Doing more in less time Some examples from Computational solid state chemistry.

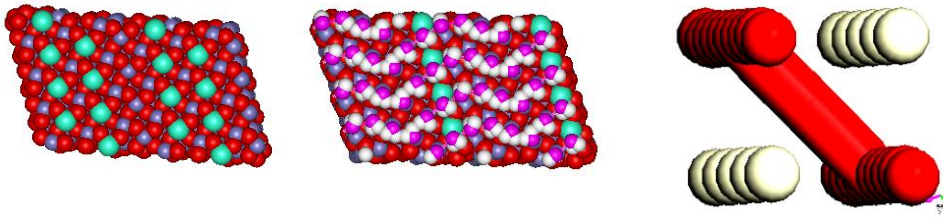
David J Cooke

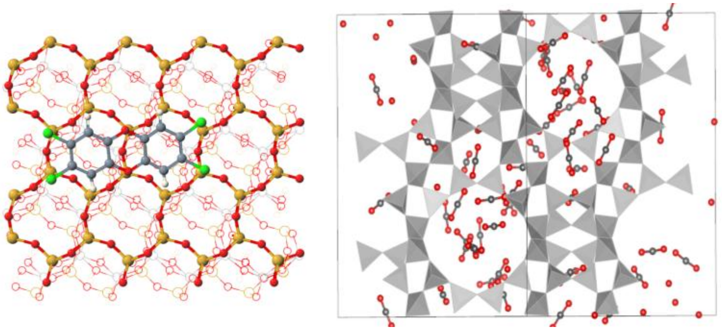
University of
HUDDERSFIELD

Materials and Catalysis
Research Centre

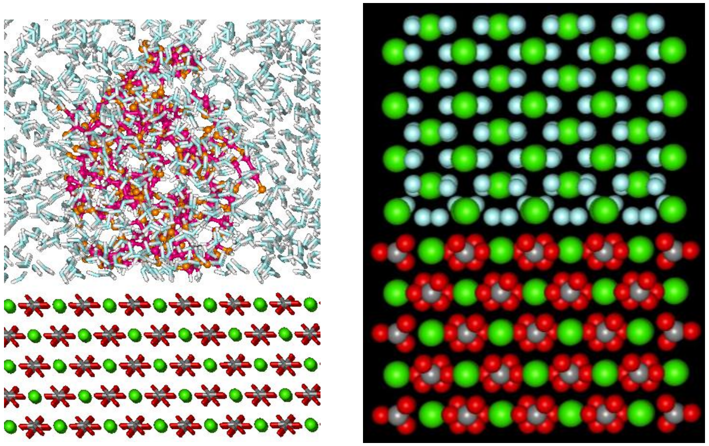
HPC FOR HUDDERSFIELD

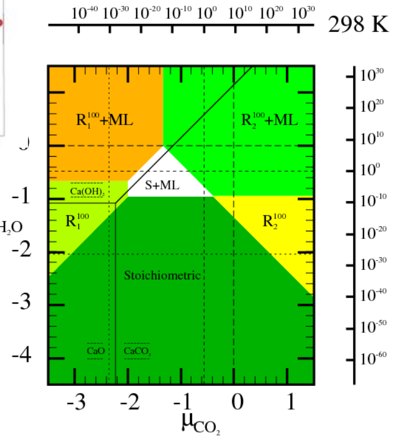
Bulk and Surface Defects



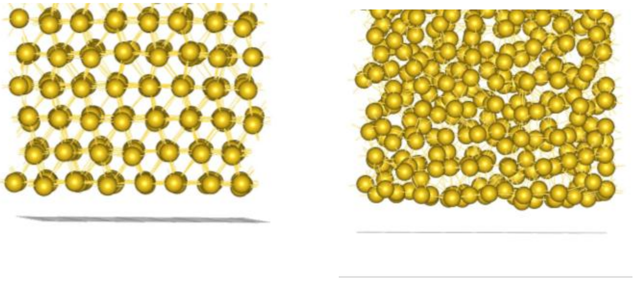


Interfaces

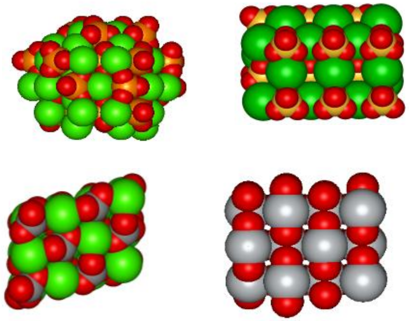




Crystal Growth



Nanoparticles



Selected Publications

J Phys Chem C 114:2462 2010 J Phys Chem B 110:7985 2006
Langmuir 26:14520 2010 J Mat. Chem., 16, 1997 2006
J Chem Phys 127:104706 2007 J Phys Cond Mat 16:S2735 2004

Desktop PCs

Used to analyse and store data
But also to do work.
Preliminary calculations.
Energy Minimisation small MD calcs.
Condor Pools allow cycle sharing

Midrange Computers

Consist of 10s – 1000s cores.
E-science enables these resources to be shared.
Production run MD calcs and simple QM calculations.

High Perform Computers

National Resources - Consists of 1000+ cores.
Access is normally as part of a consortium
Enables large *capability* calculations to be undertaken.
Allocated time counts as research income
UK machine at 32 in TOP500 list

My PhD (1999-2002)

- Considered the segregation of 10 dopents to two hematite surfaces in three chemical conditions in 256 possible surface configurations.
- All calculations were short (<2 hour) but required >20000 calculations.
- Main tool were PCs, workstations and a few small clusters shared by several research groups (usually with little disk space).

Then and Now

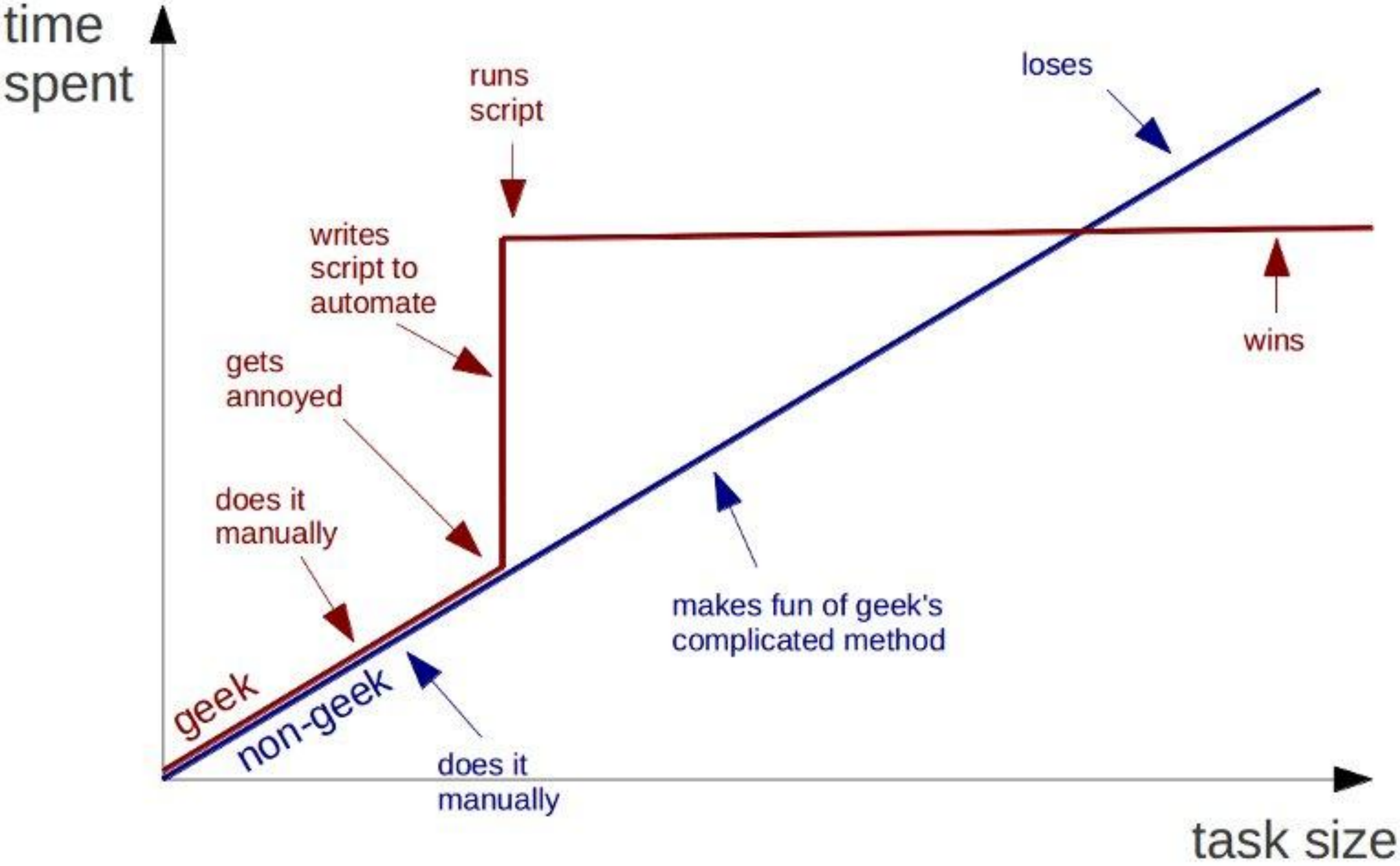
Knew relatively little scripting
PC Pentium 4 -- 256 Mb RAM
University cluster had 20 cores (upgraded to 96)

Can script in perl and sh
i7 quad core – 2 GB RAM
University clusters have combined 400 cores (plus condor pool)

Ran jobs on serial batches on PCs, workstations and clusters. Calculations took over 2 years to complete

Run calculations in parallel batch on much fast computers.
Make use of condor pools
Calculations could be done in days.

Geeks and repetitive tasks



Why not just submit them all in one go?

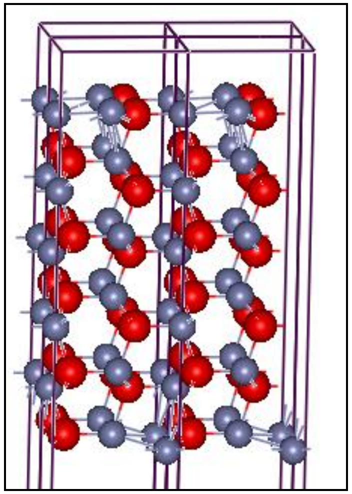
- Many systems limit the number of jobs you can submit to the queue (e.g. on Hector 8)
- System admins may define fair usage in terms of #jobs run
- Some systems require a single job to use all of a node (e.g. on Hector 32 cores).
- Annoys other users (if they take days to clear the queue)

Options

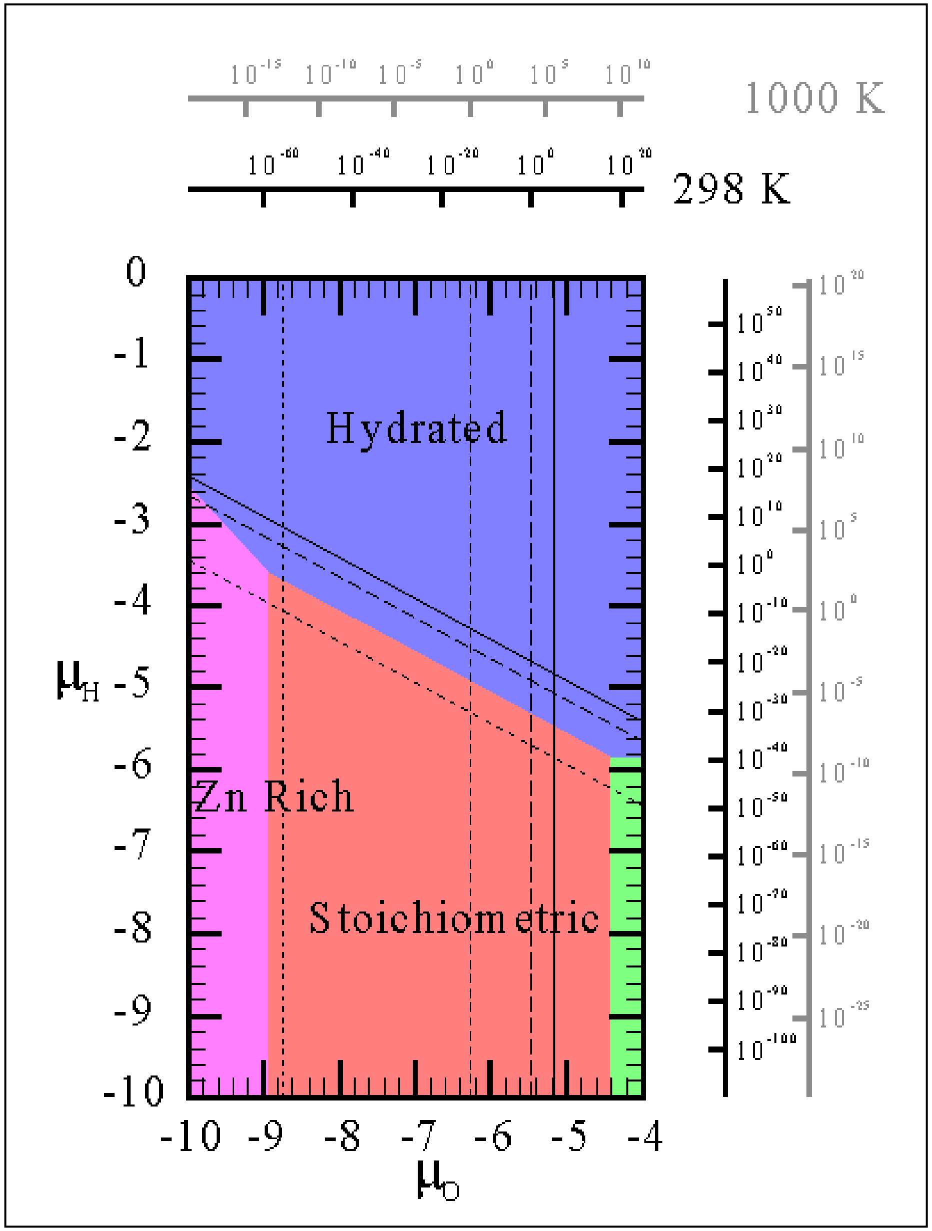
- Use a system designed for high throughput submission (e.g. CONDOR)
- Which is the best machine for my job?
- Submit Jobs as an array.
- Add an MPI wrapper to code so many serial jobs run as one MPI job.
- Investigate parallelising your code
- Are there ways of speeding each calc up?
- Submit jobs in a series of batches

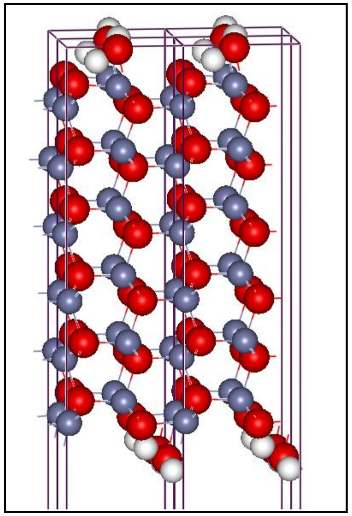
$ZnO(10.0)$ Phase Diagram

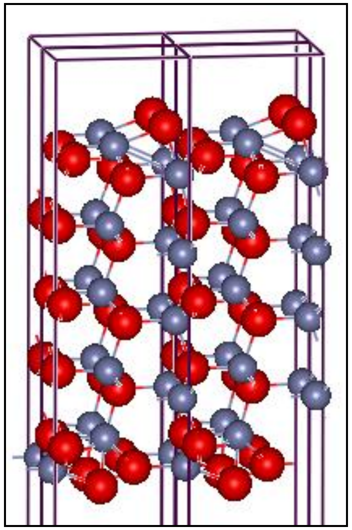
Dominated by hydrated surface and the pure stoichiometric



Zn Rich



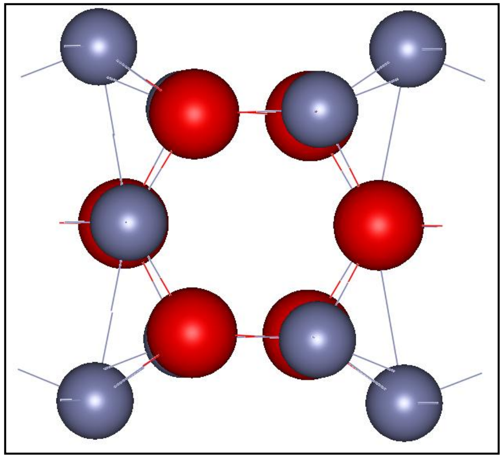


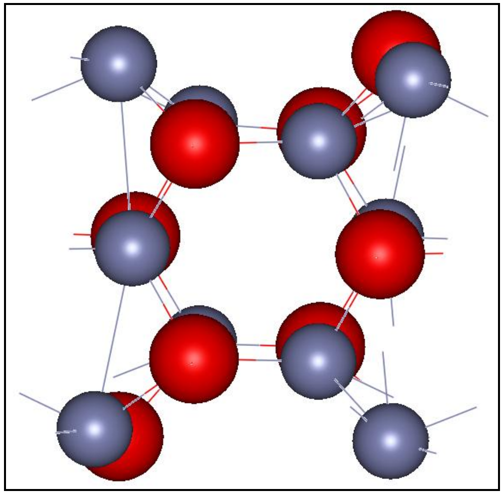


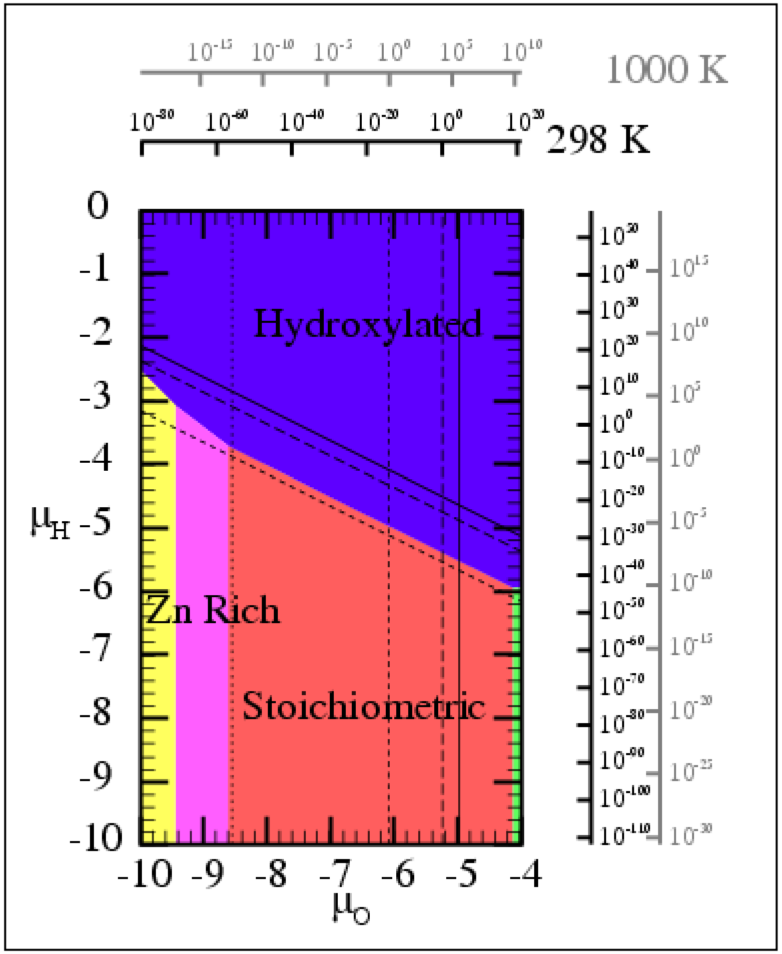
O Rich

J Phys Chem B 110: 7985 (2006)

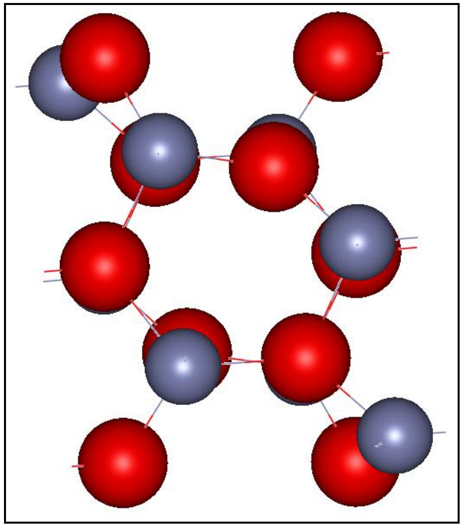
(11.0) Phase Diagram







Phase diagram now dominated by hydroxylated surface



Scanning Configuration Space

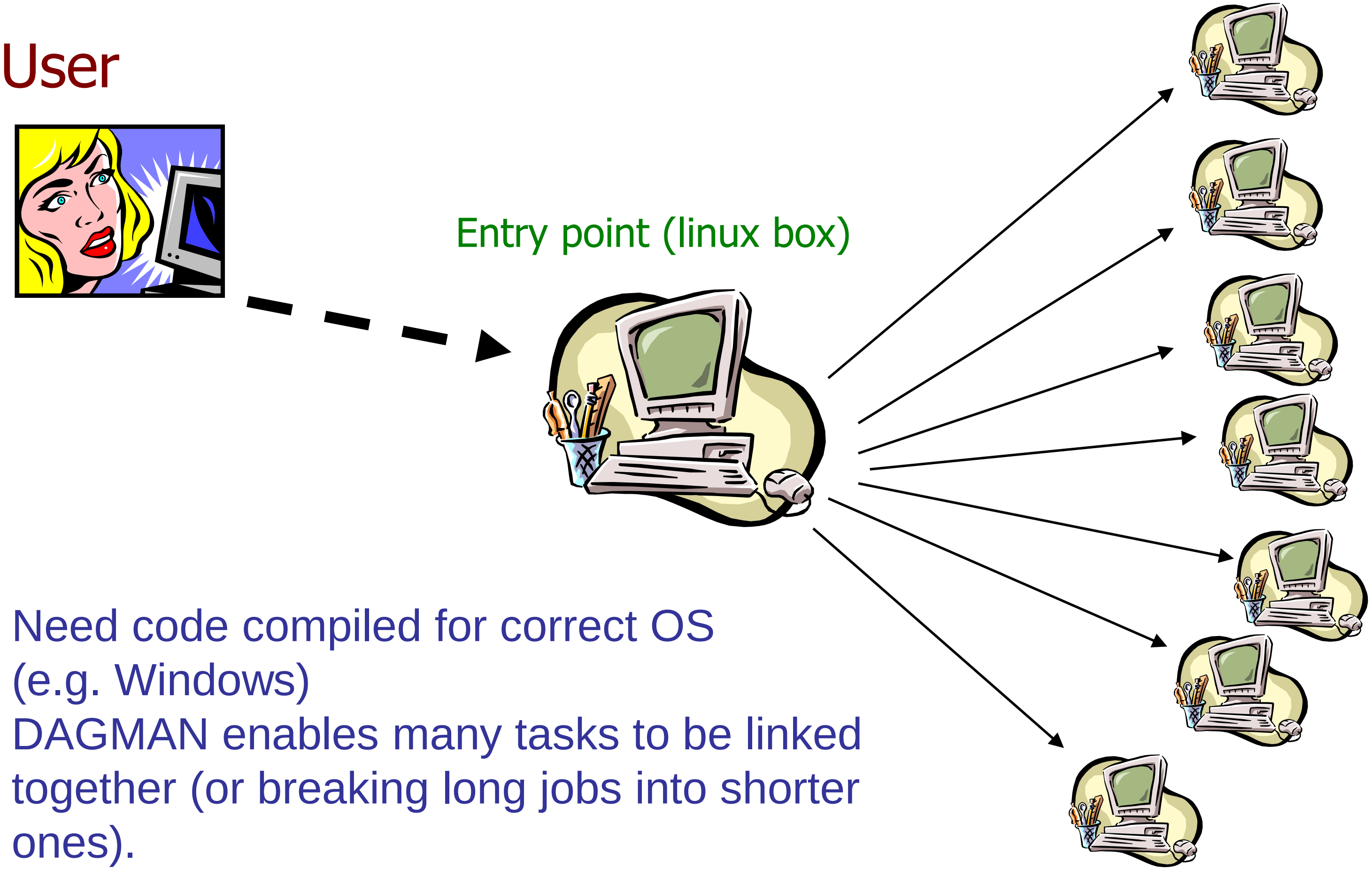
So far we have only considered a 1x1 surface cell but does concentration have an effect on whether hydration of hydroxylation is favoured?

But to do this properly is a lot of calculations
Consider 2x2 cell

(10.0)	(11.0)
4 cation sites	8 cation sites
16 Configurations for hyd	256 Configurations for hyd
70 Configurations fo hydrox	12865 Configurations for hydrox

An awful lot of calculations – would be unpractical to do in a conventional way

Principle of condor



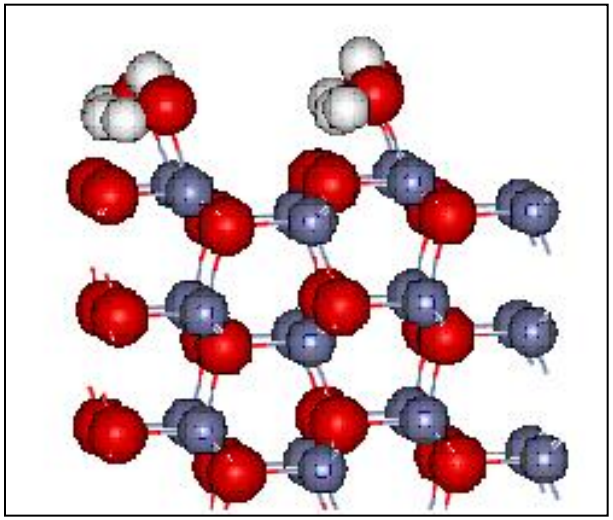
User

Entry point (linux box)

Loads of small PC (library, Labs, etc …)

Need code compiled for correct OS (e.g. Windows)
DAGMAN enables many tasks to be linked together (or breaking long jobs into shorter ones).

Metadise (10.0) Surface



	E_{hyd} / eV	**γ / J/m^2**	**E_{OH} / eV**	**γ / J/m^2**
0.25	-1.15	0.95	-1.09	0.95
0.5	-1.17	0.76	-1.12	0.78
0.75	-1.12	0.62	-0.84	0.83
1.0	-0.89	0.67	-0.64	0.93

Hydrated surface is most stable at all concentrations

Optimum coverage is below 100% but hydration energy suggests in aqueous conditions a full monolayer would form

Structure similar to DFT structure Hydrogen bond is 1.74Å

(11.0) Surface

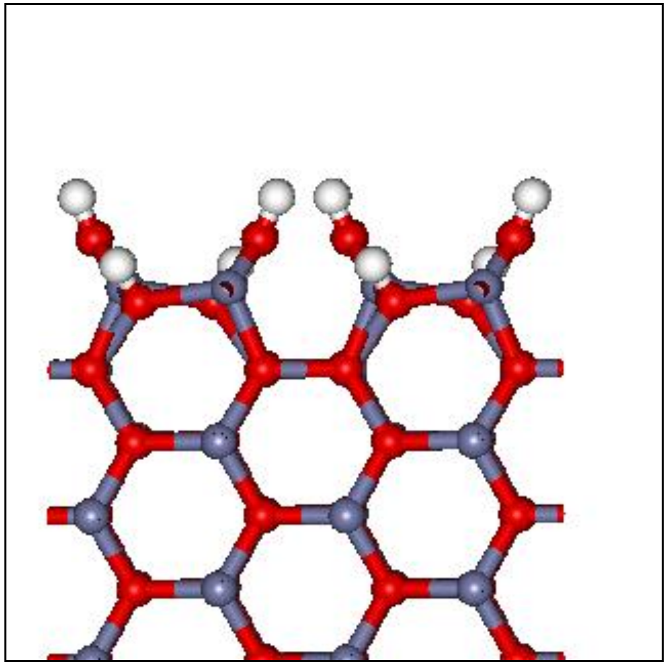
	Ehyd / eV	**γ / J/m2**	**Ehydrox / eV**	**γ / J/m2**
12.5	-1.00	1.09	-0.75	1.13
25	-1.24	0.94	-1.01	1.01
37.5	-1.34	0.79	-1.10	0.89
50	-1.40	0.63	-1.05	0.83
62.5	-1.22	0.73	-1.05	0.74
75	-1.12	0.69	-1.06	0.65
87.5	-1.16	0.45	-1.02	0.61
100	-1.02	0.52	-0.89	0.67

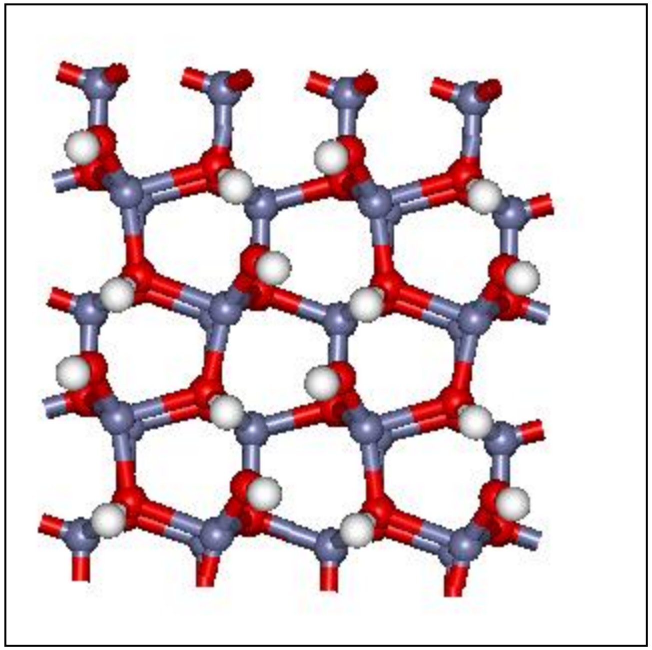
Again Hydroxation and Hydration energies are similar

But Atomistic Calculations predict it is the hydrated surface that will be more stable

Optimum coverage again just below 100%

Lack of Hydrogen Bonding





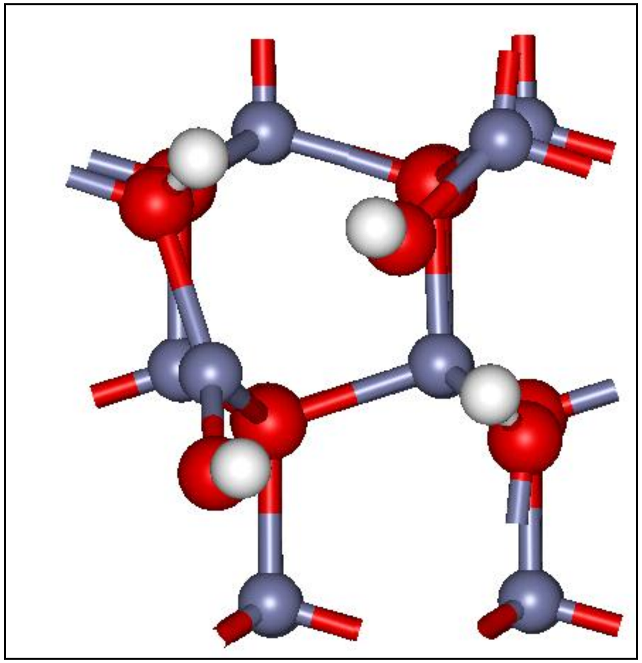
Whilst the DFT calculation was stablised by a hydrogen bonding network this in not present in the atomistic case

Potential needs refining? Calculation in local minimum?

Some More Scanning

Changed the initial position of H^+ and OH^- using two approaches

1) Using relaxed coordinates from DFT calc
2) Add small fluctuations to the initial positions of the hydrogen



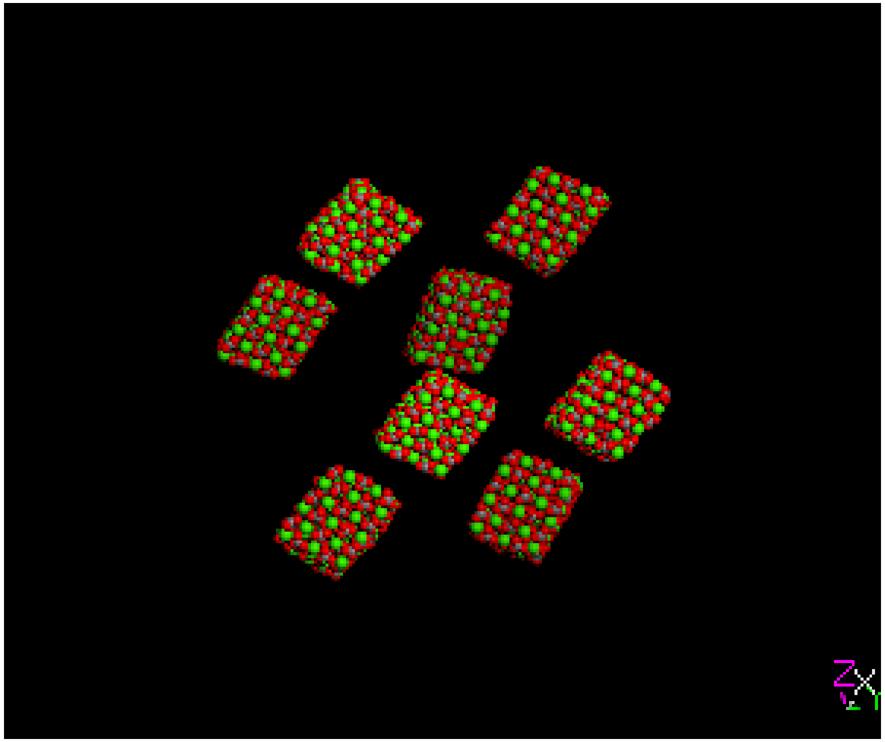
Fully Hydroxylated System
Hydrogen bonding now prediced
(bond distance ~2A)

Stabilises surface by a small amount
Surface energy reduces to 0.65J/m^2

Still less stable than hydrated surf

Nanoparticle Growth

Another import issue in crystal growth processes is understanding how small particles aggregate to form larger and potentially more stable systems



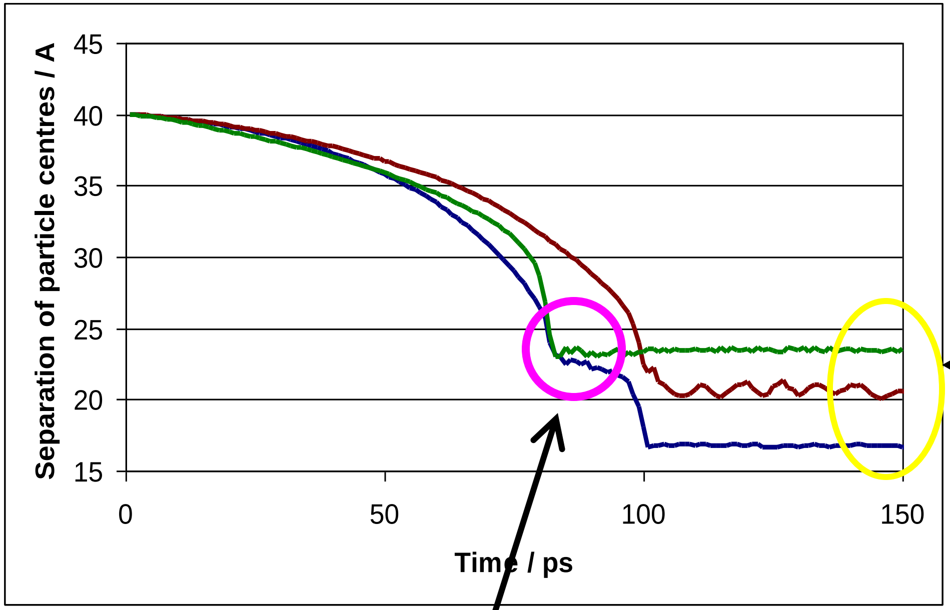
8 75$CaCO_3$ nanoparticles based on the (10.4) calcite surface
Placed on corners of a cube 30Å apart
MD run at 300K for 5ns

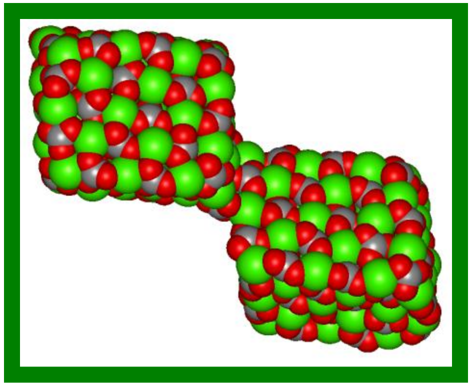
Initially particles relax internally before pairs come together
Which join to form larger particles

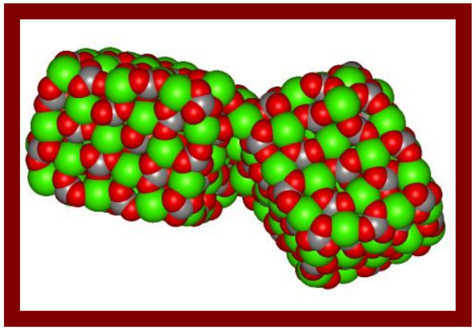
Interesting but can we be more quantitative?

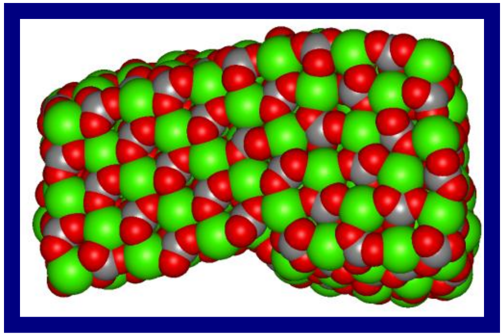
Nanoparticle Growth

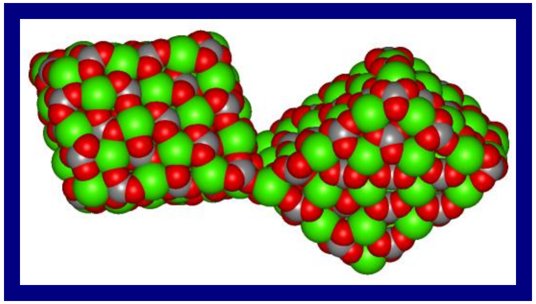
Consider Pair of particles coming together

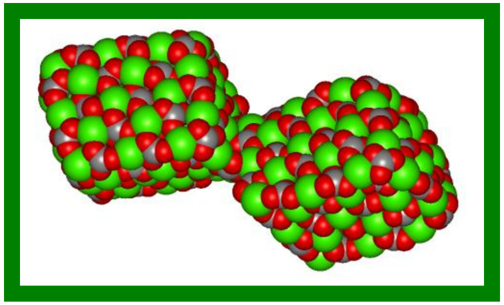


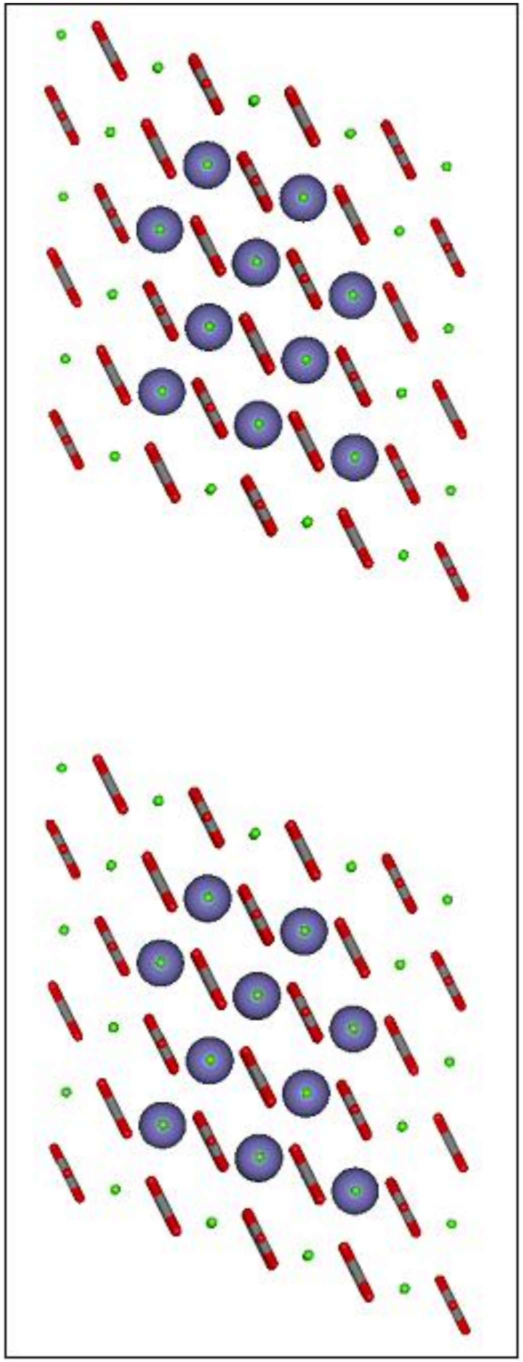












Potential of Mean Force

Particles restrained so the separation the central Ca^{2+} ions in each particle remain constant

Run 100ps MD simulation monitoring Force on the system due to the restraint

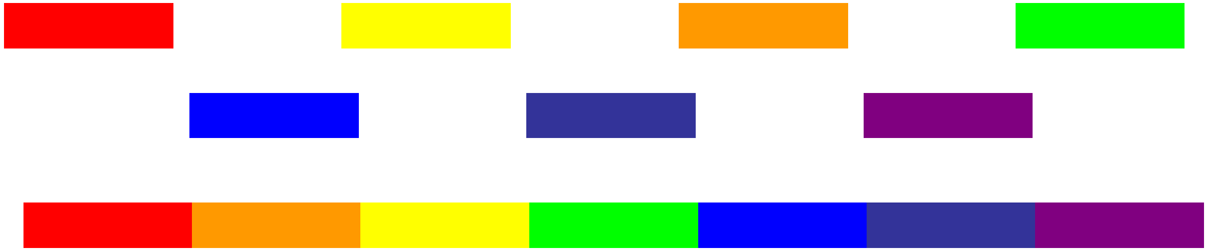
Run dozens of calcs with different separations

Integrating the PMF with respect to separation gives
Free Energy of Aggregation

Task Farming / Job Arrays

Use the method developed by Changman Moon (DLMASTER)

Allows multiple small MPI jobs to run as a single large MPI job



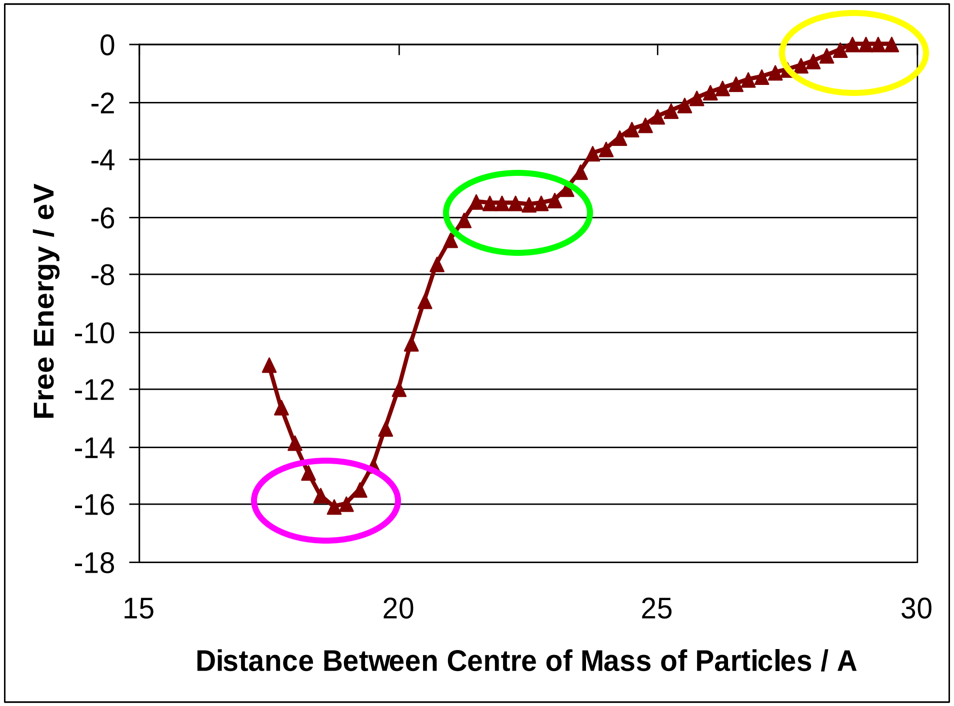
Each job runs in a separate directory and there is no communication between the jobs after the initial cpu allocations

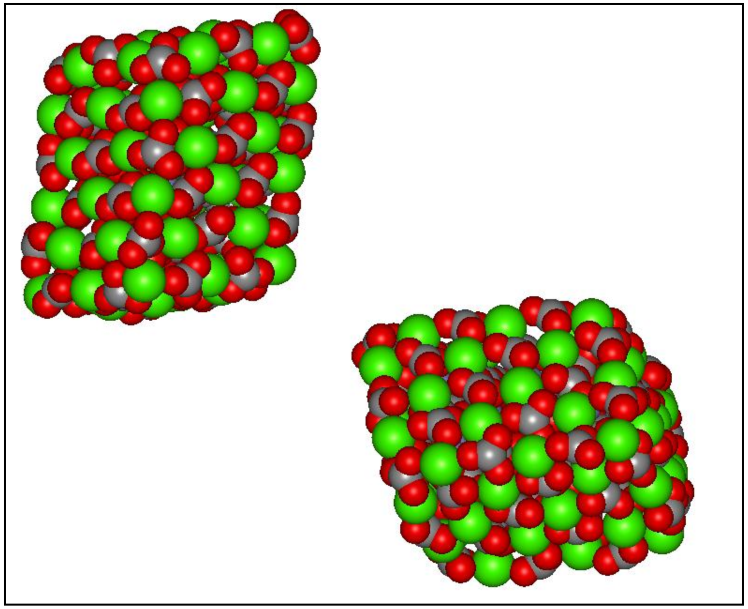
Consequently scaling is near perfect but large amounts of output data is written so disc space can become an issue

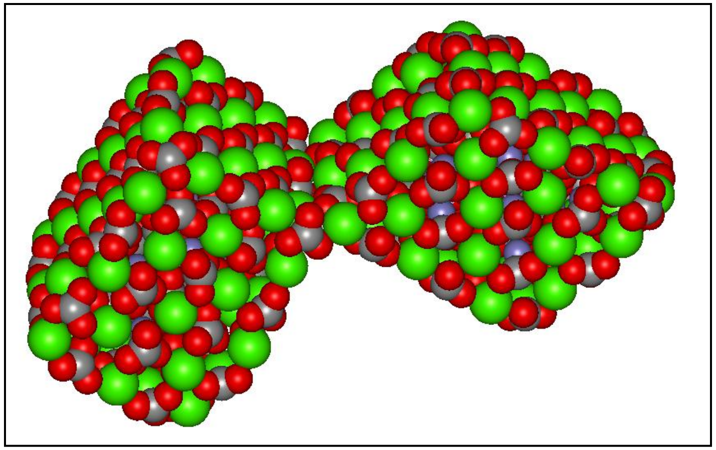
Can be applied to multiple serial jobs.

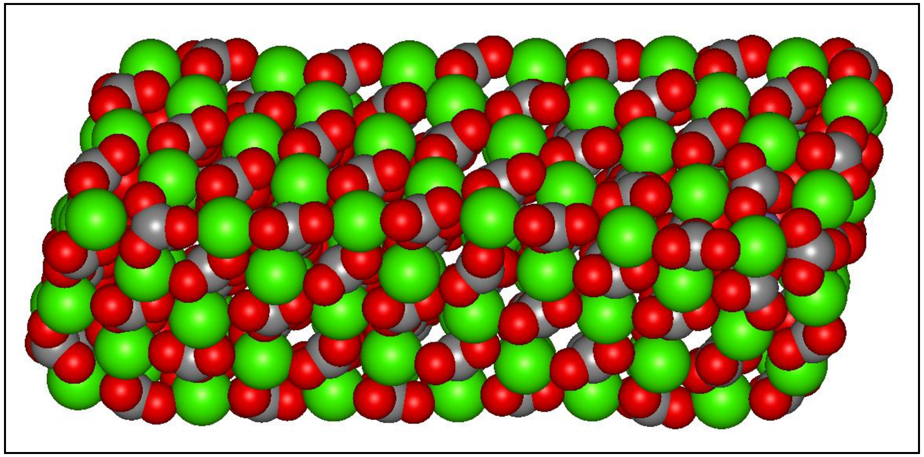
Job arrays do similar thing but use scheduler

PMF

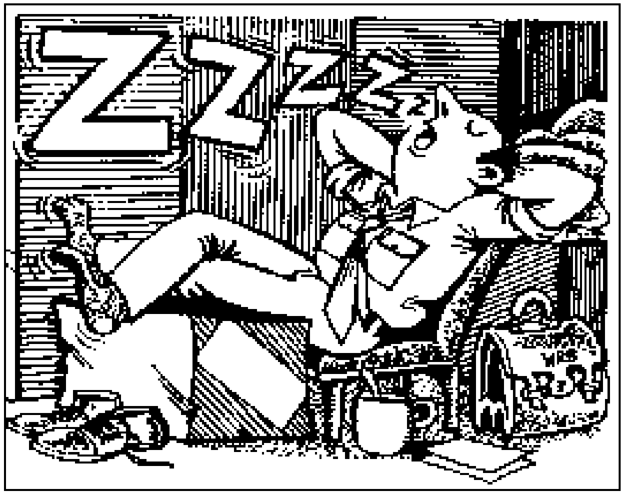

0
-2
-4
-6
-8
-10
-12
-14
-16
-18
Free Energy / eV
15
20
25
30
Distance Between Centre of Mass of Particles / A

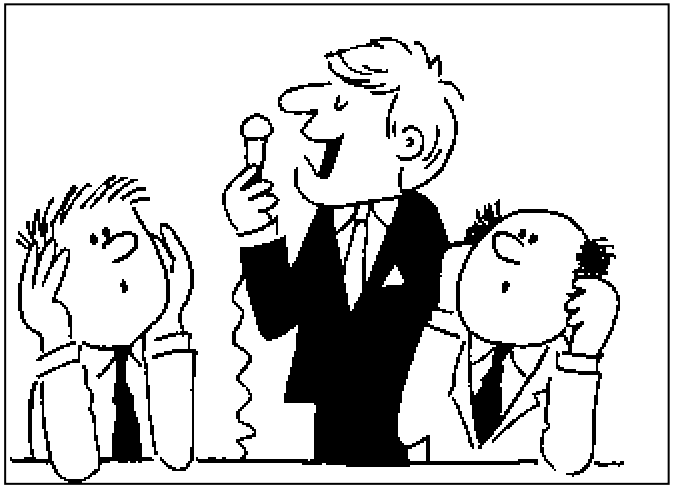






Thankyou For Your Attention





Acknowledgements

James Elliott (Cantan) Steve Parker (Bath) Arnaud Marmier (Exeter)
Various consortia, and research grants for computer time over the past 15 years