Large-scale quantum Monte Carlo electronic structure calculations on the EGEE grid

Antonio Monari Anthony Scemama Michel Caffarel

Laboratoire de Chimie et Physique Quantiques / IRSAMC,
Toulouse (France)
antonio.monari@irsamc.ups-tlse.fr
scemama@irsamc.ups-tlse.fr
caffarel@irsamc.ups-tlse.fr
http://qmcchem.ups-tlse.fr

14 May 2010

Large-scale QMC calculations on the EGEE grid

What is Quantum Monte Carlo (QMC) ?

Stochastic approach to solve the Schrödinger equation

$$\mathcal{H}\Psi = E\Psi \tag{1}$$

- The highest precision for total energies
- Favorable scaling
- Marginal in quantum chemistry : prohibitive CPU time to reach the “chemical accuracy” for large systems
- But :
  - Highly parallelizable
  - Little communication
  - Small RAM and I/O requirements
- $\Longrightarrow$ GRID !

Technical details

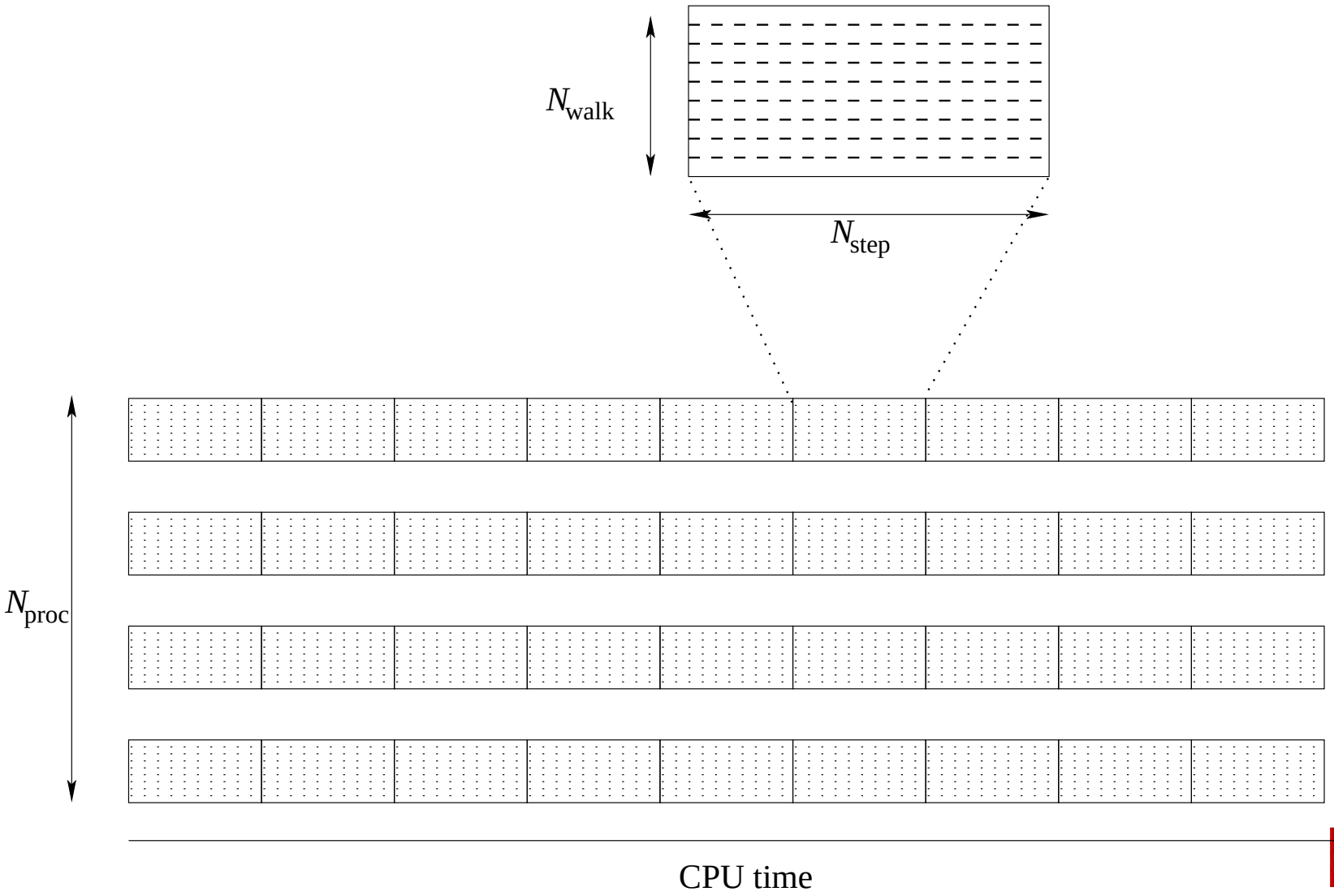
- **Walker**: vector $\mathbf{X} \in \mathbb{R}^{3N}$, coordinates of the N electrons
- **Trial wave function**: $\Psi_T(\mathbf{X})$ needed to produce a target probability density $\Pi(\mathbf{X})$
- **Markov chain Monte Carlo algorithm** : sample $\mathbb{R}^{3N}$ according to $\Pi(\mathbf{X})$
- **Energy** E: stochastic average of the local energy $E_L(\mathbf{X})$:

$$E = \langle E_L \rangle_\Pi \qquad E_L(\mathbf{X}) = \frac{\mathcal{H}\Psi_T(\mathbf{X})}{\Psi_T(\mathbf{X})} \tag{2}$$

- The energy can be calculated by averaging stochastic averages computed over independent blocks $\Longrightarrow$ Different initial conditions:
  - Initial positions of the walkers $\mathbf{X}_0$
  - Random seeds S_0

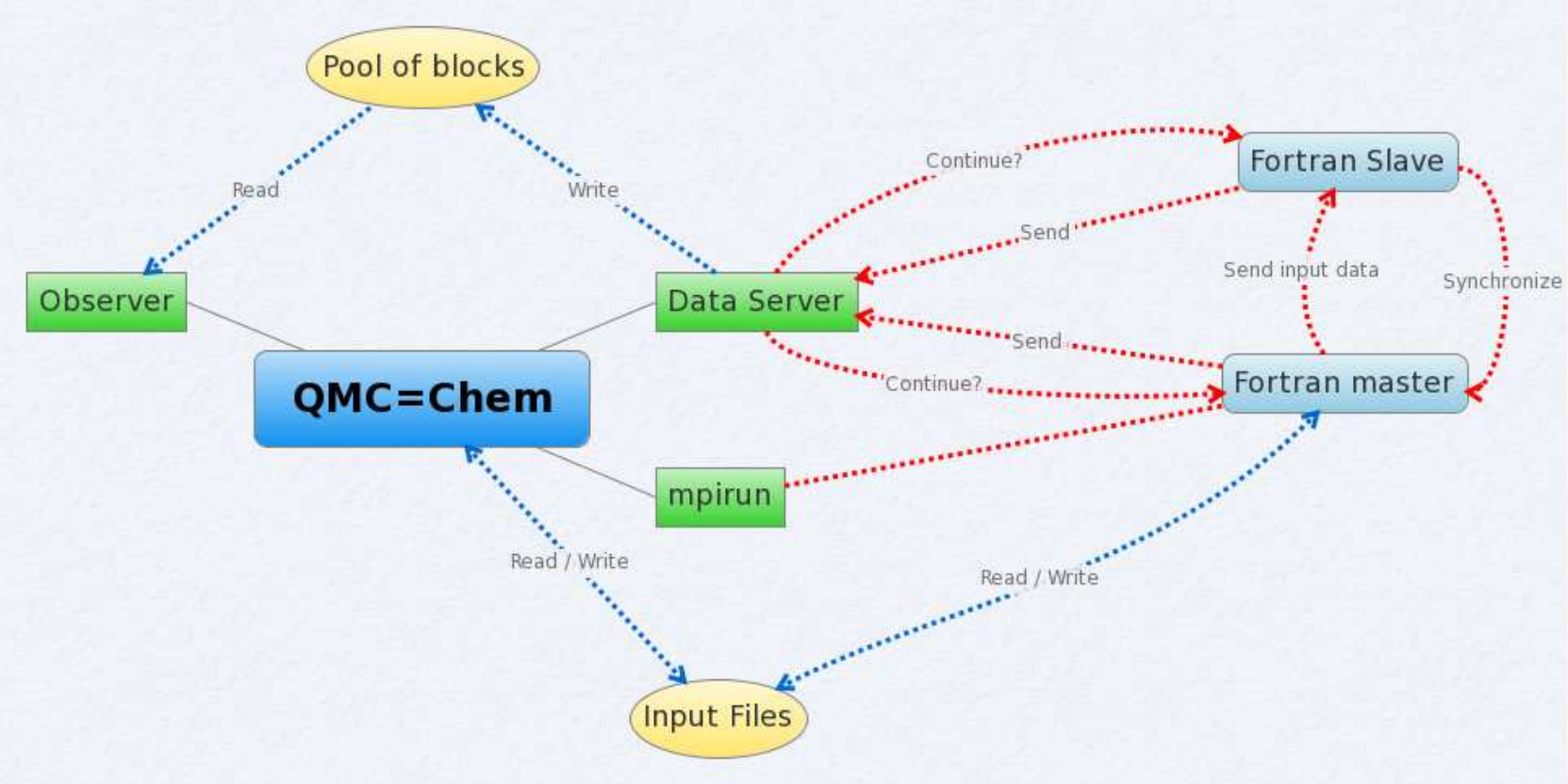
Large-scale QMC calculations on the EGEE grid

QMC=Chem produces and analyzes blocks



Design

- Fortran code which computes blocks
- Python script which forks to:
  - **Observer**: knows the current state of the calculation (running average, current number of computed blocks, . . .)
  - **Computation engine**: drives the Fortran executable
  - **Data server**: keeps the currently computed data
- The results are analyzed by post-processing the blocks


Pool of blocks
Read
Write
Observer
Data Server
Continue?
Fortran Slave
Send
Send input data
Synchronize
QMC=Chem
Send
Continue?
Fortran master
mpirun
Read / Write
Read / Write
Input Files

Adaptation to the grid

Each compute node receives:

- The static mono-processor Fortran executable
- A unique set of initial conditions $(\mathbf{X}_0, S_0)$

Each compute node sends back:

- The produced pool of blocks

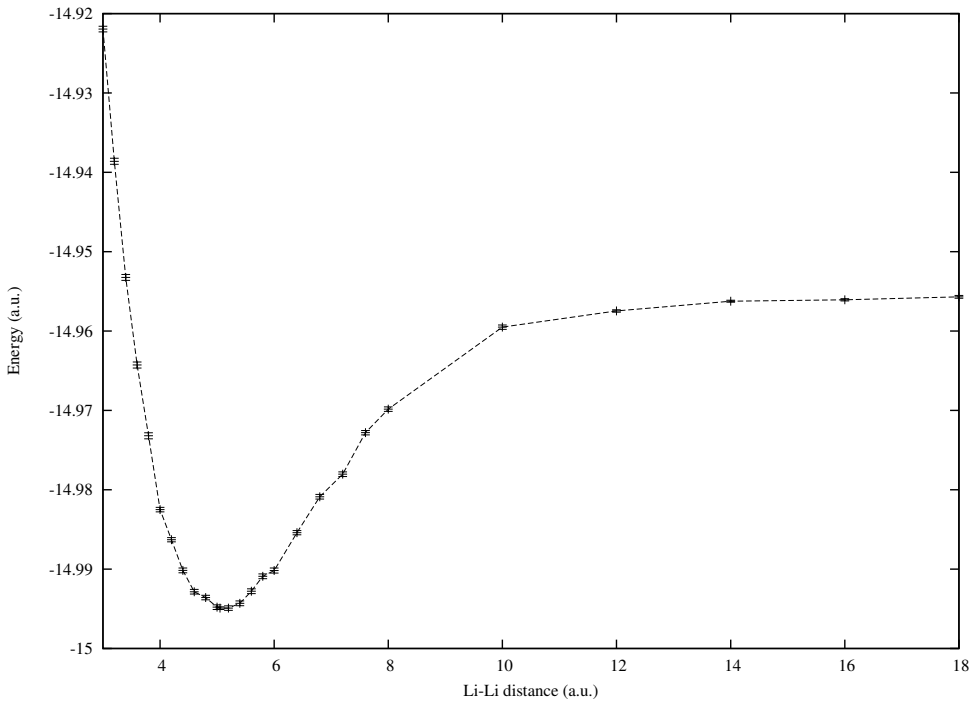
The results are post-processed by averaging the blocks.

Large-scale QMC calculations on the EGEE grid

QMC Li_2 potential energy curve

- Statistical error : $\sim 0.003\%$
- Dissociation energy of Li_2 (atomic units):

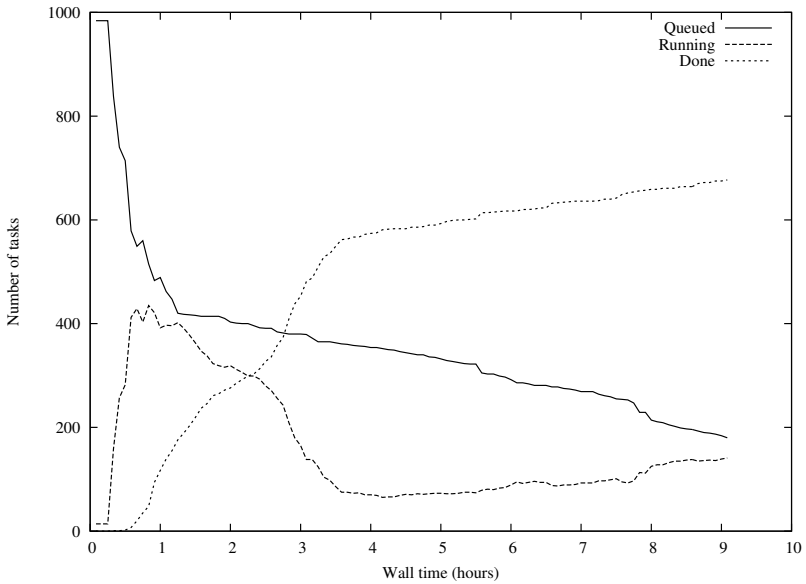
Experimental	QMC
-0.03928	-0.0395 ± 0.0002



Grid performance

For each Li-Li distance (32 points):

- 1000 submitted tasks
- Stopping condition: Wall time > 1 hour
- $\sim$ 700-800 completed jobs



Conclusions

- Distributed computation for each Li–Li distance
- Efficient exploitation of the grid architecture:
  - Independent calculations (QMC)
  - Code Flexibility
  - Very few needed resources (RAM,I/O,network)
  - Post-processing
- The most accurate Li-Li dissociation curve up to date
  - Toward accurate treatement of complex chemical systems

Acknowledgments

- EGEE-III grid organization
- CompChem virtual organization
- COST D37 “GridChem”
- PICS action 4263