



Welcome!

Virtual tutorial starts at 15:00 GMT

Please leave feedback afterwards at:
www.archer.ac.uk/training/feedback/online-course-feedback.php



Multi-resolution modelling of biological systems in LAMMPS

ARCHER Virtual Tutorial, 19th Oct 2016

Iain Bethune ibethune@epcc.ed.ac.uk

Oliver Henrich ohenrich@epcc.ed.ac.uk

EPSRC

NERC SCIENCE OF THE ENVIRONMENT



CRAY
THE SUPERCOMPUTER COMPANY

epcc



Reusing this material



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

http://creativecommons.org/licenses/by-nc-sa/4.0/deed.en_US

This means you are free to copy and redistribute the material and adapt and build on the material under the following terms: You must give appropriate credit, provide a link to the license and indicate if changes were made. If you adapt or build on the material you must distribute your work under the same license as the original.

Note that this presentation contains images owned by others. Please seek their permission before reusing these images.



Outline

- ARCHER eCSE programme
- Implementation of Dual Resolution Simulation Methodology in LAMMPS
 - ELBA force-field
 - Implementation in LAMMPS
 - Performance testing
 - Summary



ARCHER eCSE programme

- Funding for the ARCHER user community to develop software
 - Implementation of algorithmic improvements within an existing code
 - Improving the scalability of software on higher core counts
 - Improvements to code which allows new science to be carried out
 - Porting and optimising a code to run efficiently on ARCHER
 - Adding new functionalities to existing codes
 - Code development to take a code from a Tier-2 (Regional) or local university cluster to Tier-1 (National) level bringing New Communities onto ARCHER
- Projects typically 3 months – 1 year
- Next call closes 31st Jan 2017



ARCHER eCSE programme

- More information on the ARCHER website:
 - <https://www.archer.ac.uk/community/eCSE/>
 - Project Reports
 - How To Apply
 - List of funded projects
- Webinar from last month:
 - <https://youtu.be/WRGsNKWrNIc>



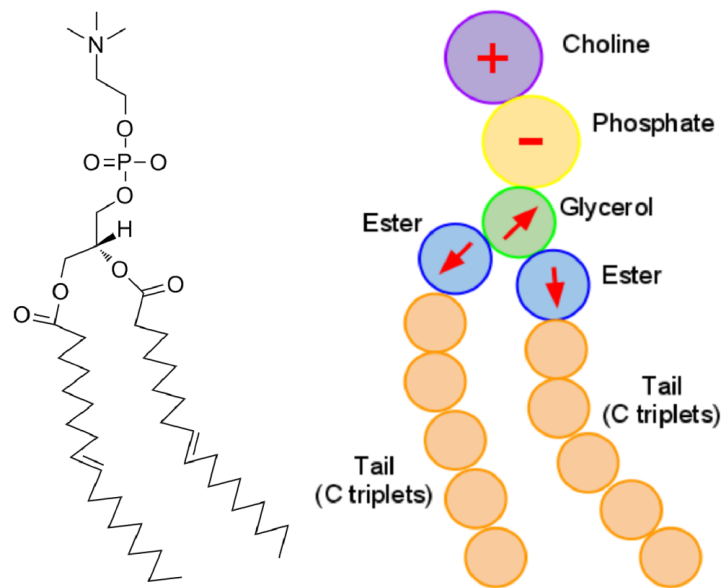
Implementation of Dual Resolution Simulation Methodology in LAMMPS

- eCSE04-7 (January 2015)
- **PI:** Prof. Jonathan Essex, Southampton
- **6 person-months funded:** August 2015 – August 2016
- **Objective:** enable fast and reliable calculations with the ELBA force-field in LAMMPS
 - New integrators
 - Parallel load balancing



ELBA Force-field

- **ELBA** = **EL**ectrostatics-**BA**sed coarse grained forcefield
 - [Orsi & Essex, PLoS ONE 6\(12\) 2011](#)
- Originally for studying lipids
 - Also applied to other biomolecules
- Explicit solvent
 - One dipolar bead per water molecule
- Allows for atomistic detail e.g.
 - Using CHARMM parameters

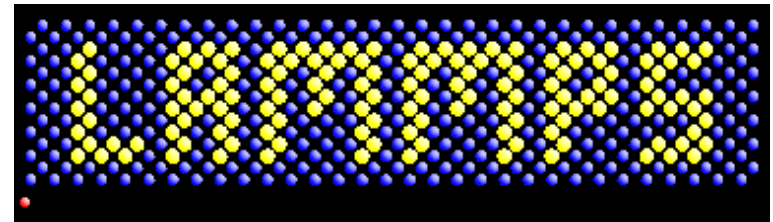


138 atoms -> 15 CG beads



ELBA Force-field

- Implemented in **BRAHMS-MD** (Biomembrane Reduced-Approach Multiresolution Simulator for Molecular Dynamics):
 - <https://code.google.com/archive/p/brahms-md/>
 - Limited user base, single developer -> not sustainable
 - No parallelisation -> small systems
- Why LAMMPS?
 - Main interaction types already implemented <http://lammps.sandia.gov>
 - Support for spherical particles
 - r-RESPA multiple timestepping
 - Flexible, scalable, large user base



Implementation in LAMMPS

- LAMMPS `fix nve/sphere` integrator does not conserve energy well
 - Better scheme to integrate rotational d.o.f. - DLM
[Dullweber, Leimkuhler and McLachlan, JCP 107\(15\) 1997](#)
1. Construct rotation matrix Q from dipole (taken as the body-fixed z-axis)
 2. In body-space, apply rotations around each local axis:

$$\omega_b = Q\omega_s$$
$$R_1 = R_x\left(\frac{\delta t}{2}\omega_1\right), \quad \omega = R_1\omega, \quad Q = R_1^T Q$$
$$R_2 = R_y\left(\frac{\delta t}{2}\omega_2\right), \quad \omega = R_2\omega, \quad Q = R_2^T Q$$
$$R_3 = R_z(\delta t\omega_3), \quad \omega = R_3\omega, \quad Q = R_3^T Q$$

$$R_4 = R_y\left(\frac{\delta t}{2}\omega_2\right), \quad \omega = R_4\omega, \quad Q = R_4^T Q$$
$$R_5 = R_x\left(\frac{\delta t}{2}\omega_1\right), \quad \omega = R_5\omega, \quad Q = R_5^T Q$$

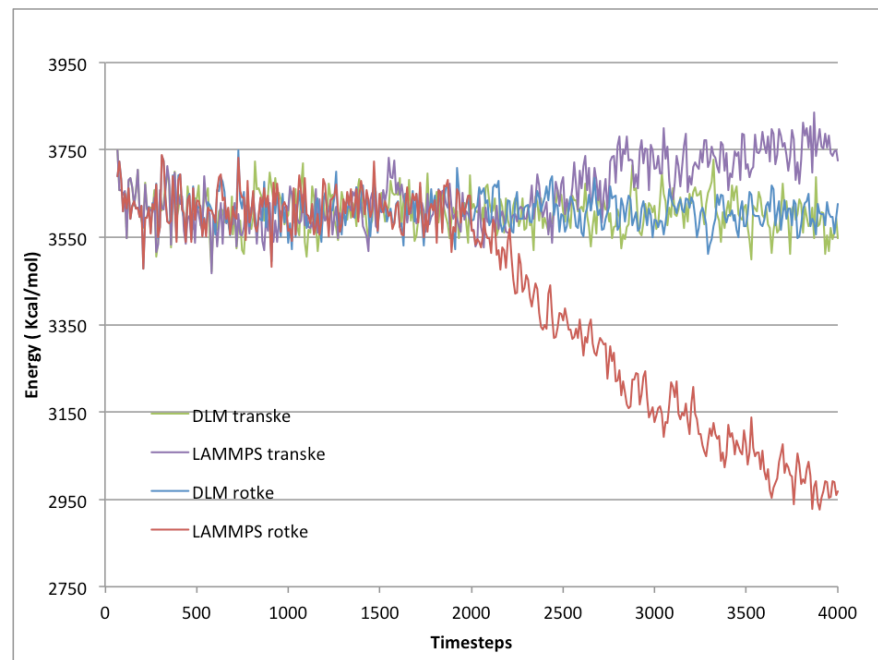
3. Finally, compute the new dipole: $\mu_s = Q^T[001] \cdot \|\mu\|$



Implementation in LAMMPS

- 4000 ELBA water beads, 10fs timestep, 20ps NVT, 20ps NVE

```
fix thermostat all langevin 303 303 200 48279 omega yes
```



Implementation in LAMMPS

- 128 DPMC molecules in water, 75ps NVT, 100ps NVE

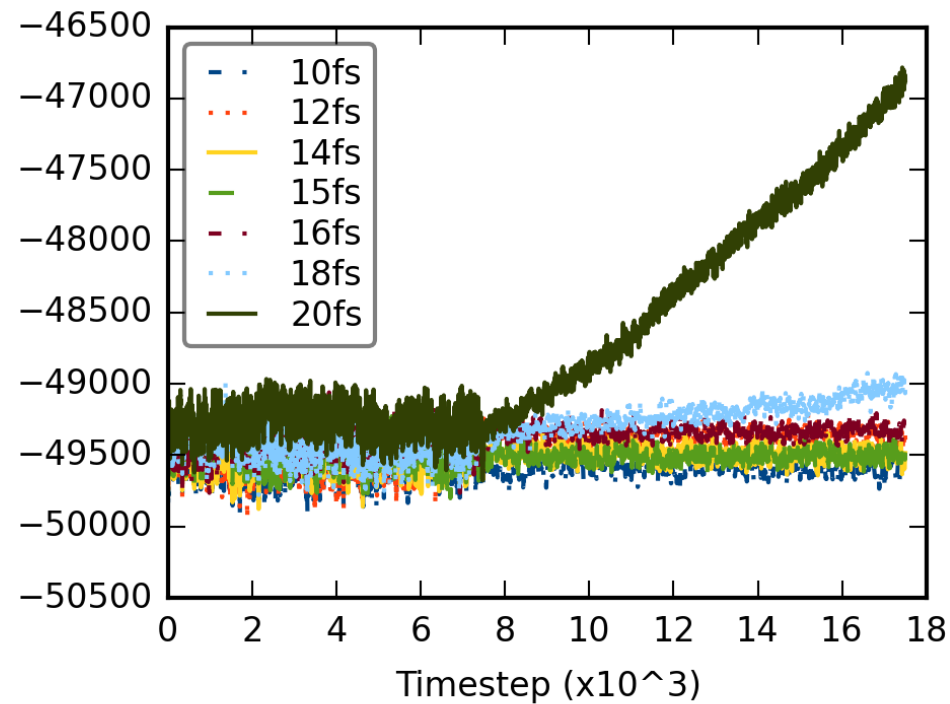


Image from Sam Genheden



Implementation in LAMMPS

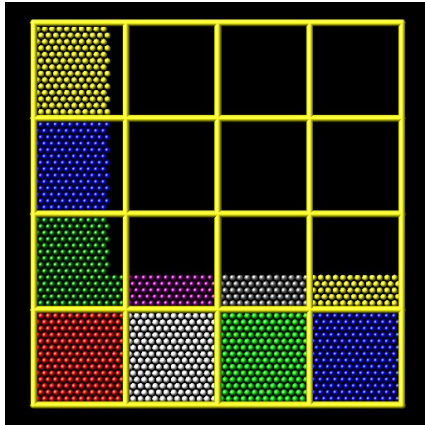
- DLM integrator enabled by an optional argument:
 - `fix nve/sphere ... update dipole/dlm`
- Also for other ensembles:
 - Constant temperature / NVT (Nosé-Hoover)
 - `fix nvt/sphere ... update dipole/dlm`
 - Isothermal-isobaric (Nosé-Hoover / Parrinello-Rahman)
 - `fix npt/sphere ... update dipole/dlm`
 - Isenthalpic (Parrinello-Rahman)
 - `fix nph/sphere ... update dipole/dlm`



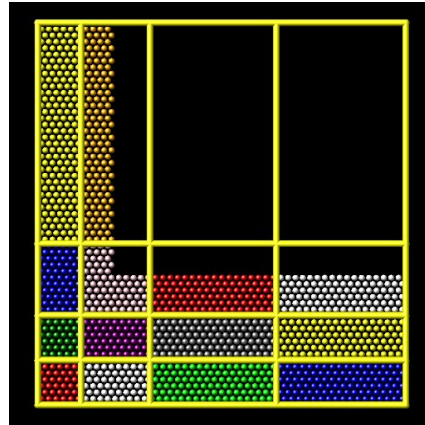
Implementation in LAMMPS

<http://lammps.sandia.gov/doc/balance.html>

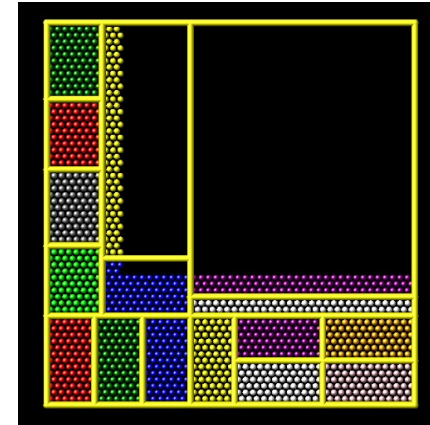
- Load balancing schemes:



None



Shift



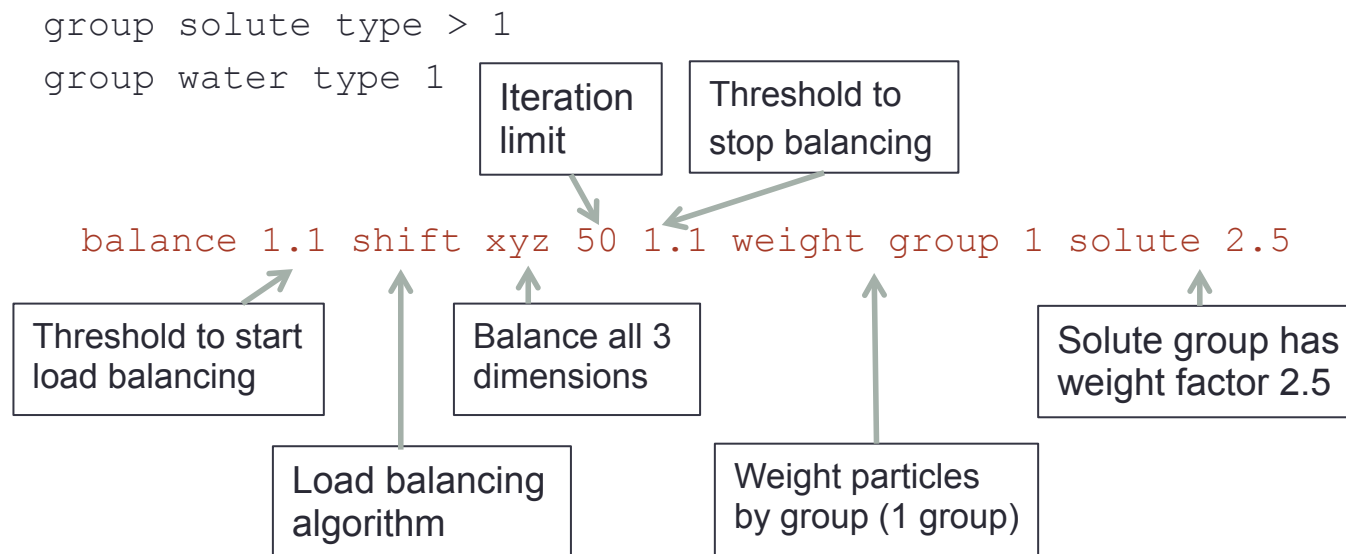
RCB

- Problem for dual-resolution simulations!
 - 90% of computational cost is force evaluation
 - Not all particles are the same
 - r-RESPA – some forces are computed more frequently than others



Implementation in LAMMPS

- New load balancing metrics:
 - Weighting by particle groups
 - Uses LAMMPS existing `group` command e.g.



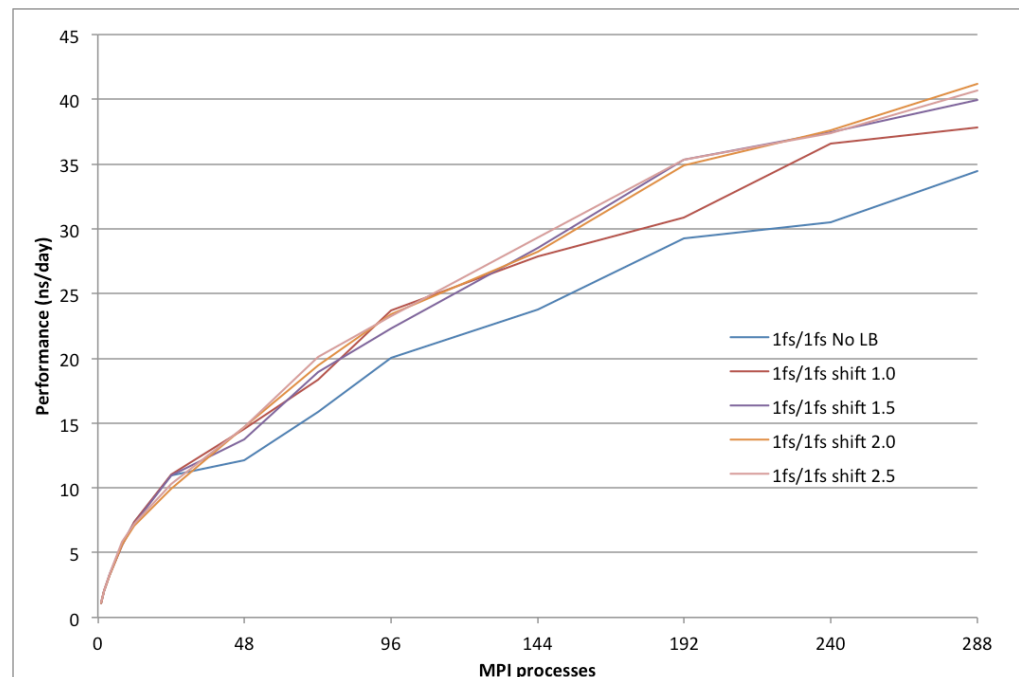
Implementation in LAMMPS

- New load balancing metrics (subsequently added by LAMMPS developers):
 - Weighting by number of neighbors – `weight neigh`
 - Weighting by compute time – `weight time`
 - Doesn't account for the different particles types contributing to different parts of the computation (pair, bond, kspace, neigh)
 - Weighting by arbitrary user-defined variables – `weight var`



Performance testing

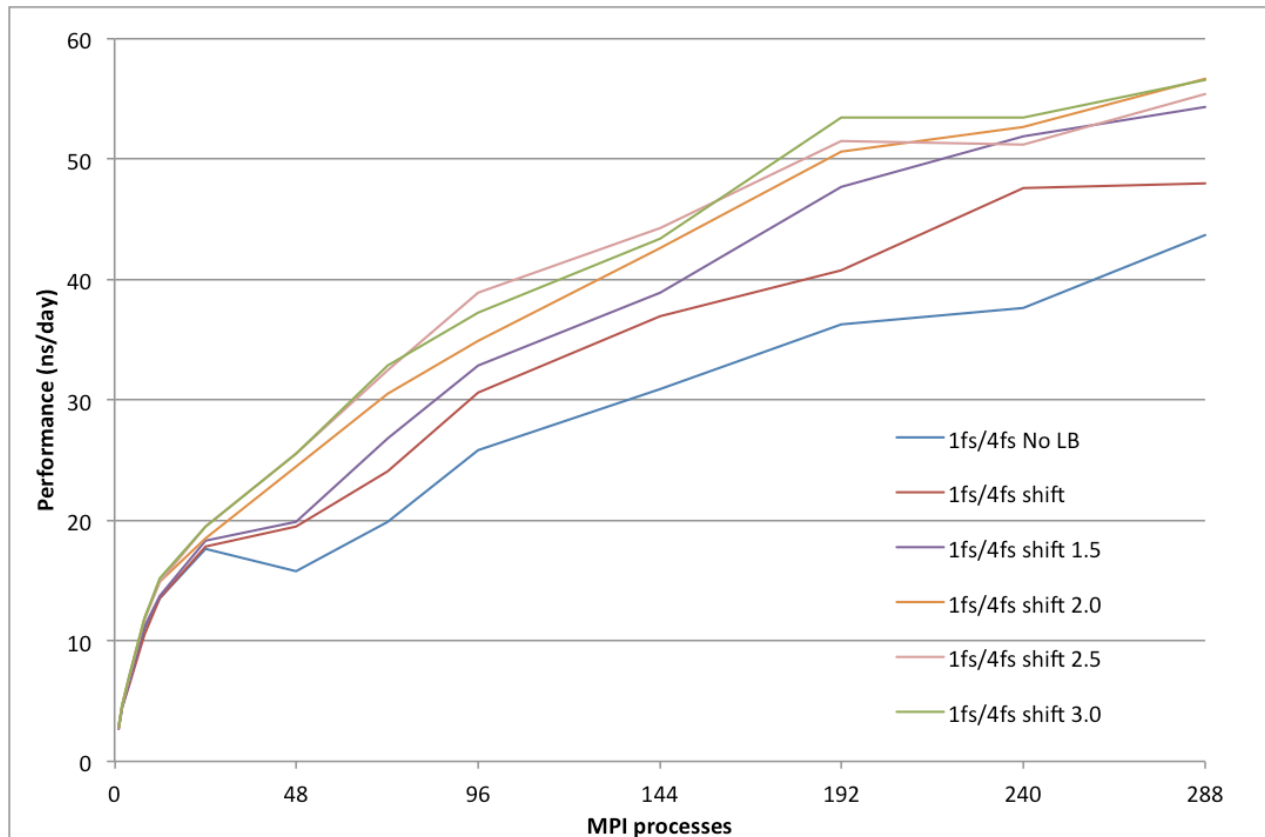
- Bovine Pancreatic Trypsin Inhibitor (BPTI) dual-resolution model:
 - 882 atoms, CHARMM force-field
 - 6136 water molecules, ELBA beads



- No r-RESPA
- 1fs timestep
- Up to **10%** speedup over non-weighted balance



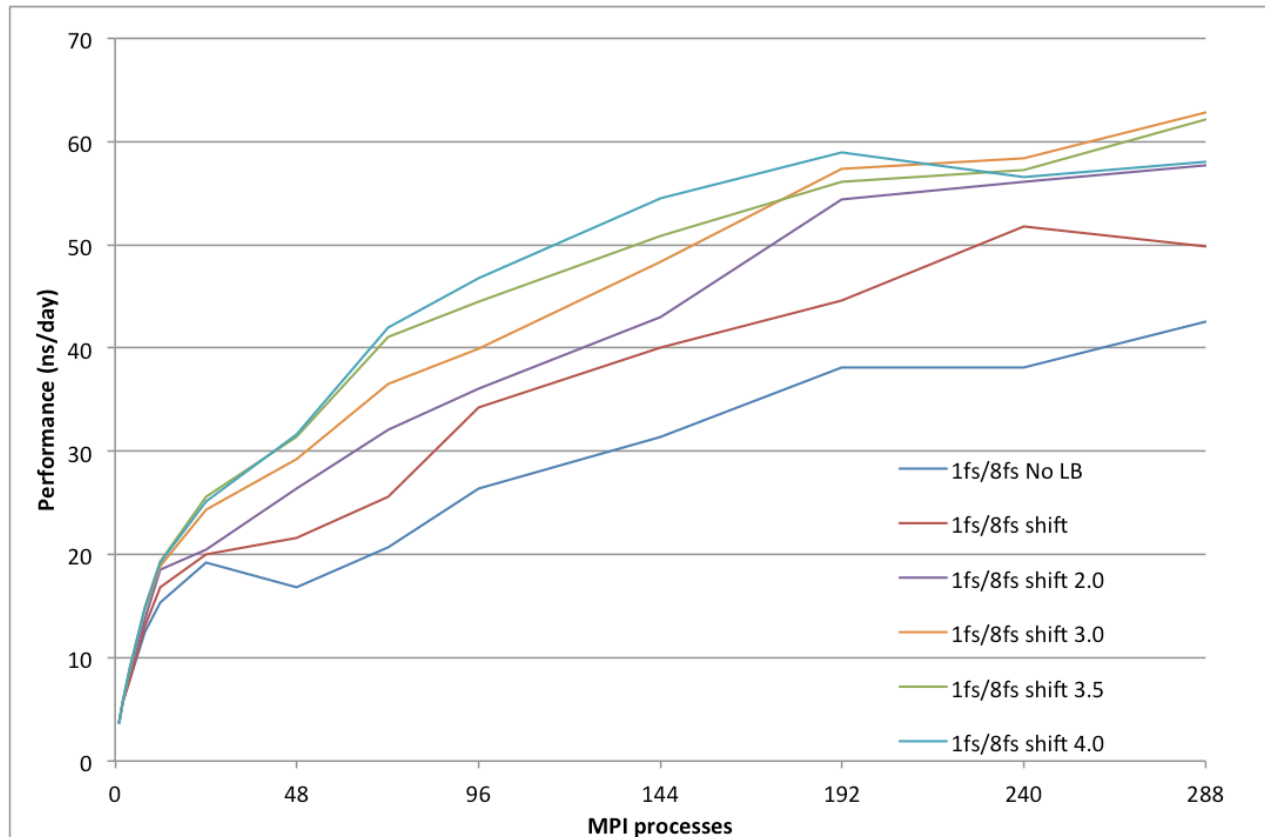
Performance testing



- 1:4 r-RESPA ratio
- Water pair forces + dihedral forces computed every fourth step
- Larger weightings better (2.5-3.0)
- Up to 36% speedup over non-weighted balance



Performance testing



- 1:8 r-RESPA ratio
- Larger weightings better (3.0-4.0)
- Up to 65% speedup over non-weighted balance



Summary

- DLM integrator for NVE/NVT/NPT/NPH dynamics
 - Stable for water up to 16fs timestep
 - Included in LAMMPS stable release **30 Jul 2016**
- New load balancing metrics
 - Better performance for r-REPSA and hybrid pair force
 - Include in LAMMPS patch release **27 Sep 2016**
- More information
 - Technical Report:
<http://www.archer.ac.uk/documentation/white-papers/lammps-elba/lammps-ecse.pdf>
 - Tutorials, references, discussion: <https://sgenheden.github.io/Elba/>

```
Installed on ARCHER  
module load  
lammps/elba
```





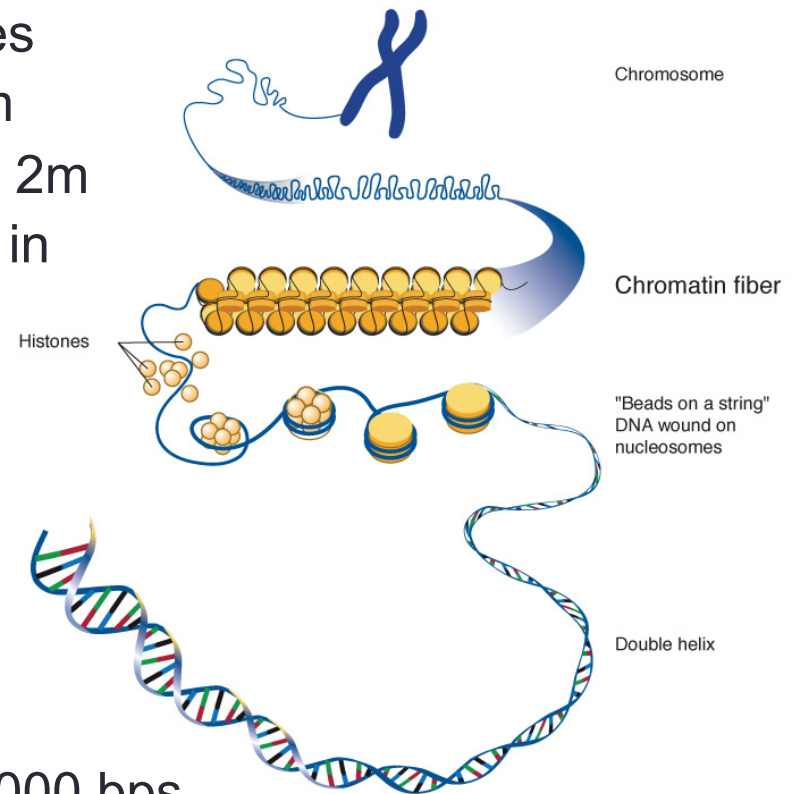
ARCHER eCSE05-10 Project

- **Adding Multiscale Models of DNA to LAMMPS** (09/2015 – 08/2016)
- Dr Oliver Henrich (PI, UoE), Prof Davide Marenduzzo (Co-I, UoE), Dr Thomas Ouldridge (Co-I, Imperial College London)
- Overview:
 - Implemented **oxDNA coarse-grained DNA model** for single- and double- stranded DNA into LAMMPS code
 - Implemented new **Langevin-type rigid-body integrators**
 - Software available from **CCPForge** (<https://ccpforge.cse.rl.ac.uk/gf/project/cgdna>) soon as **LAMMPS USER-package**
 - Currently adapting utility software of oxDNA standalone version



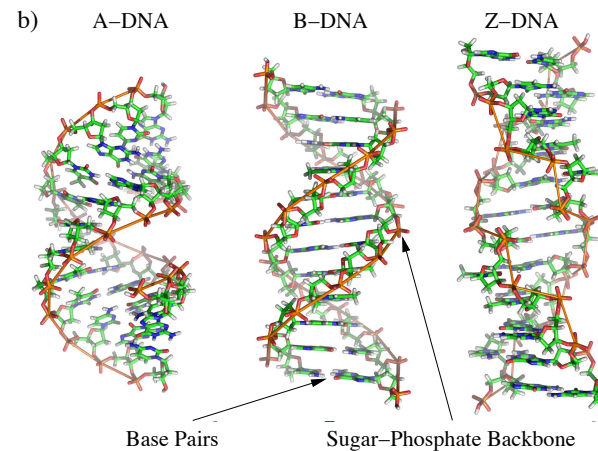
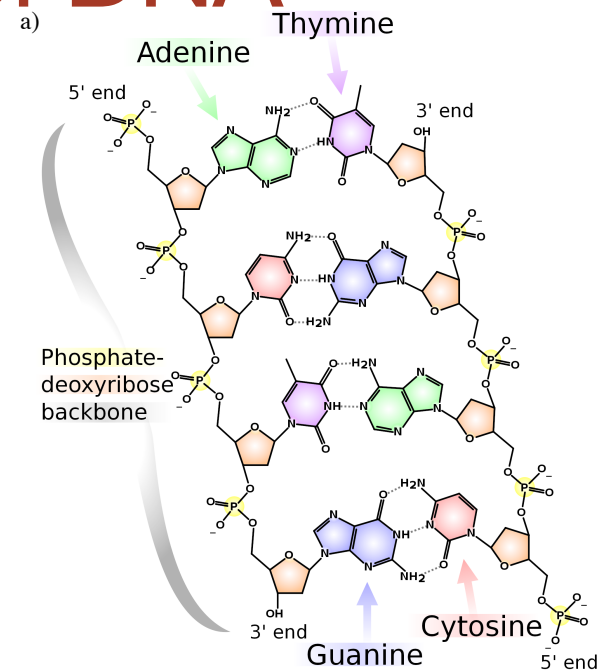
From DNA to Chromosomes

- Haploid human genome contains 3.2 billion base pairs organised in 23 chromosomes
 - Diameter of DNA strand = $2 \times 10^{-9}\text{m}$
 - Total length of DNA in human cell = 2m
 - Diameter of spherical 'blob' of DNA in human cell = $2 \times 10^{-6}\text{m}$
- Hierarchical organisation
 - Histone octamer
 - Nucleosome core particles 200 bps
 - 10 nm beads-on-a-string
 - 30 nm chromatin fibre
 - smallest loop in chromatin fibre 50,000 bps



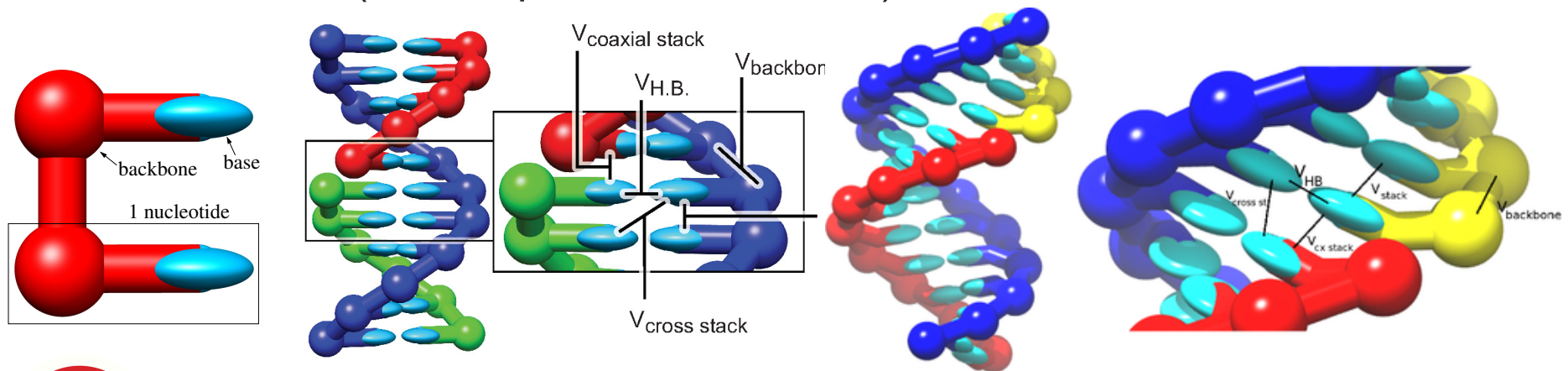
Atomistic Simulation of DNA

- Good for capturing fast conformational fluctuations and protein-DNA binding
- Usually limited to a few 1000 base pairs
- Phenomena on large time and length scales, e.g. DNA supercoiling or nucleosome positioning are permanently out of reach



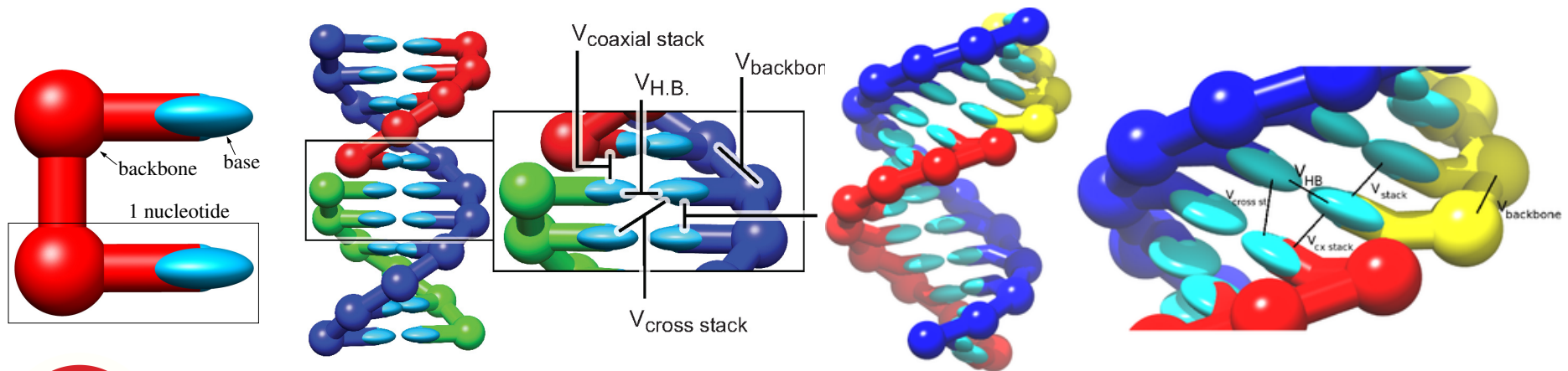
Coarse-grained simulation with oxDNA

- Must deliver correct longitudinal and torsional persistence length, electrostatics, if sequence-specific correct melting temperature, elastic constants ...
- oxDNA: top-down approach of a CG model, nucleotides modelled as rigid bodies (DOF are COM & quaternion)
- Parameterize interaction between nucleotides with 6 independent interactions (7 for implicit electrostatics)



oxDNA Force Field

- **Backbone:** FENE (finite extensible nonlinear elastic) springs
- **Excluded volume:** Lennard-Jones potential
- **Stacking:** harmonic angle \times Morse potential
- **Cross-stacking:** harmonic angle \times harmonic distance potential
- **Coaxial stacking:** harmonic angle \times harmonic distance potential
- **Hydrogen bonding:** harmonic angle \times Morse potential



oxDNA Force Field

- Smoothed, truncated and modulated forms of the above
- 1 bonded interaction (backbone), 5 pair interactions (excluded volume, stacking, hydrogen bonding, cross-stacking, coaxial stacking)

- FENE spring (used to connect backbones):

$$V_{\text{FENE}}(r, \epsilon, r^0, \Delta) = -\frac{\epsilon}{2} \ln \left(1 - \frac{(r - r^0)^2}{\Delta^2} \right)$$

- Morse potential (used for stacking and H-bonding):

$$V_{\text{Morse}}(r, \epsilon, r^0, a) = \epsilon (1 - \exp(-(r - r^0)/a))^2$$

- Harmonic potential (used for cross-stacking and coaxial stacking):

$$V_{\text{Harm}}(r, k, r^0) = \frac{k}{2} (r - r^0)^2$$

- Lennard - Jones potential (used for soft repulsion):

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

- Quadratic terms (used for modulation):

$$V_{\text{mod}}(\theta, a, \theta^0) = 1 - a(\theta - \theta^0)^2$$

- Quadratic smoothing terms for truncation:

$$V_{\text{smooth}}(x, b, x^0) = b(x^0 - x)^2$$

- The radial part of the stacking and hydrogen-bonding potentials:

$$f_1(r) = \begin{cases} V_{\text{Morse}}(r, \epsilon, r^0, a) - V_{\text{Morse}}(r^*, \epsilon, r^0, a) & \text{if } r^{\text{low}} < r < r^{\text{high}}, \\ eV_{\text{smooth}}(r, b^{\text{low}}, r^{\text{low}}) & \text{if } r^{\text{low}} < r < r^{\text{low}}, \\ eV_{\text{smooth}}(r, b^{\text{high}}, r^{\text{high}}) & \text{if } r^{\text{high}} < r < r^{\text{high}}, \\ 0 & \text{otherwise.} \end{cases}$$

- The radial part of the cross-stacking and coaxial stacking potentials:

$$f_2(r) = \begin{cases} V_{\text{Harm}}(r, k, r^0) - V_{\text{Harm}}(r^*, k, r^0) & \text{if } r^{\text{low}} < r < r^{\text{high}}, \\ kV_{\text{smooth}}(r, b^{\text{low}}, r^{\text{low}}) & \text{if } r^{\text{low}} < r < r^{\text{low}}, \\ kV_{\text{smooth}}(r, b^{\text{high}}, r^{\text{high}}) & \text{if } r^{\text{high}} < r < r^{\text{high}}, \\ 0 & \text{otherwise.} \end{cases}$$

- The radial part of the excluded volume potential:

$$f_3(r) = \begin{cases} V_{\text{LJ}}(r, \epsilon, \sigma) & \text{if } r < r^*, \\ eV_{\text{smooth}}(r, b, r^*) & \text{if } r^* < r < r^*, \\ 0 & \text{otherwise.} \end{cases}$$

- The angular modulation factor used in stacking, hydrogen-bonding, cross-stacking and coaxial stacking:

$$f_4(\theta) = \begin{cases} V_{\text{mod}}(\theta, a, \theta^0) & \text{if } \theta^0 - \Delta\theta^* < \theta < \theta^0 + \Delta\theta^*, \\ V_{\text{smooth}}(\theta, b, \theta^0 - \Delta\theta^*) & \text{if } \theta^0 - \Delta\theta^* < \theta < \theta^0 - \Delta\theta^*, \\ V_{\text{smooth}}(\theta, b, \theta^0 + \Delta\theta^*) & \text{if } \theta^0 + \Delta\theta^* < \theta < \theta^0 + \Delta\theta^*, \\ 0 & \text{otherwise.} \end{cases}$$

- Another modulating term which is used to impose right-handedness (effectively a one-sided modulation):

$$f_5(\phi) = \begin{cases} 1 & \text{if } x > 0, \\ V_{\text{mod}}(x, a, 0) & \text{if } x^* < x < 0, \\ V_{\text{smooth}}(x, b, x^*) & \text{if } x^* < x < x^*, \\ 0 & \text{otherwise.} \end{cases}$$

$$V_{\text{backbone}} = V_{\text{FENE}}(\delta r_{\text{backbone}}, \epsilon_{\text{backbone}}, \delta r_{\text{backbone}}^0, \Delta_{\text{backbone}}).$$

$$V_{\text{exc}} = f_3(\delta r_{\text{backbone}}, \epsilon_{\text{exc}}, \sigma_{\text{backbone}}, \delta r_{\text{backbone}}^0) + f_3(\delta r_{\text{base}}, \epsilon_{\text{base}}, \sigma_{\text{base}}, \delta r_{\text{base}}^0) + f_3(\delta r_{\text{back-base}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^0) + f_3(\delta r_{\text{base-back}}, \epsilon_{\text{exc}}, \sigma_{\text{back-base}}, \delta r_{\text{back-base}}^0).$$

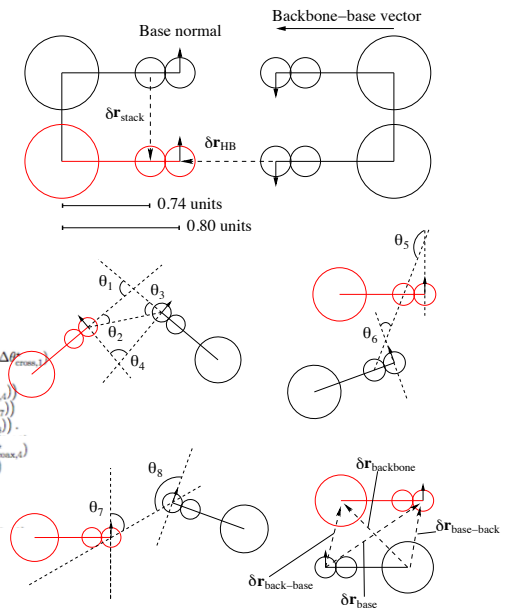
$$V_{\text{stack}} = f_2(\delta r_{\text{stack}}, \epsilon_{\text{stack}}, a_{\text{stack}}, \delta r_{\text{stack}}^0, \delta r_{\text{stack}}^{\text{low}}, \delta r_{\text{stack}}^{\text{high}}, \delta r_{\text{stack}}^0, \delta r_{\text{stack}}^{\text{high}}) \times f_1(\theta_{\text{stack}}, a_{\text{stack},1}, \theta_{\text{stack},1}^0, \Delta\theta_{\text{stack},1}^*) \times f_1(\theta_{\text{stack}}, a_{\text{stack},2}, \theta_{\text{stack},2}^0, \Delta\theta_{\text{stack},2}^*) \times f_5(-\cos(\phi_1), a_{\text{stack},1}, -\cos(\phi_1), a_{\text{stack},1}, -\cos(\phi_2), a_{\text{stack},2}, -\cos(\phi_2), a_{\text{stack},2}).$$

$$V_{\text{HB}} = f_1(\delta r_{\text{HB}}, \epsilon_{\text{HB}}, a_{\text{HB}}, \delta r_{\text{HB}}^0, \delta r_{\text{HB}}^{\text{low}}, \delta r_{\text{HB}}^{\text{high}}, \delta r_{\text{HB}}^0, \delta r_{\text{HB}}^{\text{high}}) \times f_1(\theta_1, a_{\text{HB},1}, \theta_{\text{HB},1}^0, \Delta\theta_{\text{HB},1}^*) \times f_1(\theta_2, a_{\text{HB},2}, \theta_{\text{HB},2}^0, \Delta\theta_{\text{HB},2}^*) \times f_1(\theta_3, a_{\text{HB},3}, \theta_{\text{HB},3}^0, \Delta\theta_{\text{HB},3}^*) \times f_1(\theta_4, a_{\text{HB},4}, \theta_{\text{HB},4}^0, \Delta\theta_{\text{HB},4}^*) \times f_1(\theta_5, a_{\text{HB},5}, \theta_{\text{HB},5}^0, \Delta\theta_{\text{HB},5}^*) \times f_1(\theta_6, a_{\text{HB},6}, \theta_{\text{HB},6}^0, \Delta\theta_{\text{HB},6}^*).$$

$$V_{\text{cross-stack}} = f_2(\delta r_{\text{cross-stack}}, \epsilon_{\text{cross-stack}}, \delta r_{\text{cross-stack}}^0, \delta r_{\text{cross-stack}}^{\text{low}}, \delta r_{\text{cross-stack}}^{\text{high}}, \delta r_{\text{cross-stack}}^0, \delta r_{\text{cross-stack}}^{\text{high}}) \times f_1(\theta_1, a_{\text{cross-stack},1}, \theta_{\text{cross-stack},1}^0, \Delta\theta_{\text{cross-stack},1}^*) \times f_1(\theta_2, a_{\text{cross-stack},2}, \theta_{\text{cross-stack},2}^0, \Delta\theta_{\text{cross-stack},2}^*) \times f_1(\theta_3, a_{\text{cross-stack},3}, \theta_{\text{cross-stack},3}^0, \Delta\theta_{\text{cross-stack},3}^*) \times f_1(\theta_4, a_{\text{cross-stack},4}, \theta_{\text{cross-stack},4}^0, \Delta\theta_{\text{cross-stack},4}^*) \times f_1(\theta_7, a_{\text{cross-stack},7}, \theta_{\text{cross-stack},7}^0, \Delta\theta_{\text{cross-stack},7}^*) \times f_1(\theta_8, a_{\text{cross-stack},8}, \theta_{\text{cross-stack},8}^0, \Delta\theta_{\text{cross-stack},8}^*) \times f_1(\pi - \theta_8, a_{\text{cross-stack},8}, \theta_{\text{cross-stack},8}^0, \Delta\theta_{\text{cross-stack},8}^*).$$

$$V_{\text{coax-stack}} = f_2(\delta r_{\text{coax-stack}}, \epsilon_{\text{coax-stack}}, \delta r_{\text{coax-stack}}^0, \delta r_{\text{coax-stack}}^{\text{low}}, \delta r_{\text{coax-stack}}^{\text{high}}, \delta r_{\text{coax-stack}}^0, \delta r_{\text{coax-stack}}^{\text{high}}) \times f_1(\theta_1, a_{\text{coax-stack},1}, \theta_{\text{coax-stack},1}^0, \Delta\theta_{\text{coax-stack},1}^*) \times f_1(\theta_2, a_{\text{coax-stack},2}, \theta_{\text{coax-stack},2}^0, \Delta\theta_{\text{coax-stack},2}^*) \times f_1(\theta_3, a_{\text{coax-stack},3}, \theta_{\text{coax-stack},3}^0, \Delta\theta_{\text{coax-stack},3}^*) \times f_1(\theta_4, a_{\text{coax-stack},4}, \theta_{\text{coax-stack},4}^0, \Delta\theta_{\text{coax-stack},4}^*) \times f_1(\theta_5, a_{\text{coax-stack},5}, \theta_{\text{coax-stack},5}^0, \Delta\theta_{\text{coax-stack},5}^*) \times f_1(\theta_6, a_{\text{coax-stack},6}, \theta_{\text{coax-stack},6}^0, \Delta\theta_{\text{coax-stack},6}^*) \times f_1(\theta_7, a_{\text{coax-stack},7}, \theta_{\text{coax-stack},7}^0, \Delta\theta_{\text{coax-stack},7}^*) \times f_1(\theta_8, a_{\text{coax-stack},8}, \theta_{\text{coax-stack},8}^0, \Delta\theta_{\text{coax-stack},8}^*) \times f_1(\pi - \theta_8, a_{\text{coax-stack},8}, \theta_{\text{coax-stack},8}^0, \Delta\theta_{\text{coax-stack},8}^*).$$

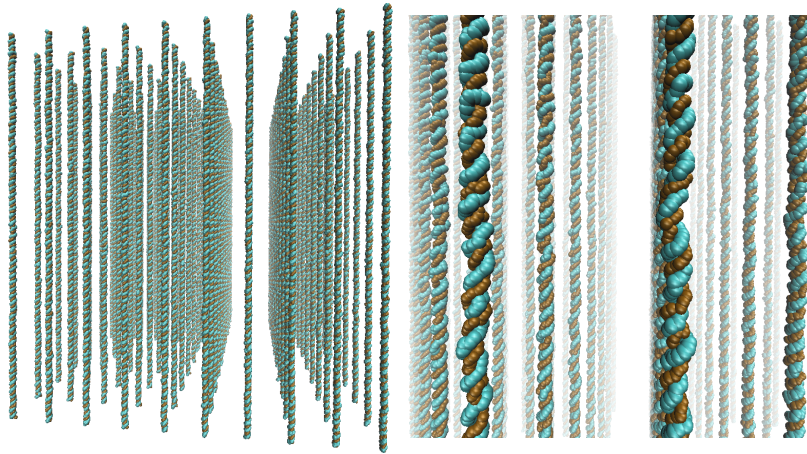
$$V = \sum_{\text{all}} (V_{\text{backbone}} + V_{\text{stack}} + V_{\text{exc}}) + \sum_{\text{other pairs}} (V_{\text{HB}} + V_{\text{cross-stack}} + V_{\text{coax-stack}} + V_{\text{exc}}).$$



For full details see Thomas Ouldridge, Coarse-grained modelling of DNA and DNA self-assembly, DPhil, University of Oxford, 2011.

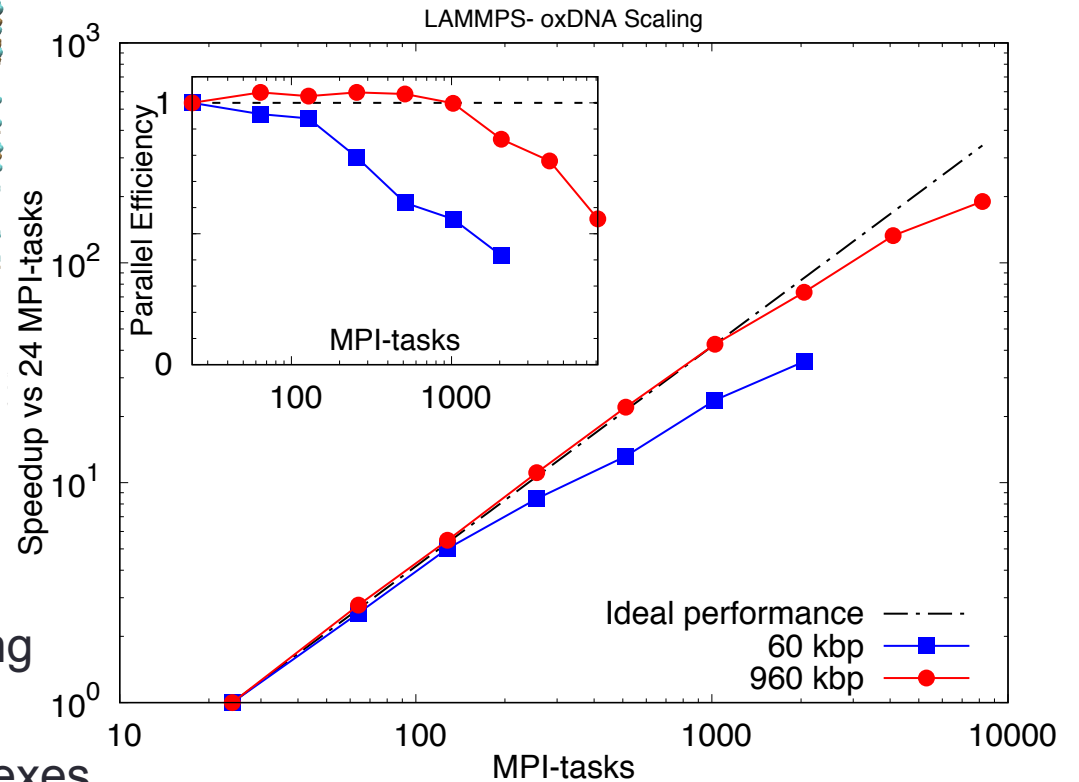


Parallel Performance



Benchmarks:

- Low density: array of 100 DNA duplexes with 600 base pairs long (see above) = 60 kbp
- High density: array of 1600 duplexes = 960 kbp



Strong scaling for 60kbp and 960 kbp
up to 8192 MPI-tasks

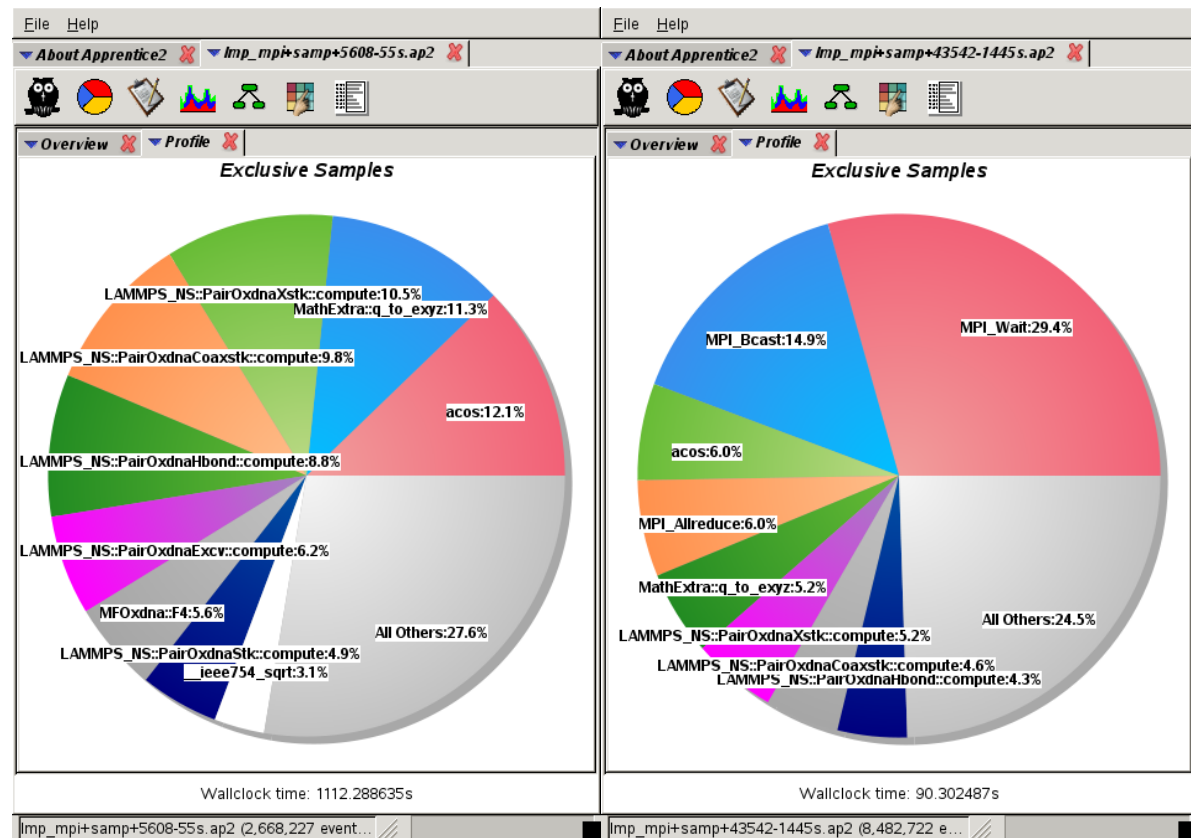


Craypat Performance Analysis

60 kbp benchmark

- Single node
 - Nlocal 5000
 - Nghost 1300
 - MPI < 5% (LMP)
 - compute 86% (LMP)
 - acos 12%
 - q_to_exyz 11%

- 2048 MPI-tasks
 - Nlocal 60
 - Nghost 225
 - MPI > 50%
 - compute 43% (LMP)



Single node
(24 MPI-tasks)

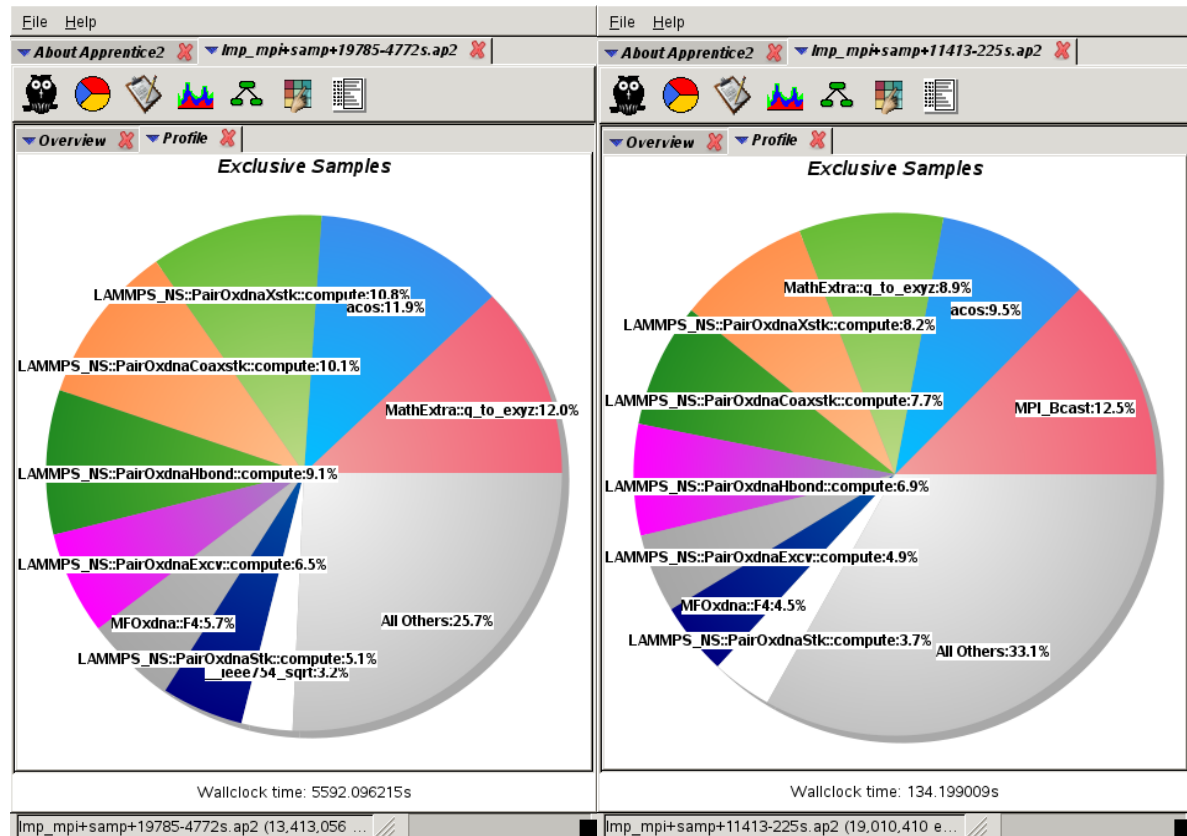
2048 MPI-tasks



Craypat Performance Analysis

960 kbp benchmark

- Single node
 - Nlocal 80000
 - Nghost 8300
 - MPI < 3%
 - compute 88% (LMP)
 - acos 12%
 - q_to_exyz 12%
- 2048 MPI-tasks
 - Nlocal 940
 - Nghost 480
 - MPI < 13%
 - compute 82% (LMP)

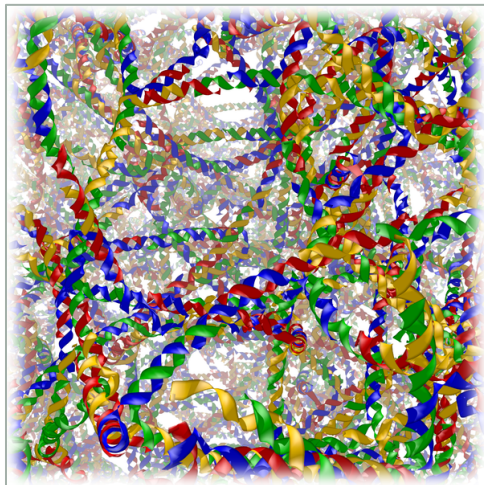


Single node
(24 MPI-tasks)

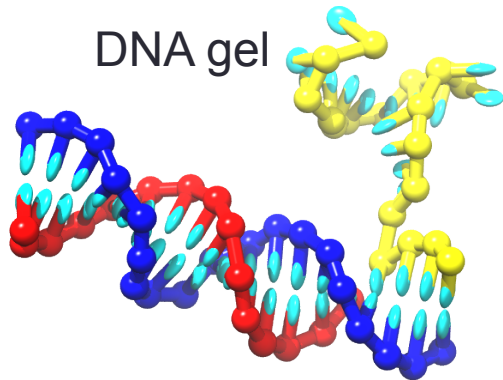
2048 MPI-tasks



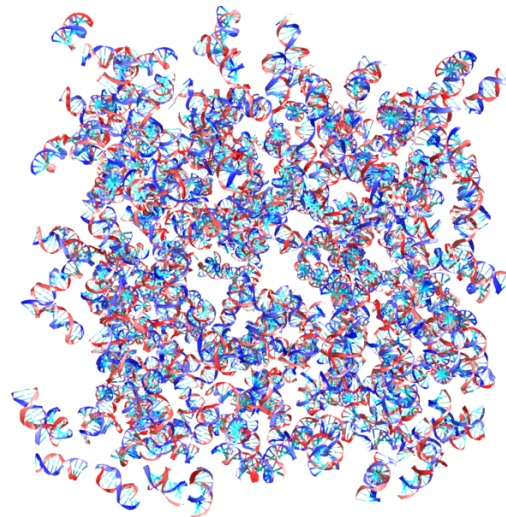
Applications



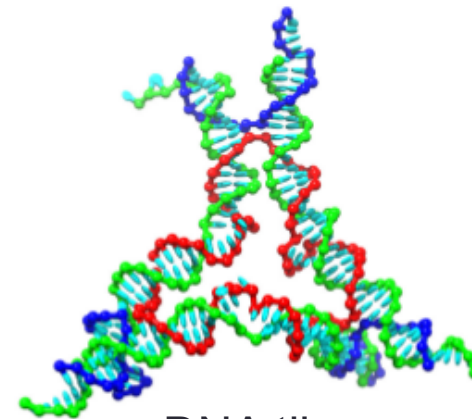
DNA gel



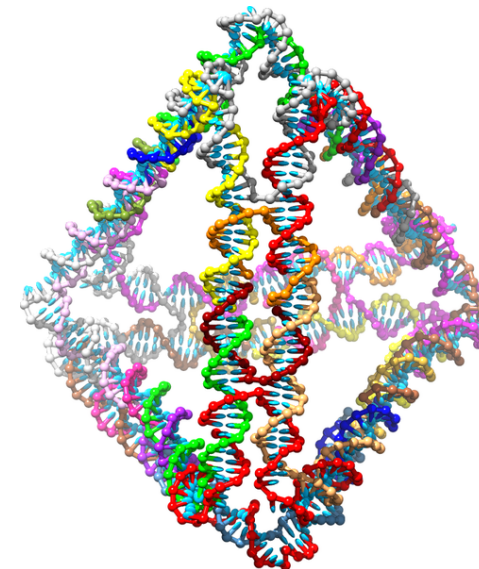
Stand displacement



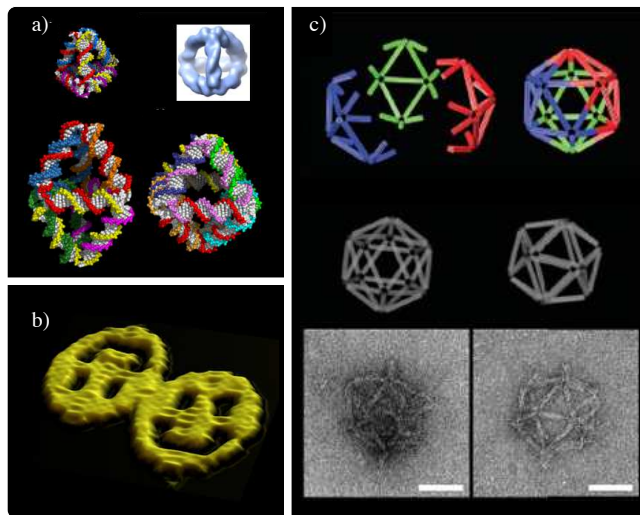
Liquid-crystalline states of DNA



DNA tiles



DNA tetrahedra



DNA nanostructures



Code Distribution

- LAMMPS version via **CCPForge**
 - <https://ccpforge.cse.rl.ac.uk/gf>
 - Project: **Coarse-Grained DNA Simulation** (cgdna)
 - Anonymous subversion access

svn checkout <https://ccpforge.cse.rl.ac.uk/svn/cgdna>

- In the near future also as **LAMMPS USER-package** with extended documentation
- Standalone version from <https://dna.physics.ox.ac.uk>



<http://www.archer.ac.uk/training/>

- Face-to-face courses
 - timetable, information and registration
 - material from all past courses
- Virtual tutorials & webinars
 - <https://www.archer.ac.uk/training/virtual/>
 - timetable plus slides and recordings
 - please leave feedback on previous tutorials after viewing material





Goodbye!

Thanks for attending

Please leave feedback at:

www.archer.ac.uk/training/feedback/online-course-feedback.php

