Adaptation of chemical dynamics codes to the GRID

György Lendvay

Institute of Structural Chemistry, Chemical Research Center
Hungarian Academy of Sciences
Budapest, Hungary

Our experience
- parallel and grid computing
- QCT on grid
- Parallel quantum dynamics
- On-the-fly QCT

Codes we have adapted

Codes we plan to adapt

Our GRID experience

GRID is not a cluster or supercomputer
- is not always available
- is not homogeneous
- communication may be slow
- some connections get disconnected

conventional parallel execution is not necessarily the best solution

EGEE
- Software installation
  - not always smooth
  - versions that are “self-installing” are easier,
    - but even then we needed help from SZTAKI
  - keeping a cluster available for EGEE requires significant human effort
    - not easy – not everyone can have a person to devote a lot of time to maintain such a connection
  - it would be helpful if the LCG software reached a stable level and would be upgraded VERY rarely

Our GRID experience

P-GRADE

a parallelization and running environment developed
at SZTAKI (Inst. of Computational and Automation Research)
recently extended to provide GRID portal services

Operates in HUNGRID, a virtual organization within EGEE

We plan to maintain a server with a P-GRADE portal for the COMPCHEM VO

We have used P-GRADE for

parallelizing codes

quasiclassical trajectory calculations

quantum scattering: the time-independent 3-atom code
abc by Manolopoulos et al.

running our QCT code on HUNGRID

Á. Bencsura, G. Lendvay, Parallelization of reaction dynamics codes using P-GRADE: a case study, LNCS **3044**, 290-299, Springer, 2004.

Á. Bencsura, G. Lendvay, Parallelization of a quantum scattering code using P-GRADE: a case study, in: Distributed and parallel systems: clusters and grid computing, Kluwer International Series in Engineering and Computer Science, Vol, 777, Z. Juhász, P. Kacsuk, D. Kranzlmüller, Eds, pp. 121-128, Springer, 2004.

Properties of the P-GRADE Portal

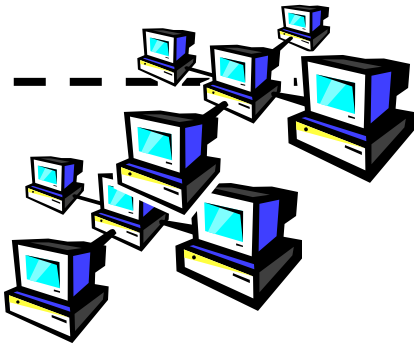
- General purpose, graphical, workflow-oriented Grid portal
  Supports the development and execution of workflow-based Grid applications

- Grid services supported by the portal
  - MyProxy – proxy credential management
  - GridFTP – file transfer
  - GT2/GT3 GRAM – job execution
  - Mercury – job monitoring
  - PROVE – workflow & job execution visualization
  - BDII and MDS-2 – obtain information about resources
  - LCG-2 broker – resource selection
  - GEMLCA – invoke legacy codes

- Support for multi-grid workflows

- GridSphere based
  - Easy to expand with new portlets
  - Easy to tailor to end-user needs

P-GRADE AS A GRID PORTAL



Workflow editor

Web browser

Portal server

Gridsphere portal framework

Workflow manager subsystem

Execution visualization subsystem (PROVE)

MyProxy

LCG-2 broker

Globus GridFTP

Globus GRAM

Mercury

BDII

Globus MDS-2

GEMLCA

HunGrid: the Hungarian virtual organization within EGEE

KFKI	SZTAKI	ELTE	CRC
▪ 250 processors ▪ 3.4 TB HD	▪ 26 processors ▪ 2 TB HD	▪ 5 processors ▪ 1.5 TB HD	▪ 12 processors ▪ 1 TB HD

Planned: Uni. Veszprém (6), Univ. Miskolc (30), Univ. Szeged (50)

Parallel and GRID application of a quasiclassical trajectory code

Embarrasingly parallel problem

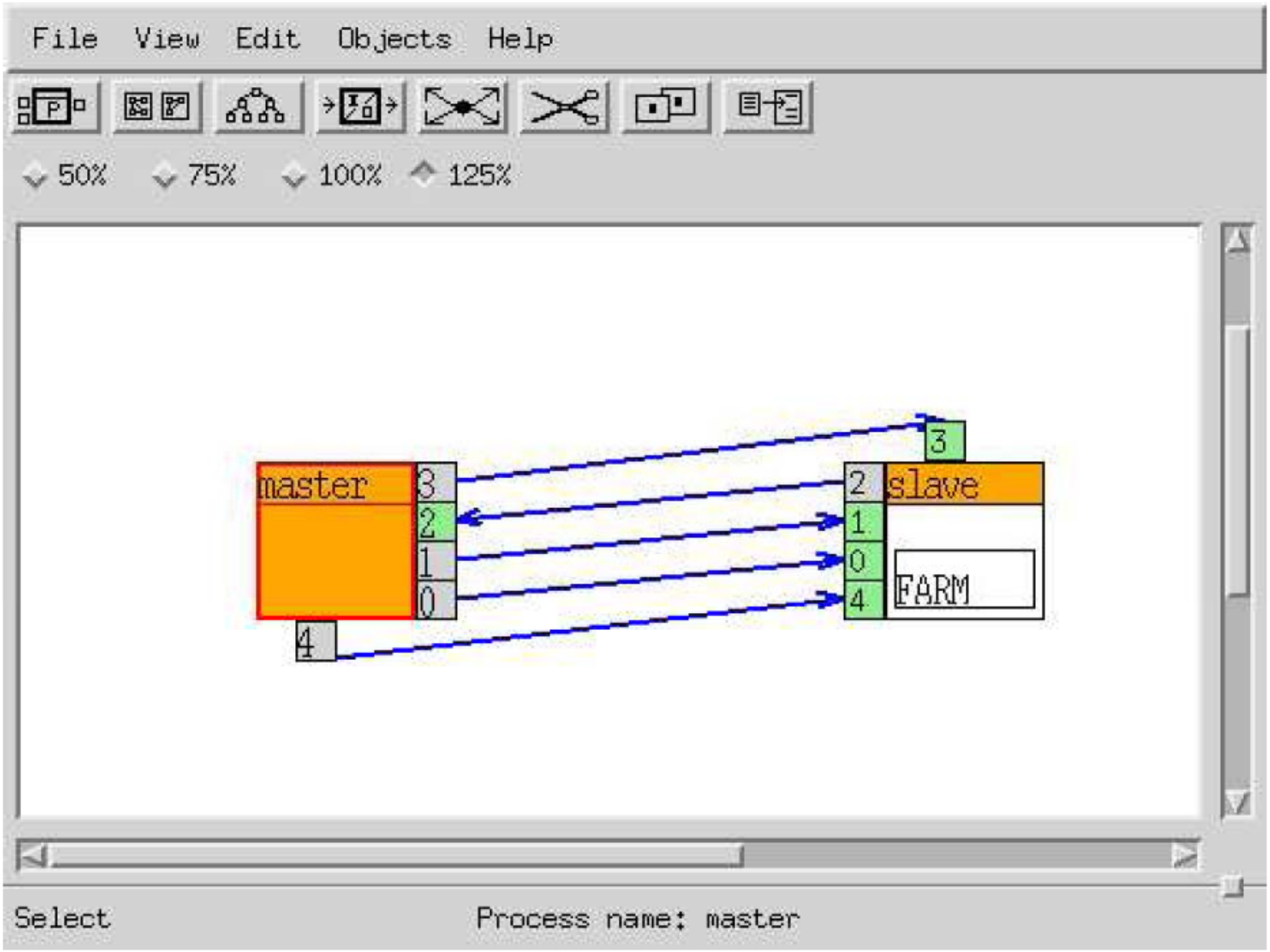
calculate rate coefficients for elementary chemical reactions
by simulating individual molecular collisions

integrate a set a ordinary differential equations starting from many different initial conditions selected by Monte Carlo sampling

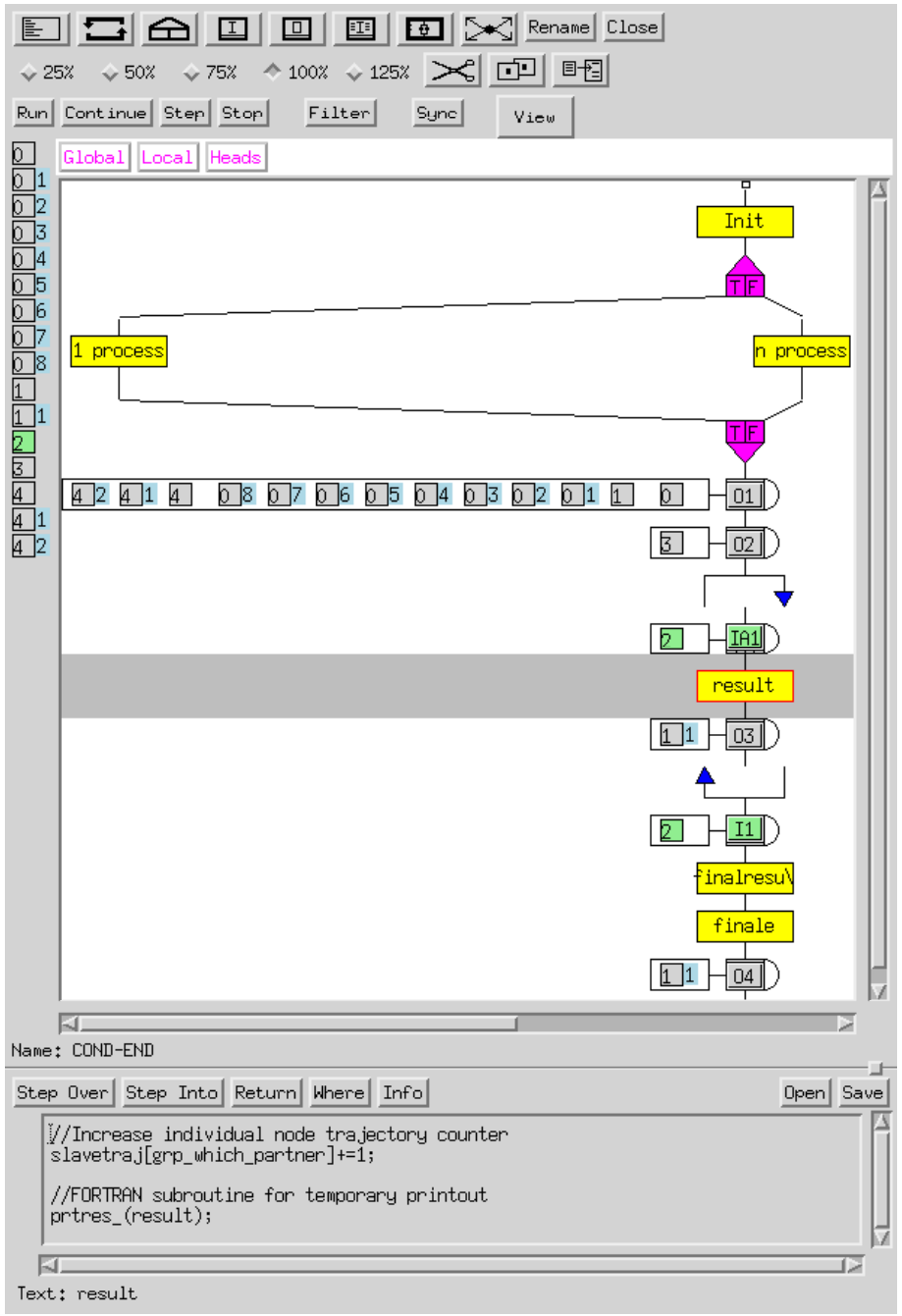
Parallel version - master-worker paradigm

master	worker
-reads system parameters	
-forwards them to workers	-receives parameters
-starts random number sequence, generates a seed for each worker	-receives seed
LOOP	
-assigns the number of trajectories to each worker	-receives # of traj.s to be done
-listens	-integrates the set of trajs.
-receives data	-sends results to master
-assigns a new set of traj.s if needed	

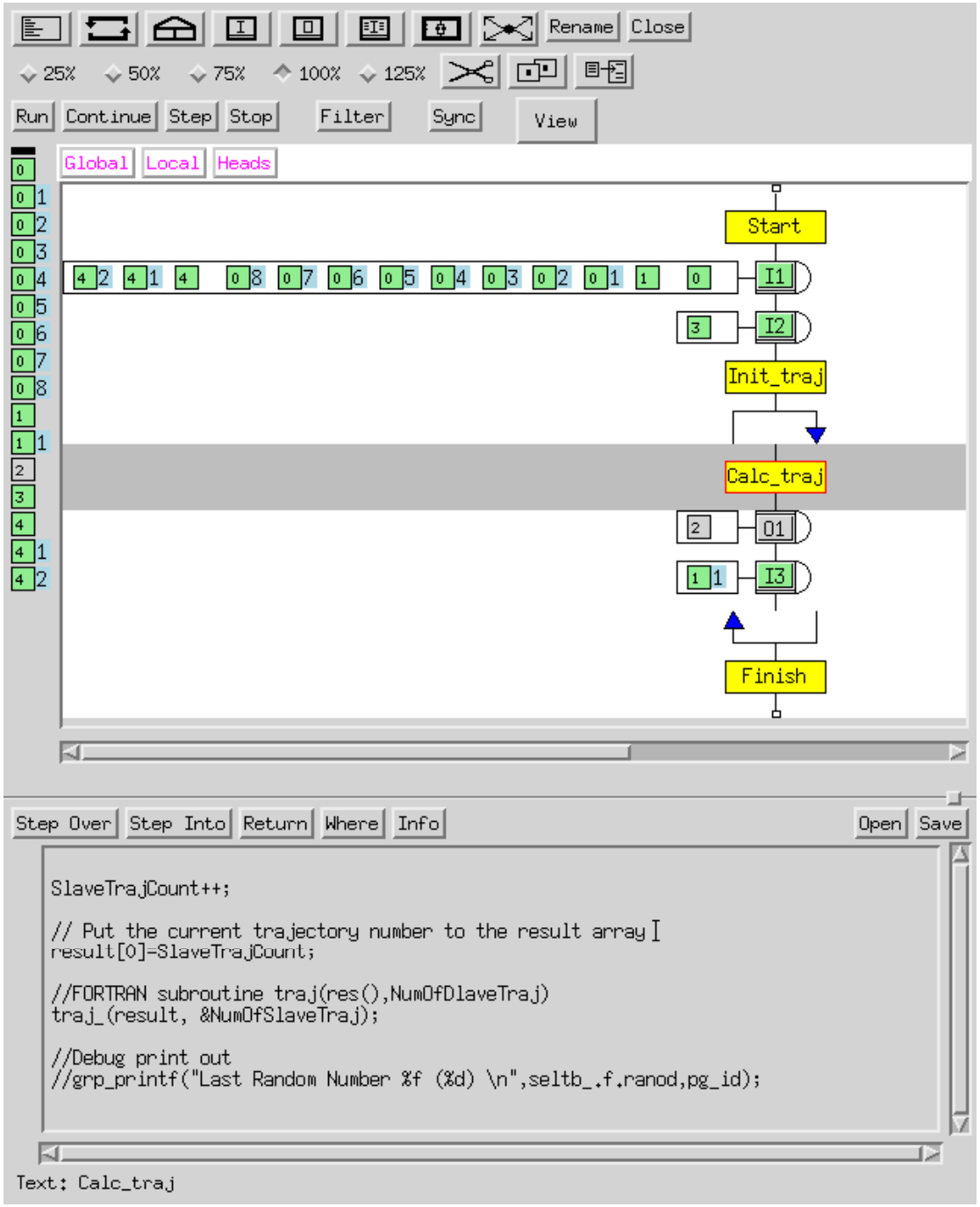
Representation of the master-worker communication in P-GRADE



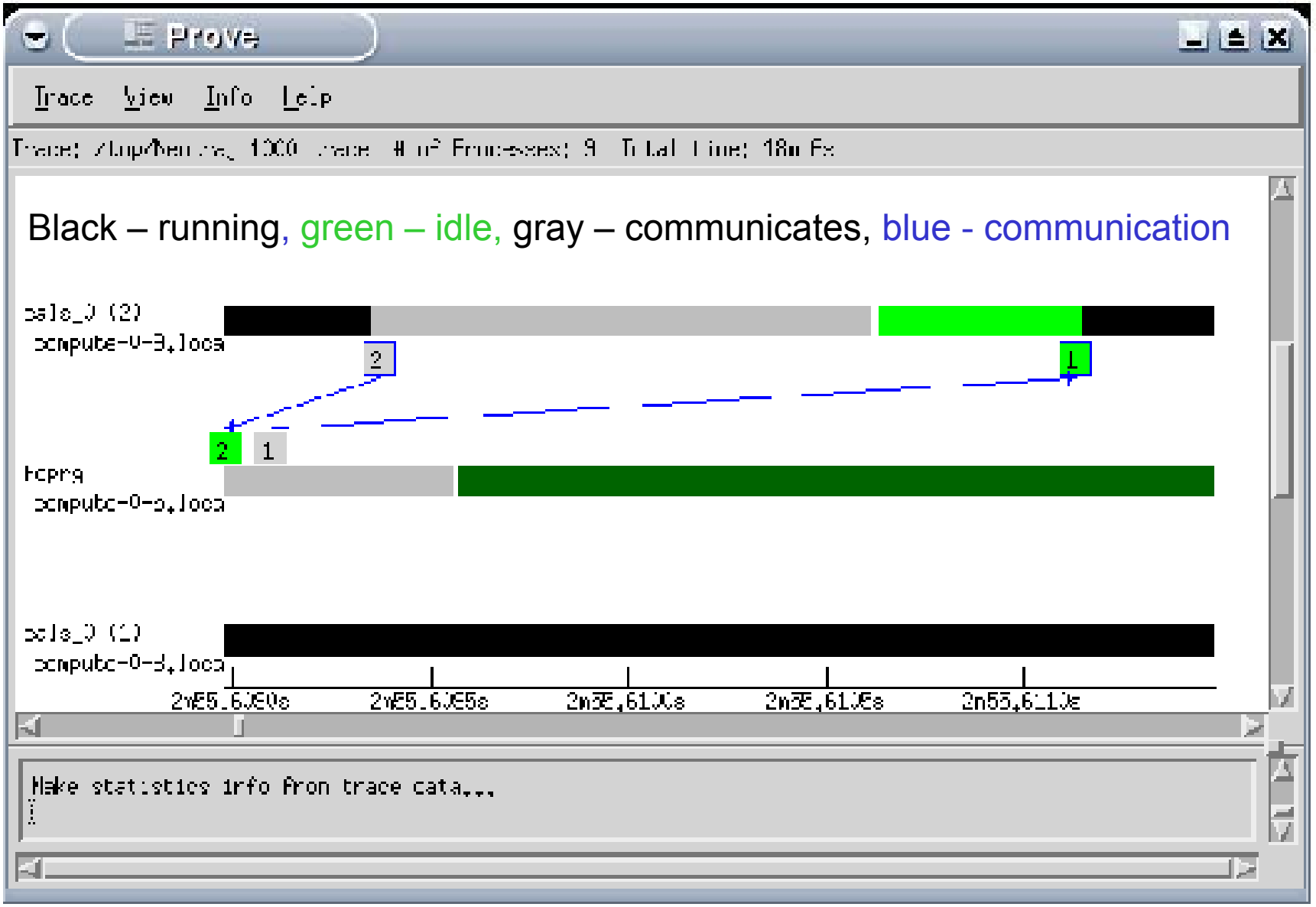
P-GRADE graphical representation of the master loop



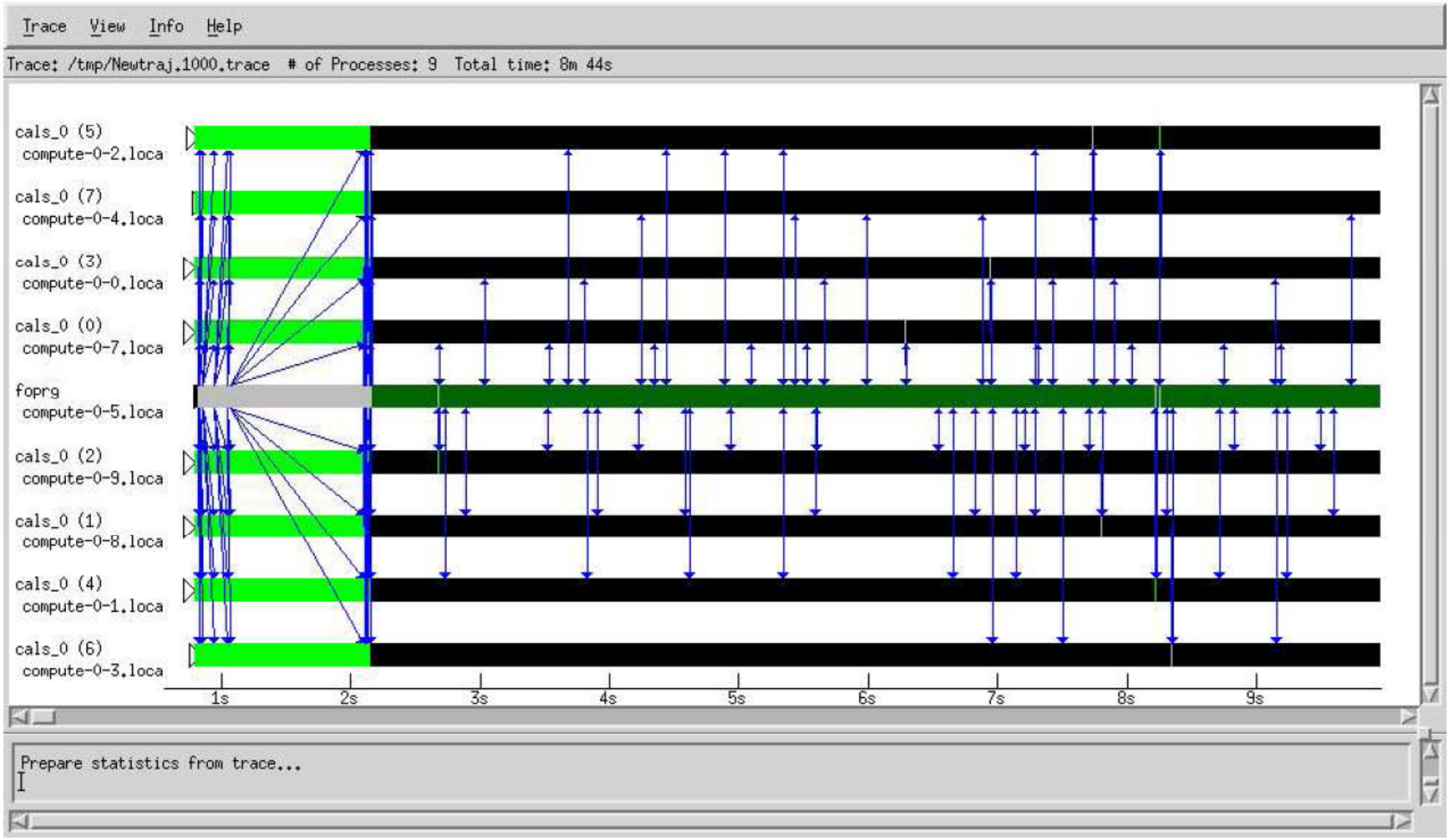
P-GRADE graphical representation of the slave process



Processor utilization and communication when the job runs

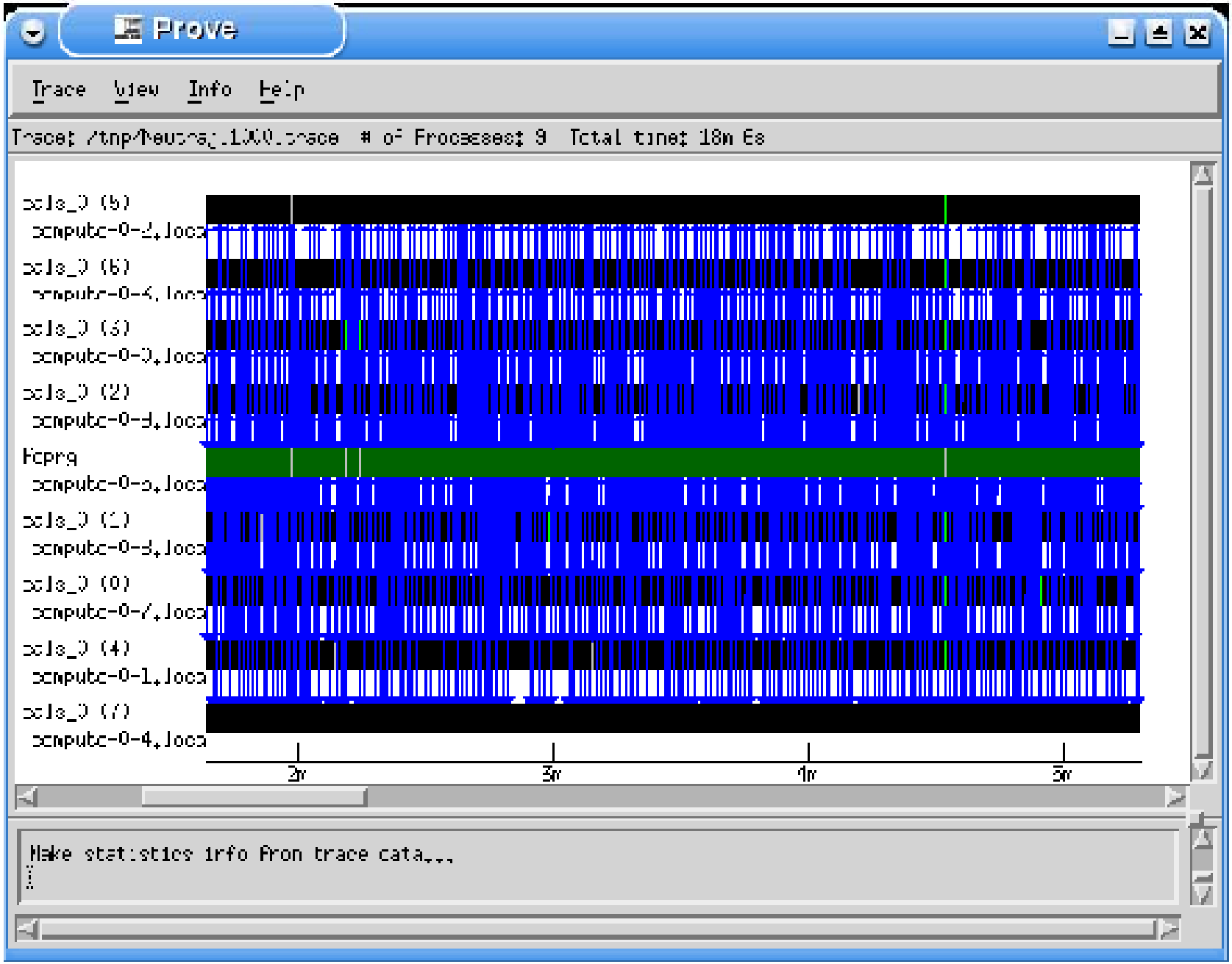


The first 10 seconds of a run on 8 processors
master is in the center row



Black – running, green – idle, gray – communicates, blue - communication

A longer time-segment



Parallel and GRID application of a QCT code

GRID version – in principle the parallel version could be submitted

more efficient (less human effort) is if several independent copies of the code are submitted

each submitted job has its own input

monitoring of progress is done manually

we have some experience now on how to design a code to do maintain progress status

Potential surfaces for reaction dynamics

Traditional way:

calculate many points, generally over a regular grid,
fit to analytical or spline-type function
Advantage: can use high-level ab initio
Disadvantage: tedious, hard to generalize to polyatomics, each channel need to be taken care of
difficult to ensure symmetry of equivalent atoms
very hard to find analytical functions for multi-channel reactions in polyatomic systems

Direct dynamics:

calculate energies and derivatives "on the fly" using electronic structure theory at every geometry the dynamical calculation requires
Advantage: automatic, no dimensional constraint, includes all reaction channels
Disadvantage: computationally demanding
- can only be done with "cheap" electronic structure theory

Possible remedy:

1. reparametrized semiempirical methods
2. introduction of corrections to match high-level ab initio

Parallel version of a triatomic quantum scattering code

(abc – time-independent method using hyperspherical coordinates)

Calculate S-matrix for a chemical reaction

$|S_{ij}|^2=P_{ij}$

Coupled-channel equations solved for many energies

set of second-order partial differential equations

log-derivative propagation

Original setup of the abc code

1 calculate basis

2 do the propagation for many energies simultaneously

a) in each sector calculate potential matrix elements

b) step the wavefunction from sector to sector

- simultaneously for each energy

Parallel version of a triatomic quantum scattering code

Modified algorithm devoted for parallel run on a cluster (is now coded)

1 set up basis

2 calculate matrix elements for each sector (done in parallel)

3 distribute potential matrix elements

4 do the propagation – parallel: each node does a certain number of energies

A simplified version has been completed

GRID version (development in progress)

separate matrix element calculation from propagation

First run - preparation

1 set up basis

2 calculate matrix elements for each sector

3 create a file of potential matrix elements

Subsequent runs – production calculation

1 set up basis

2 read potential matrix elements

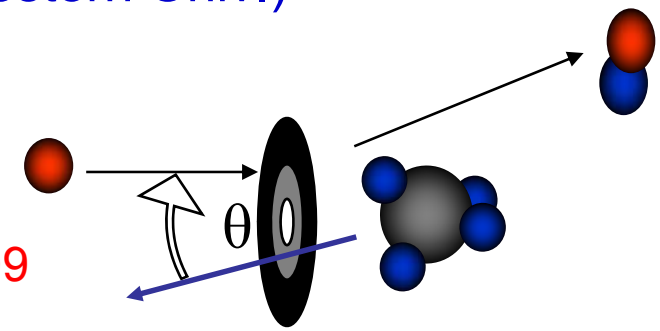
3 do the propagation – each unit does a certain number of energies

On-the-fly QCT dynamics of the H + $CD_4 \rightarrow HD + CD_3$ reaction with the Schatz group (Northwestern Univ.)

Zare lab experiments (2003-2006)

Detect preferred stripping

E_{coll} (eV)	1.2	1.95
<cos**θ**>	−0.07 ± 0.10	−0.20 ± 0.09



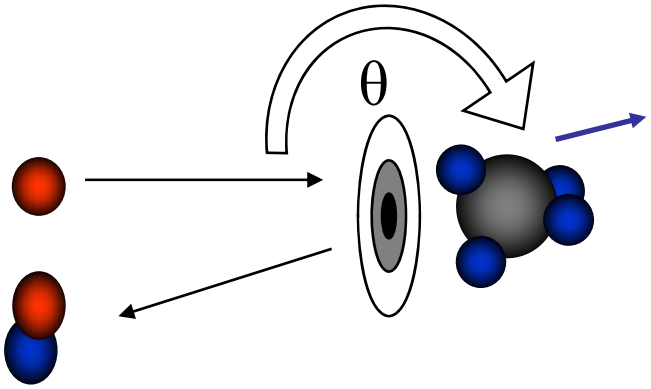
CD_3 backward scattered

Earlier trajectory studies using an empirical potential surface (1974-75)

Predicted rebound mechanism

E_{coll} (eV)	1.73	~2
<cos**θ**>	+0.28/+0.57	+0.5

Small impact parameters preferred



CD_3 forward scattered

Can we understand the difference?

Comparison of several PES helps us to reveal the connection of dynamics and features of the underlining potential surface

Potential surfaces for $H + CH_4 \rightarrow H_2 + CH_3$

Analytical PES

Jordan and Gilbert (JG): attempt to make all four H atoms equivalent – (see Manthe)

Espinoza-Garcia (EG): JG PES + high-level ab initio data

Canonical variational transition state theory to match thermal rate coefficients

centered on the minimum energy path

Numerical PES – get energy on the fly from electronic structure theory at every geometry

MSINDO1: INDO-type semiempirical quantum chemistry method (K. Jug et al.)

too high barrier

Reparametrized MSINDO1 (MSINDO1-SRP):

the atom-pair parameters of MSINDO1 are modified to the best match of the ab initio barrier height, saddle point geometry and reaction enthalpy

B3LYP/6-31G** (B3LYP):

standard electronic structure method based on density functional theory

is known to give very good equilibrium geometries for organic molecules

barrier heights are generally in error of 3-10 kcal/mol

Methods used in the quasiclassical trajectory calculations

Quasiclassical initial conditions

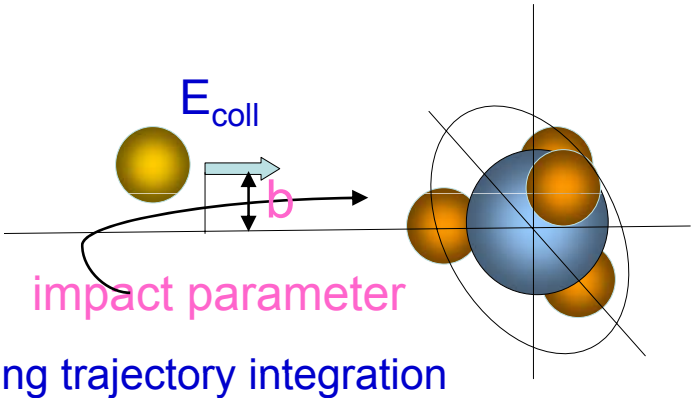
CD_4 in vibrational ground state

an intramolecular trajectory with one quantum of energy in each mode is generated starting from the harmonic normal mode assumption

when collisions are initialized, random phases of this trajectory are selected

Trajectory code: separate custom-designed codes for

1) generating initial conditions
2) trajectory integration
3) analysis of trajectory results



Direct dynamics trajectories:

electronic structure code is called at every time step during trajectory integration

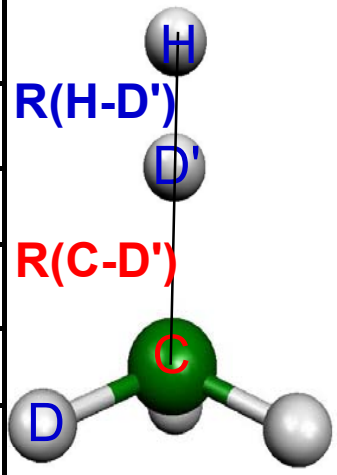
10,000 trajectories at each collision energy to get a few hundred reactive

B3LYP: 4-6 trajectory/day/processor

Calculate reactive cross sections, angular distributions, product state distributions

The properties of the saddle point on five PESs

Parameter	RCCSD(T)/ aug-cc-pVTZ[a]	B3LYP /6-31G**	EG	MSINDO	reparamet-rized MSINDO
R(C-D)/Å	1.0854	1.087	1.094	1.077	1.078
R(C-D')/Å	1.3991	1.412	1.331[b]	1.272	1.327
R(H-D')/Å	0.8970	0.894	0.931[b]	0.932	0.834
<DCD'/°	103.12	103.4	107.4	107.1	106.0
Imaginary frequency /cm^{-1}	1437i	1132i	1293i	2261i	720i
ΔE_{barr}/kcal mol^{-1}	14.78	9.4	12.9	28.6	15.4
ΔE_{react}/kcal mol^{-1}	3.0	1.9	2.8	–1.1	2.8

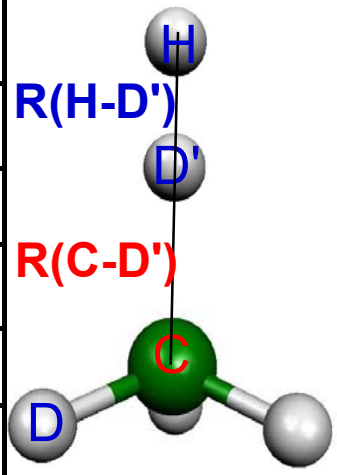


[a]Bowman et al, J.Chem.Phys. 124, 021104 (2006)

Barrier is late, C-D'-H collinear – vibrationally unexcited products expected

The properties of the saddle point on five PESs

Parameter	RCCSD(T)/ aug-cc-pVTZ[a]	B3LYP /6-31G**	EG	MSINDO	reparamet-rized MSINDO
R(C-D)/Å	1.0854	1.087	1.094	1.077	1.078
R(C-D')/Å	1.3991	1.412	1.331[b]	1.272	1.327
R(H-D')/Å	0.8970	0.894	0.931[b]	0.932	0.834
<DCD'/°	103.12	103.4	107.4	107.1	106.0
Imaginary frequency /cm^{-1}	1437i	1132i	1293i	2261i	720i
ΔE_{barr}/kcal mol^{-1}	14.78	9.4	12.9	28.6	15.4
ΔE_{react}/kcal mol^{-1}	3.0	1.9	2.8	–1.1	2.8



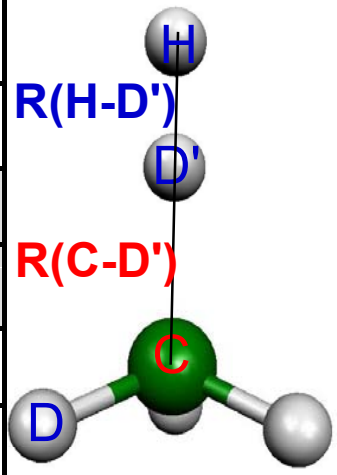
[a]Bowman et al, J.Chem.Phys. 124, 021104 (2006)

Barrier is late, C-D'-H collinear – vibrationally unexcited products expected

Both the EG and B3LYP barrier is too low, B3LYP is worse – enhanced reactiv

The properties of the saddle point on five PESs

Parameter	RCCSD(T)/ aug-cc-pVTZ[a]	B3LYP /6-31G**	EG	MSINDO	reparamet-rized MSINDO
R(C-D)/Å	1.0854	1.087	1.094	1.077	1.078
R(C-D')/Å	1.3991	1.412	1.331	1.272	1.327
R(H-D')/Å	0.8970	0.894	0.931	0.932	0.834
<DCD'/°	103.12	103.4	107.4	107.1	106.0
Imaginary frequency /cm^{-1}	1437i	1132i	1293i	2261i	720i
ΔE_{barr}/kcal mol^{-1}	14.78	9.4	12.9	28.6	15.4
ΔE_{react}/kcal mol^{-1}	3.0	1.9	2.8	–1.1	2.8



[a]Bowman et al, J.Chem.Phys. 124, 021104 (2006)

Barrier is late, C-D'-H collinear – vibrationally unexcited products expected

Both the EG and B3LYP barrier is too low, B3LYP is worse – enhanced reactivity?

The location of the barrier is better on B3LYP

CD_3 angular distributions at two collision energies

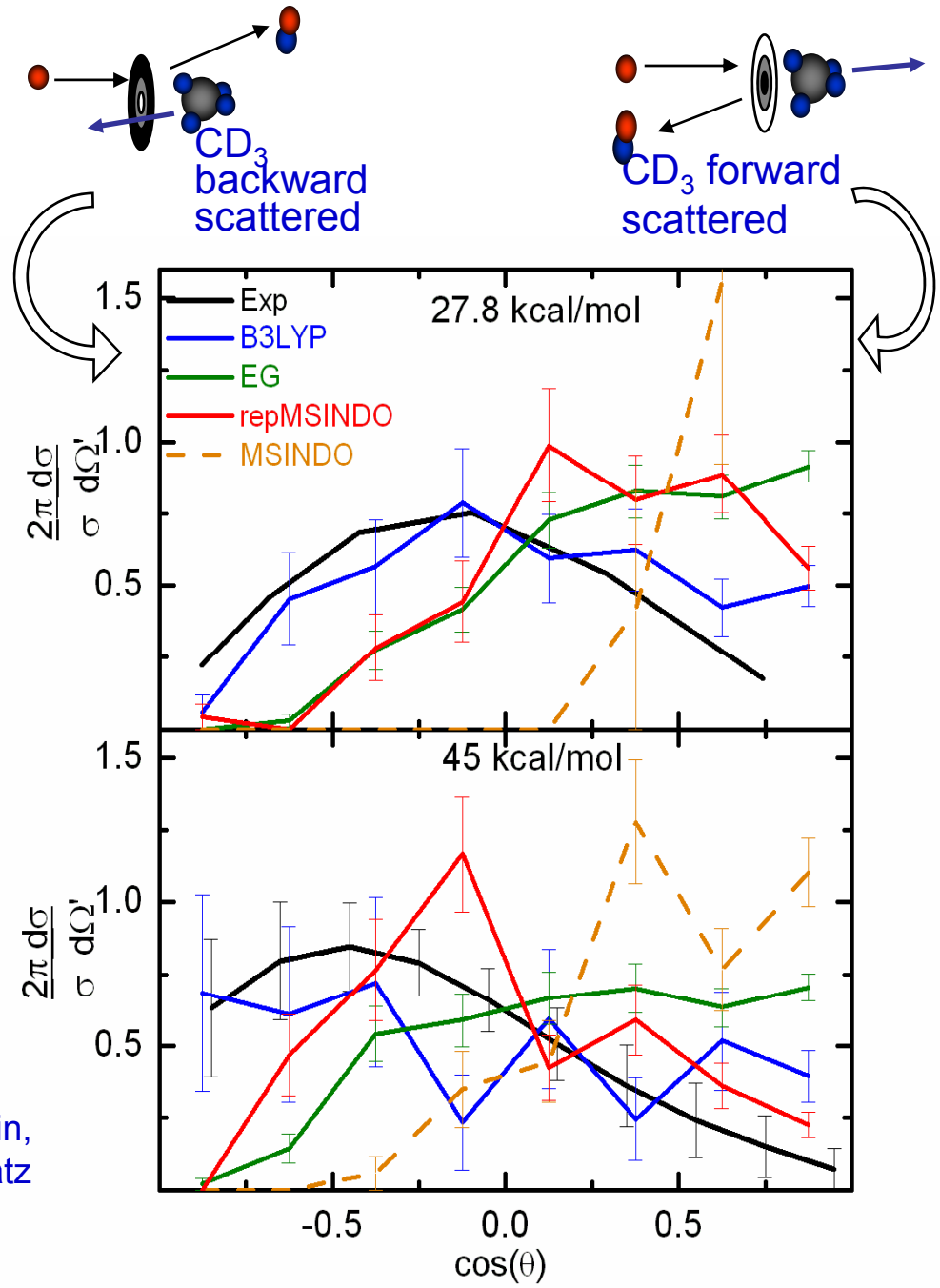
Experiment
sideways and backward scattering

EG
mostly forward scattered

B3LYP
close to experiment

MSINDO
badly fails

repMSINDO is better



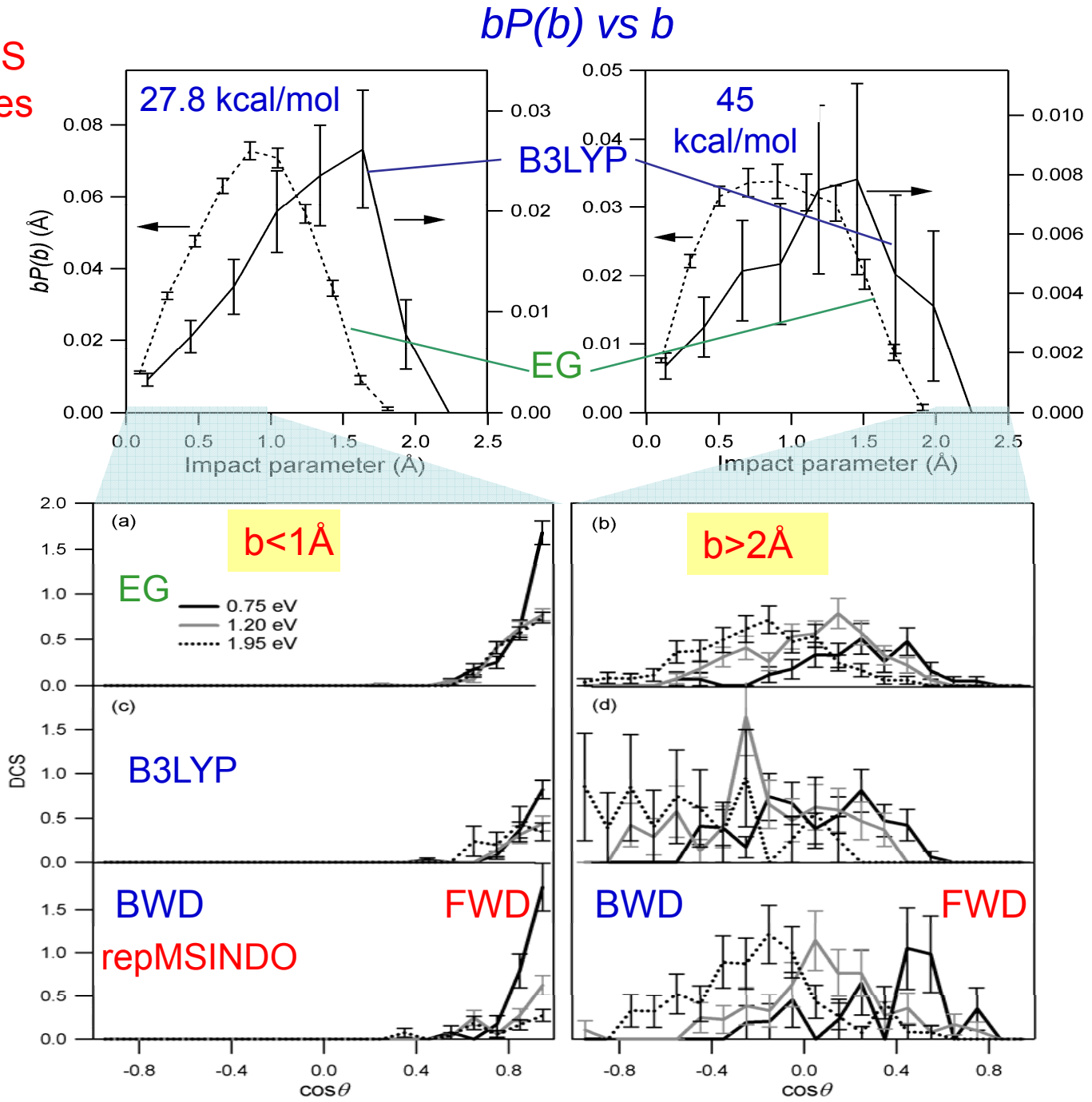
J. P. Camden, H.A. Bechtel, D.J. Brown, M.R. Martin, R. N. Zare, W. Hu,G. Lendvay, D. Troya, G.C. Schatz
J. Am. Chem. Soc. **127**, 11898 (2005).

OPACITY FUNCTIONS
at two collision energies

B3LYP:
large-impact
parameter collisions
dominate

Angular distributions
at small and large
impact parameters

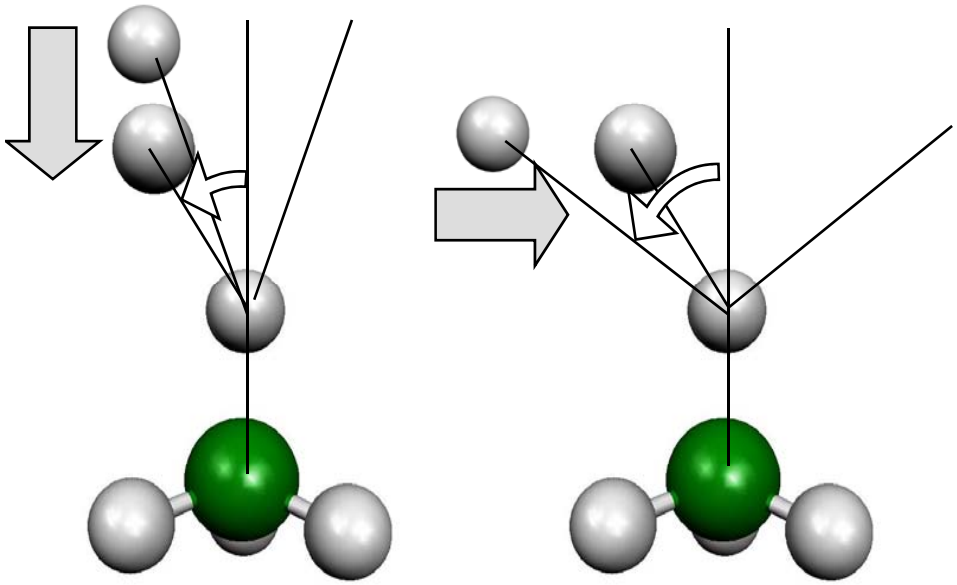
Small b:
forward scattering
Large b:
sideways
and on B3LYP
backward scattering



Reactive collisions at large impact parameters and the cone of acceptance

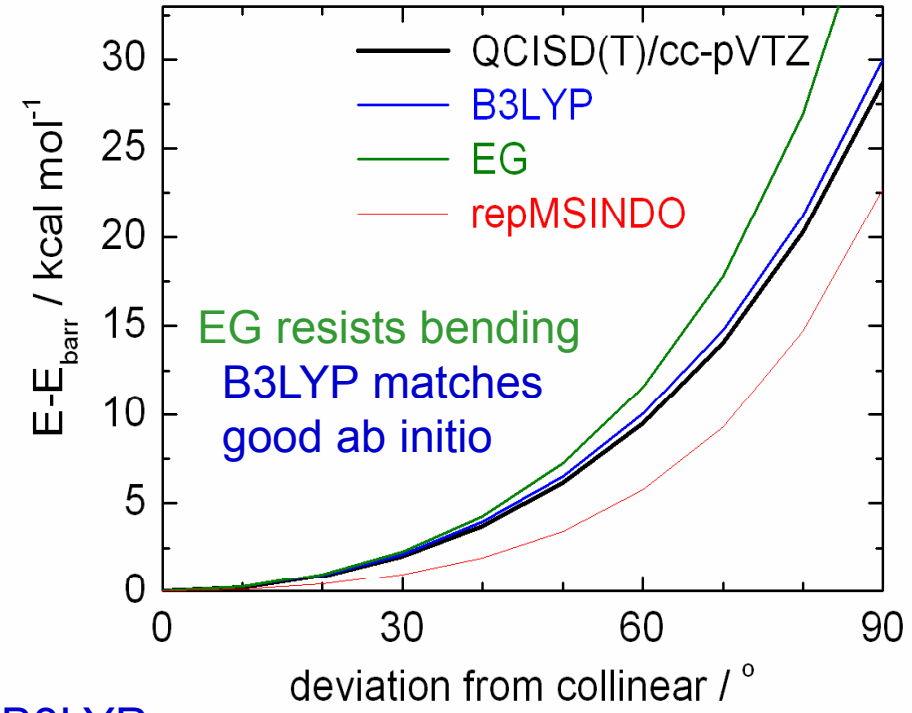
small impact parameters: the H-D'-C angle remains close to 180°

large impact parameters: approach via large deviation from collinear



small-angle approach needs to be allowed

Potential energy change when the H-D'-C angle changes at the saddle point



B3LYP allows bent approach

On EG H can proceed only close to the collinear "channel"

W. Hu, G. Lendvay, D. Troya, G.C. Schatz, J. P. Camden, H.A. Bechtel, D.J. Brown, M.R. Martin, R. N. Zare, J. Phys. Chem. A110, 3017 (2006).

SUMMARY

Codes we have

- trajectory code for grid
  - can do direct dynamics if the electronic structure code is available on the grid
- parallel triatomic quantum scattering code
  - has run parallel on a cluster

Experience we have

- code parallelization
  - for clusters
  - for supercomputers
  - for grid
- use P-GRADE for all this
- EGEE software

What we plan to do

- routine connection to the COMCHEM portal
- install the most recent LCG2 software
- set up a P-GRADE server
- adapt a chemical reaction mechanism reduction code on the grid

Acknowledgments

Parallel, grid and computer issues
Ákos Bencsura
Institute of Structural Chemistry, Hungarian Academy of Sciences

Direct dynamics QCT project
Wenfang Hu, Diego Troya, and George C. Schatz
Northwestern University

Dick Zare and his group – experiments and motivation