Science & Technology
Facilities Council

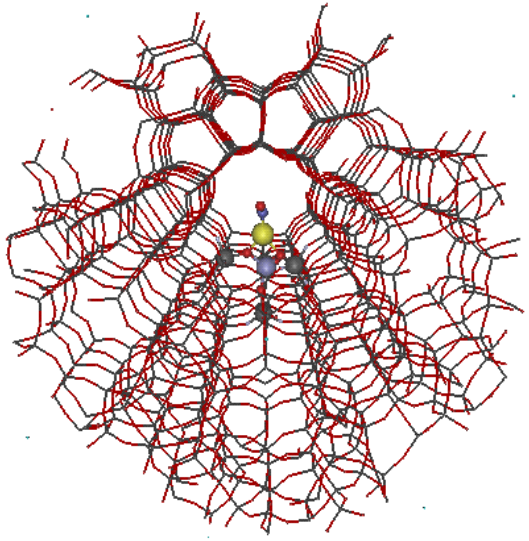
Task-farming parallelism in ChemShell

Tom Keal
STFC Daresbury Laboratory

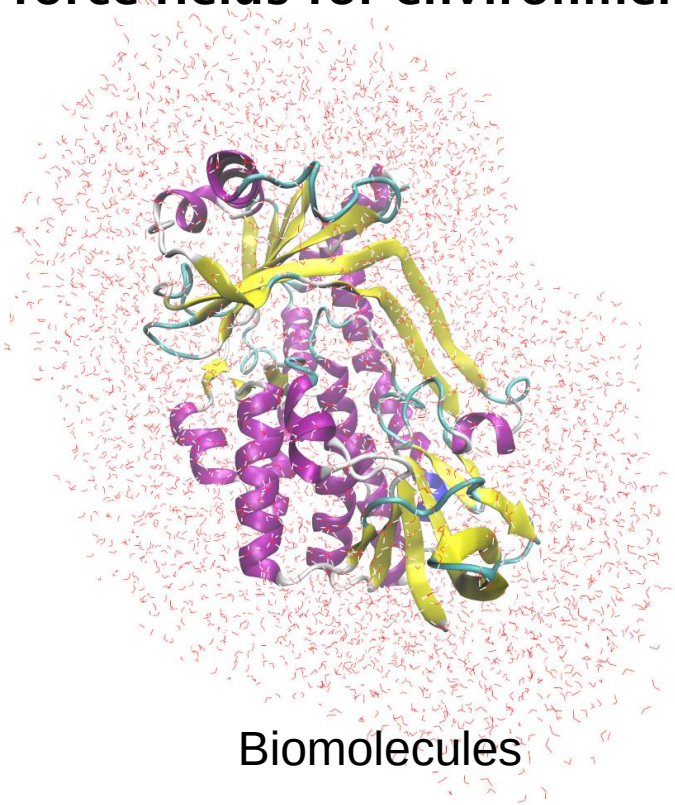
8 May 2013

QM/MM modelling

- **ChemShell combines high-level quantum mechanical calculations for reactive region with classical force fields for environment**

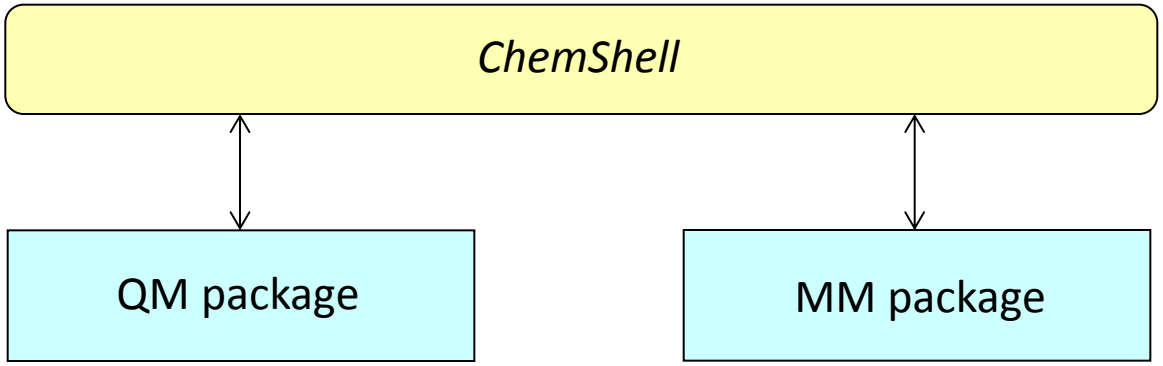


Materials



Biomolecules

QM/MM calculations

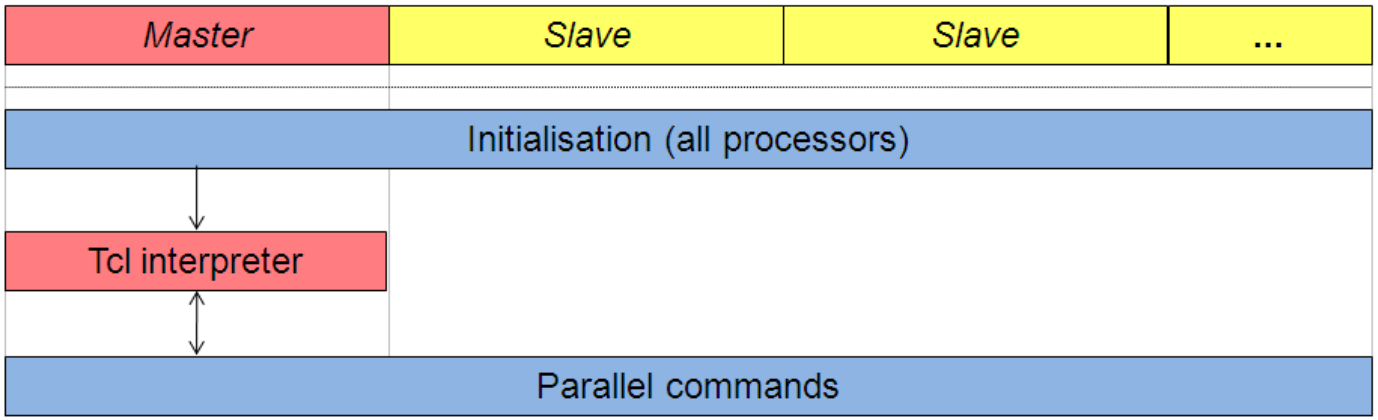


- User controls ChemShell using Tcl commands
- ChemShell modules written in C/Fortran

- External programs called to obtain energy & gradient
- ChemShell forms combined QM/MM gradient

ChemShell in parallel

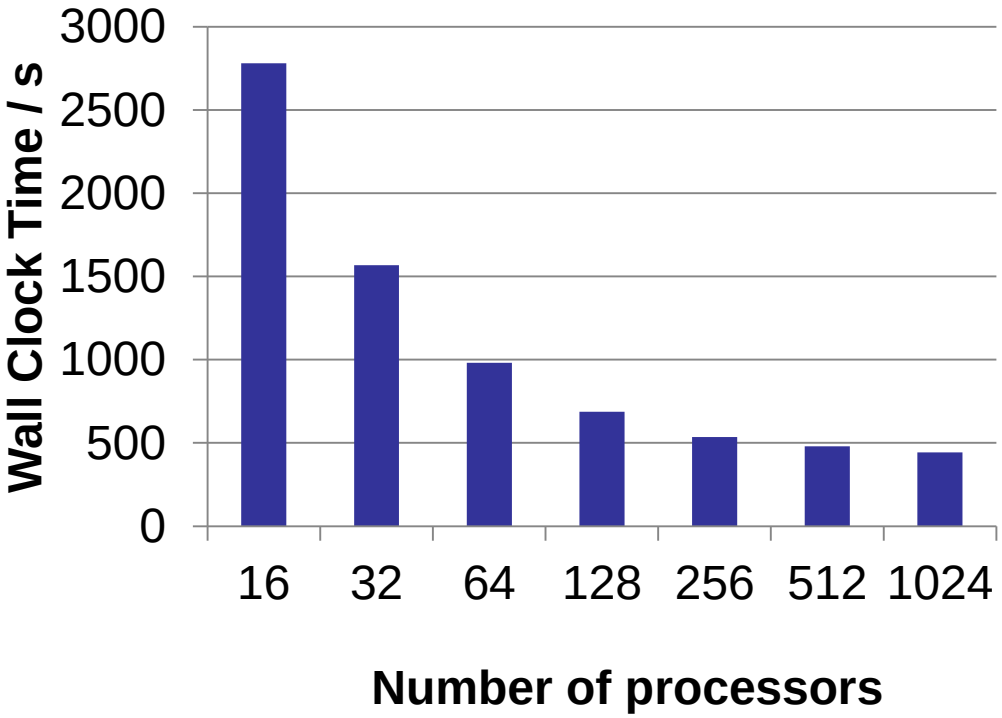
- **ChemShell can run in parallel using MPI**



- **Takes advantage of parallel external programs**
  - e.g. parallel energy/gradient evaluation in GAMESS-UK
- **However, this approach does not scale well to large numbers of processors**

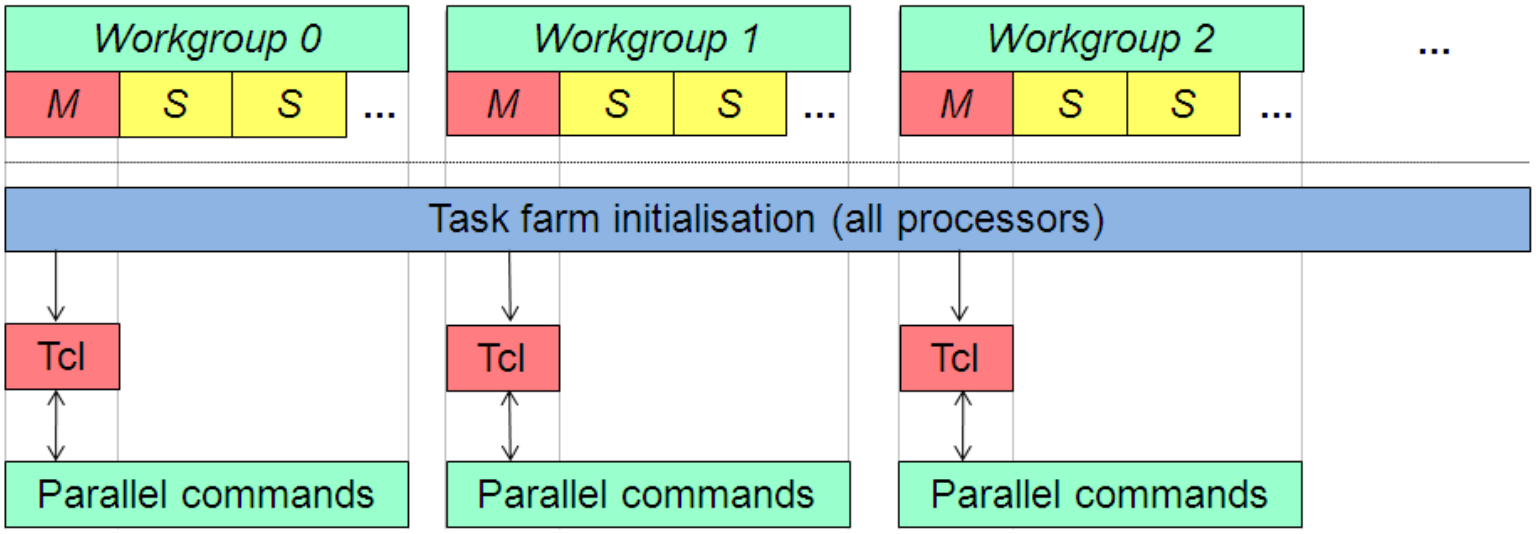
Typical scaling behaviour

- **Single-point GAMESS-UK timings (57 atom silicate cluster, TZVP)**



Task-farming parallelism

- **Aim is to parallelise ChemShell algorithms as well**
  - e.g. parallel Hessian evaluation, optimisations, etc.
- **Task-farming approach: divide up processors into workgroups working independently on tasks**

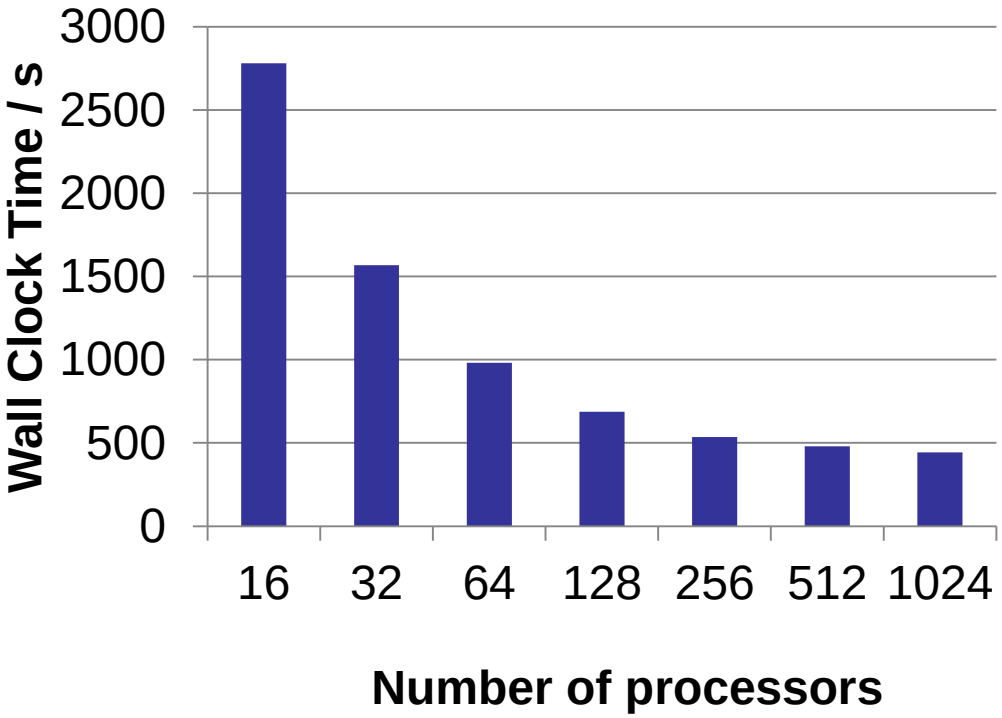


Task-farming parallelism

- **Workgroups are essentially independent**
  - Separate outputs, working directories
- **All lower-level parallelism (e.g. external calculations) occurs within a single workgroup.**
  - GAMESS-UK, GULP, etc. modified to work within workgroups
  - MPI communicator split by ChemShell and passed to external program
- **All workgroups are controlled via a single Tcl input script**
  - Tcl commands to allow workgroup-specific tasks, synchronise workgroups, share data, etc.

Example: finite-difference Hessian

- **Single-point GAMESS-UK timings (57 atom silicate cluster, TZVP)**



Example: finite difference Hessian

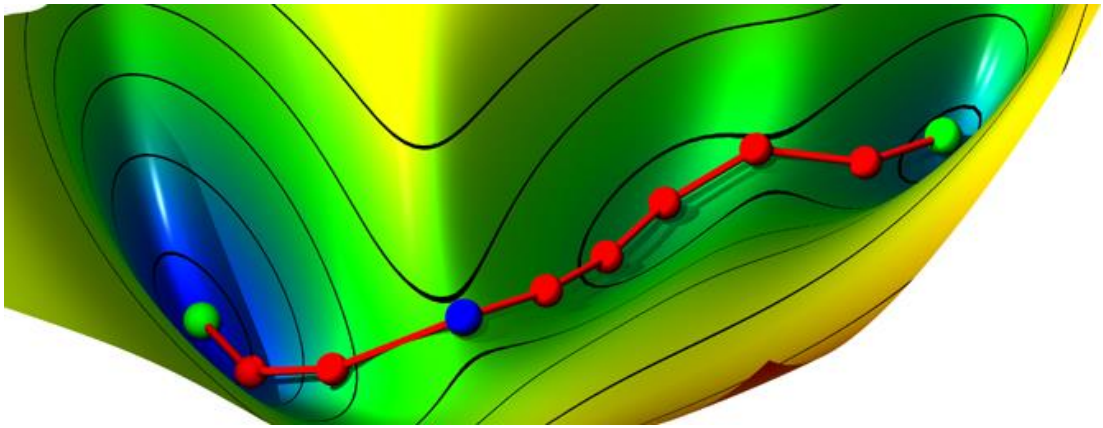
- **Performance can be improved by dividing up the set of 1024 processors into workgroups:**

Workgroups	Procs / workgroup	Time / s	Speed-up
1	1024	52762	
64	16	7812	6.8

- **Optimal number of workgroups depends on system size / total processor count / algorithm**

Example 2: nudged elastic band method

- Optimising reaction paths: finds the minimum energy path

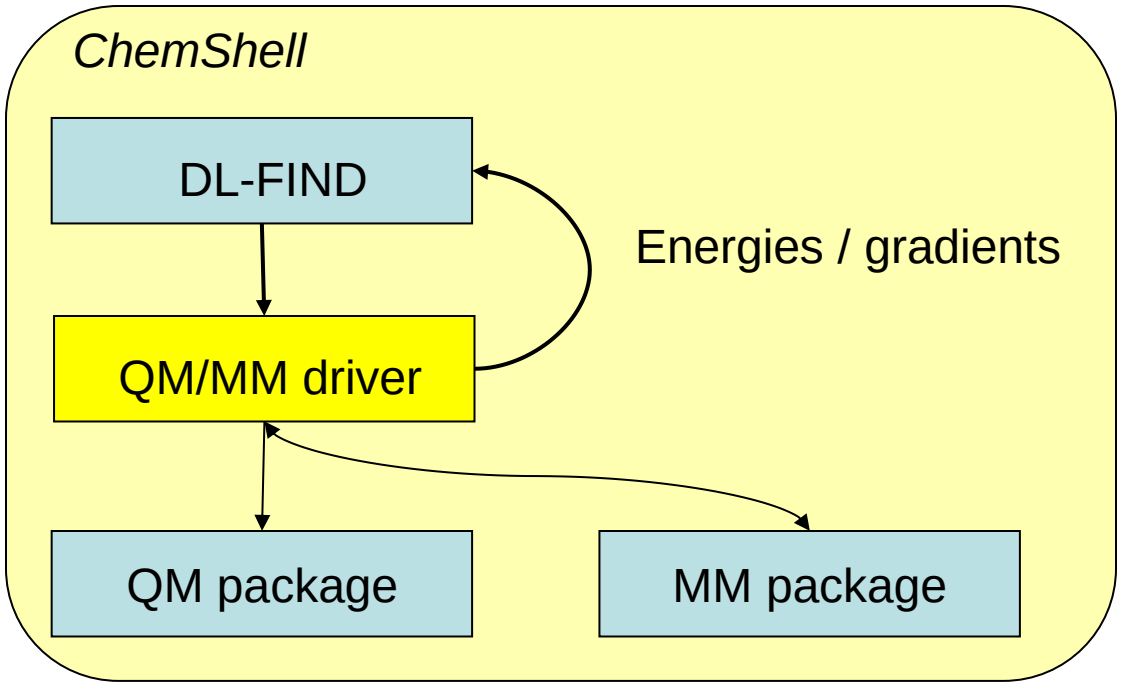


- Multiple images, connected by spring forces.
- Climbing image to find transition state
- Image E/g evaluations are independent, so can be parallelised

G. Henkelman, B.P. Uberuaga, H. Jónsson, *J. Chem. Phys.*, **113**, 9901, (2000)

DL-FIND

- An open-source geometry optimisation library
- Interface to ChemShell for QM/MM optimisations
- NEB routines implemented in DL-FIND



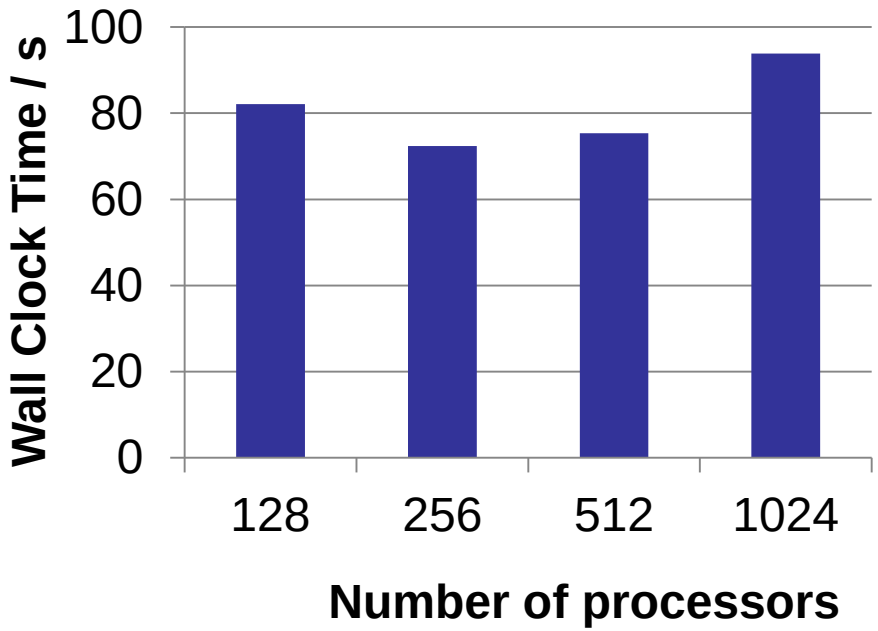
J. Kästner, J.M. Carr, T.W. Keal, W. Thiel, A. Wander, P. Sherwood, *J. Phys. Chem. A,* **113**, 11856, (2009)

Parallel NEB in DL-FIND

- **ChemShell/DL-FIND parallel interface implemented**
  - Pass the relevant workgroup information
- **Each workgroup runs DL-FIND**
  - Images allocated according to workgroup ID
  - Images remain with same workgroup throughout calculation (benefit from restart information)
  - Energies/gradients shared between workgroups at the end of each cycle

Parallel NEB timings

- **Test system: hydrogen interchange on Al-doped ZnO (see later)**
  - 3207-atom QM/MM cluster (32 QM atoms, PVDZ)
- **Number of workgroups is limited by number of NEB images**
  - Test system: 10 images (2 fixed)
- **Single point calculations:**



Parallel NEB timings

- **Test over 50 cycles of NEB**
- **Again divide up set of 1024 processors into workgroups:**

Workgroups	Procs / workgroup	Time / s	Speed-up vs 1024	Speed-up vs 256
1	1024	26404		
1	*256*	*23536*		
2	512	14673	**1.8**	**1.6**
4	256	7089	**3.7**	**3.3**
8	128	3110	**8.5**	**7.6**

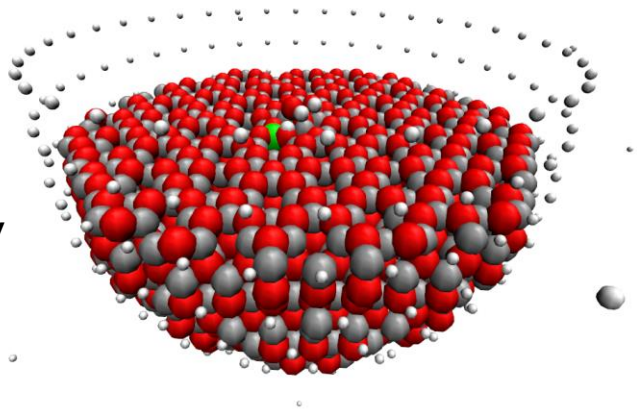
- **Task-farming is essential to make use of > 256 cores**

Science & Technology Facilities Council

Hydrogen dissociation study

- **Al-doped ZnO important in industrial catalysis (methanol synthesis)**
  - Important to understand interaction of hydrogen with ZnO
  - Catalytically important polar oxygen-terminated surface studied
  - Active site postulated to be vacant oxygen interstitial surface site

- **Embedded cluster model: 3200-atom system**
  - 61 QM atoms around a VOISS centre with nearby Al impurity
  - DFT ionic pseudopotentials (e.g. Zn^{2+}) associated with 63 atoms in the boundary region (cations only)
  - Polarised MM forcefield (shell model)

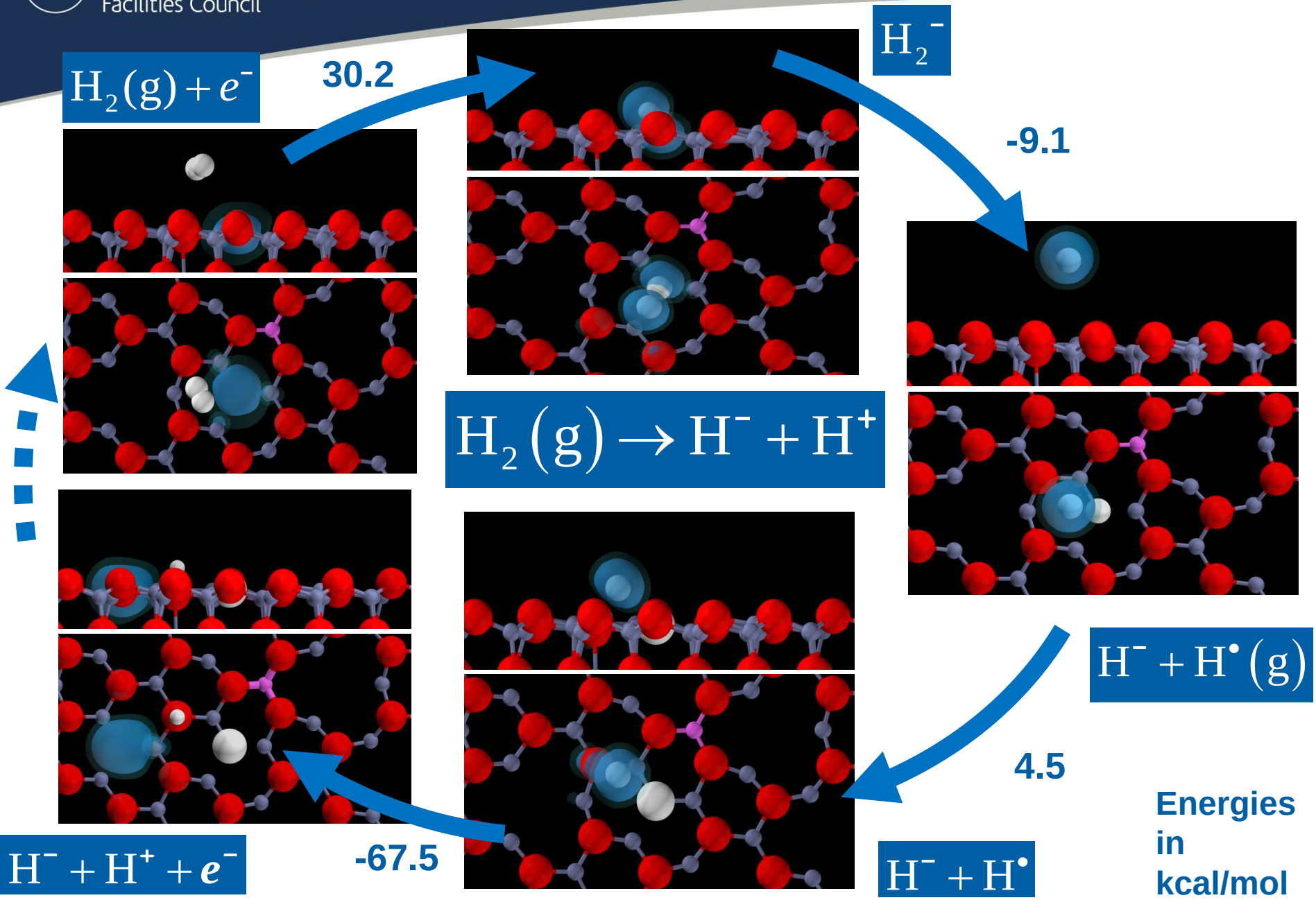


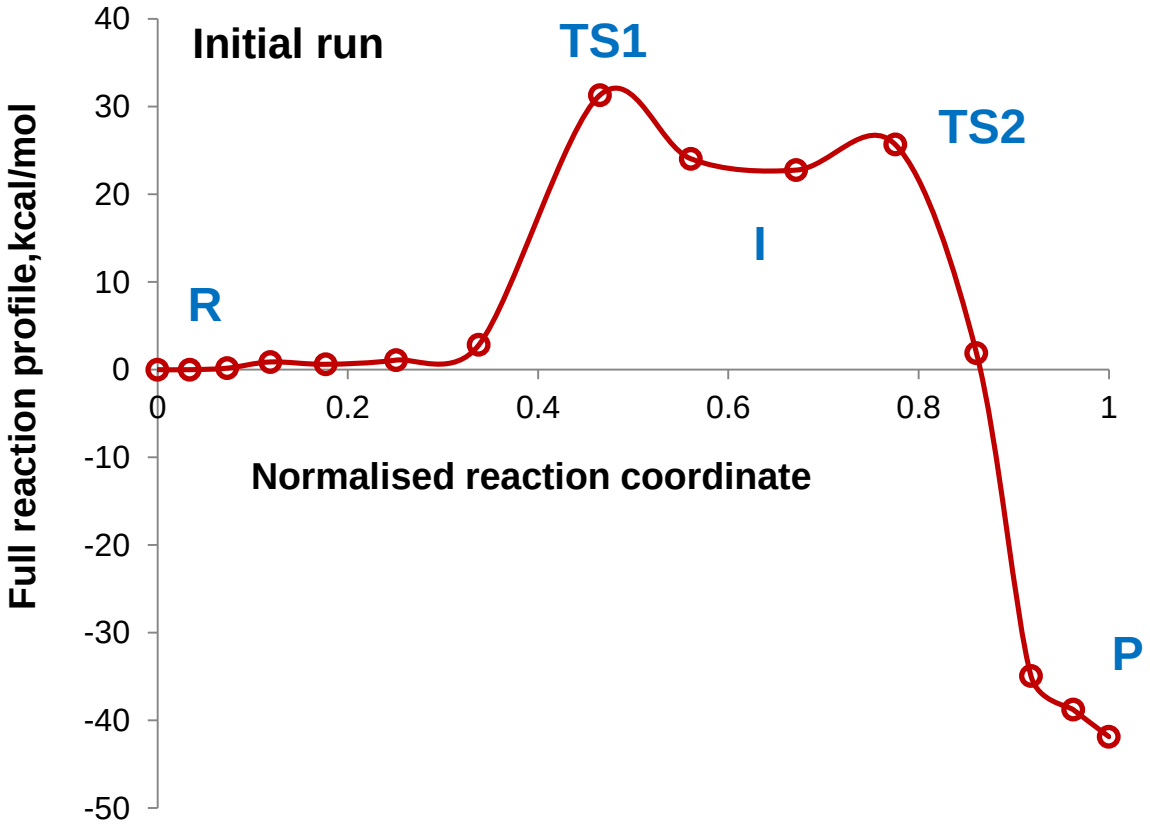
T.W. Keal, P. Sherwood, G. Dutta, A.A. Sokol, C.R.A. Catlow, *Proc. Roy. Soc. A,* **467**, 1900, (2011)

Exploring the potential energy surface

- **DFT: B97-2 functional, TZVP basis set**
- **Starting point: physisorption of hydrogen molecule on surface**
- **Endpoint: previous studies showed dissociation is heterolytic**
- **NEB method used to calculate barrier for dissociation**
  - First run indicated a stable intermediate, so two stage reaction
  - Two NEB runs then required: R->I and I->P

Science & Technology Facilities Council




Initial run
TS1
TS2
I
R
P
Full reaction profile,kcal/mol
Normalised reaction coordinate
40
30
20
10
0
-10
-20
-30
-40
-50
0
0.2
0.4
0.6
0.8
1

Comparison with experiment

- **Total activation energy of H_2 desorption = E(TS1) - E(P)**
  - 72.1 kcal/mol

- **Bowker et al: 71.7 kcal/mol**
- **Kunat et al: 77.9 kcal/mol**

- **Consistent with irreversible reaction**

- **Dissociated surface H's useful for further catalytic reactions over Al/ZnO (methanol synthesis)**

M. Bowker, H. Houghton, K. C. Waugh, *J. Chem. Soc., Faraday Trans. 1,* **77**, 3023, (1981)
M. Kunat, U. Burghaus, C. Wöll, *Phys. Chem. Chem. Phys.,* **5**, 4962, (2003)

Applicability of task-farmed ChemShell

- **Only relevant to large scale calculations**
- **External programs generally need to be modified and linked in**
  - QM codes: GAMESS-UK, DALTON, LSDALTON, FHI-AIMS
  - MM codes: GULP, DL_POLY
- **Works with parallel diagonalisers: PEIGS, ScaLAPACK (MPIBLACS)**
- **Not yet compatible with Global Arrays**
  - Required to support NWChem
  - Possible in principle?

ChemShell on BlueGene/Q

- **ChemShell ported successfully with GAMESS-UK and GULP**
  - Some modifications required (released in v3.6)
  - GAMESS-UK needs optimising
- **FHI-AIMS also working**
- **NWChem awaits a Global Arrays implementation on BGQ**

Outlook

- Microiterative optimisation in DL-FIND
- New embedding methods (NWChem)
- New QM/MM techniques, e.g. adaptive QM regions
- Replacing Tcl user interface with Python

Acknowledgements

- Paul Sherwood, Johannes Kästner
- Gargi Dutta, Alexey Sokol, Richard Catlow – H_2 dissociation study
- HECToR dCSE, EPSRC SLA, PRACE for funding

www.chemshell.org