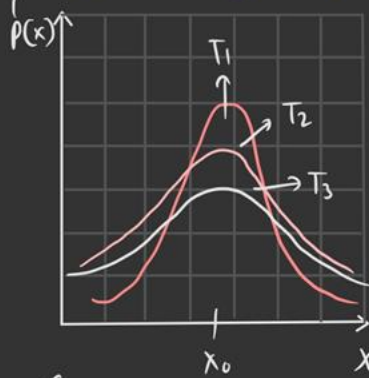


The system in boltzmann sampling is restricted to only sampling this chunk.



Histograms of processes above

probability as a functⁿ of x .



$$\gamma, F \propto T$$

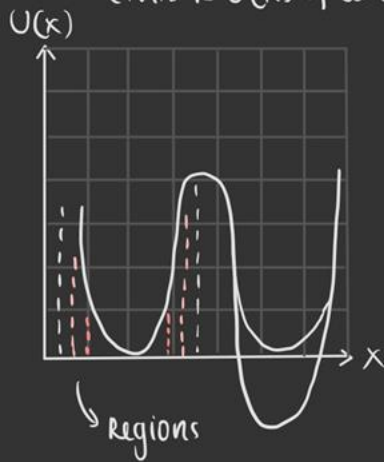
• Here $T_2 > T_1$, width of gaussian curve $\propto T$.

• This variation is a thermal fluctuation

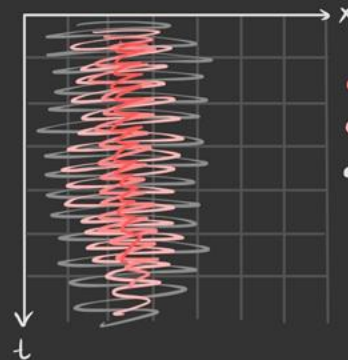
(Notice the same centre of both the curves)

(This is $U(x)$ of above)

eg. kick a molecular ball at 2 temperatures. It'll move further w/ same force at $\uparrow T$.



Bistable 1D Potential.



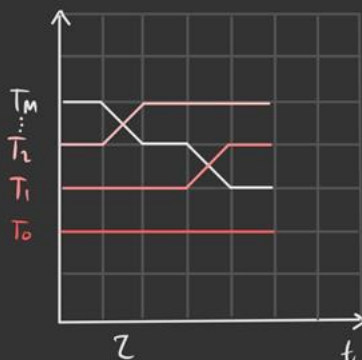
- least temperature
- medium temperature
- highest temperature

• < • < •

↓
 T_2 Area Sampled at that T , More probability of escaping local minima.

- We perform 'M' simulations at diff T 's. (T_1, \dots, T_M)
- Now we consider a temperature, eg. T_1 .
- Let's assume we wish to calculate rate of binding of ligand to protein

↗ 360 K → fast



Each coloured line represents a simulation.

↘
A → simulation A changed temperature of occurrence

- At random, we exchange the simulation temperatures.
- This is done at ' Z ' time intervals.
- Z is of the order $\approx 10^5 \Delta t$. $\Delta t = 1$ femtosecond