

Hands-on Molecular Dynamics with GROMACS on Discoverer: running efficient jobs and building GROMACS from source

Stephen Farr, EPCC & BioExcel

s.farr@epcc.ed.ac.uk

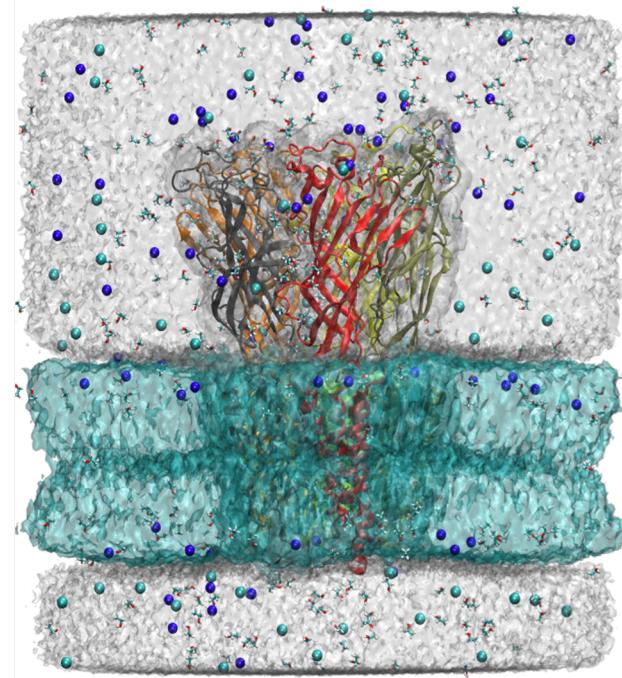
Molecular simulation

Aim:

Generate enough representative confirmations of the molecular system in such a way that accurate values of a property can be obtained.

One Method:

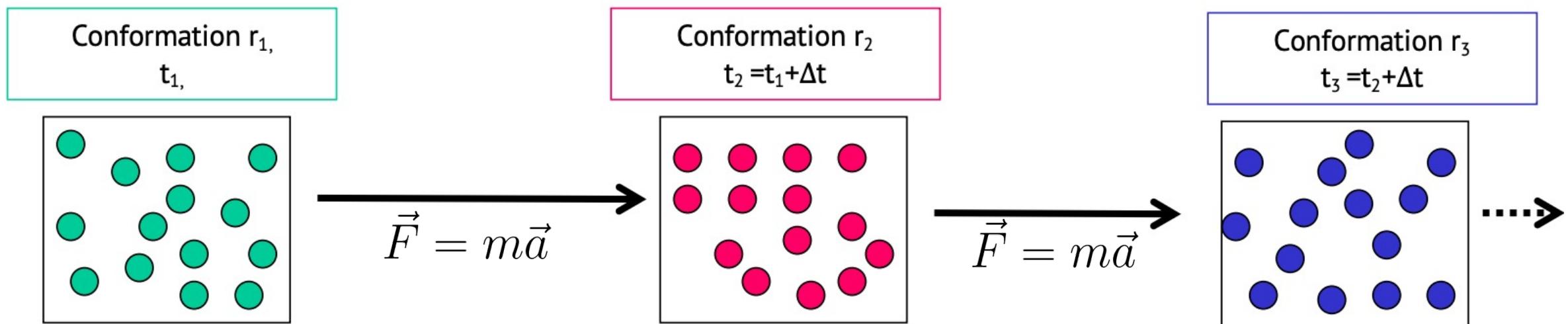
Molecular Dynamics



- How fast do ions pass through channels?
- What happens when protein residues are altered?

Molecular Dynamics

Generates configurations (a trajectory) by applying Newton's equations of motion for atoms of a molecular system.



Molecular Dynamics

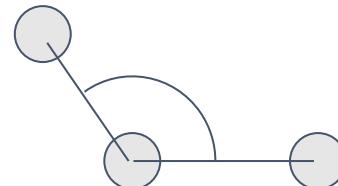
Classical Physics at Atomic resolution

- Atoms are point masses

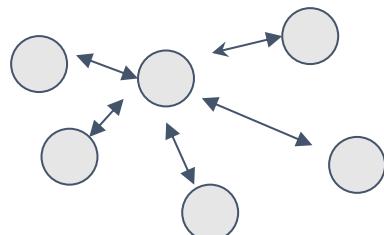


- Bonds are springs

- 3-body angle interactions



- Pairwise interactions for Van der Waals and Electrostatics



Potential energy function
(Force field)

$$\vec{F} = -\nabla U$$

$$U(R) = \sum_{bonds} k_r (r - r_{eq})^2 + \sum_{angles} k_\theta (\theta - \theta_{eq})^2 + \sum_{i < j}^{atoms} \epsilon_{ij} \left[\left(\frac{r_m}{r_{ij}} \right)^{12} - 2 \left(\frac{r_m}{r_{ij}} \right)^6 \right] + \sum_{i < j}^{atoms} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}$$

Molecular Dynamics

Equation of motion

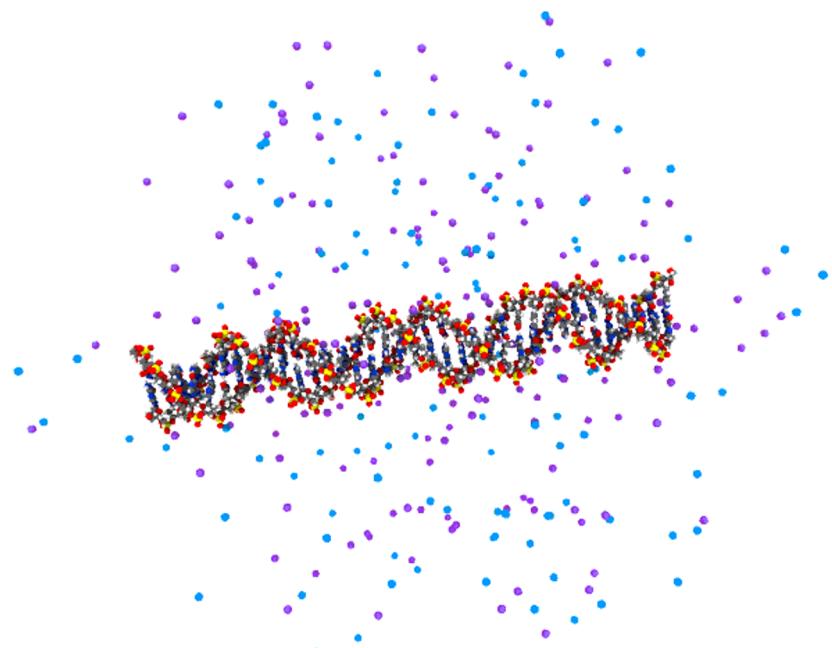
$$m \frac{d^2\vec{r}}{dt^2} = \vec{F} \quad \vec{F} = -\nabla U$$

Solve equation using explicit time stepping numerical method ,
typically **Velocity-Verlet or Leap-Frog**

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t) \Delta t + \frac{1}{2} \vec{a}(t) \Delta t^2,$$

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2} \Delta t.$$

Computationally expensive part is evaluating F at each timestep for all atoms,
 $O(N^2)$ without certain methods



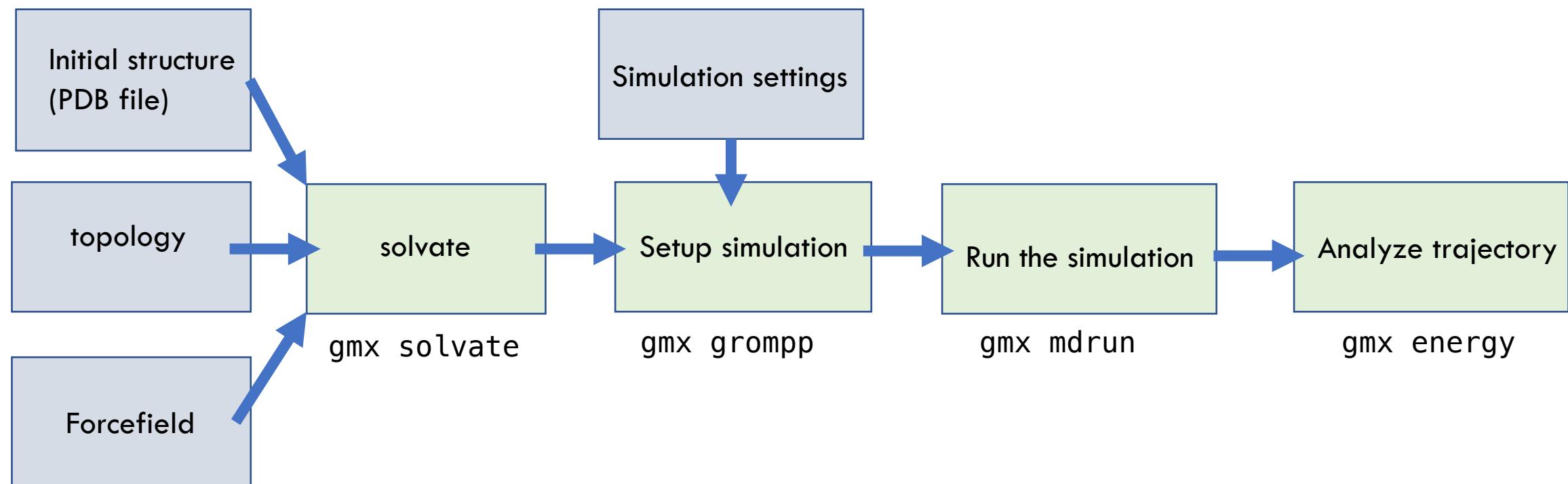
Typical simulation timescales

- Interesting biology at microsecond time-scale
- Simulation step at 2-4fs
- → need 10^{10} time steps
- Biomolecular processes are stochastic
- → average over simulation ensembles
- → for solid answers, need up to 10^{15} simulation steps
- Best performance is at 100 microseconds / step
- → typical simulations run days to months
- The more atoms the slower the simulation

GROMACS

GROMACS is a MD package which can perform most aspects of a MD workflow

Many command line tools ``gmx <tool>``



GROMACS

- In general all parts of the workflow and tools are system independent, can learn more from various places:

<https://manual.gromacs.org/documentation/current/user-guide/index.html>

<http://www.mdtutorials.com/gmx/>

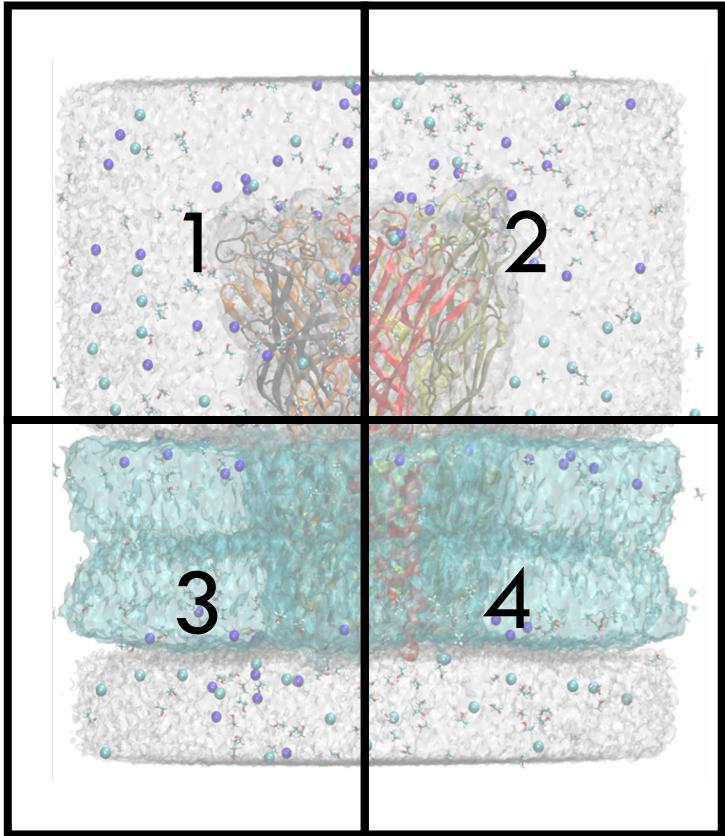
- mdrun is the part that uses the most compute time

GROMACS mdrun

Uses various methods to speed up the expensive calculations

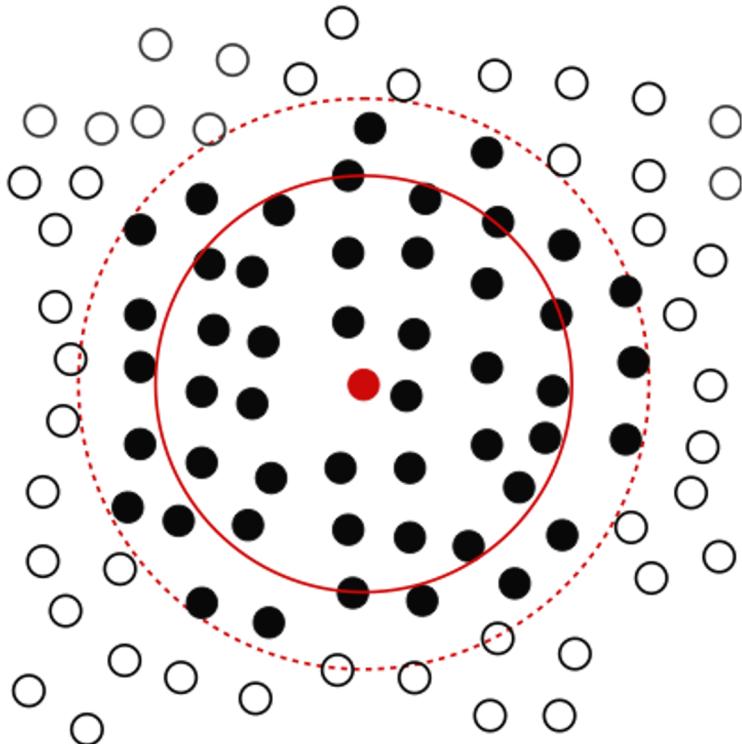
- SIMD (single instruction multiple data). Small C++ SIMD layer with highly optimised math functions for: SSE2/4, AVX2-128, AVX(2)-256, AVX-512, ARM, VSX, HPC-ACE
- Parallelisation: MPI, OpenMP, CUDA
- Domain decomposition (with dynamic load balancing)
- Neighbor lists for short range interactions $O(N^2) \rightarrow O(N)$
- Non-rectangular unit cells
- Particle Mesh Ewald (PME) for long range interactions $O(N^2) \rightarrow O(N \log N)$
- Constraints to enable increased timesteps (LINCS)

Domain Decomposition



- Make use of parallel computing
- Split simulation up into parts
- Simulation each part on its own CPU
- Dynamic load balancing, change the size of the regions based on how expensive they are
(regions with only water less expensive, can be larger)

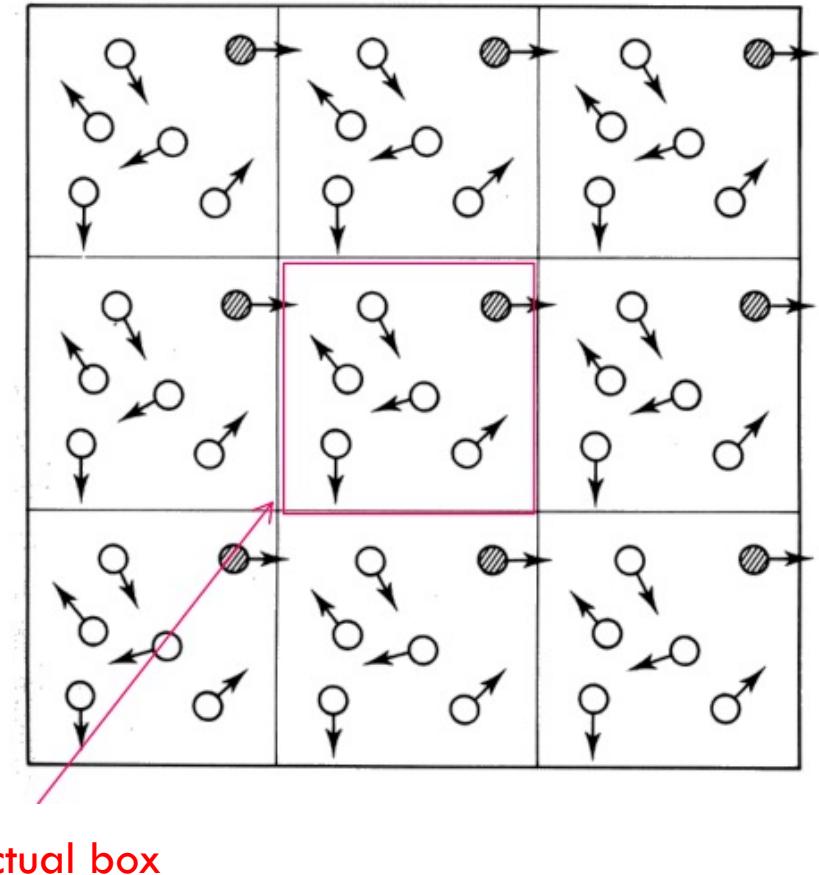
Neighbor lists



- Atoms don't move that far in one timestep
- Can keep track of close ones via a neighbor list,
- Only calculate short range interactions between atoms in each other lists
- Update the neighbor list less frequently

Boundary conditions

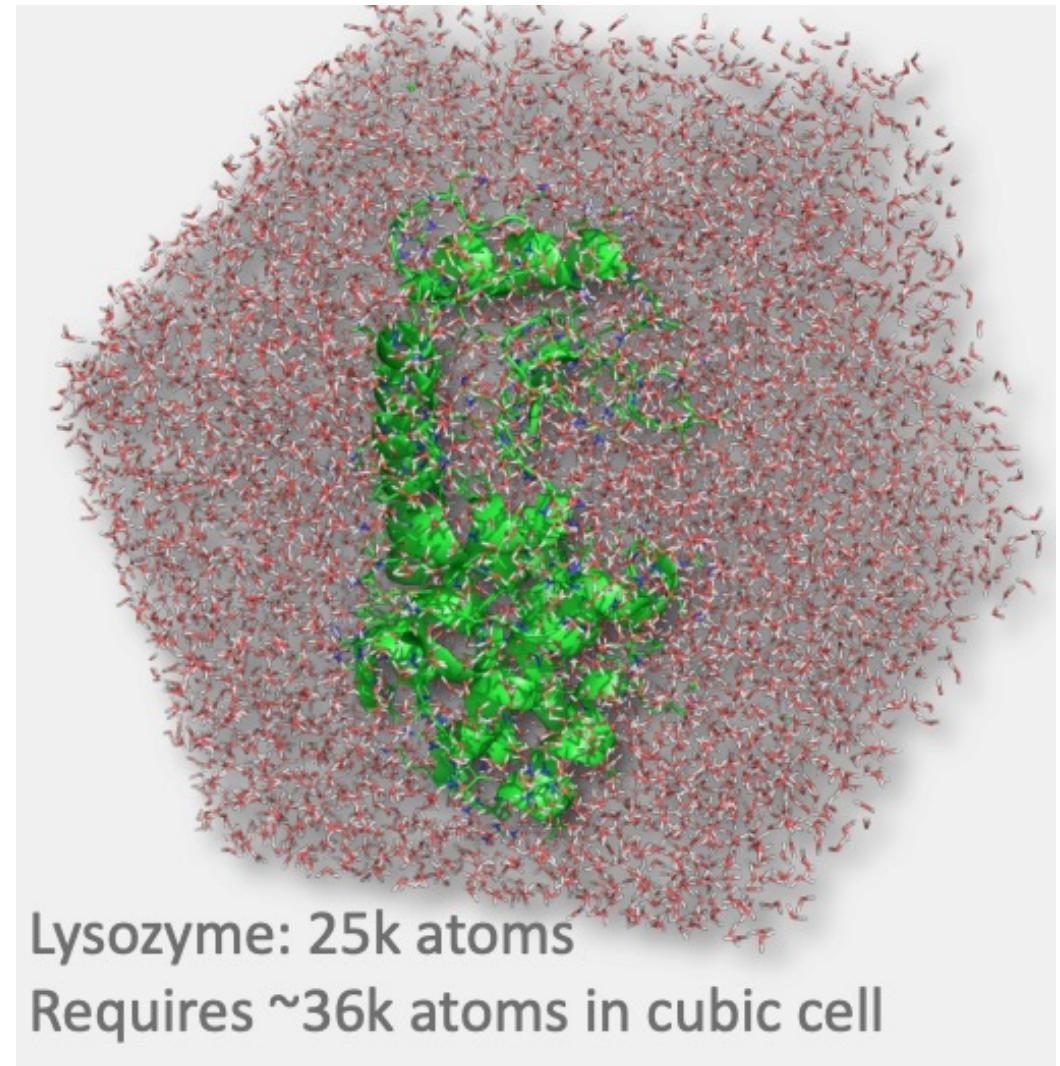
- Simulations are done with periodic boundary conditions
- Simulation box is replicated in all dimensions
- Particles leaving the central box are reintroduced on the opposite side
- Interactions are calculated with particles in neighboring cells
- No vacuum interface



Actual box

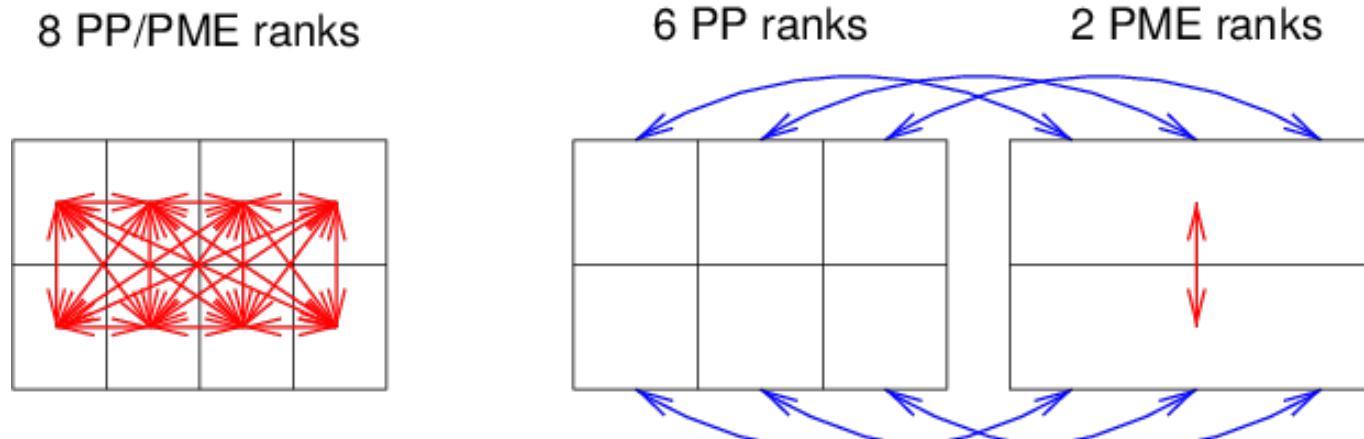
Non-rectangular unit cells

- Often you are simulating an approximately spherical macromolecule in solution
- If we use a box shape closer to a sphere less solvent molecules are needed to fill the box
- GROMACS uses triclinic unit cells than can represent hexagonal as well as standard rectangular boxes



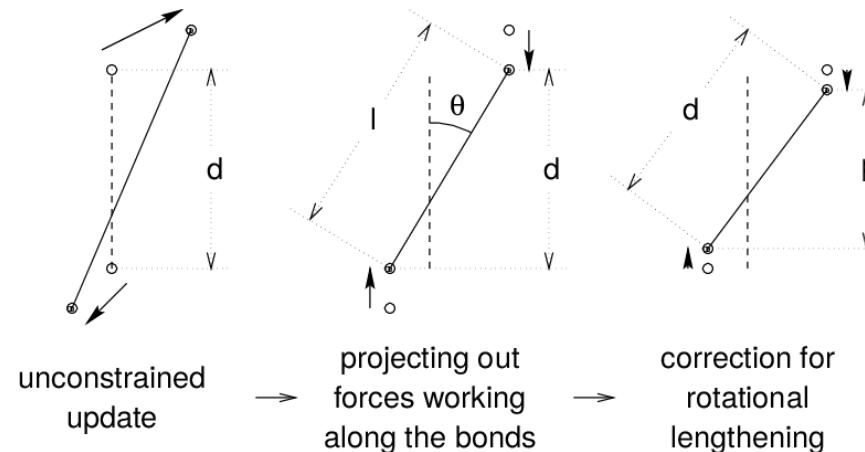
Particle Mesh Ewald (PME)

- Long range electrostatics must sum over all atoms, in all images, this is very slow
- Assign charges to a grid Mesh and do Ewald summation - this involves 3D Fourier transforms (FFT), scales as $N \log(N)$
- GROMACS uses Multiple-Program, Multiple-Data PME parallelisation
- PME ranks have to do global communication, Particle-Particle (PP) ranks only need to communicate with nearby ranks



Constraints

- The timestep is limited by the fastest motions: these are bond vibrations $\sim 1\text{fs}$
- These are often not important for the phenomenon being studied
- Can constrain them to stay at a fixed length
- This removes the fastest motion in the system, can now use a timestep of 2fs
- GROMACS uses the LINCS (LINear Constrain Solver) which can be done in parallel (P-LINCS) for bonds that cross domain decomposition cells.

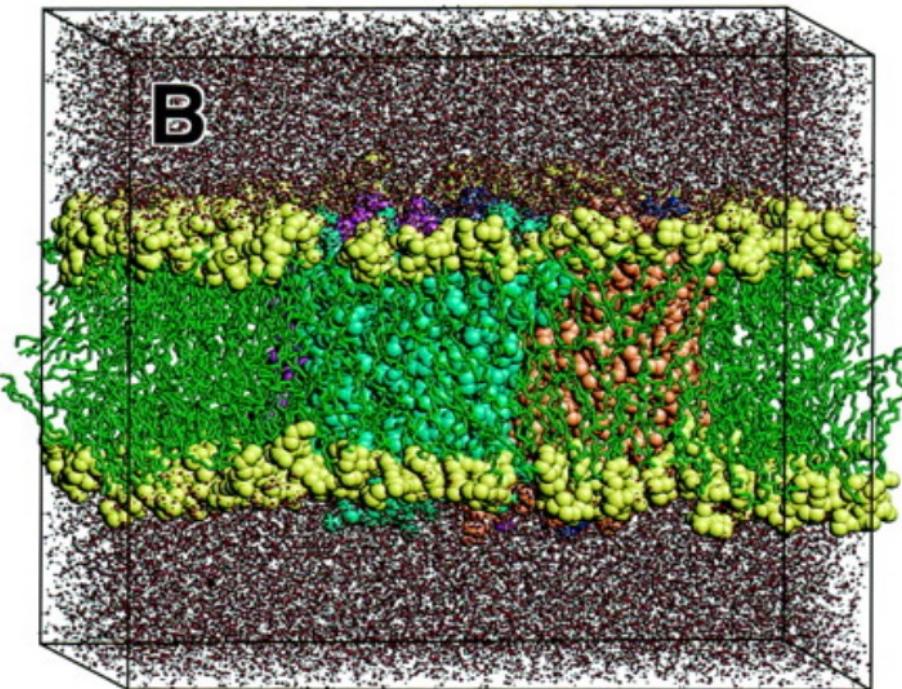


Hands-on: Running GROMACS

- We will use a standard GROMACS benchmark simulation, benchMEM.tpr, from here: <https://www.mpinat.mpg.de/grubmueller/bench>
- A .tpr file is a binary file which contains all information needed to run the simulation
- You can learn how to prepare them elsewhere:
<https://manual.gromacs.org/documentation/current/user-guide/index.html>
<http://www.mdtutorials.com/gmx/>
- We will look at benchmarking this system on Discoverer

benchMEM.tpr

- <https://www.mpinat.mpg.de/grubmueller/bench>
- Publication:
de Groot, B. L.; Grubmueller, H.: Water permeation across biological membranes: Mechanism and dynamics of Aquaporin-1 and GlpF. *Science* **294**, pp. 2353 - 2357 (2001)
<https://doi.org/10.1126/science.1066115>

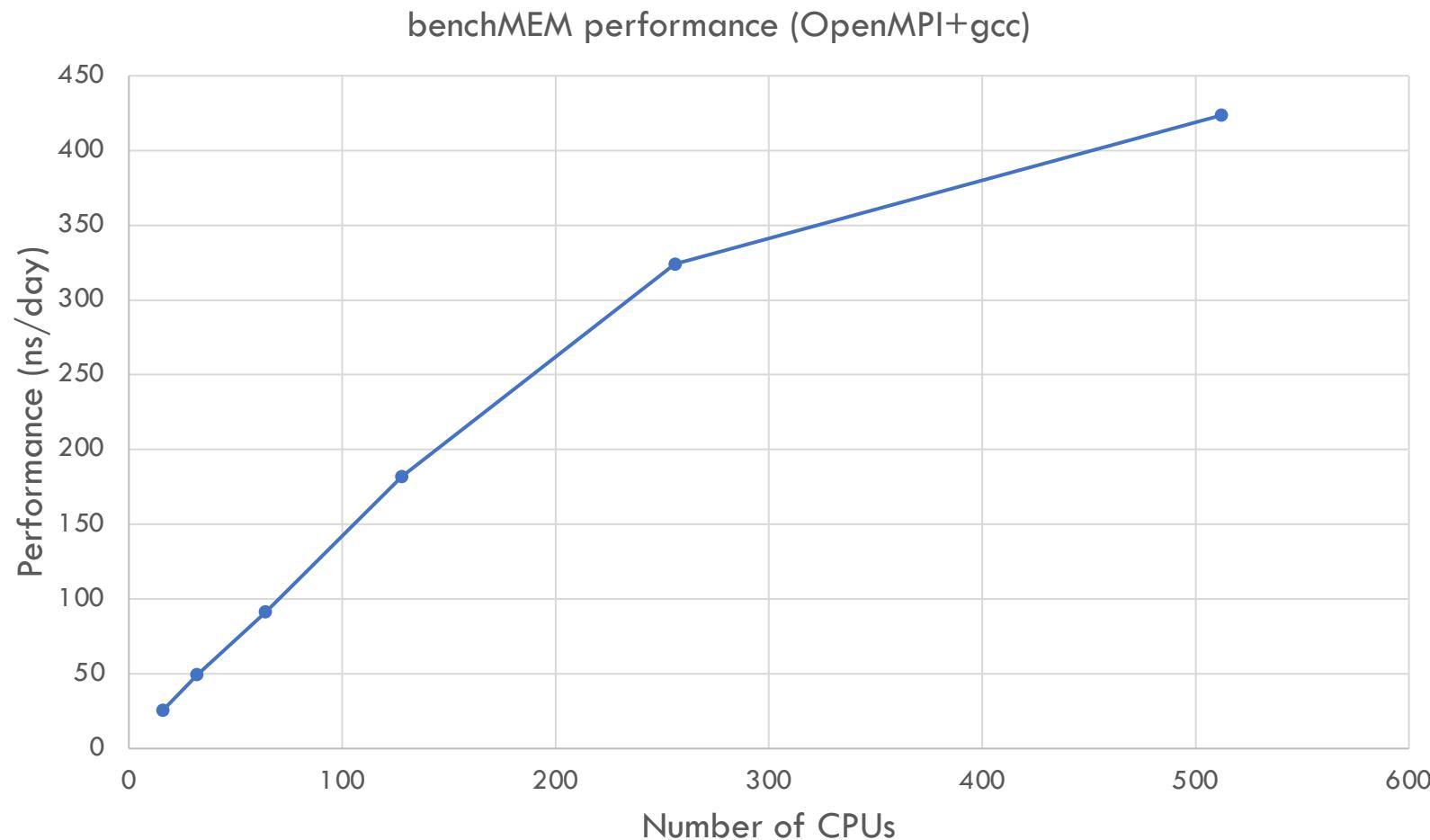


Hands-on: Running GROMACS

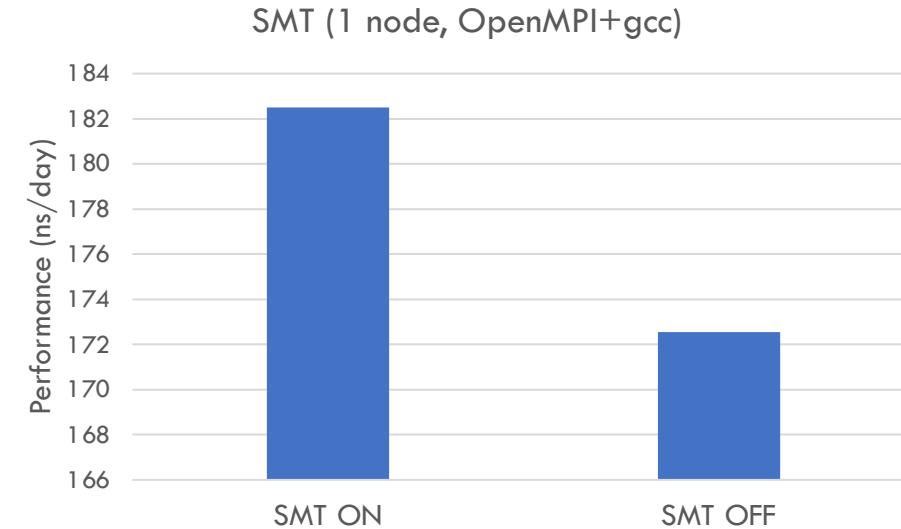
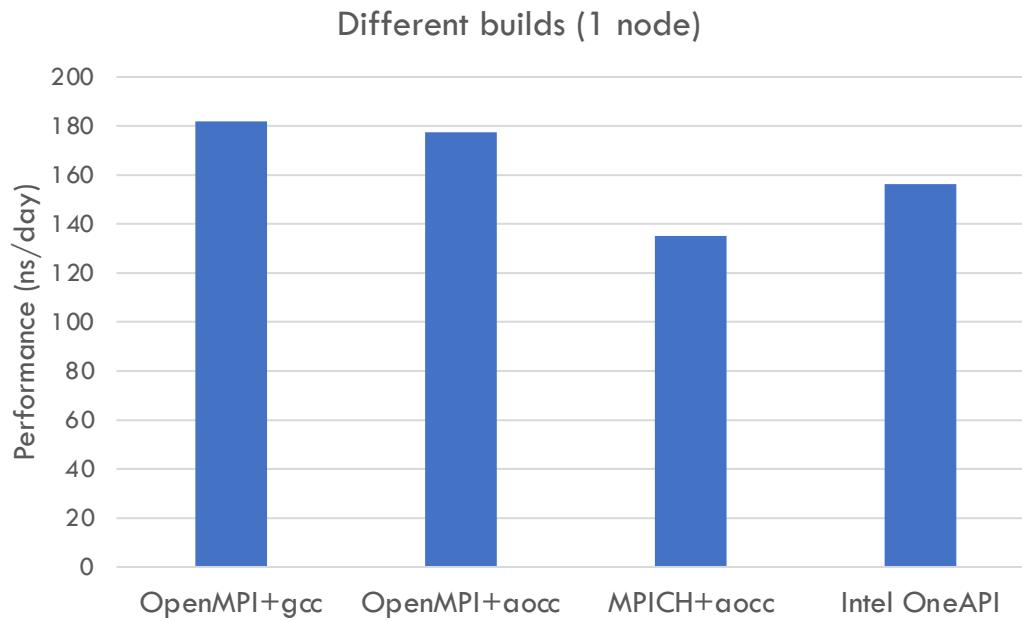
- All material for this can be found here:

https://github.com/sfarr-epcc/gromacs_on_discoverer

benchMEM performance results



benchMEM performance results



OpenMPI+gcc: module load gromacs/2021/2021.4-intel-nogpu-openmpi-gcc

OpenMPI+aocc: module load gromacs/2021/2021.4-intel-nogpu-openmpi-aocc

MPICH+aocc: module load gromacs/2021/2021.4-intel-nogpu-mpi

Intel OneAPI: module load gromacs/2021/2021.4-oneapi-nogpu-mpi

Hands-on: Building GROMACS from source

- All material for this can be found here:

https://github.com/sfarr-epcc/gromacs_on_discoverer

- General information that is applicable to all systems can be found here:

<https://manual.gromacs.org/documentation/current/install-guide/index.html>



BioExcel Partners



Ian Harrow Consulting

NORMAN
CONSULTING

Utrecht University



acrosslimits

EMBL-EBI



MAX-PLANCK-GESELLSCHAFT



MANCHESTER
1824



UNIVERSITY OF EDINBURGH



UNIVERSITY OF JYVÄSKYLA



JÜLICH
FORSCHUNGSZENTRUM
NBD
NOSTRUM BIODISCOVERY



IRB
BARCELONA

INSTITUTO DE INVESTIGACIÓN BIOMÉDICA



Horizon 2020
European Union Funding
for Research & Innovation

BioExcel is funded by the European Union
Horizon 2020 program under grant
agreements 675728 and 823830.