

# Classical molecular dynamics

Lecture 4

**Alexander Kvashnin**

# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo

# Outline

- **Born-Oppenheimer approximation**
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo

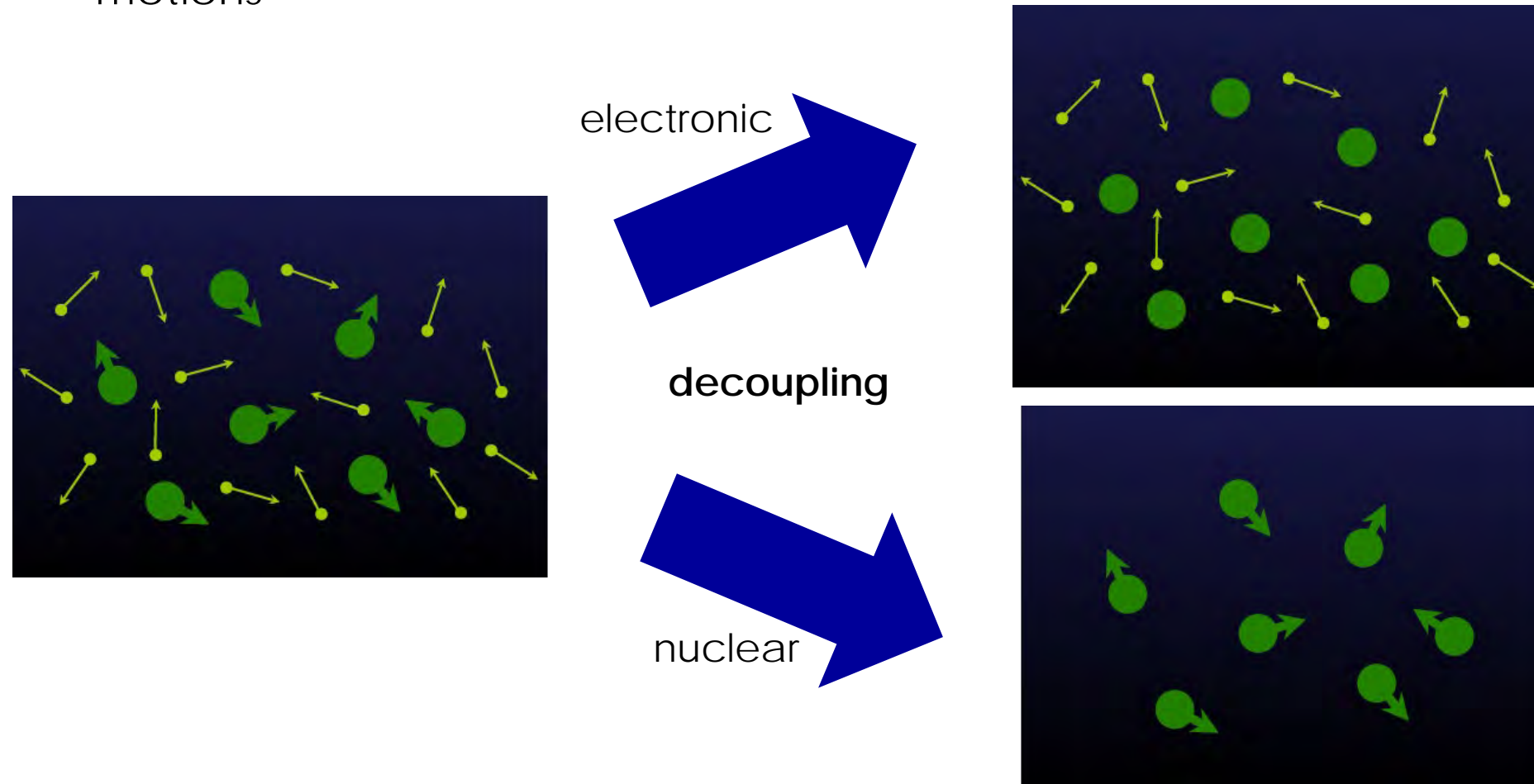
# Born-Oppenheimer approximation

- Nuclei are much heavier than electrons
- Nuclei move much slower than electrons
- We can decouple electronic and nuclear motions



Max  
Born

Robert  
Oppenheimer



# Born-Oppenheimer approximation

- Nuclei are much heavier than electrons
- Nuclei move much slower than electrons
- We can decouple electronic and nuclear motions

## Hamiltonian

$$\hat{H}\psi(\vec{r}, \vec{R}) = E\psi(\vec{r}, \vec{R}) \quad \Rightarrow \quad \hat{H} = \hat{H}_1(\vec{r}) + \hat{H}_2(\vec{R}) \quad \Rightarrow \quad \psi(\vec{r}, \vec{R}) = \psi_1(\vec{r})\psi_2(\vec{R})$$

Separable Hamiltonian

$$E = E_1 + E_2$$

$$\hat{H} = \hat{T}_N(\vec{R}) + \hat{T}_e(\vec{r}) + \hat{V}_{NN}(\vec{R}) + \hat{V}_{eN}(\vec{r}, \vec{R}) + \hat{V}_{ee}(\vec{r})$$

For this case

$$\Psi_T(\vec{r}, \vec{R}) = \Psi_e(\vec{r}, \vec{R})\Psi_N(\vec{R}) \quad \text{Approximately Separable!}$$

$$\hat{H}\Psi_e(\vec{r}, \vec{R})\Psi_N(\vec{R}) = E_{tot}\Psi_e(\vec{r}, \vec{R})\Psi_N(\vec{R})$$

## Electronic problem

$$\hat{H}_e = \hat{T}_e(\vec{r}) + \hat{V}_{eN}(\vec{r}, \vec{R}) + \hat{V}_{ee}(\vec{r})$$

$$\left\{ - \sum_i \frac{\hbar^2}{2m} \nabla_i^2 + \frac{1}{4\pi} \sum_{i,n} \frac{Z_n e^2}{|r_i - R_n|} + \frac{1}{4\pi} \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|r_i - r_j|} \right\} \Psi_e(\vec{r}, \vec{R}) = E_{tot} \Psi_e(\vec{r}, \vec{R})$$

## Nuclei problem

$$\{\hat{T}_N(\vec{R}) + \hat{E}_e(\vec{R}) + \hat{V}_{NN}(\vec{R})\} \Psi_N(\vec{R}) = E_{tot} \Psi_N(\vec{R})$$

$$\hat{H}_N \Psi_N(\vec{R}) = \left\{ - \sum_n \frac{\hbar^2}{2M} \nabla_n^2 + E_e(\vec{R}) + \frac{1}{4\pi} \frac{1}{2} \sum_{n \neq m} \frac{Z_n Z_m e^2}{|R_n - R_m|} \right\} \Psi_N(\vec{R}) = E_{tot} \Psi_e(\vec{R})$$

The nuclei move in a potential created by the electrons

# Outline

- Born-Oppenheimer approximation
- **Molecular mechanics**
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo

## “exactly” in Born-Oppenheimer approximation

Molecular Mechanics is based on several assumptions:

- It treats the electrons around a nucleus and the nucleus itself as a perfect sphere.
- The bonds between molecules are treated as springs.
- Potential functions rely on experimental parameters such as force constants and equilibrium volumes.
- The potential energy function is the sum of individual functions for bond stretching, angle bending, torsional energies, and non-bonding interactions.

So,

- We are not interested in electronic properties
- Electronic state remains unchanged
- Electronic dynamics is irrelevant

**Think about examples when MM does not work**



**Challenge** is to approximate exact Potential Energy Surface (PES) by interatomic potentials

**Grand challenge** is to make such potentials transferable

**Think about examples when transferability is hardly possible**





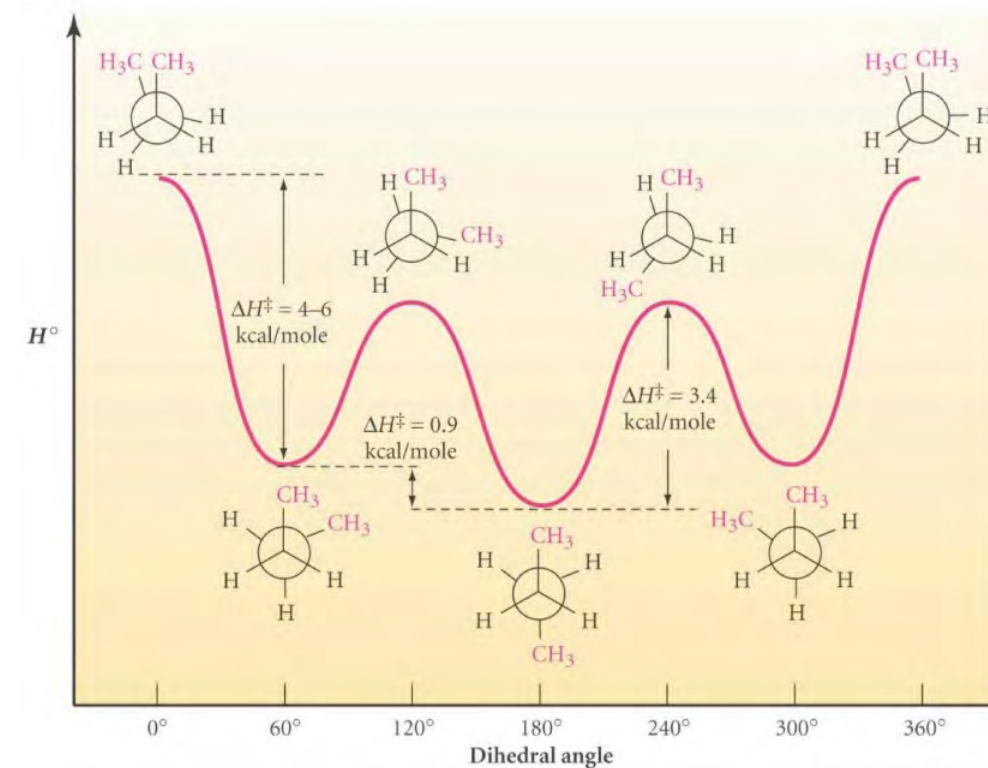
# Molecular mechanics

**Molecular Mechanics** is a computational method that computes the potential energy surface for a particular arrangement of atoms using potential functions that are derived using classical physics.

**Molecular Mechanics** methods use classical type models (no quantum mechanics) to predict the energy of a system as a function of atomic coordinates, i.e. approximate PES.

## Can be used to:

- Optimize geometry (of minima or transition state)
- Calculate relative energies between conformers or polymorphs



**Molecular Mechanics** is a computational method that computes the potential energy surface for a particular arrangement of atoms using potential functions that are derived using classical physics.

**Molecular Mechanics** methods use classical type models (no quantum mechanics) to predict the energy of a system as a function of atomic coordinates, i.e. approximate PES.

## Can be used to:

- Optimize geometry (of minima or transition state)
- Calculate relative energies between conformers or polymers

## But mostly used to:

- Run molecular dynamics
- Sample PES
- Large-scale simulations

## Why can we do it ?

- Chemical bonding is local
- Number of local structural patterns is limited
- Nonlocal interactions are pairwise



# Outline

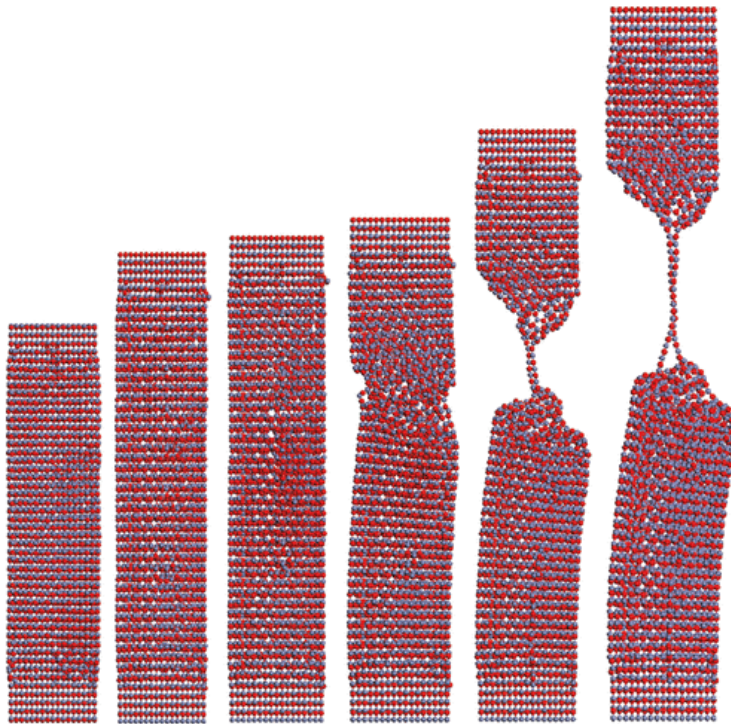
- Born-Oppenheimer approximation
- Molecular mechanics
- **Potentials of interatomic interaction**
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo

# Potentials

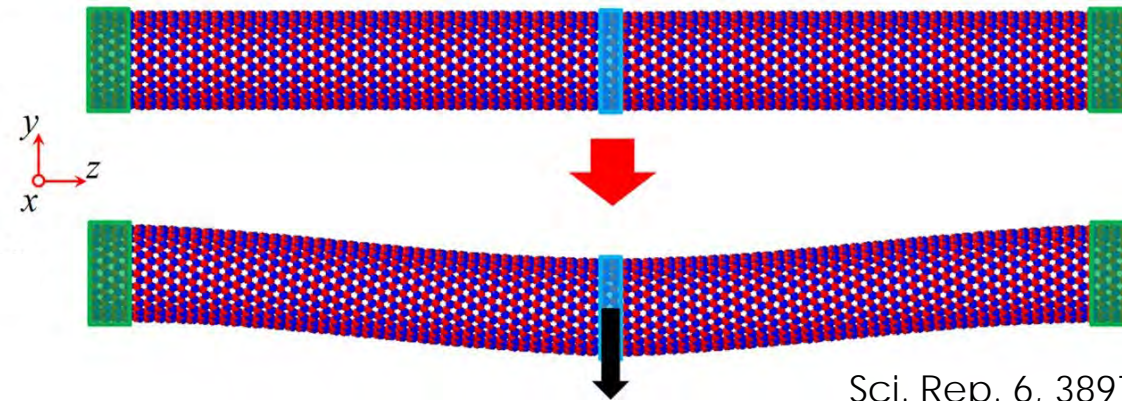
**Interatomic potentials** are mathematical functions to calculate the potential energy of a system of atoms with given positions in space

Examples of quantitative properties and qualitative phenomena that are explored with interatomic potentials include

lattice parameters, surface energies, interfacial energies, adsorption, cohesion, thermal expansion, and elastic and plastic material behavior, as well as chemical reactions



Langmuir 10, 26, 2, 1165-1171 (2010)



Sci. Rep. 6, 38974 (2016)

Interatomic potentials can be divided into several **types**

- Pair potentials
- Force fields
- ML potentials

# Potentials Pair potential

is a function that describes the potential energy of two interacting objects solely as a function of the distance between them

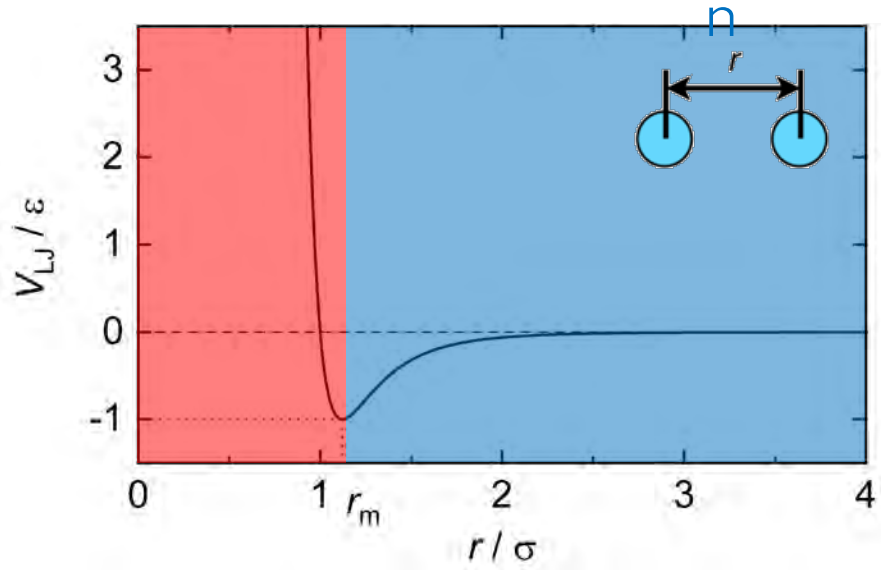
Mostly used for gases, closely packed lattices (metals, ionic crystals), but not for covalent compounds

## Lennard-Jones potential

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

repulsion

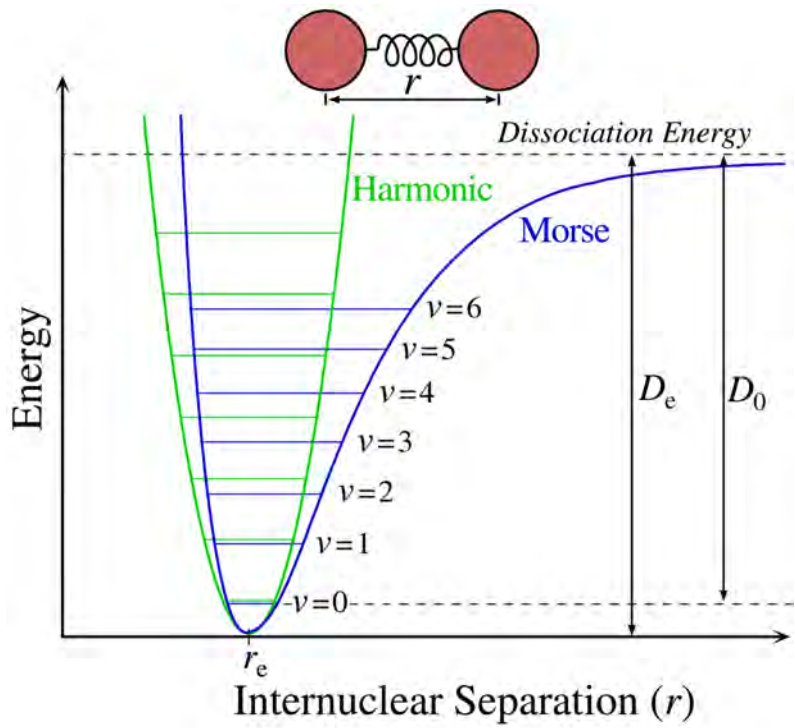
attraction



$\sigma$  is the distance at which the potential energy is 0  
 $\epsilon$  is the depth of potential well

## Morse potential

WHY?



$$V_M = D_e \left(1 - e^{-\alpha(r-r_e)}\right)^2$$

$$\alpha = \sqrt{\frac{k_e}{2D_3}}$$

$k_e$  is the force constant at the minimum of the well (stiffness of the bond)

$D_e$  is the depth of potential well,  $\alpha$  controls the width of the potential

EAM is an approximation describing the energy between atoms and is a type of interatomic potential  
The energy is a function of a sum of functions of the separation between an atom and its neighbors

$$E_i = F_\alpha \left( \sum_{j \neq i} \rho_\beta(r_{ij}) \right) + \frac{1}{2} \sum_{j \neq i} \phi_{\alpha\beta}(r_{ij})$$

Embedding energy

Electron cloud contribution

Pair interaction

$r_{ij}$  is the distance between atoms  $i$  and  $j$ ,  $\phi_{\alpha\beta}$  is a pair-wise potential function,  $\rho_\beta$  is the contribution to the electron charge density from atom  $j$  of type  $\beta$  at the location of atom  $i$ ,  $F$  is an embedding function that represents the energy required to place atom  $i$  of type  $\alpha$  into the electron cloud

For a binary alloy, the EAM potential requires seven functions:

three pair-wise interactions (A-A, A-B, B-B),  
two embedding functions, and  
two electron cloud contribution functions.

Generally, these functions are provided in a tabularized format and interpolated by cubic splines

a **force field** is a computational method that is used to estimate the forces between atoms within molecules, between molecules, and between atoms in solids

The parameters for a chosen energy function may be derived from experiments in physics and chemistry, calculations in quantum mechanics, or both

### General contributions

- Bonding interactions = stretching + bending + torsion
- Non-bonding interactions = van der Waals + electrostatic
- Cross-terms

$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

### Mostly used force fields

AMBER (Assisted Model Building and Energy Refinement)

CFF (Consistent Force Field)

CHARMM (Chemistry at HARvard Molecular Mechanics)

IFF (Interface Force Field)

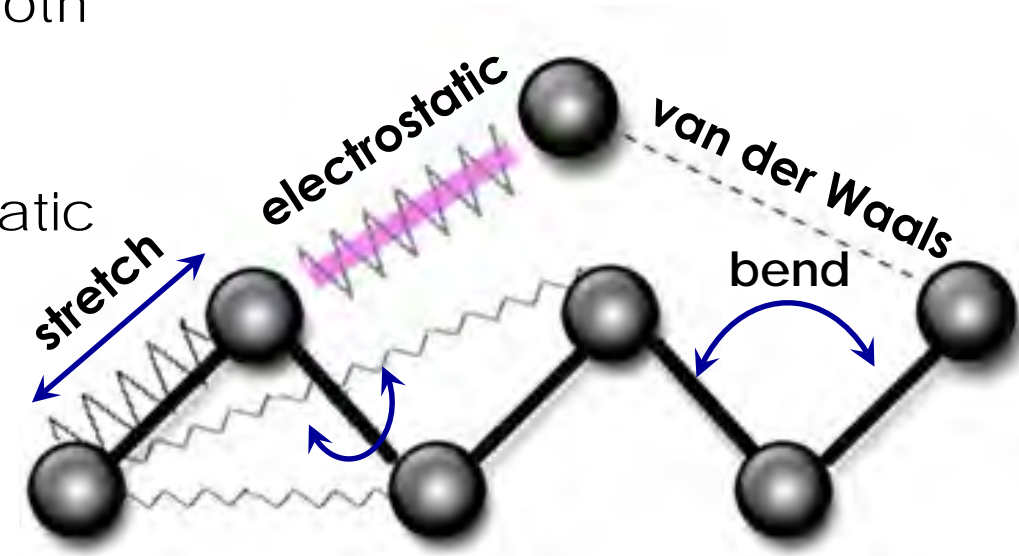
MMFF (Merck Molecular Force Field)

OPLS (Optimized Potential for Liquid Simulations)

UFF (Universal Force Field)

EVB (Empirical valence bond)

ReaxFF – reactive force field (interatomic potential)



## ML potentials

Will be discussed later  
and in a separate lecture

# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- **Energy contributions**
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo



# The stretch energy

Start with Taylor expansion of energy  $E(R)$  near the minimum

$$E_{str}(R^{ij} - R_0^{ij}) = E(0) + \frac{dE}{dR}(R^{ij} - R_0^{ij}) + \frac{1}{2} \frac{d^2E}{dR^2}(R^{ij} - R_0^{ij})^2$$

Can be set to 0

Vanishes at minimum



$$E_{str}(\Delta R^{ij}) = \underbrace{k_2^{ij}}_{\text{P2 term}} (\Delta R^{ij})^2 + \underbrace{k_3^{ij}}_{\text{P3 term}} (\Delta R^{ij})^3 + \underbrace{k_4^{ij}}_{\text{P4 term}} (\Delta R^{ij})^4 + \dots$$

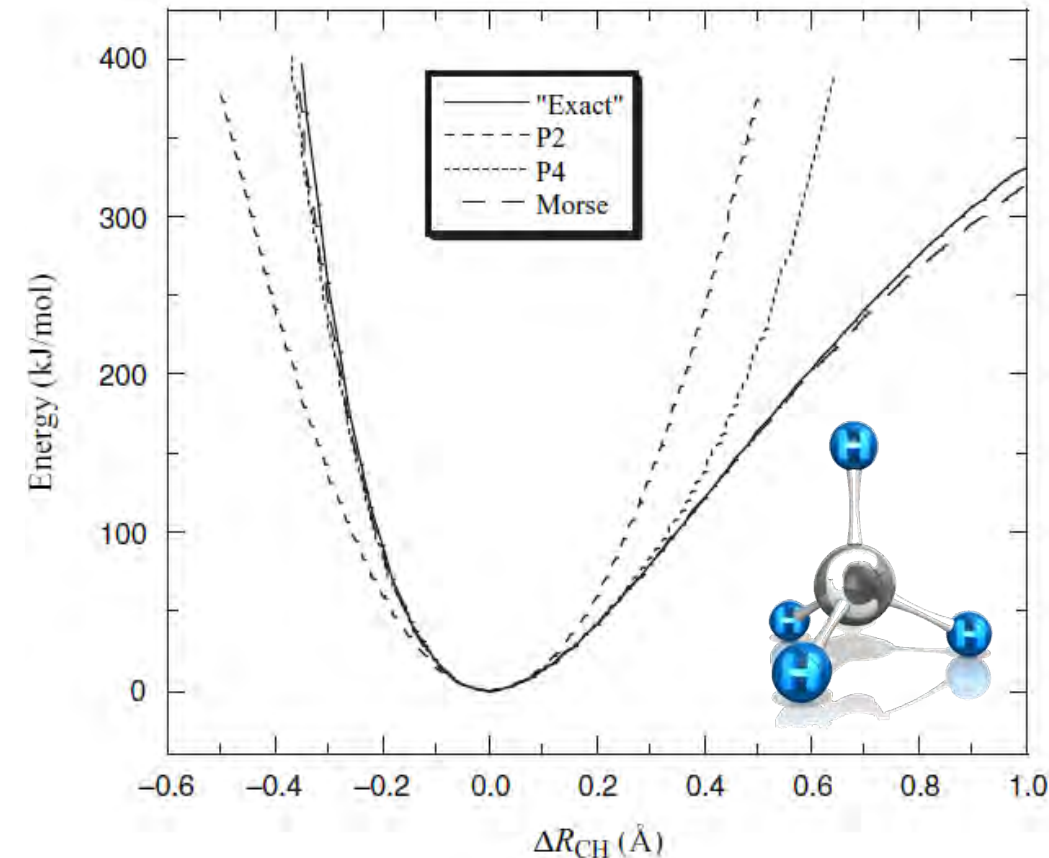
Alternative forms include Morse potential

$$E_M(\Delta R^{ij}) = D_e \left(1 - e^{-\alpha \Delta R^{ij}}\right)^2$$

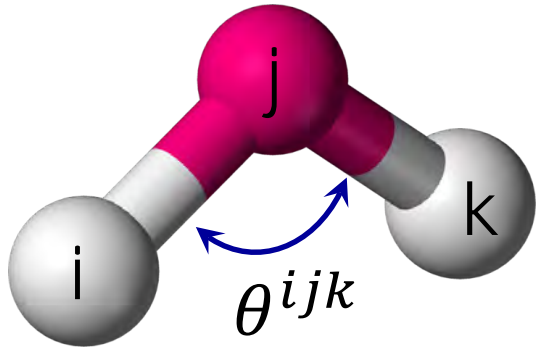
However, numerically friendly polynomial expansion is usually used instead of Morse potential

Every pair of atoms (i,j) requires at least 2 parameters:

$$k_2^{ij} \text{ and } R_0^{ij}$$

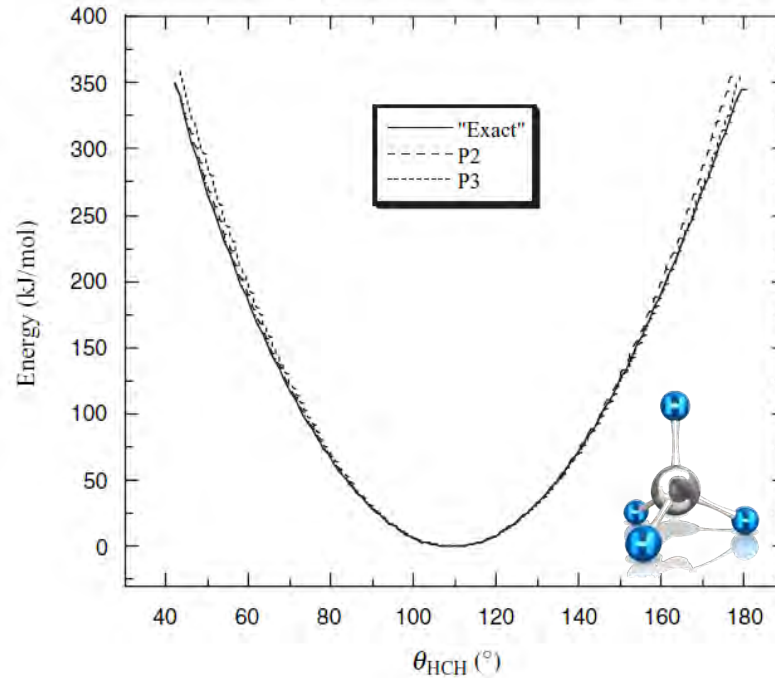
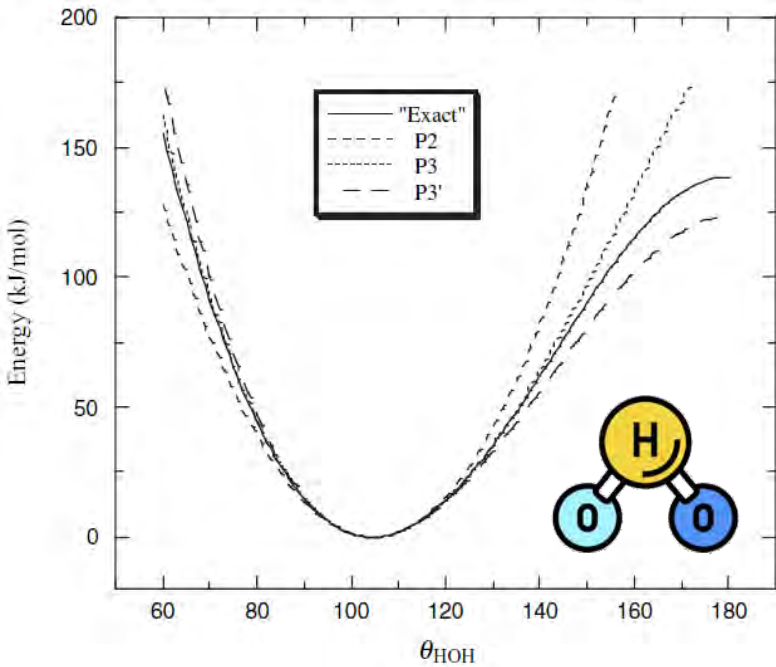
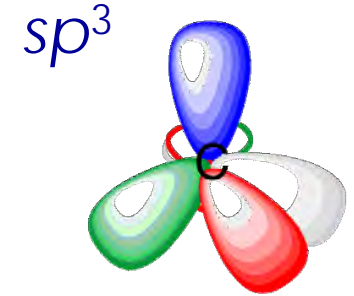
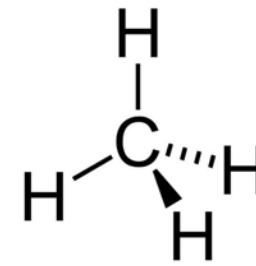


# The bending energy



$$E_{bend}(\theta^{ijk} - \theta_0^{ijk}) = k_2^{ijk} (\theta^{ijk} - \theta_0^{ijk})^2$$

Why bending change energy?



Must be accurate in region of few  $kT$  above the minimum (accessible conformational space)

Problems with  $\theta = 180^\circ$ , multiple minima, out-of-plane bending

Every triple of atoms (i,j,k) requires 2 parameters:

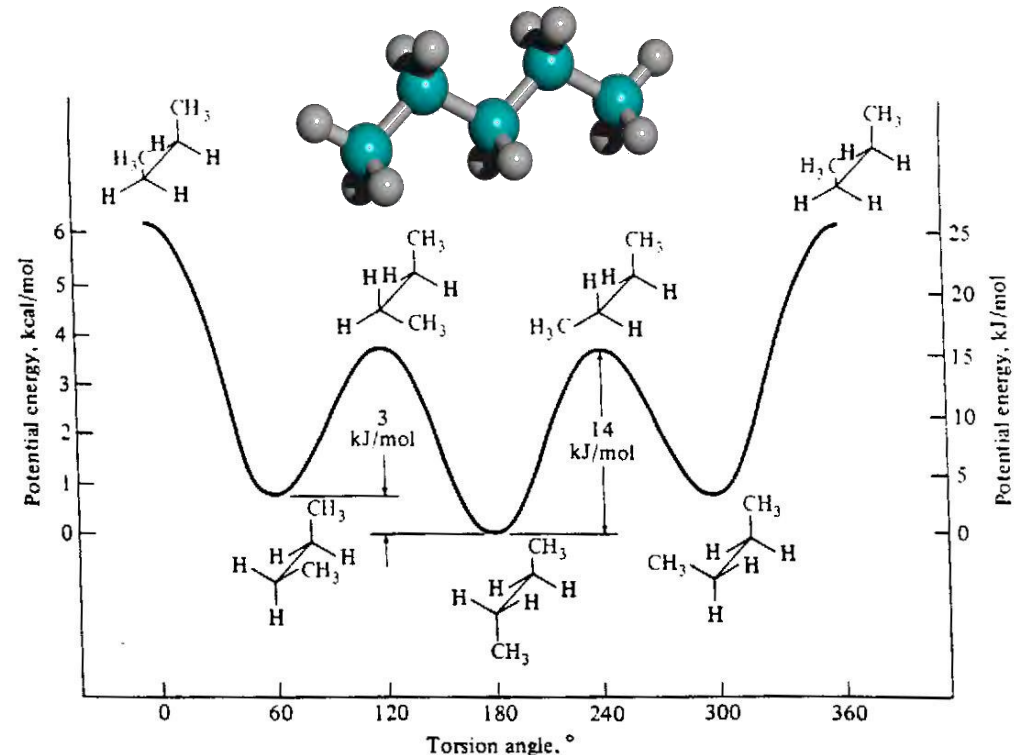
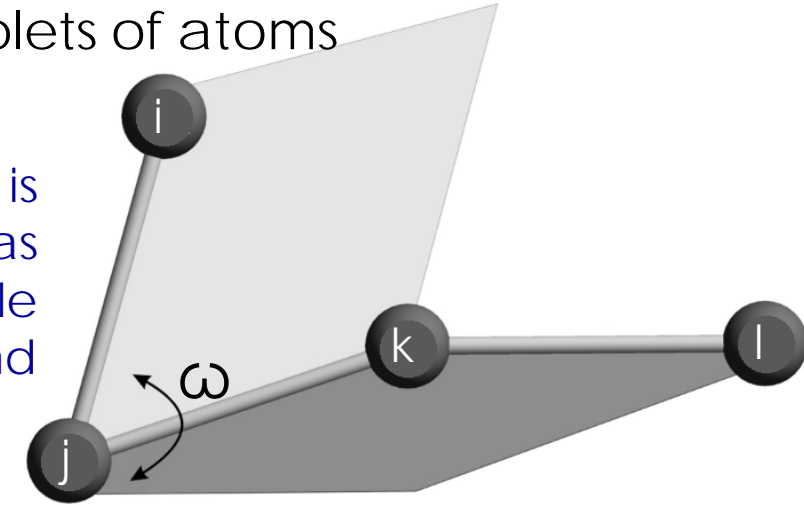
$$k_2^{ij} \text{ and } \theta_0^{ijk}$$

# The torsion energy

Torsion energy depends on the dihedral angle between quadruplets of atoms

$$E_{tors}(\omega^{ijkl}) = \frac{1}{2}V_1^{ijkl}(1 + \cos(\omega^{ijkl})) + \frac{1}{2}V_2^{ijkl}(1 - \cos(2\omega^{ijkl})) + \frac{1}{2}V_3^{ijkl}(1 + \cos(3\omega^{ijkl}))$$

Dihedral angle is defined as signed angle between  $ijk$  and  $jkl$  planes

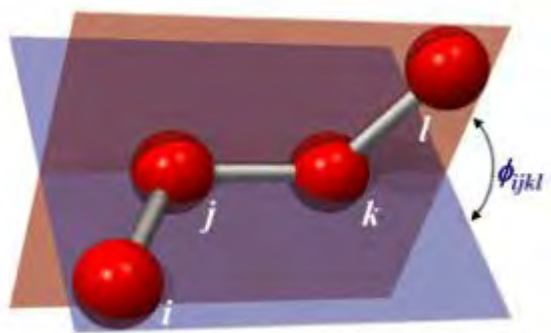


- Dihedral angles are usually flexible and correspond to large-amplitude motions
- Energy scale is 1-2 orders of magnitude smaller than for stretching and bending
- Includes non-bonding interactions
- Every quadruplet of atoms ( $i,j,k,l$ ) requires a set of parameters:

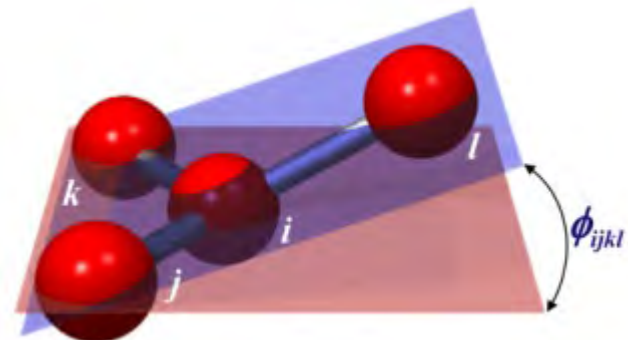
$$V_1^{ijkl}, V_2^{ijkl}, V_3^{ijkl}, \dots$$

# Improper torsion

An improper torsion angle is one where the torsion angle is not defined by four angles connected sequentially. It defines improper interactions between quadruplets of atoms



Proper torsion angle

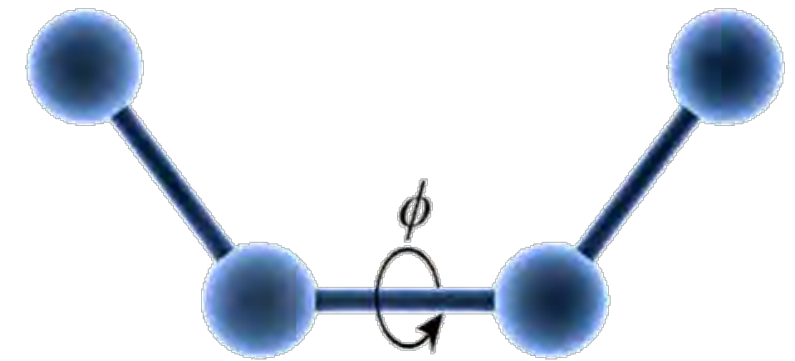


Improper torsion angle

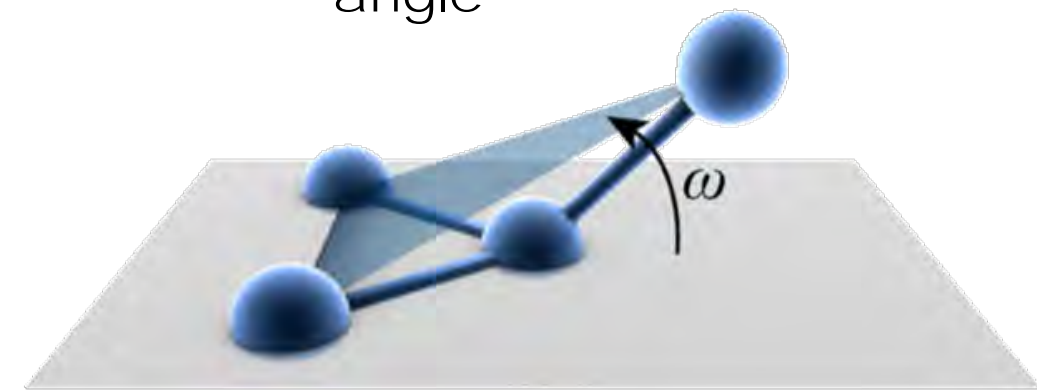
$$E_{imp}(\phi^{ijkl}) = k_{ijkl}(\phi_{ijkl} - \phi_0)^2$$

- Correspond to small-amplitude motions, so the harmonic potential is good enough
- Every planar quadruplet of atoms (i,j,k,l) requires only a single parameter:

$$k_{ijkl}$$



Proper angle torsion



Improper angle torsion

# The van der Waals energy

- Accounts for all non-electrostatic interaction between non-bonded atoms
- Repulsive at short distance due to steric interaction (Pauli exchange)
- Attractive at large distances due to dispersion interaction, induced dipoles  $\epsilon^{AB} = (\epsilon^{AA}\epsilon^{BB})^{1/2}$
- The Lennard-Jones (LJ) potential is commonly used
- For LJ potential every pair of atoms (A,B) requires 2 parameters:  $\epsilon$  and  $\sigma$
- To minimize number of parameters one uses
- Usually, vdW interactions are excluded between bonded atoms

## Some other forms

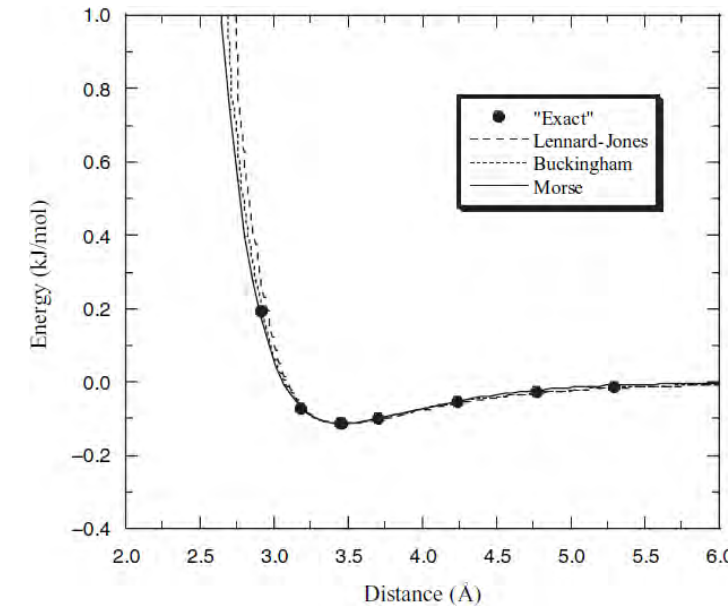
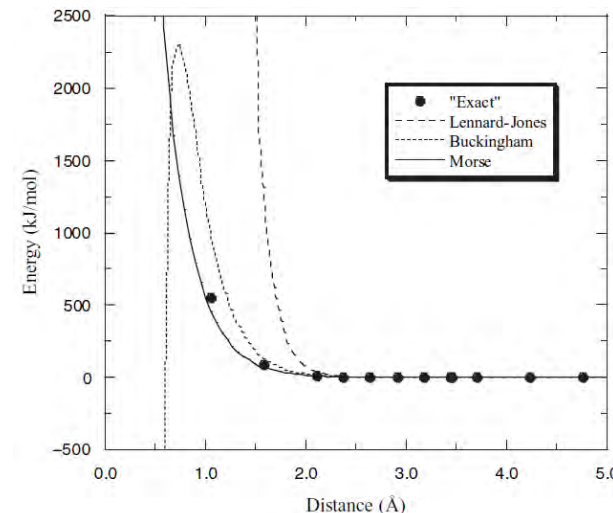
The Buckingham potential

$$E_{buck}(r) = \epsilon \left[ \frac{6}{\alpha - 6} e^{\alpha(1-r/R_0)} - \frac{\alpha}{\alpha - 6} \left( \frac{R_0}{r} \right)^6 \right]$$

Hydrogen bonding potential

$$E_{H-bond}(r) = \epsilon \left[ 5 \left( \frac{R_0}{r} \right)^{12} - 6 \left( \frac{R_0}{r} \right)^{10} \right]$$

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



$$E_{FF} = E_{str} + E_{bend} + E_{tors} + E_{vdw} + E_{el} + E_{cross}$$

The various terms in the total energy are not independent, and cross-terms should be considered

$$u_i(\vec{x}) = u_i(0) + \sum_j x_j \left( \frac{\partial u_i}{\partial x_j} \right)_0 + \sum_{k \leq j} \sum_j x_j x_k \left( \frac{\partial^2 u_i}{\partial x_j \partial x_k} \right)_0$$

For example, stretching-bending:

$$E_{str/bend} = k^{ABC} (\theta^{ABC} - \theta_0^{ABC}) [(R^{AB} - R_0^{AB}) - (R^{BC} - R_0^{BC})]$$

Other terms

$$E_{str/str} = k^{ABC} (R^{AB} - R_0^{AB})(R^{BC} - R_0^{BC})$$

$$E_{bend/bend} = k^{ABCD} (\theta^{ABC} - \theta_0^{ABC})(\theta^{BCD} - \theta_0^{BCD})$$

$$E_{str/tors} = k^{ABCD} (R^{AB} - R_0^{AB}) \cos(n\omega^{ABCD})$$

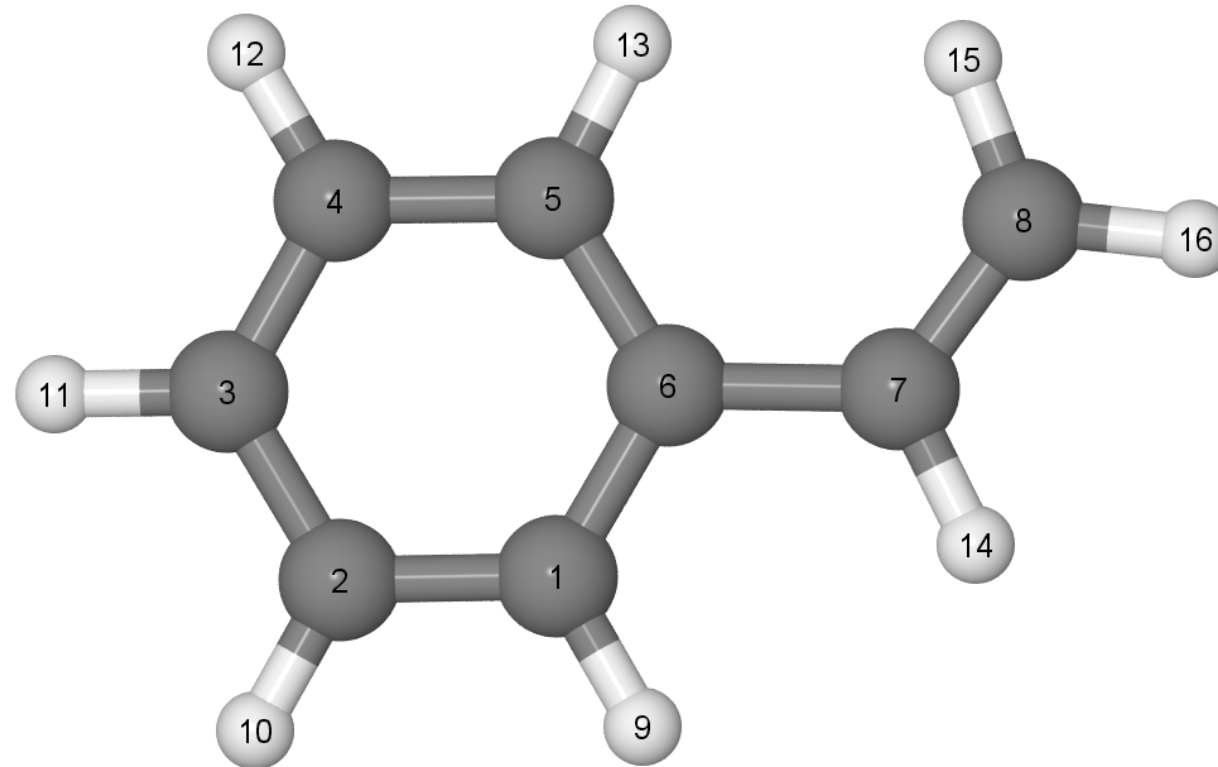
$$E_{bend/tors} = k^{ABCD} (\theta^{ABC} - \theta_0^{ABC}) \cos(n\omega^{ABCD})$$

$$E_{bend/tors/bend} = k^{ABCD} (\theta^{ABC} - \theta_0^{ABC})(\theta^{BCD} - \theta_0^{BCD}) \cos(n\omega^{ABCD})$$

**Cross-terms are rarely used because of problems with their parametrization**

# Discussion

Illustrate all elements of a force field using this molecule:



# MM2 force field

Type	Symbol	Description	Type	Symbol	Description
1	C	sp <sup>3</sup> -carbon	28	H	enol or amide
2	C	sp <sup>2</sup> -carbon, alkene	48	H	ammonium
3	C	sp <sup>2</sup> -carbon, carbonyl, imine	36	D	deuterium
4	C	sp-carbon	20	lp	lone pair
22	C	cyclopropane	15	S	sulfide (R <sub>2</sub> S)
29	C·	radical	16	S <sup>+</sup>	sulfonium (R <sub>3</sub> S <sup>+</sup> )
30	C <sup>+</sup>	carbocation	17	S	sulfoxide (R <sub>2</sub> SO)
38	C	sp <sup>2</sup> -carbon, cyclopropene	18	S	sulfone (R <sub>2</sub> SO <sub>2</sub> )
50	C	sp <sup>2</sup> -carbon, aromatic	42	S	sp <sup>2</sup> -sulfur, thiophene
56	C	sp <sup>3</sup> -carbon, cyclobutane	11	F	fluoride
57	C	sp <sup>2</sup> -carbon, cyclobutene	12	Cl	chloride
58	C	carbonyl, cyclobutanone	13	Br	bromide
67	C	carbonyl, cyclopropanone	14	I	iodide
68	C	carbonyl, ketene	26	B	boron, trigonal
71	C	ketonium carbon	27	B	boron, tetrahedral
8	N	sp <sup>3</sup> -nitrogen	19	Si	silane
9	N	sp <sup>2</sup> -nitrogen, amide	25	P	phosphine
10	N	sp-nitrogen	60	P	phosphor, pentavalent
37	N	azo or pyridine (-N=)	51	He	helium
39	N <sup>+</sup>	sp <sup>3</sup> -nitrogen, ammonium	52	Ne	neon
40	N	sp <sup>2</sup> -nitrogen, pyrrole	53	Ar	argon
43	N	azoxy (-N=N-O)	54	Kr	krypton
45	N	azide, central atom	55	Xe	xenon
46	N	nitro (-NO <sub>2</sub> )	31	Ge	germanium
72	N	imine, oxime (=N-)	32	Sn	tin
6	O	sp <sup>3</sup> -oxygen	33	Pb	lead
7	O	sp <sup>2</sup> -oxygen, carbonyl	34	Se	selenium
41	O	sp <sup>2</sup> -oxygen, furan	35	Te	tellurium
47	O <sup>-</sup>	carboxylate	59	Mg	magnesium
49	O	epoxy	61	Fe	iron (II)
69	O	amine oxide	62	Fe	iron (III)
70	O	ketonium oxygen	63	Ni	nickel (II)
5	H	hydrogen, except on N or O	64	Ni	nickel (III)
21	H	alcohol (OH)	65	Co	cobalt (II)
23	H	amine (NH)	66	Co	cobalt (III)
24	H	carboxyl (COOH)			



# MM2 force field

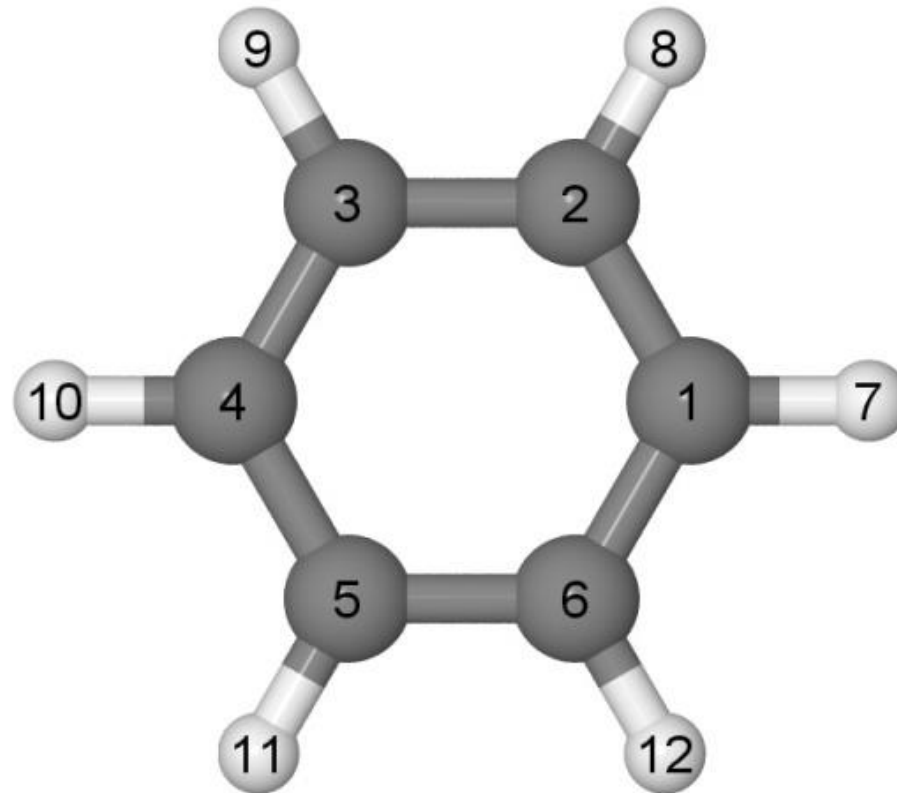
- Each of the 71 atom types has two van der Waals parameters,  $R_0^A$  and  $\epsilon^A$ , giving **142** parameters
- There are approximately  $1/2 \times 30 \times 30 = 450$  different  $E_{str}$  terms, each requiring at least two parameters  $R_0^{AB}$  and  $k^{AB}$ , for a total of at least **900** parameters
- There are approximately  $1/2 \times 30 \times 30 \times 30 = 13500$  different  $E_{bend}$  terms, each requiring at least two parameters  $k^{ABC}$  and  $\theta_0^{ABC}$ , for a total of at least **27000** parameters
- There are approximately  $1/2 \times 30 \times 30 \times 30 \times 30 = 405000$  different  $E_{tors}$  terms, each requiring at least three parameters  $V_1^{ABCD}$ ,  $V_2^{ABCD}$ ,  $V_3^{ABCD}$  for a total of at least **1215000** parameters
- Cross-terms may add another million possible parameters

**Table 2.3** Comparison of possible and actual number of MM2(91) parameters

Term	Estimated number of parameters	Actual number of parameters
$E_{vdw}$	142	142
$E_{str}$	900	290
$E_{bend}$	27000	824
$E_{tors}$	1 215 000	2466

# Reduce number of parameters

- Assign atomic charges (e.g. from DFT) instead of fitting them
- Use element-wise LJ parametrization instead of pair-wise
- Group atom types into **atom classes** for bonded interactions
- Omit unnecessary dihedral and minimize number of impropers



# Parametrization

How to do it? Define functional form, objective function, and fit parameters

$$ErrF(parameters) = \sum_i^{data} weight_i \cdot (reference\ value - calculated\ value)_i^2$$

1) How to choose **functional form**:

- The choice is not unique and depends on material, scales, and goals
- Linear optimization is preferable
- Systematic improvement of accuracy is challenging for non-ML potentials

2) What are **reference values**:

- Experimental data (vibrational frequencies, X-ray structures, heats of formation) were commonly used in past, but they are often incomplete and have large error bars (especially lab-to-lab and batch-to-batch variations)
- Today high-throughput DFT calculations are used, including semi-automated parametrization routines (e.g. AMBER-DFT or ML-potentials)

3) How to define and generate **fitting (training) dataset**:

- The dataset must be representative (cover required part of PES with proper weights) and suitable for fitting (confidence intervals for all parameters must be reasonable, parameter dependencies must be resolved)

**Parametrization is a difficult process requiring good knowledge of underlying physical phenomena/interactions, chemical intuition, use of experimental data and utilization of electronic structure codes**

# Various types of potentials

## Different classes of materials require different approaches:

- Simple metals (Al – yes, Pu – no) – embedded-atom (EAM)
- Ionic solids (NaCl) – embedded-ion model
- Complex bonding – charge optimized many body potential (COMB)
- Tetrahedral semiconductors/insulators (Si, SiO<sub>2</sub>) – Tersoff potentials
- Water – TIP3/4/5P models
- $\pi$ -conjugated molecules – MM3 force field + Huckel model
- Biomolecules – well parameterized AMBER, CHARMM, GROMOS with residue-based parametrization (proteins) and united atoms
- Attempts to make universal force field for molecules – OPLS, UFF, DREIDING
- Polarizable force– DRUDE, AMOEBA
- Chemical reactions – reactive force (ReaxFF)

## Different states of the same molecule require different values of parameters

Anion, cation, excited singlet  $S_1$ , etc

**Fitting-friendly parametrization** (allow for black-box fitting and systematic improvement by ML approaches):

$$E = \sum_i E(\text{local geometry of } i^{\text{th}} \text{ atom}) + E_{\text{long-range}}$$

# Computational efficiency

- Make no mistake: molecular mechanics is MUCH faster compared to any electronic structure model based on quantum mechanics!
- Evaluation of the non-bonded energy is the most time-consuming step growing  $\sim N^2$

Can be reduced to  $\sim N$  using cut-off distances or advanced summation techniques

**Table 2.6** Number of terms for each energy contribution in  $\text{CH}_3(\text{CH}_2)_{n-2}\text{CH}_3$

$n$	$N_{\text{atoms}}$	$E_{\text{str}}$	$E_{\text{bend}}$	$E_{\text{tors}}$	$E_{\text{vdw}}$
10	32	31 (5%)	30 (10%)	81 (14%)	405 (70%)
20	62	61 (3%)	60 (6%)	171 (8%)	1710 (83%)
50	152	151 (1%)	300 (3%)	441 (4%)	11025 (93%)
100	302	301 (1%)	600 (1%)	891 (2%)	44550 (96%)
	$N$	$(N - 1)$	$2(N - 2)$	$3(N - 5)$	$\frac{1}{2}N(N - 1) - 3N + 5$

# Benchmarks by LAMMPS

Potential	System	# Atoms	Timestep	Neighs/atom	Memory	CPU	LJ Ratio
<u>Granular</u>	chute flow	32000	0.0001 tau	7.2	33 Mb	2.08e-7	0.26x
<u>FENE bead/spring</u>	polymer melt	32000	0.012 tau	9.7	8.4 Mb	2.86e-7	0.36x
<u>Lennard-Jones</u>	LJ liquid	32000	0.005 tau	76.9	12 Mb	8.01e-7	1.0x
<u>DPD</u>	pure solvent	32000	0.04 tau	41.3	9.4 Mb	1.22e-6	1.53x
<u>EAM</u>	bulk Cu	32000	5 fmsec	75.5	13 Mb	1.87e-6	2.34x
<u>REBO</u>	polyethylene	32640	0.5 fmsec	149	33 Mb	3.18e-6	3.97x
<u>Stillinger-Weber</u>	bulk Si	32000	1 fmsec	30.0	11 Mb	3.28e-6	4.10x
<u>Tersoff</u>	bulk Si	32000	1 fmsec	16.6	9.2 Mb	3.74e-6	4.67x
<u>ADP</u>	bulk Ni	32000	5 fmsec	83.6	25 Mb	5.58e-6	6.97x
<u>EIM</u>	crystalline NaCl	32000	0.5 fmsec	98.9	14 Mb	5.60e-6	6.99x
<u>Peridynamics</u>	glass fracture	32000	22.2 nsec	422	144 Mb	7.46e-6	9.31x
<u>SPC/E</u>	liquid water	36000	2 fmsec	700	86 Mb	8.77e-6	11.0x
<u>CHARMM + PPPM</u>	solvated protein	32000	2 fmsec	376	124 Mb	1.13e-5	14.1x
<u>MEAM</u>	bulk Ni	32000	5 fmsec	48.8	54 Mb	1.32e-5	16.5x
<u>Gay-Berne</u>	ellipsoid mixture	32768	0.002 tau	140	21 Mb	2.20e-5	27.5x
<u>BOP</u>	bulk CdTe	32000	1 fmsec	4.4	74 Mb	2.51e-5	31.3x
<u>AIREBO</u>	polyethylene	32640	0.5 fmsec	681	101 Mb	3.25e-5	40.6x
<u>ReaxFF/C</u>	PETN crystal	32480	0.1 fmsec	667	976 Mb	1.09e-4	136x
<u>COMB</u>	crystalline SiO2	32400	0.2 fmsec	572	85 Mb	2.00e-4	250x
<u>eFF</u>	H plasma	32000	0.001 fmsec	5066	365 Mb	2.16e-4	270x
<u>ReaxFF</u>	PETN crystal	16240	0.1 fmsec	667	425 Mb	2.84e-4	354x
<u>VASP/small</u>	water	192/512	0.3 fmsec	N/A	320 procs	26.2	17.7e6
<u>VASP/medium</u>	CO2	192/1024	0.8 fmsec	N/A	384 procs	252	170e6
<u>VASP/large</u>	Xe	432/3456	2.0 fmsec	N/A	384 procs	1344	908e6

- Blind use of empirical potentials is very dangerous and can lead to errors.
- Well-parameterized potentials can give more accurate geometries and relative energies than low-level QM methods
- Often FF are bound to specific MD code (AMBER, CHARMM etc.) so that parameters are not easily transferable between codes.

# Typical workflow

- 1) Choose the **functional form** of empirical potentials:
  - define atom types;
  - decide on type of potentials and number of parameters (e.g. not all dihedral angles are usually needed, parametrization of LJ interactions can be chosen element-wise or pair-wise)
  - which parameters are fitted, and which are assigned (e.g. often it is bad idea to fit charges in force fields, for intramolecular dynamics LJ parameters can be assigned from generic tables)
- 2) Force fields require **topology** – define it
- 3) Provide initial parametrization  
(from generic force field like OPLS or from DFT calculations)
- 4) Parametrize



# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- **ML potentials**
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- Monte-Carlo

# ML as interpolation

... data driven and multidimensional

Problem

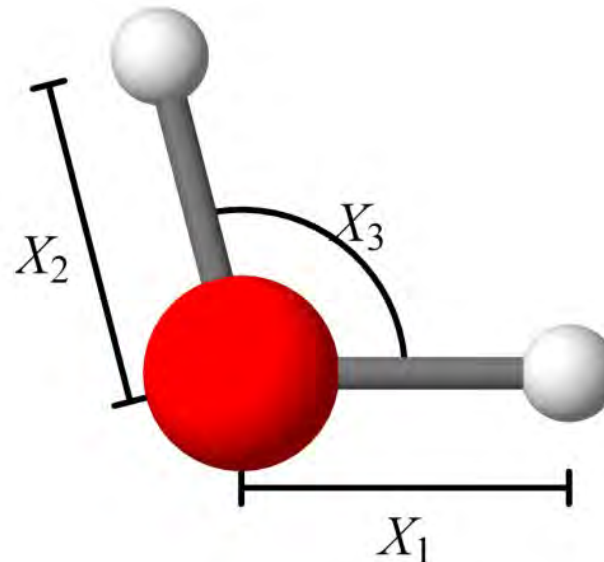
Given  $E^{qm}(X)$ , interpolate with  $E(X)$

Issue

No transferability w.r.t. the number of atoms

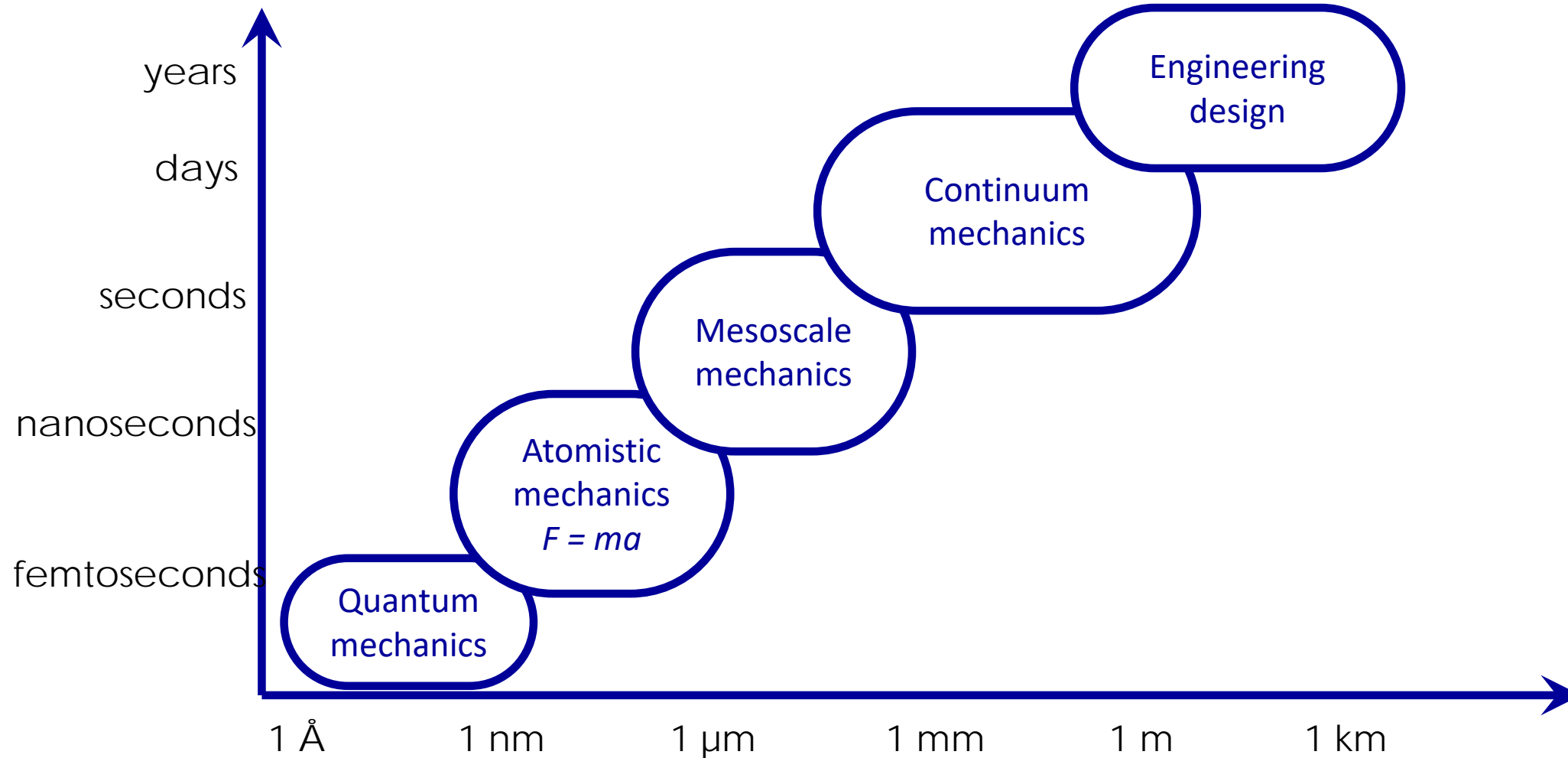
Solution

**Use locality!** An atom interacts with only 10-100 neighbor



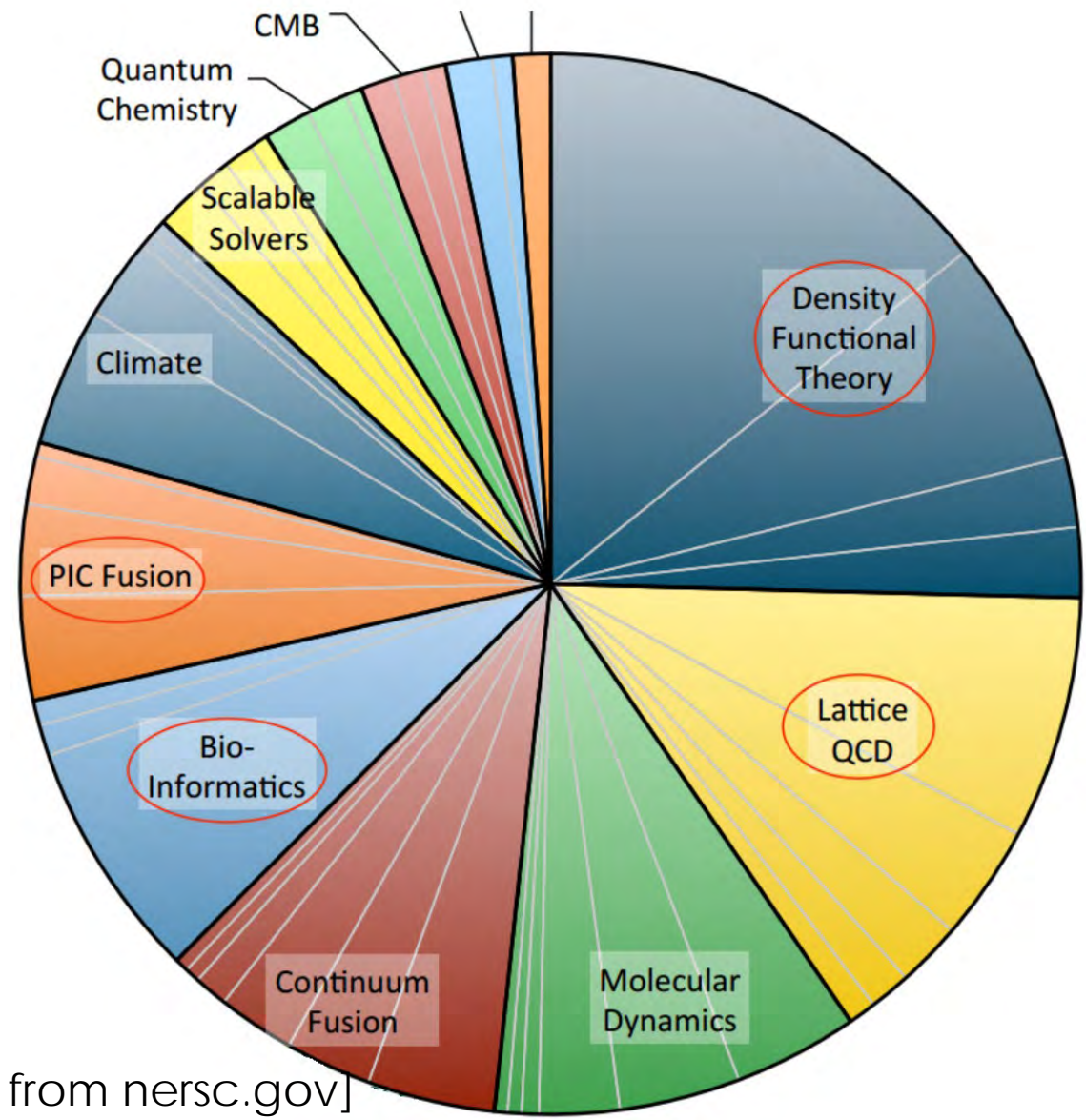
# Why is this important?

A dream of computational materials scientist



# Molecular modeling

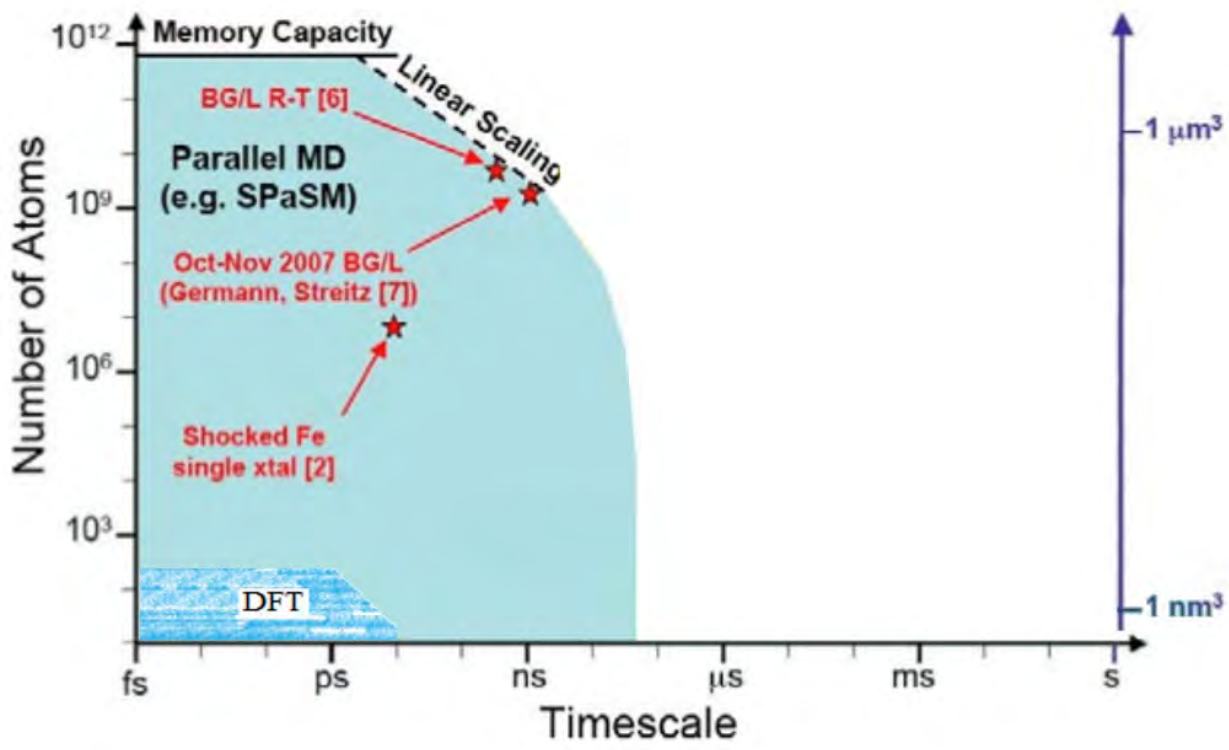
~40% of supercomputing time is spent on **Molecular Modeling**



[Adopted from nersc.gov]

# Molecular dynamics scales

Number of tasks is limited by the available computational resources depending on the number of atoms in the system.



The goal is to increase timescale of simulations using the same number of computational resources

[D. Perez, LANL]

# Machine learning ideology

1. Choose a (machine-learning) model  $E = E(X)$ ,  $X$  is atomic configuration
2. We want to minimize  $|E - E^{qm}|$

so we

Generate data:  $x^{(1)}, x^{(2)}, \dots; E^{qm}(x^{(1)}), E^{qm}(x^{(2)}), \dots; f^{qm}(x^{(1)}), f^{qm}(x^{(2)}), \dots$

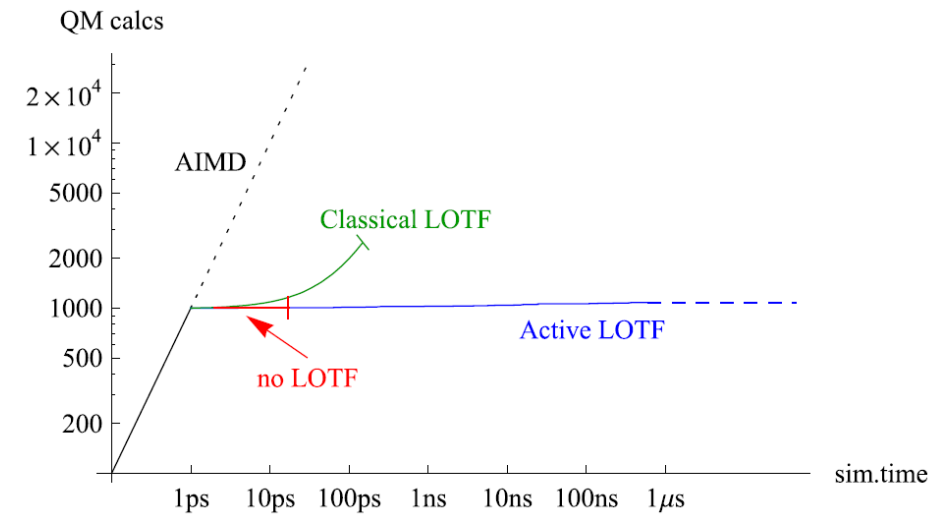
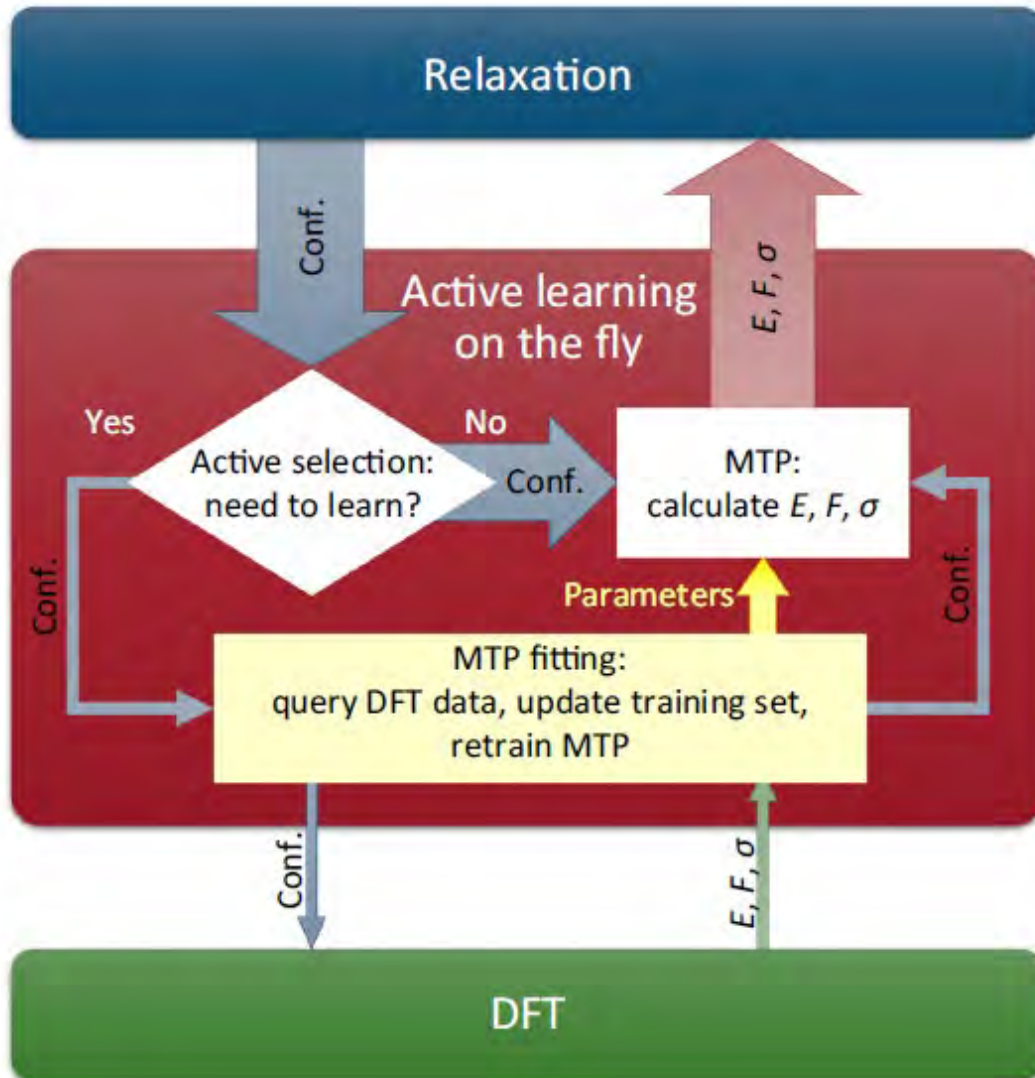
Minimize on data:  $\sum_i |E(x^{(i)}) - E^{qm}(x^{(i)})|^2 + (\mathbf{forces}) + \dots$

But what if sampling the right  $x^{(i)}$  is a part of the problem?

# Active learning / learning on-the-fly

A crucial and often time-consuming part is the construction of the training set.

**Active learning is a machine-learning technique allowing one to entrust these training set refinement iterations to a computer, thus completely automating the training set construction.**



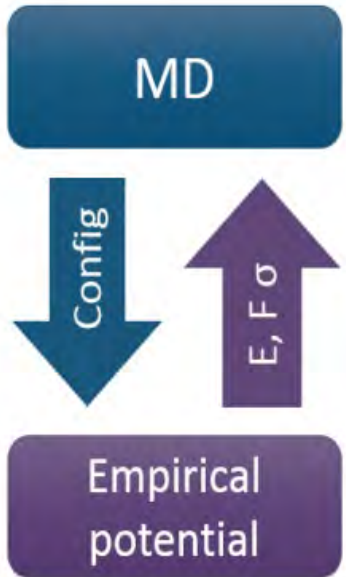
The no-learning and classical LOTF MD are not completely reliable: on average every 15 ps the no-learning MD fails, i.e., escapes into an unphysical region in the phase space.

The classical LOTF ten times more reliable (failure time of 150 ps) at the expense of extra 1500 QM calculations.

The active LOTF makes MD **completely reliable** as measured over the first 0.5 μs of simulation time.

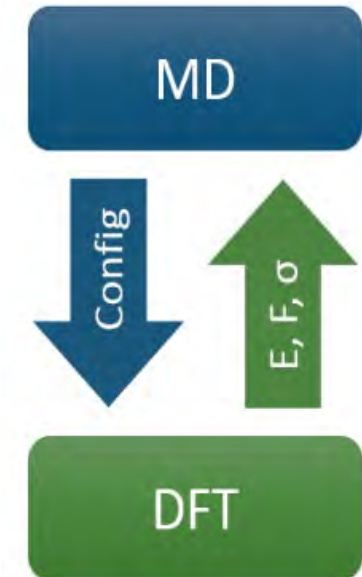
# Active learning

## Classic MD



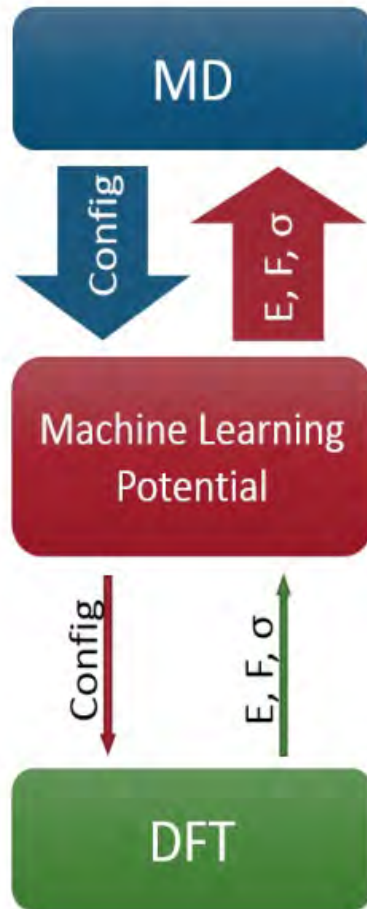
- + Fast
- Qualitative accuracy only

## AIMD



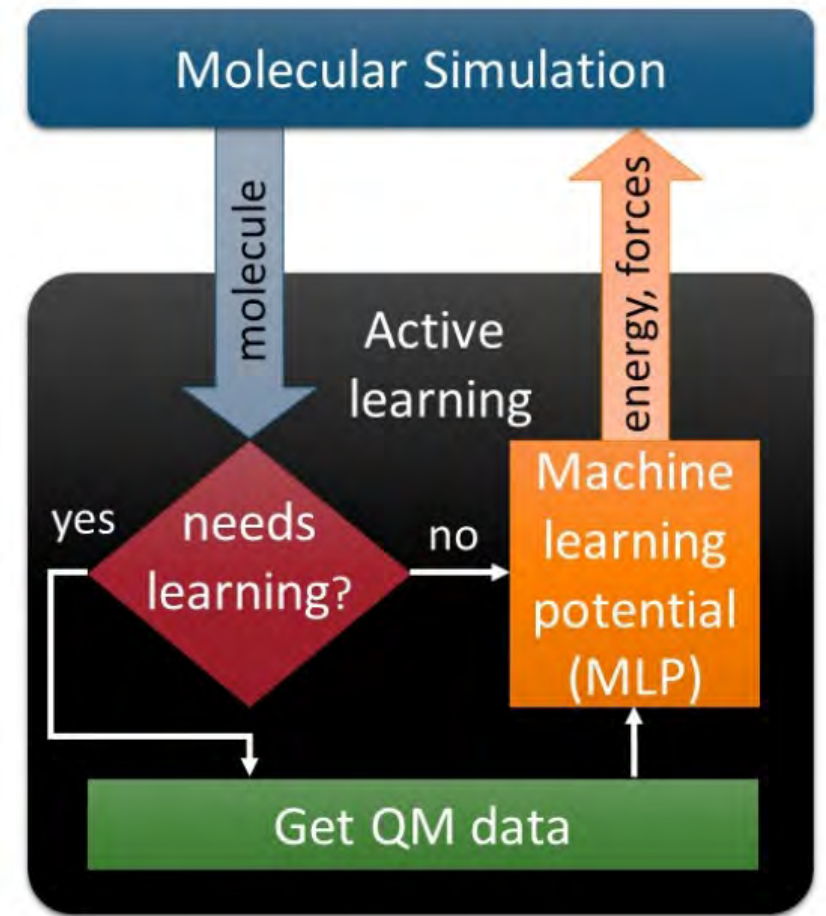
- + Accurate
- Time consuming

## Hybrid MD



- + Fast
- + Accurate (hopefully)

## Active learning



- + Fast
- + Accurate (for sure)<sup>41</sup>

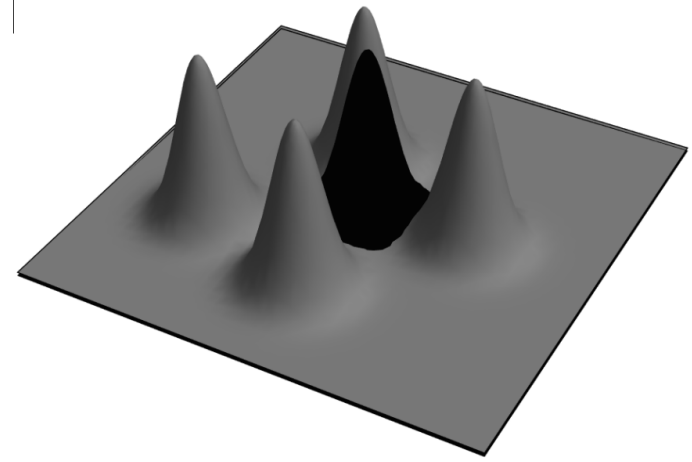


# MLIPs

Nowadays there are several types of machine learning interatomic potentials (MLIPs) are known and used

**GAP**

Gaussian Approximation

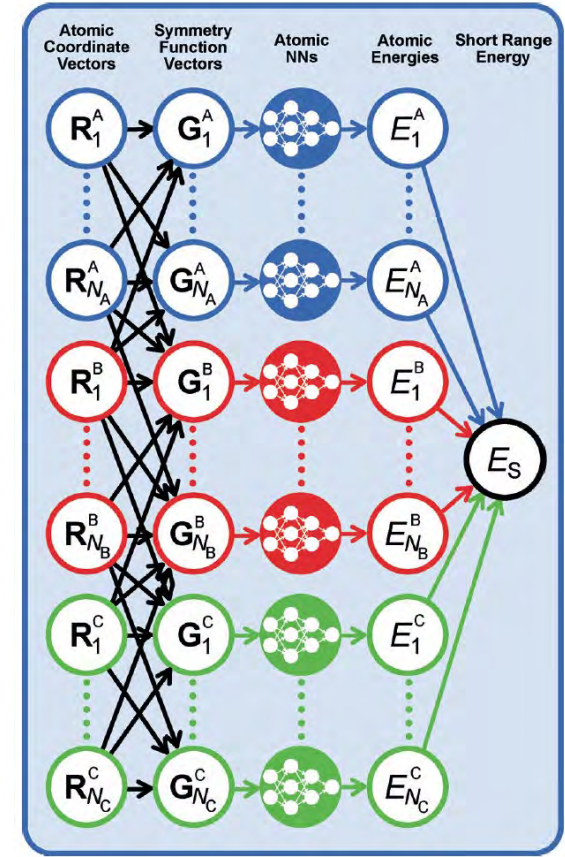


A 2D illustration of the atomic neighbor density function used in GAP

Bartók et al.

**NNP**

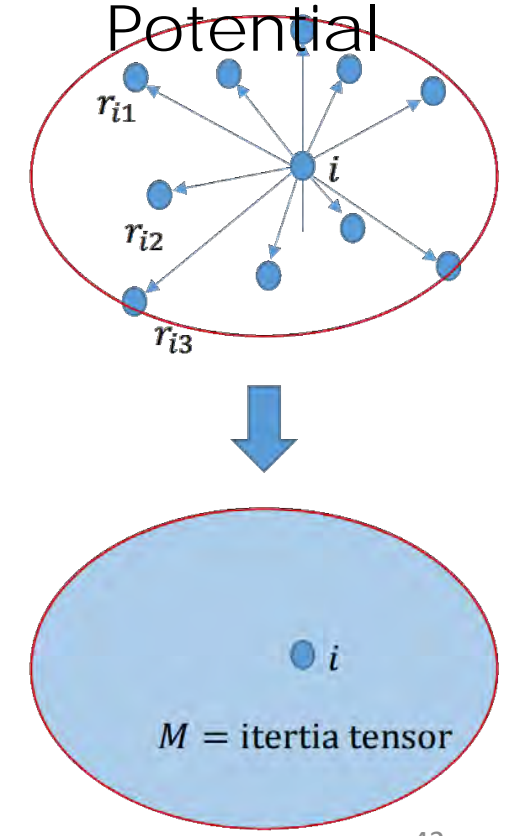
Neural Network



Behler et al.

**MTP**

Moment Tensor Potential



Shapeev et al.<sup>42</sup>

# Comparison of MLIPs

THE JOURNAL OF PHYSICAL CHEMISTRY A

pubs.acs.org/JPCA

Article

## Performance and Cost Assessment of Machine Learning Interatomic Potentials

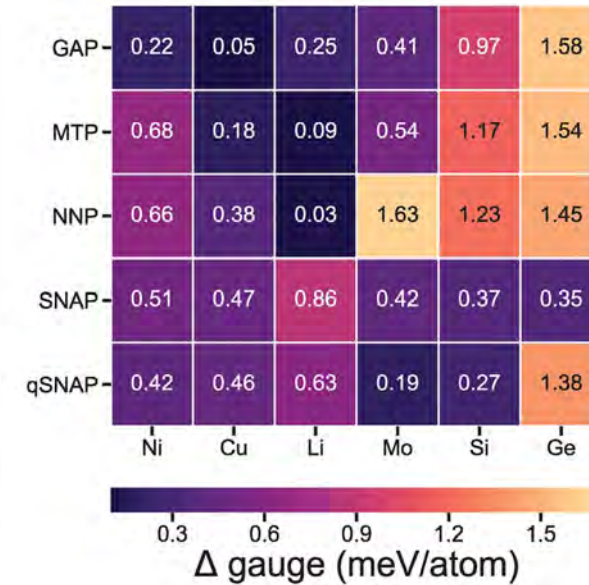
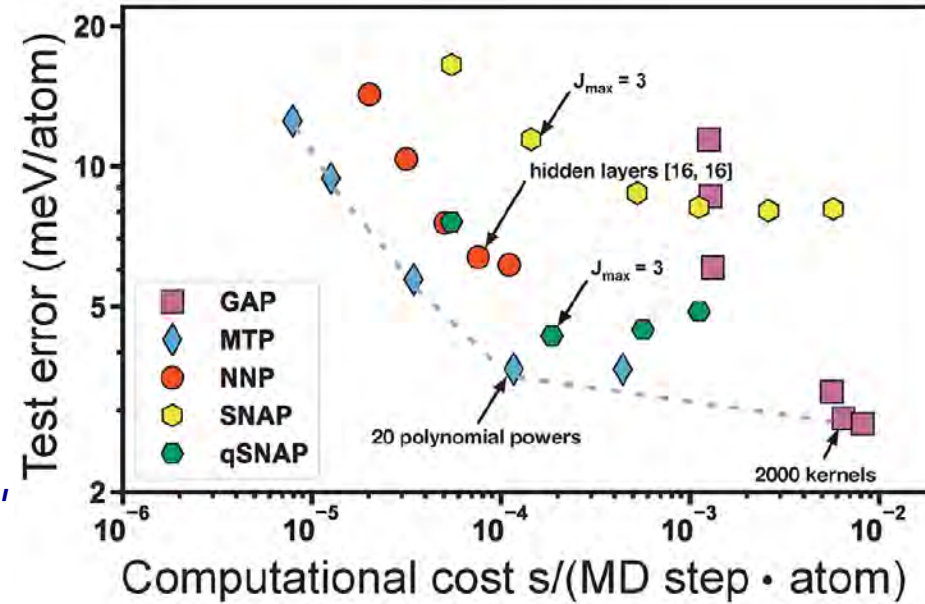
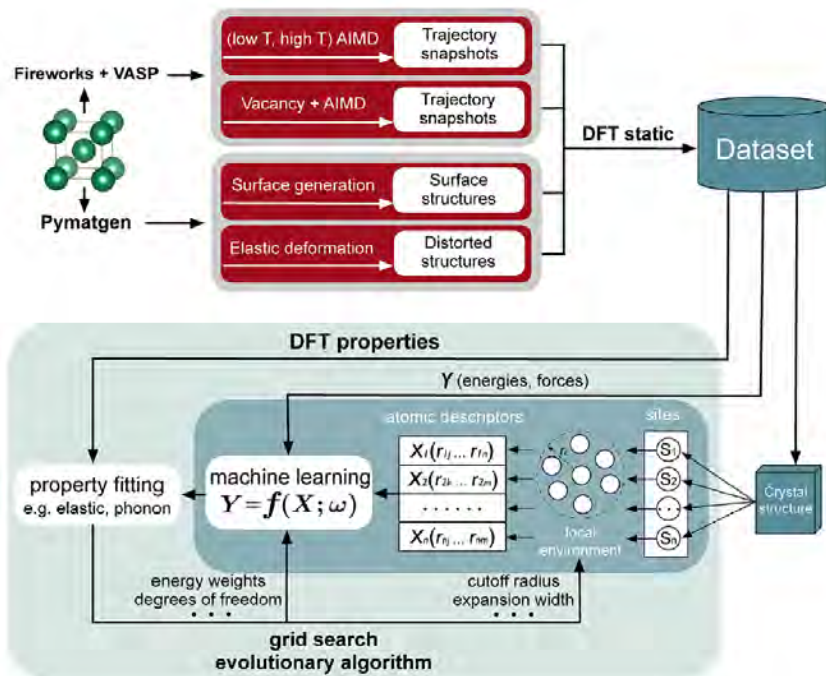
Published as part of *The Journal of Physical Chemistry virtual special issue "Machine Learning in Physical Chemistry"*.

Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V. Shapeev, Aidan P. Thompson, Mitchell A. Wood, and Shyue Ping Ong\*

Cite This: *J. Phys. Chem. A* 2020, 124, 731–745

Read Online


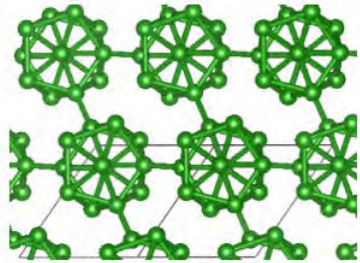
A comprehensive DFT data set was generated for six elements - **Li, Mo, Ni, Cu, Si**



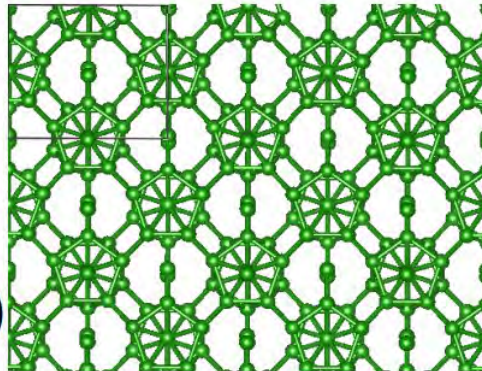
Test error versus computational cost for the Mo system. The gray dashed line indicates an approximate Pareto front.

This Pareto frontier represents an optimal trade-off between accuracy and computational cost.

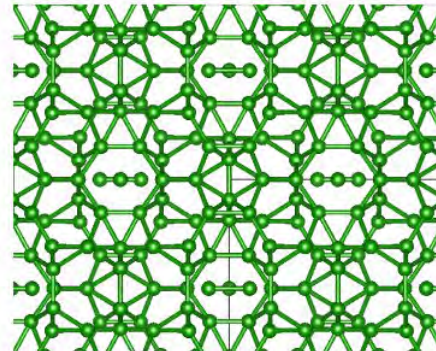
$\Delta$  gauge comparison provides quantitative estimate of deviation between the EOS curve from each ML-IAP with that of DFT.

PHYSICAL REVIEW B **99**, 064114 (2019)**Accelerating crystal structure prediction by machine-learning interatomic potentials with active learning**Evgeny V. Podryabinkin,<sup>1,\*</sup> Evgeny V. Tikhonov,<sup>1,2,3</sup> Alexander V. Shapeev,<sup>1</sup> and Artem R. Oganov<sup>1,4</sup><sup>1</sup>Skolkovo Institute of Science and Technology, Skolkovo Innovation Center, Nobel St. 3, Moscow 143026, Russia<sup>2</sup>Sino-Russian Joint Center for Computational Materials Discovery, State Key Laboratory of Solidification Processing, School of Material Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China<sup>3</sup>International Center for Materials Discovery, School of Material Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China<sup>4</sup>Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny, Moscow Region 141701, Russia (Received 22 February 2018; revised manuscript received 29 November 2018; published 27 February 2019)

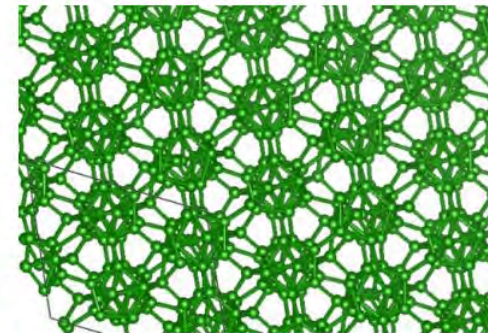
B-12 (6.7058eV/atom)  
10 days with DFT  
3 days with MLIP



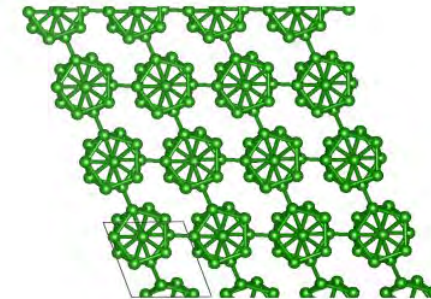
B-28 (6.678eV/atom)  
2 months with DFT  
5 days with MLIP



B-54 (6.667eV/atom)  
2 year on DFT  
8 days on MLIP



B-106  
Best Found on MLIP  
within 2 weeks



B-108 = B12 x9  
10 years with DFT  
2 weeks with MLIP

**Boron structure prediction challenges:**

- A lot of allotropes
- Some allotropes has more than 100 atoms (impossible for DFT)
- Small energy/atom difference between structures with PES minima

## Nanohardness from First Principles with Active Learning on Atomic Environments

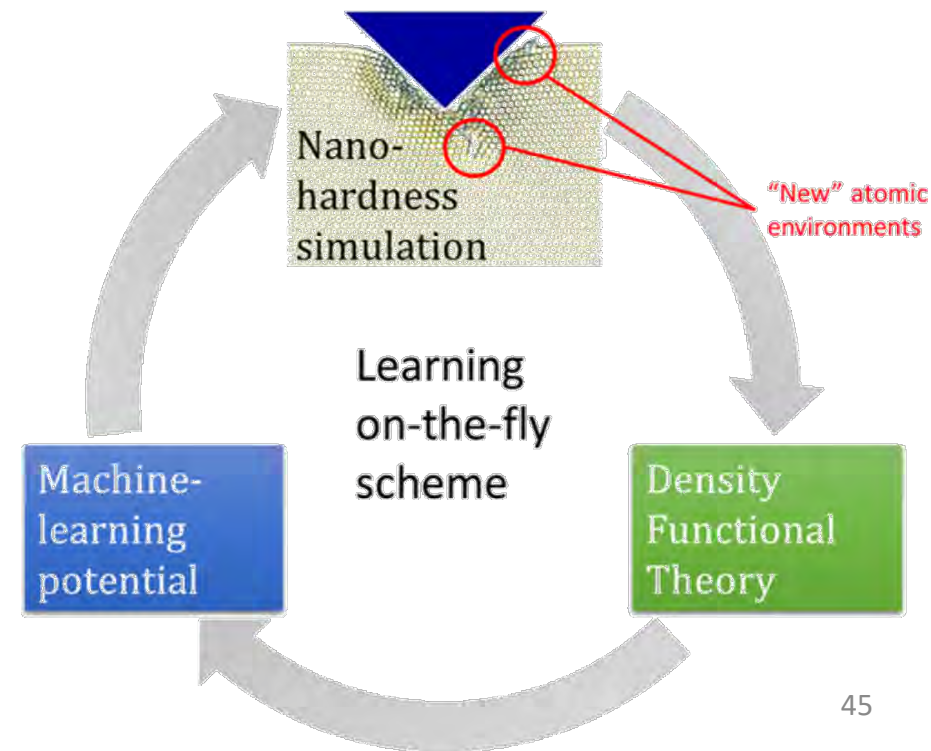
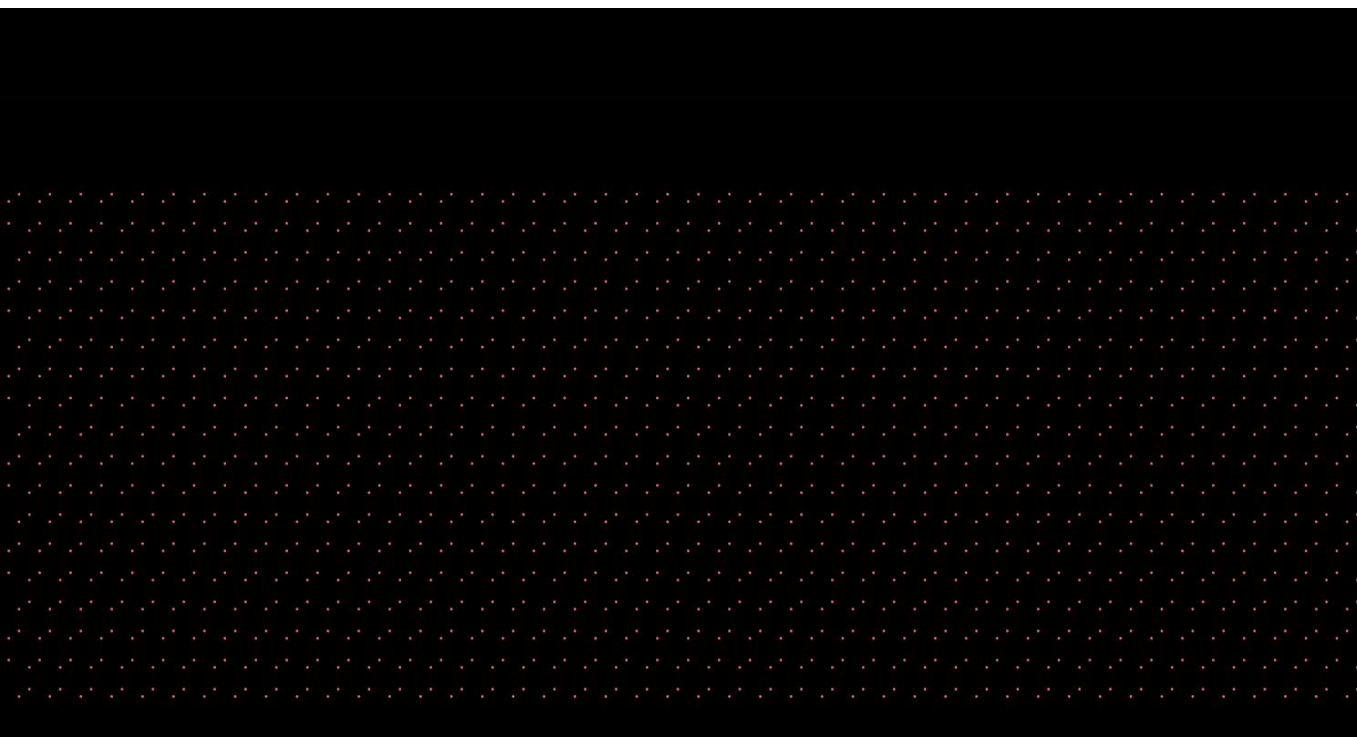
Evgeny V. Podryabinkin,\* Alexander G. Kvashnin, Milad Asgarpour, Igor I. Maslenikov, Danila A. Ovsyannikov, Pavel B. Sorokin, Mikhail Yu Popov, and Alexander V. Shapeev

Cite This: <https://doi.org/10.1021/acs.jctc.1c00783>

Read Online

Hardness is a very difficult characteristic for atomistic simulations

Simulation of nanoindentation of materials by using MTP via direct measurements of forces and imprints from the indenter (direct simulation of experimental setup)



# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- **Molecular dynamics**
  - Examples
- Extend molecular dynamics
- Monte-Carlo

- Solve Newton's equation of motion for N classical particles (3N coupled equations)
- For now, let limit our selves by natural NVE ensemble
- The force depends on positions only (not velocities)
- Each particle is allowed to interact simultaneously with every other particle and can experience an additional external potential
- A single point in a 6N-dimensional phase space  $(p,r)$  represents our dynamical system

**Our dimensionality:** N particles, 3N-dim

vectors

$$\vec{r} = \vec{q} = (x_1, y_1, z_1, x_2, \dots)$$

- Coordinates

$$v = d\vec{r}/dt$$

- Velocities

$$\vec{p} = m\vec{v}$$

- Momenta

$$\vec{a} = d\vec{v}/dt$$

- Accelerations

$$V(\vec{r})$$

- Potential energy

$$K(\vec{p}) = \sum \frac{1}{2} m \vec{v}_i^2$$

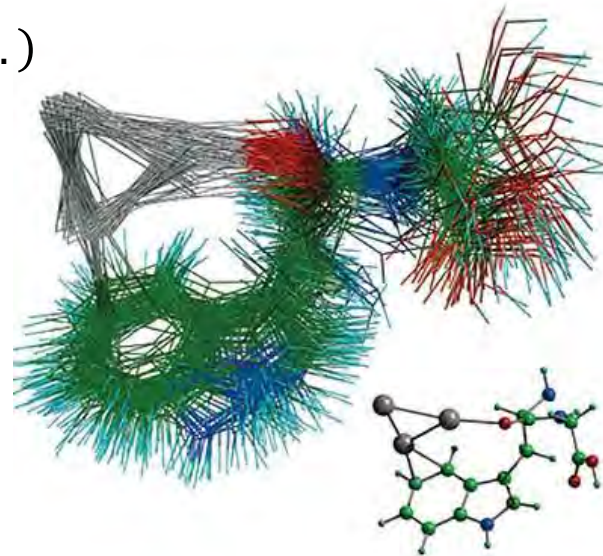
- Kinetic energy

$$\vec{F}(\vec{p}) = d\vec{V}/d\vec{r}$$

- Forces

$$m_i$$

- Particle masses



Ensemble of possible structures of a tryptophan-glycine-silver cluster cation complex, obtained by superimposing the configurations sampled from MD trajectory simulation  
Phys Rev Lett 101, 213001 (2008)

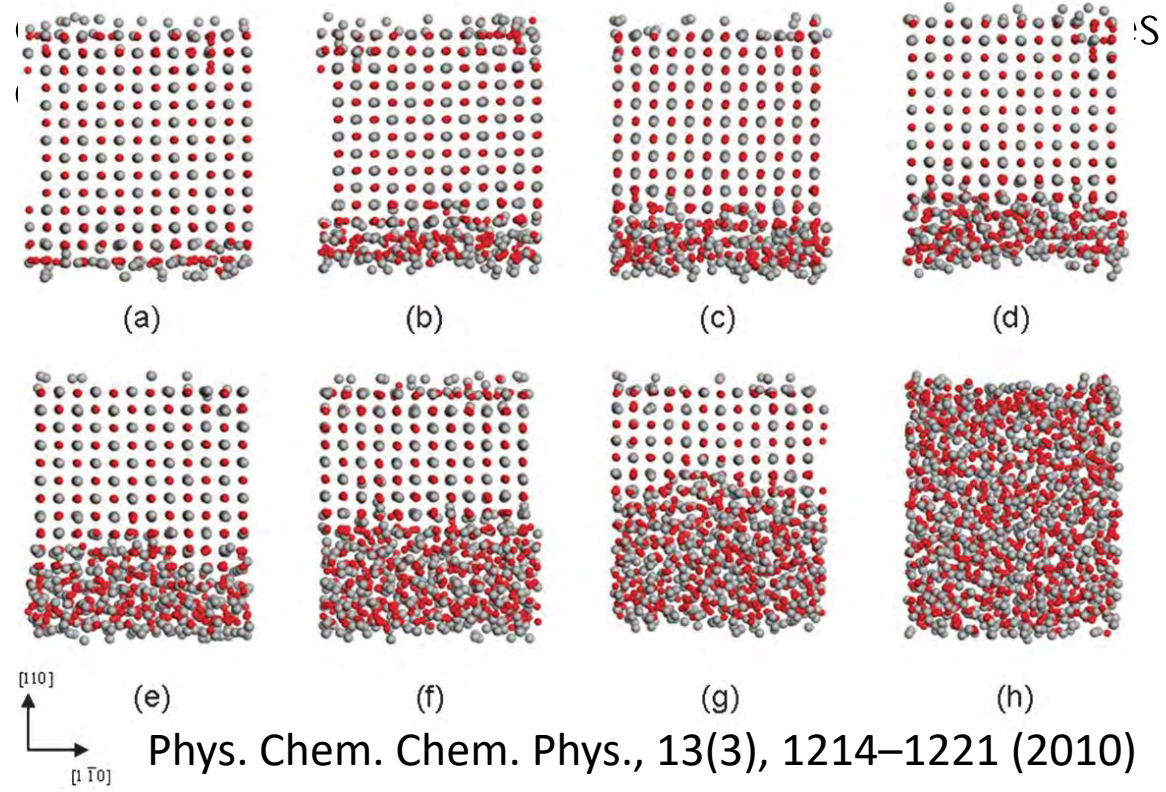
$$\vec{F} = m\vec{a}$$

$$E_{tot} = \sum_{i=1}^N \frac{1}{2} m \vec{v}_i^2 + V(\vec{r})$$

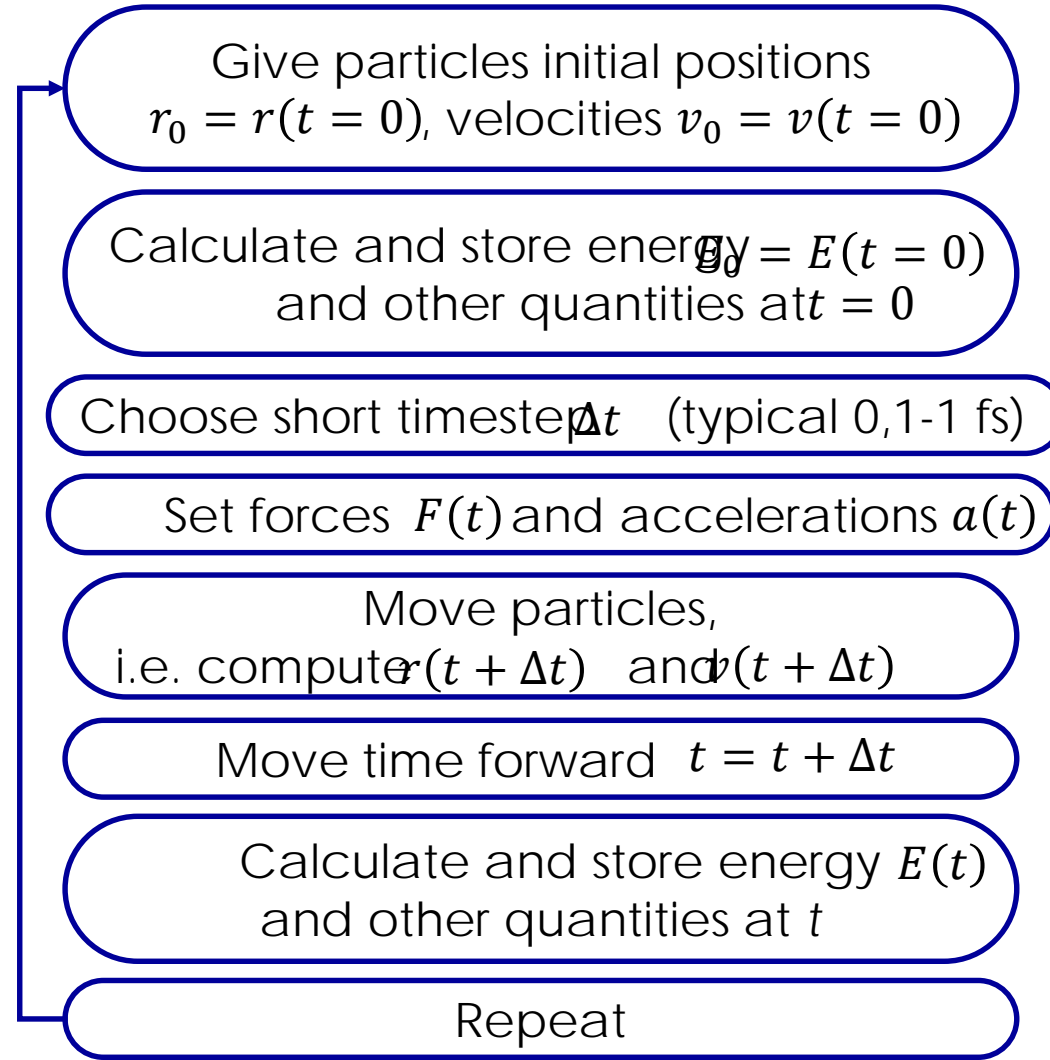
$$-\frac{dV}{d\vec{r}} = m \frac{d^2\vec{r}}{dt^2}$$

# Molecular dynamics Typical algorithm

- **Initialize**: select position and velocities
- **Propagate**: compute all forces, and then determine new positions
- **Equilibrate**: let the system reach equilibrium for a given thermodynamic ensemble (NVE, NPT etc.)
- **Sample (average)**: accumulate long



## Algorithm



### Selecting initial positions:

- Interconnection topology should agree with chemical structure
- Avoid short distances – huge energy penalty ( $\sim 1/r^{12}$  for Lenard-Jones)
- Avoid highly non-equilibrium conditions

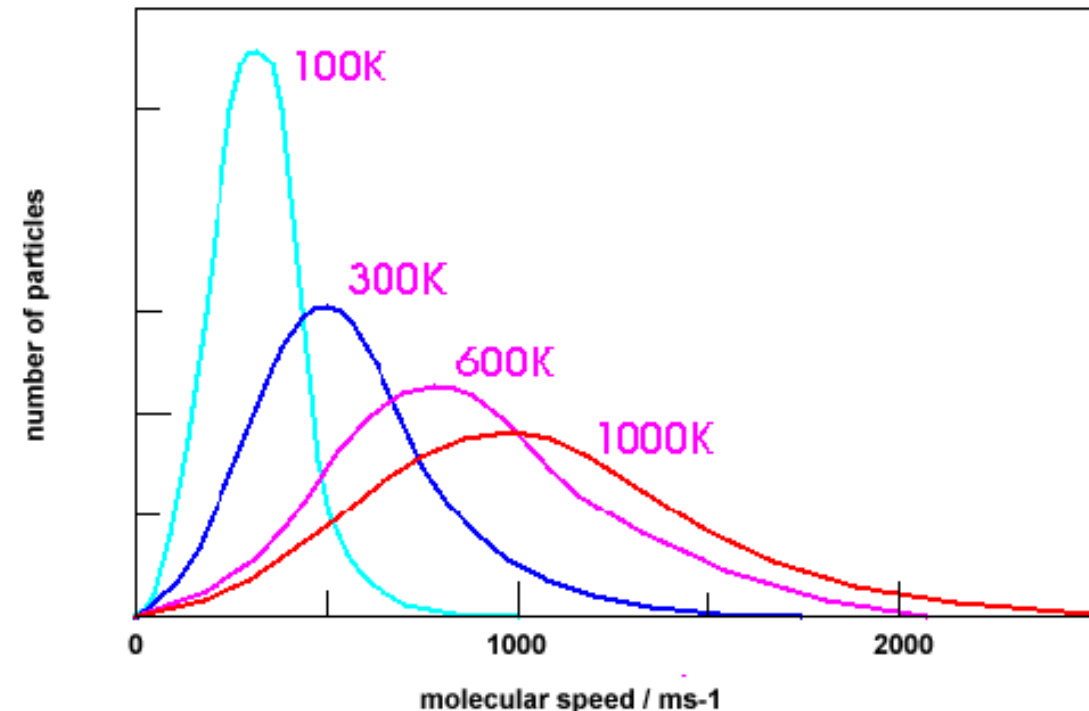
### Selecting initial velocities

- Start with  $v=0$ , then allow to equilibrate/thermalize with an increase of temperature OR
- Start with some distribution (e.g. Maxwell-Boltzmann distribution)

$$f(v) = \sqrt{\left(\frac{m}{2\pi k_B T}\right)^3} 4\pi v^2 \exp\left(-\frac{mv^2}{2k_B T}\right)$$

$$\frac{dN}{N} = f(v)dv \quad v_p = \sqrt{\frac{2k_B T}{m}}$$

$$\langle v \rangle = \sqrt{\frac{8k_B T}{\pi m}} \quad v_{rms} = \sqrt{\frac{3k_B T}{m}}$$





**Verlet type integration (typical accuracy up to  $O(t^4)$ ):** Most common in the MD land!  
Simple to calculate, well preserves the energy along the trajectory (i.e. time-reversible)

**As usual, start with Taylor expansion**

$$\vec{r}_{i+1} = \vec{r}_i + \frac{\partial \vec{r}}{\partial t} (\Delta t) + \frac{1}{2} \frac{\partial^2 \vec{r}}{\partial t^2} (\Delta t)^2 + \frac{1}{6} \frac{\partial^3 \vec{r}}{\partial t^3} (\Delta t)^3 + \dots$$

**or**  $\vec{r}_{i+1} = \vec{r}_i + \vec{v}_i(\Delta t) + \frac{1}{2} \vec{a}_i(\Delta t)^2 + \frac{1}{6} \vec{b}_i(\Delta t)^3 + \dots$  **and**  $\vec{r}_{i-1} = \vec{r}_i - \vec{v}_i(\Delta t) + \frac{1}{2} \vec{a}_i(\Delta t)^2 - \frac{1}{6} \vec{b}_i(\Delta t)^3 + \dots$

### The original Verlet

$$\vec{r}_{i+1} = (2\vec{r}_i - \vec{r}_{i-1}) + \vec{a}_i(\Delta t)^2 + \dots$$

$$\vec{a}_i = \frac{\vec{F}_i}{m_i} = -\frac{1}{m_i} \frac{dV}{d\vec{r}_i}$$

To initiate  $\vec{r}_{-1} = \vec{r}_0 - \vec{v}_0 \Delta t$

Note velocities ( $\vec{v}(t) = [\vec{r}(t + \Delta t) - \vec{r}(t - \Delta t)]/2\Delta t$ ) are not necessary but useful. Also  $2\vec{r}_i - \vec{r}_{i-1}$  is a large difference! Require 9N variables for storage for  $i=1, \dots, N$  (compact!)

### The leap-frog Verlet

$$\vec{r}_{i+1} = \vec{r}_i + \vec{v}_{i+\frac{1}{2}} \Delta t$$

$$\vec{v}_{i+\frac{1}{2}} = \vec{v}_{i-\frac{1}{2}} + \vec{a}_i \Delta t$$

$$\vec{a}_i = \frac{\vec{F}_i}{m_i} = -\frac{1}{m_i} \frac{dV}{d\vec{r}_i}$$

To initiate  $\vec{r}_{-1} = \vec{r}_0 - \vec{v}_0 \Delta t$

Note velocities ( $\vec{v}_{i+\frac{1}{2}}$ ) 'leap' over coordinates half-step. Explicit velocities  $\vec{v}$  is a plus. But  $\vec{r}$  and are out of phase

### The Velocity Verlet

$$\vec{r}_{i+1} = \vec{r}_i + \vec{v}_i \Delta t + \frac{1}{2} \vec{a}_i \Delta t^2$$

$$\vec{v}_{i+1} = \vec{v}_i + \frac{1}{2} \{\vec{a}_i + \vec{a}_{i+1}\} \Delta t$$

$$\vec{a}_i = \frac{\vec{F}_i}{m_i} = -\frac{1}{m_i} \frac{dV}{d\vec{r}_i}$$

To initiate  $\vec{r}_{-1} = \vec{r}_0 - \vec{v}_0 \Delta t$

**Best numerical performance and compact storage make it method of choice for MD codes!**

**NVE (microcanonical) ensemble is natural:** energy is conserved being redistributed along potential and kinetic parts along the trajectory

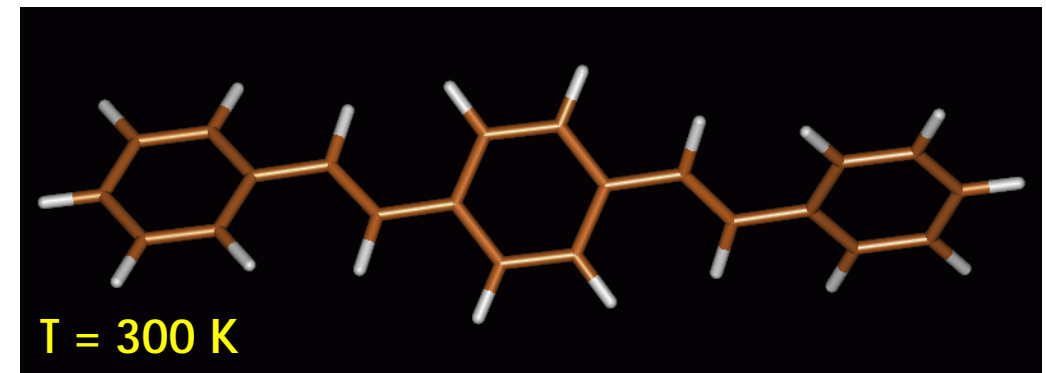
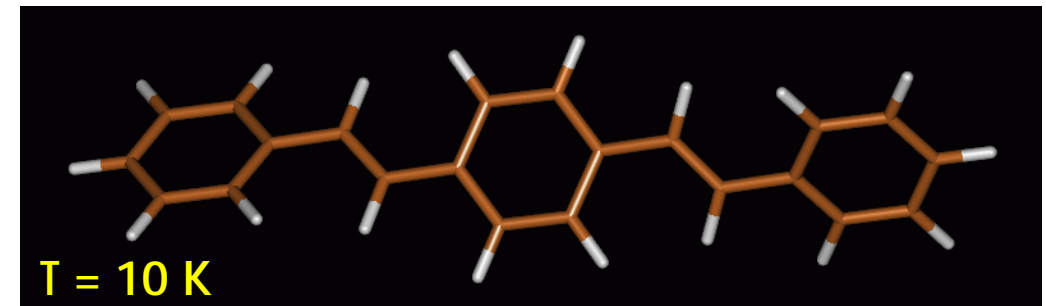
**What about NVT (canonical) ensemble?** Need thermostat (e.g. Anderson, Nose-Hoover, Berendsen, Langevin etc.) allowing an exchange of the energy with a bath (e.g. solvent, buffer-gas)

**Langevin equation of motion**

$$m \frac{d^2 \vec{r}}{dt^2} = -\zeta \frac{d\vec{r}}{dt} + \vec{F}_{intra} + \vec{F}_{random}$$

**Compared to the standard Newton's equation we have new terms:**

- 1)  $\zeta \vec{v}$  - friction term with a friction coefficient ( $\sim 2\text{ps}^{-1}$ ), which removes energy from the system
- 2)  $\vec{F}_{random}$  - random force associated with the temperature, which balances energy from the system



## Molecular dynamics with coupling to an external bath

H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola,<sup>a)</sup> and J. R. Haak  
*Laboratory of Physical Chemistry, The University of Groningen, Nijenborgh 16, 9747 Ag Groningen, The Netherlands*

(Received 30 April 1984; accepted 27 June 1984)

In molecular dynamics (MD) simulations the need often arises to maintain such parameters as temperature or pressure rather than energy and volume, or to impose gradients for studying transport properties in nonequilibrium MD. A method is described to realize coupling to an external bath with constant temperature or pressure with adjustable time constants for the coupling. The method is easily extendable to other variables and to gradients, and can be applied also to polyatomic molecules involving internal constraints. The influence of coupling time constants on dynamical variables is evaluated. A leap-frog algorithm is presented for the general case involving constraints with coupling to both a constant temperature and a constant pressure bath.

J. Chem. Phys. **81** (8), 15 October 1984

## Thermostat

$$\sum_i \frac{|\vec{p}_i|^2}{m_i} = 3Nk_B T$$

## Barostat

$$\sum_i \left( \frac{|\vec{p}_i|^2}{m_i} + \sum_{j>i} \vec{r}_{ij} \vec{F}_{ij} \right) = 3pV$$

Rescale velocities by  $\lambda$  and coordinates (or unit cell) by  $\mu$ , where

$$\lambda^2 = 1 + \gamma \Delta t \left( \frac{T_0}{T} - 1 \right), \quad \mu^3 = 1 - \beta \Delta t (p_0 - p),$$

Advanced thermostat, barostat, and integrator see in program manuals

LAMMPS users see `fix nvt/npt/nph` command

# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- **Molecular dynamics**
  - **Examples**
- Extend molecular dynamics
- Monte-Carlo

Simulation of indentation of nanometer-thick carbon membranes by using classical MD with Tersoff potential (direct simulation of experimental setup)

### Lonsdaleite Films with Nanometer Thickness

Alexander G. Kvashnin<sup>\*,†,‡</sup> and Pavel B. Sorokin<sup>†,‡,||</sup>

<sup>†</sup>Technological Institute of Superhard and Novel Carbon Materials, 7a Centralnaya Street, Troitsk, Moscow 142190, Russian Federation

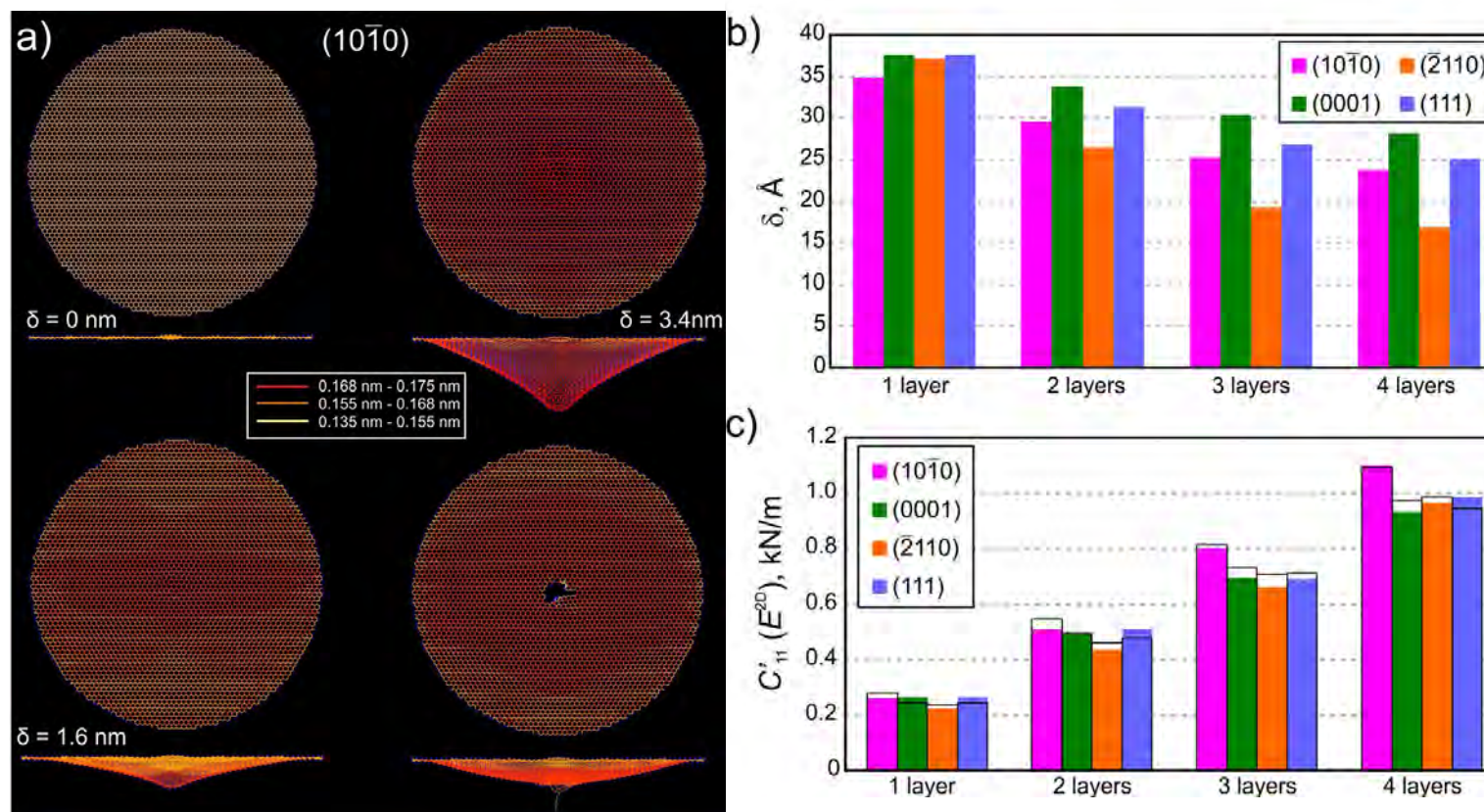
<sup>‡</sup>Moscow Institute of Physics and Technology, 9 Institutsky lane, Dolgoprudny 141700, Russian Federation

<sup>||</sup>National University of Science and Technology "MISIS", Leninsky Avenue 4, Moscow 119049, Russian Federation

✉ Cite this: *J. Phys. Chem. Lett.* 2014, 5, 3, 541–548

Publication Date: January 24, 2014

<https://doi.org/10.1021/jz402528q>

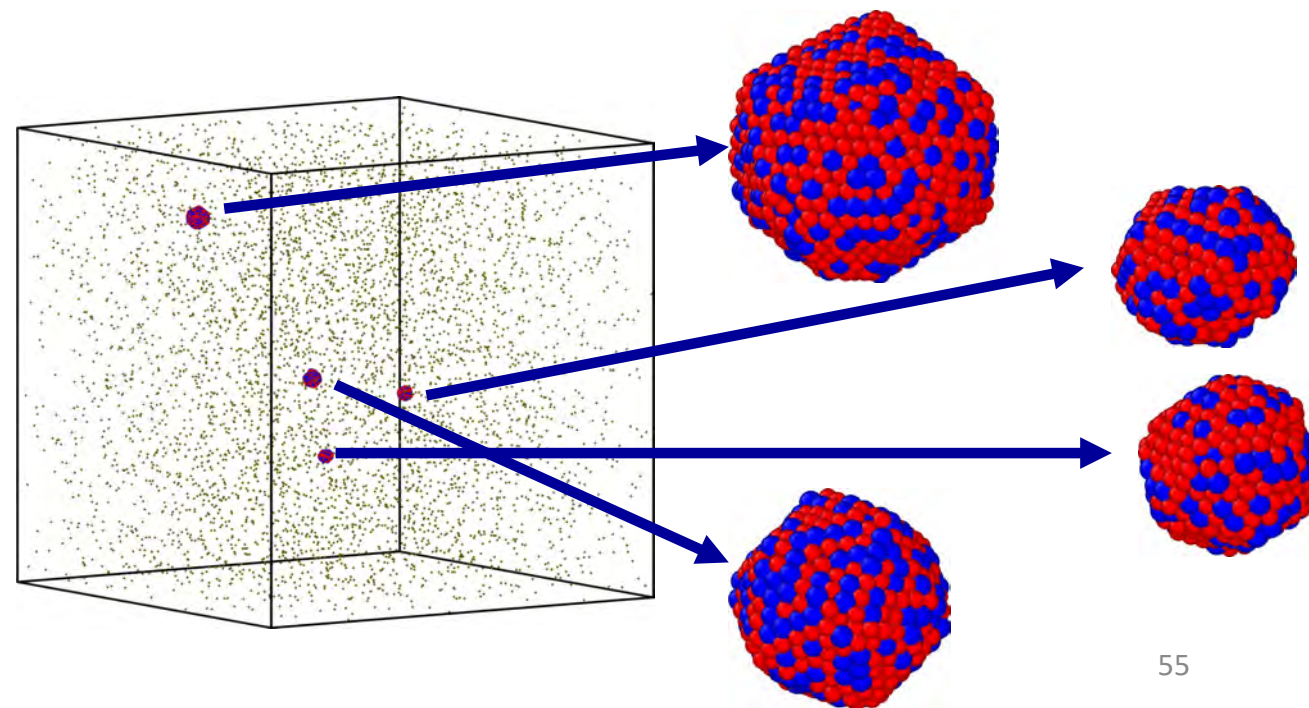
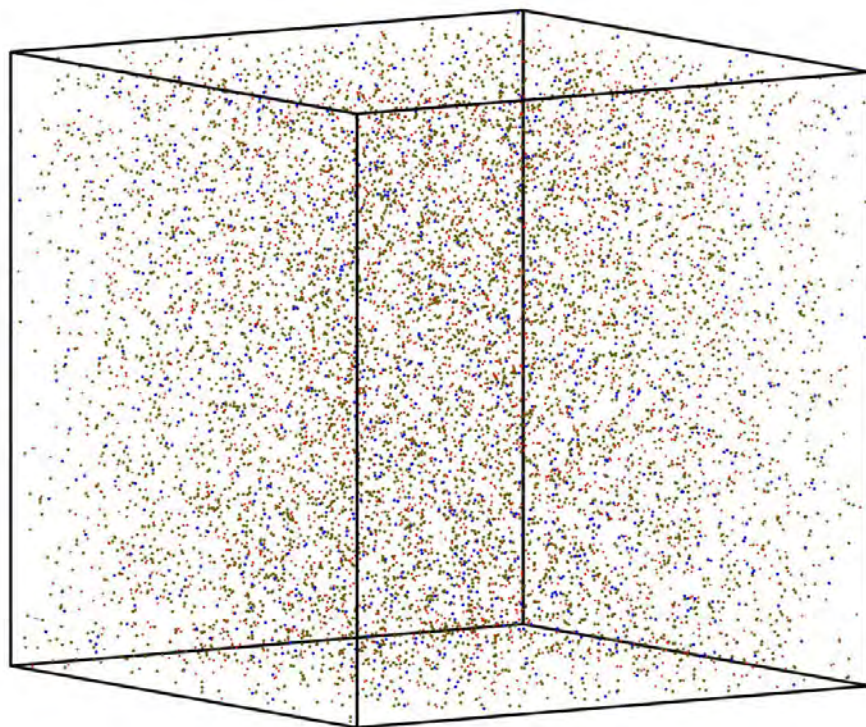


## Cu–Au nanoparticles produced by the aggregation of gas-phase metal atoms for CO oxidation

Ilya V. Chepkasov<sup>1</sup> | Viktor S. Baidyshev<sup>2</sup> | Artem A. Golubnichiy<sup>2</sup> | Ivan S. Zamulin<sup>2</sup> | Alexander G. Kvashnin<sup>1</sup> | Sergey M. Kozlov<sup>3</sup> 

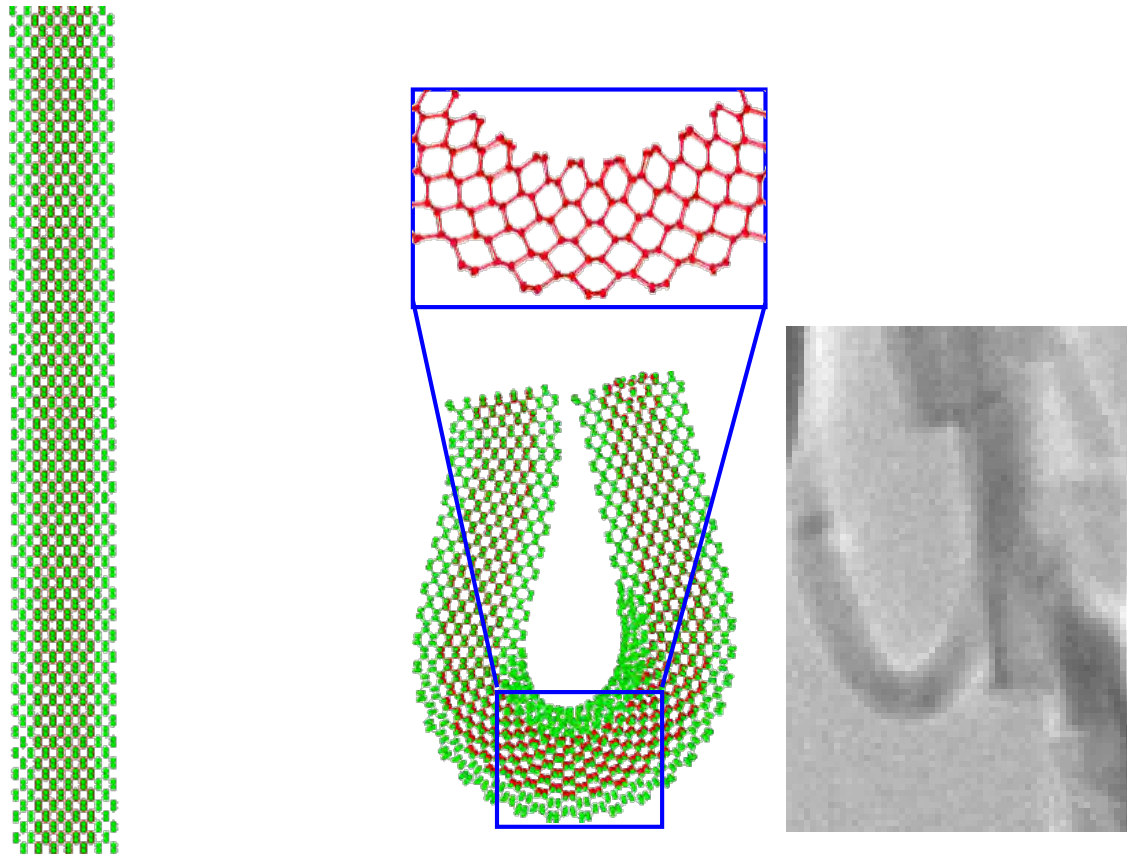
## EAM potential

Molecular dynamics to simulate the PVD synthesis of Cu–Au nanoalloys through the non-isothermal aggregation of Cu and Au atoms at a 3:1 ratio in the Ar atmosphere to obtain realistic structures of Cu–Au nanoparticles



### Mechanical, Electrical, and Crystallographic Property Dynamics of Bent and Strained Ge/Si Core–Shell Nanowires As Revealed by *in situ* Transmission Electron Microscopy

Chao Zhang,<sup>\*,†</sup> Dmitry G. Kvashnin,<sup>‡,⊥</sup> Laure Bourgeois,<sup>§</sup> Joseph F. S. Fernando,<sup>†</sup> Konstantin Firestein,<sup>†</sup> Pavel B. Sorokin,<sup>‡,⊥</sup> Naoki Fukata,<sup>||</sup> and Dmitri Golberg<sup>\*,†,||</sup>



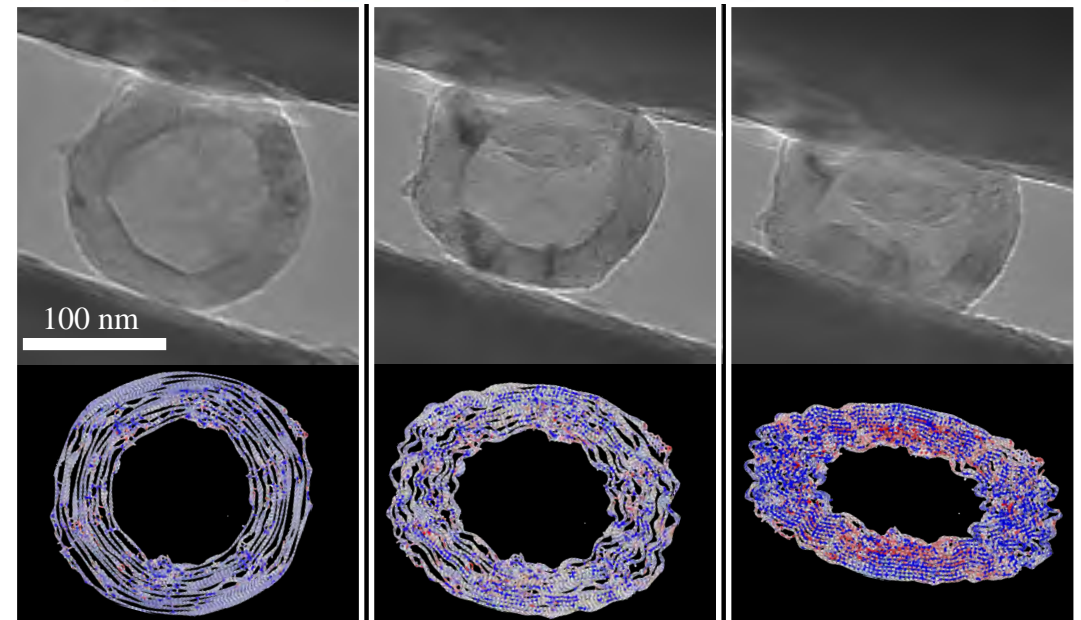
### Tersoff many-body potential

*Nanoscale*, 10, 8099 (2018)

### Compressive properties of hollow BN nanoparticles: theoretical modeling and testing using a high-resolution transmission electron microscope



Konstantin L. Firestein,<sup>ID \*a,b</sup> Dmitry G. Kvashnin,<sup>ID a</sup> Andrey M. Kovalskii,<sup>a</sup> Zakhar I. Popov,<sup>ID a</sup> Pavel B. Sorokin,<sup>ID \*a,c</sup> Dmitri V. Golberg<sup>ID b,d</sup> and Dmitri V. Shtansky<sup>ID \*a</sup>



Nanoscale



PAPER

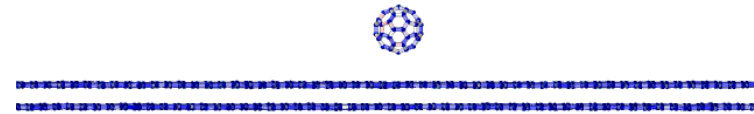
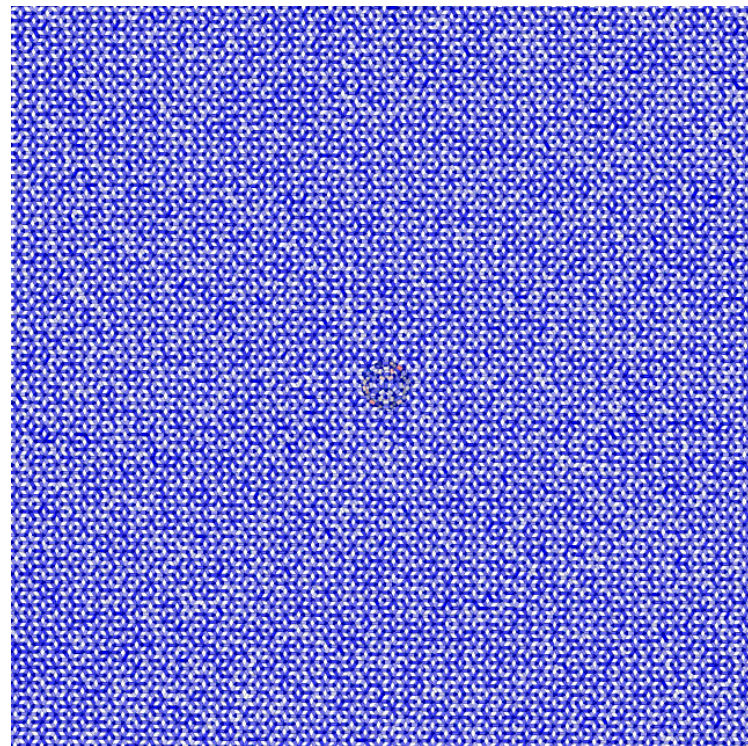
View Article Online  
View Journal | View Issue

Check for updates

### Nanostructuring few-layer graphene films with swift heavy ions for electronic application: tuning of electronic and transport properties†

N. A. Nebogatikova,<sup>1</sup> I. V. Antonova,<sup>2,3</sup> S. V. Erohin,<sup>4,5</sup> D. G. Kvashnin,<sup>6,7</sup> A. Olejniczak,<sup>8,9</sup> V. A. Volodin,<sup>10,11</sup> A. V. Skuratov,<sup>12,13</sup> A. V. Krashennnikov,<sup>14,15</sup> P. B. Sorokin<sup>16,17,18</sup> and L. A. Chernozatonskii<sup>19</sup>

Cite this: *Nanoscale*, 2018, 10, 14499

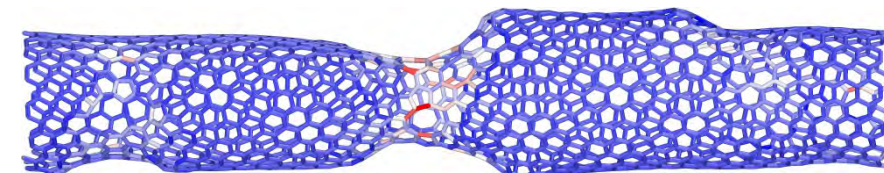
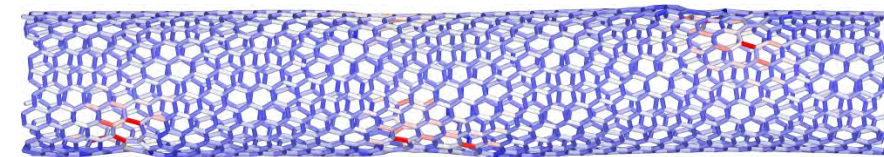
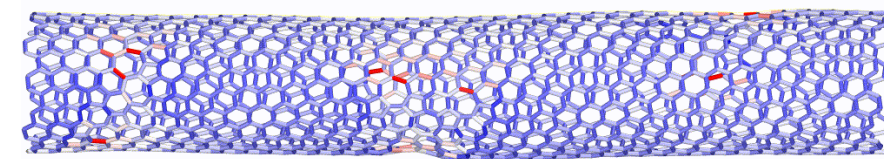


Application of classical potentials and MD simulations to describe irradiation and changes of structure of carbon nanomaterials

## Tersoff many-body potential Semiconductor nanochannels in metallic carbon nanotubes by thermomechanical chirality alteration

Dai-Ming Tang<sup>1\*</sup>, Sergey V. Erohin<sup>2</sup>, Dmitry G. Kvashnin<sup>2,3</sup>, Victor A. Demin<sup>3</sup>, Ovidiu Cretu<sup>4</sup>, Song Jiang<sup>5</sup>, Lili Zhang<sup>5</sup>, Peng-Xiang Hou<sup>5</sup>, Guohai Chen<sup>6</sup>, Don N. Futaba<sup>6</sup>, Yongjia Zheng<sup>7</sup>, Rong Xiang<sup>7</sup>, Xin Zhou<sup>1</sup>, Feng-Chun Hsia<sup>1</sup>, Naoyuki Kawamoto<sup>4</sup>, Masanori Mitome<sup>1</sup>, Yoshihiro Nemoto<sup>8</sup>, Fumihiko Uesugi<sup>8</sup>, Masaki Takeguchi<sup>8</sup>, Shigeo Maruyama<sup>7</sup>, Hui-Ming Cheng<sup>5,9,10</sup>, Yoshio Bando<sup>11,12</sup>, Chang Liu<sup>5\*</sup>, Pavel B. Sorokin<sup>2,13\*</sup>, Dmitri Golberg<sup>1,14\*</sup>

Tang *et al.*, *Science* **374**, 1616–1620 (2021)





# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- **Extend molecular dynamics**
- Monte-Carlo

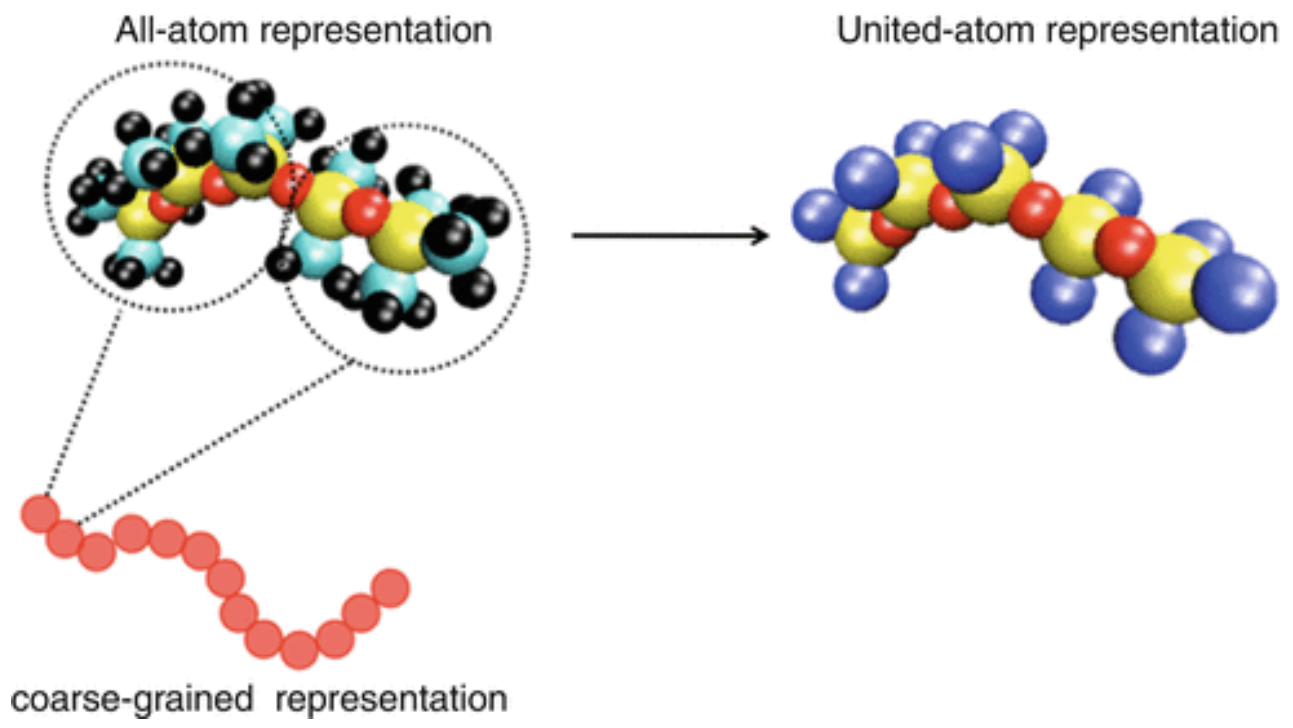
# Extend MD **Coarse graining**

**Aim** is to simulate the behavior of complex systems using their coarse-grained (simplified) representation.

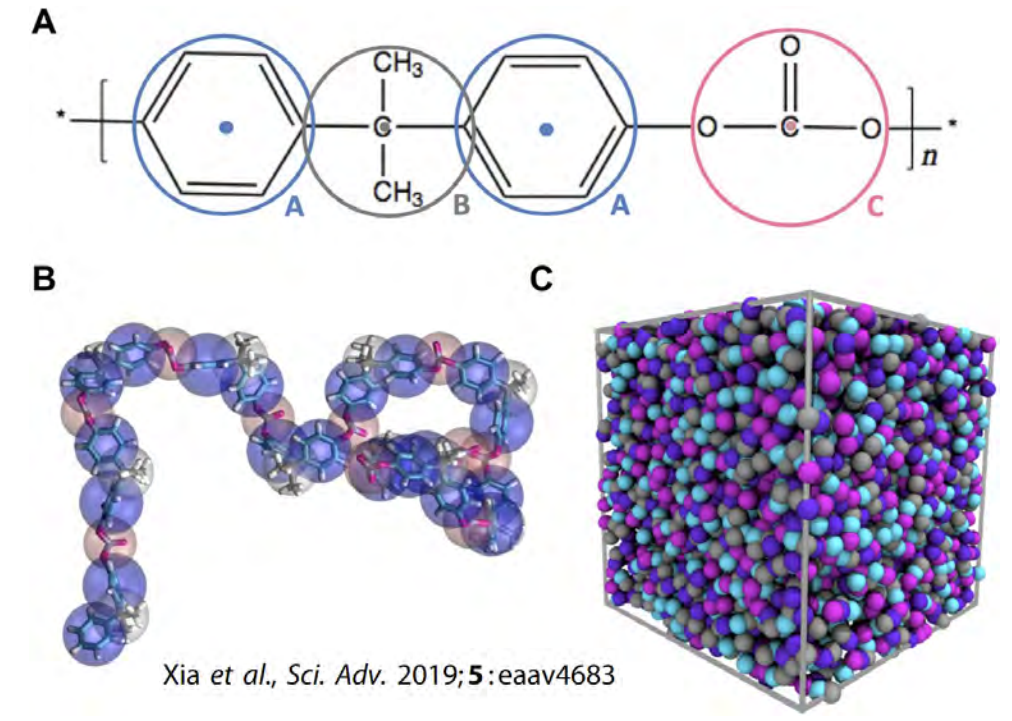
Coarse-grained models are widely used for molecular modeling of biomolecules at various granularity levels.

**Idea** is to simulate coarse grained model, then get back to all-atom description to refine local geometry

## Multiscale coarse-grained (CG)



Bisphenol-A polycarbonate chemical structure and the corresponding CG



# Statistical mechanics

**Thermodynamics variables:** volume ( $V$ ), pressure ( $p$ ), temperature ( $T$ ). Number of particles ( $N$ ), energy ( $E$ ), chemical potential ( $\mu$ )

**Statistical mechanics:** connection between properties of a microscopic system and a macroscopic sample

**Ensemble:** collection of microscopic states consistent with thermodynamic boundary conditions; defined by 3 variables ( $NVT$ ) or ( $NVE$ ) or ( $NPT$ ) or ( $\mu VT$ )

**Boltzmann probability:** relative probability to be in a state with an energy  $E$  at a temperature  $T$  and is proportional to  $\exp(-E/k_B T)$

**Partition function** is a key quantity in statistical mechanics, a normalization factor for the Boltzmann probability distribution:

$$Q = \sum_i^{\infty} e^{-E_i/k_B T} = \int e^{-E(\vec{r}, \vec{p})/k_B T} d\vec{r} d\vec{p}$$

**Everything can be expressed through  $Q$ , e.g., for a canonical ensemble ( $NVT$ )**

**- Internal energy**

$$U = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V$$

**- Enthalpy**

$$H = U + PV = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V + k_B T V \left( \frac{\partial \ln Q}{\partial V} \right)_T$$

**- Entropy**

$$S = \frac{U - A}{T} = k_B T^2 \left( \frac{\partial \ln Q}{\partial T} \right)_V + k_B T \ln Q$$

**- Gibbs free energy**

$$G = H - TS = k_B T V \left( \frac{\partial \ln Q}{\partial V} \right)_T - k_B T \ln Q$$

**- Helmholtz free energy**

$$A = U - TS = -k_B T \ln Q$$

# Thermodynamic ensembles

**Microcanonical ( $NVE$ ):** Newtonian system ( $N = \text{const}$ ) in box ( $V = \text{const}$ ) with elastic walls (or PBC)

**Canonical ( $NVT$ ):** Newtonian system ( $N = \text{const}$ ) in box ( $V = \text{const}$ ) with non-elastic walls (walls are equilibrated with  $T = \text{const}$  - thermostat)

**Isothermal-isobaric ( $NPT$ ):** Newtonian system ( $N = \text{const}$ ) in box with varying volume (keeping  $P = \text{const}$  - barostat) and non-elastic walls (keeping  $T = \text{const}$  - thermostat)

**Grand-canonical ( $\mu VT$ ):** Open system (number of particles is not conserved but their energy in the reservoir is fixed at  $\mu$ )

Constants in different ensembles, and corresponding equilibrium states

$N$	$P$	$V$	$T$	$E$	$\mu$	Acronym	Equilibrium	Name
×		×	×			$NVT$	$A$ has minimum	Canonical
×		×		×		$NVE$	$S$ has maximum	Micro-canonical
×	×		×			$NPT$	$G$ has minimum	Isothermal-isobaric
		×	×		×	$VE\mu$	$(PV)$ has maximum	Grand canonical

**Ergodicity:** Time-average of a property can be replaced by a suitable average over collection of possible microscopic states (**ensemble average**)

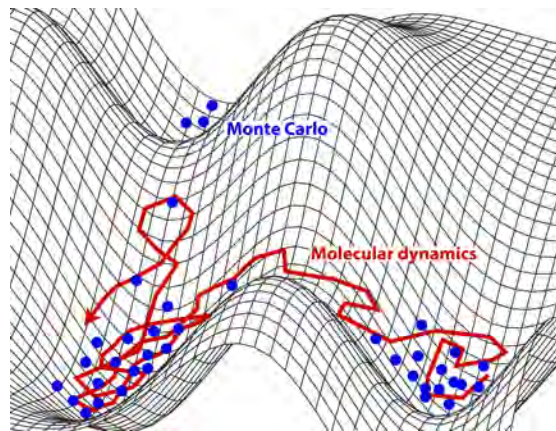
Given a potential energy surface  $E_p(r)$ , what are the possible modeling routes for generating ensembles?

**Molecular dynamics (MD):** propagate Newton's equation of motion, analyze trajectories – generally gives rates and time constants

**Monte-Carlo (MC):** 'flip a coin' statistical approach  
( $T$  is an essential component) – gives information on the thermodynamics, no rates and time scales

Differences between Monte Carlo and molecular dynamics methods

Property	MC	MD
Basic information needed	Energy	Gradient
Particles moved in each step	One	All
Coordinates	Any	Cartesian
Constraints	Easy	Difficult
Atomic velocities	No	Yes
Time dimension	No	Yes
Deterministic	No	(Yes)
Sampling	Non-physical	Physical
Natural ensemble	$NVT$	$NVE$



# Outline

- Born-Oppenheimer approximation
- Molecular mechanics
- Potentials of interatomic interaction
- Energy contributions
- Force fields
- ML potentials
- Molecular dynamics
  - Examples
- Extend molecular dynamics
- **Monte-Carlo**

## Replace strategy

“first pick, then weight” with  
“first weight, the pick”

by penalizing states with high energy  
by a Boltzmann factor  $\exp(-\Delta E/k_B T)$

## How?

Say on  $(i-1)$  step, the energy of accepted  
configuration is  $E_{i-1}$

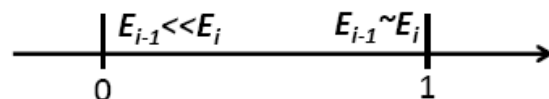
on  $i^{\text{th}}$  step, the energy of new  
configuration is  $E_i$

**Case1:**  $E_i < E_{i-1}$ : Good! We are going lower,  
accept  $i$

**Case2:**  $E_i > E_{i-1}$ : We are going higher,  
calculate  $\exp(-\Delta E/k_B T) = \exp(-(E_i - E_{i-1})/k_B T)$ ,  
pick a random number  $\xi$  from  $[0, 1]$

If  $\xi > \exp(-\Delta E/k_B T)$  reject it

If  $\xi < \exp(-\Delta E/k_B T)$  accept it



## Metropolis (+Fermi, Ulam, von Neuman)

**Monte-Carlo (1952):** “walks” through phase space  
(Markov chain of states) visiting each state with  
proper probability (in the infinite time limit)

## Algorithm

Generate trial configuration  $\mathbf{q}_1$ , its energy  
 $E_1 = E(\mathbf{q}_1)$  and store the desired property  
 $A_1 = A(\mathbf{q}_1)$

For  $i = 2, \dots, M$  do

Perturb the system, get new configurations

Calculate energy of  $\mathbf{q}_i$ :  $E_i = E(\mathbf{q}_i)$

Test  $E_i$  vs  $E_{i-1}$  for acceptance

If rejected, discard  $\mathbf{q}_i$

If accepted, calculate and store  $A_i = A(\mathbf{q}_i)$

If  $i < M$

Calculate  $\langle A(M) \rangle = \frac{1}{M} \sum_{i=1}^M A_i$

- 1) The natural Monte-Carlo ensemble is (NVT)
- 2) The “Devil is in details”: the art of running an MC calculation lies in defining the perturbation step(s)  
If the steps are very small, then volume of phase space sampled will increase only slowly over time, and it is computationally expensive. If the steps are too large, then the rejection rate will grow so high that again resources will be wasted by an inefficient phase space.

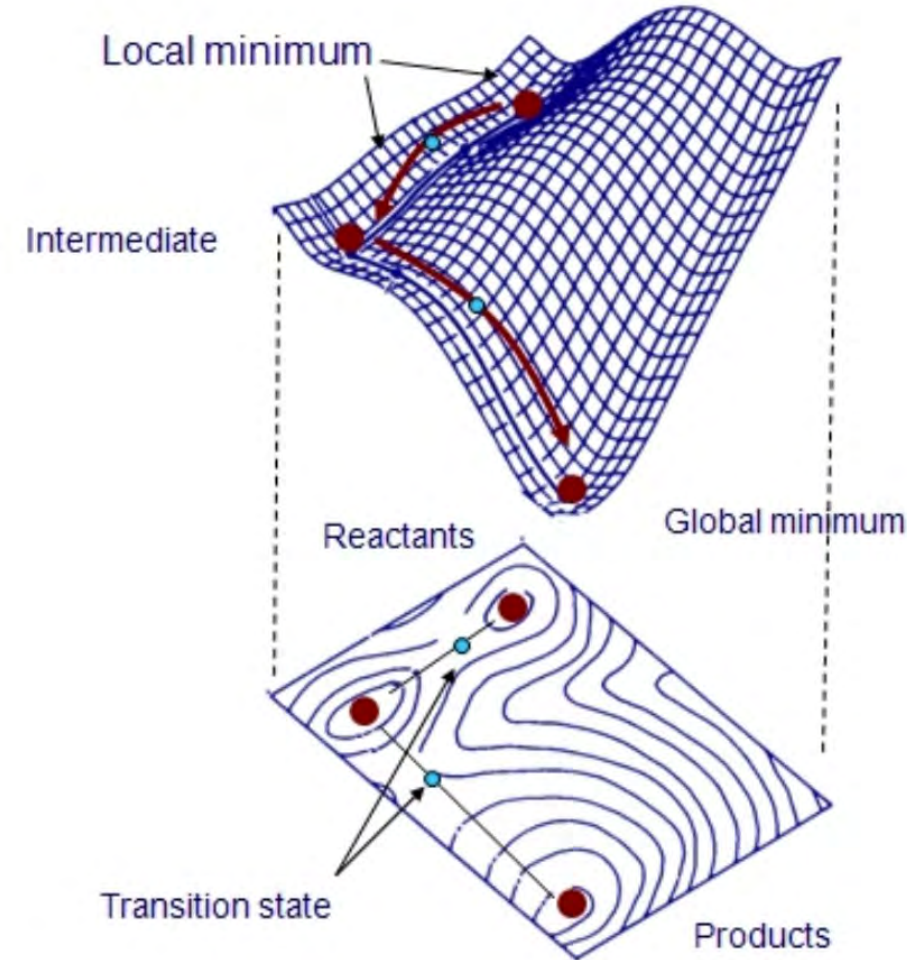
### 3) Monitoring convergence of $\langle A(M) \rangle$ and possibly $\langle E(M) \rangle$ with M is necessary

This will tell you a lot about convergence rate and possible transitions to a different PES valley

### 4) Multiple MC variations were developed, including different ensembles, annealing (varying T) algorithms, simulating timescales.

For example, (NPT) isothermal-isobaric ensemble, G (Gibbs free energy) at min (frequent in experimental conditions)

$$\Delta E \rightarrow \Delta E + P\Delta V - Nk_B T \ln(1 + \Delta V/V)$$



Example of PES



**Thx!**