Conceptual Aspects of

Parallel Electronic Structure Codes

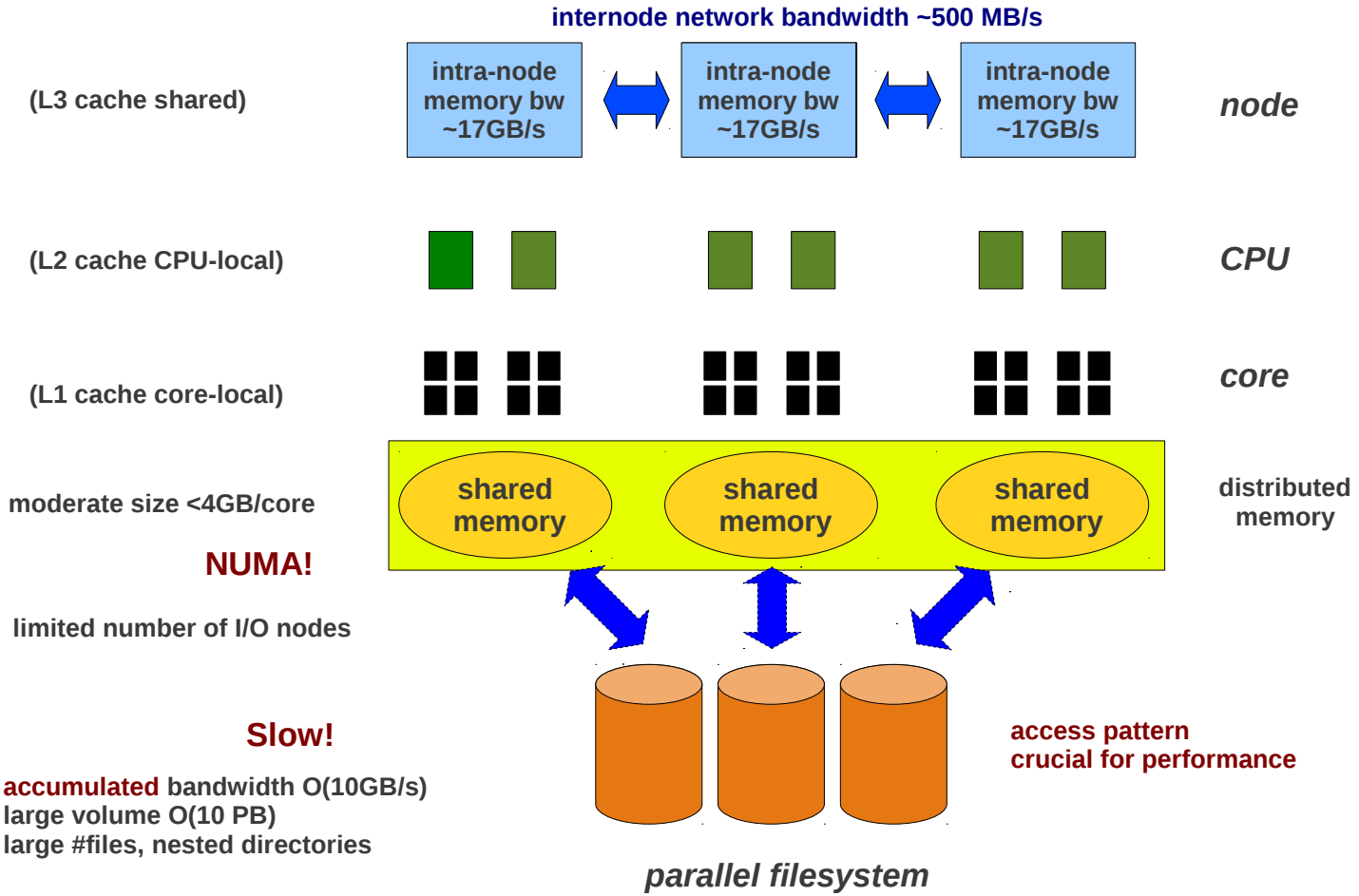
Thomas Müller
Jülich Supercomputer Centre

Mai, 8, 2013

Overview

- Major limitations of current hardware
- Programming Models & Elements
- Characteristics of Parallel Electronic Structure Codes
- Summary

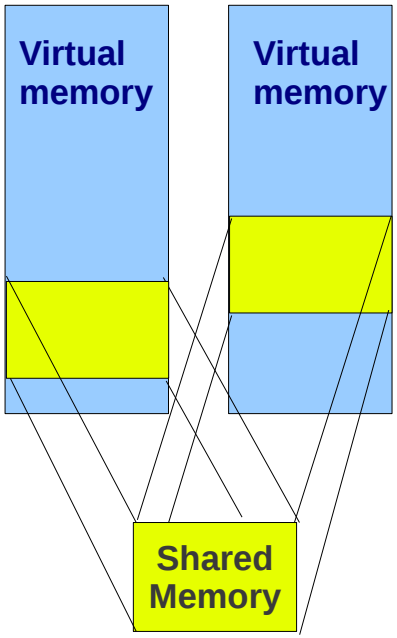
Cluster Configuration (qualitatively)



Processes versus Threads

process-based

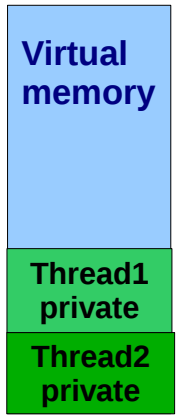
all data private
shared memory public
NOT necessarily redundant



Control:
SMP: all kernel utils through system libs
Inter-node: MPI, TCP/IP
large overhead for process creation

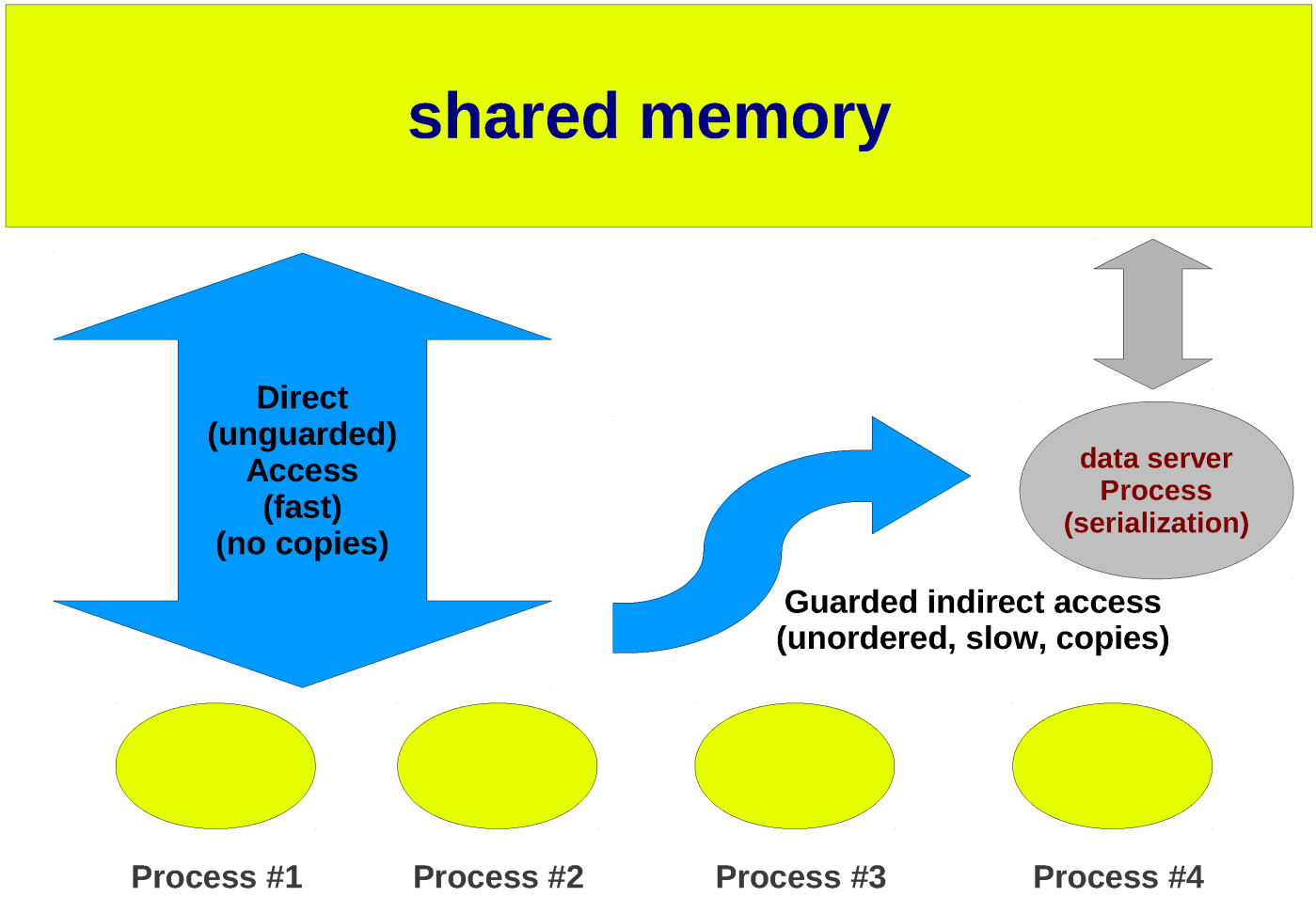
thread-based

all data public
thread private data
always kept redundant

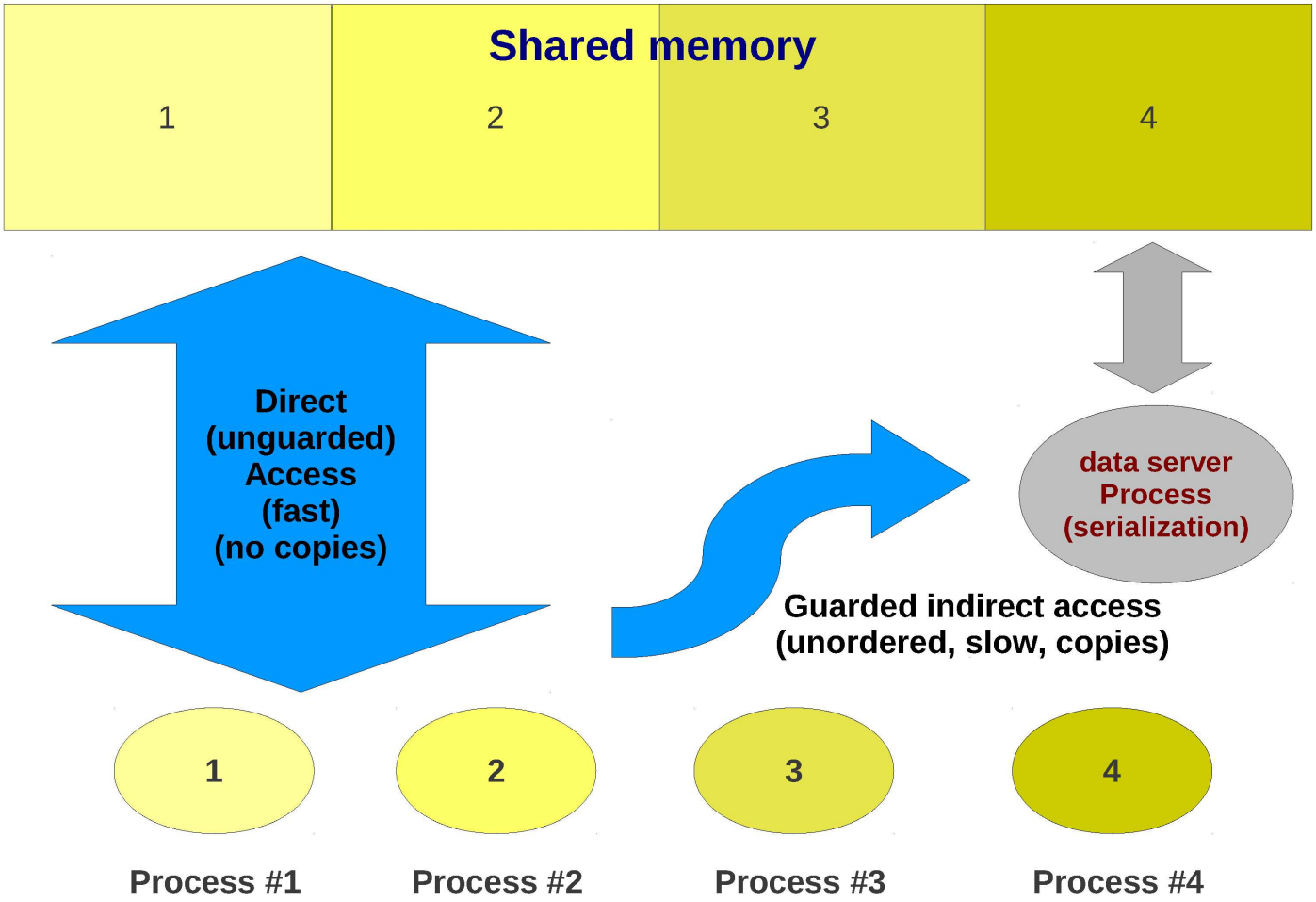


Control:
SMP: (kernel utils through system libs)
pthreads, OMP
small overhead for thread creation

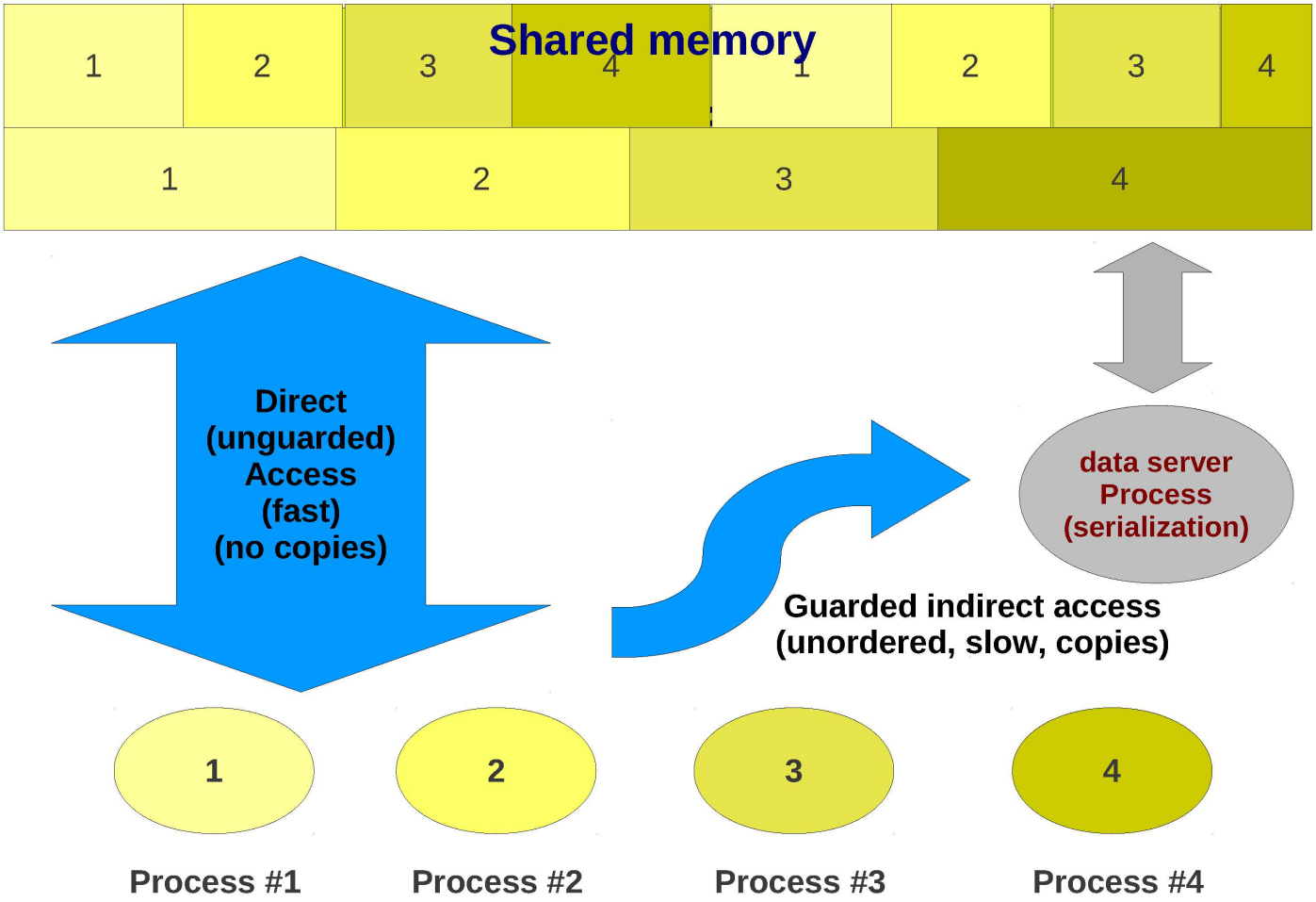
Shared-Memory Access



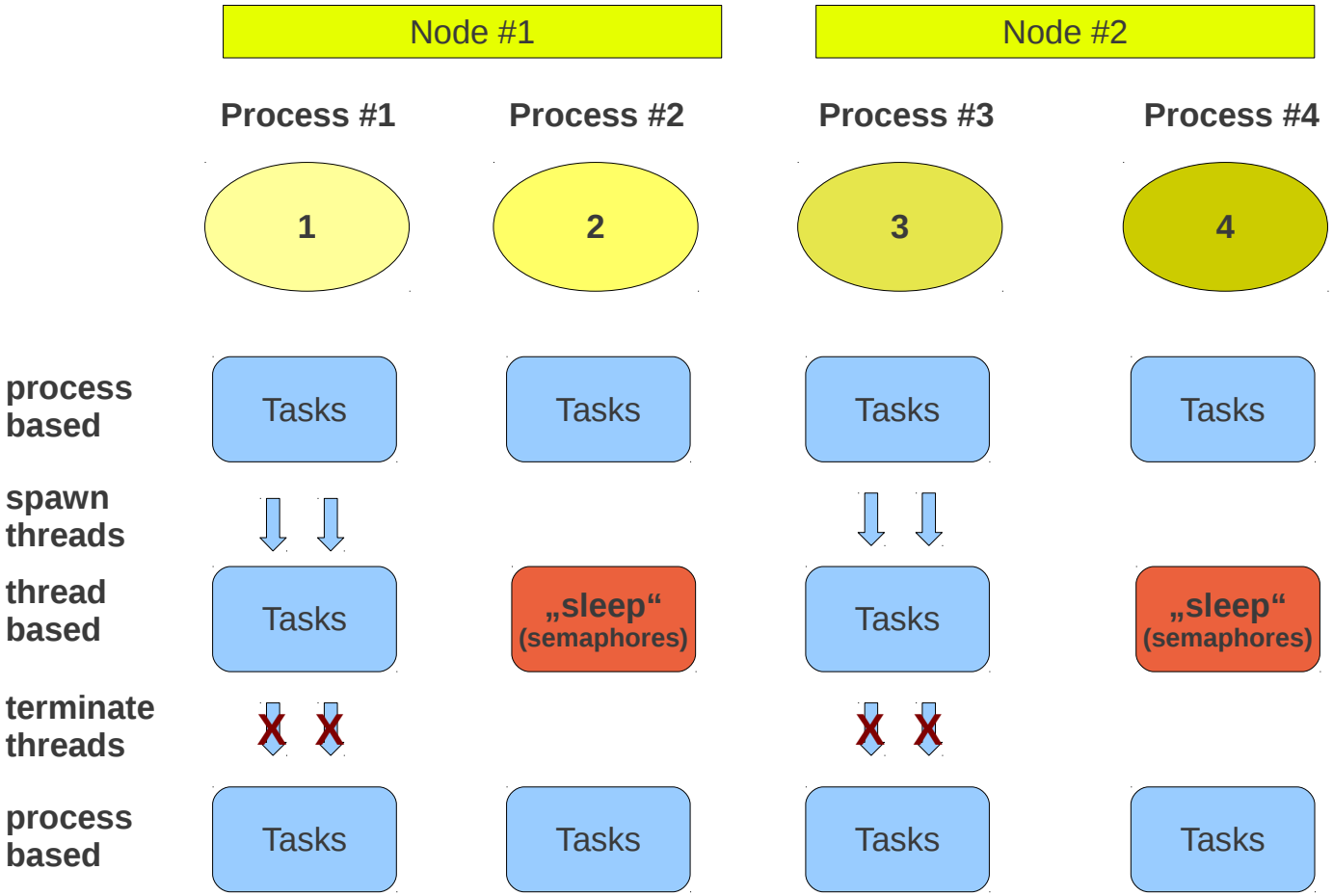
Shared-Memory Access



Shared-Memory Access

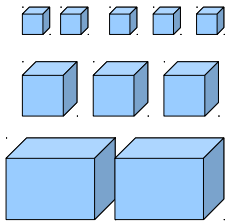


Hybrid-Parallelization

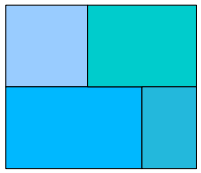


Load-Balancing

constraints:
- *** memory consumption**
- *** data locality**
- *** load balance issues**
- *** task granularity**
- *** counter saturation**



enumerated, ordered tasks



data/domain decomposition

dynamic assoc. task-proc

(multiple) shared counters

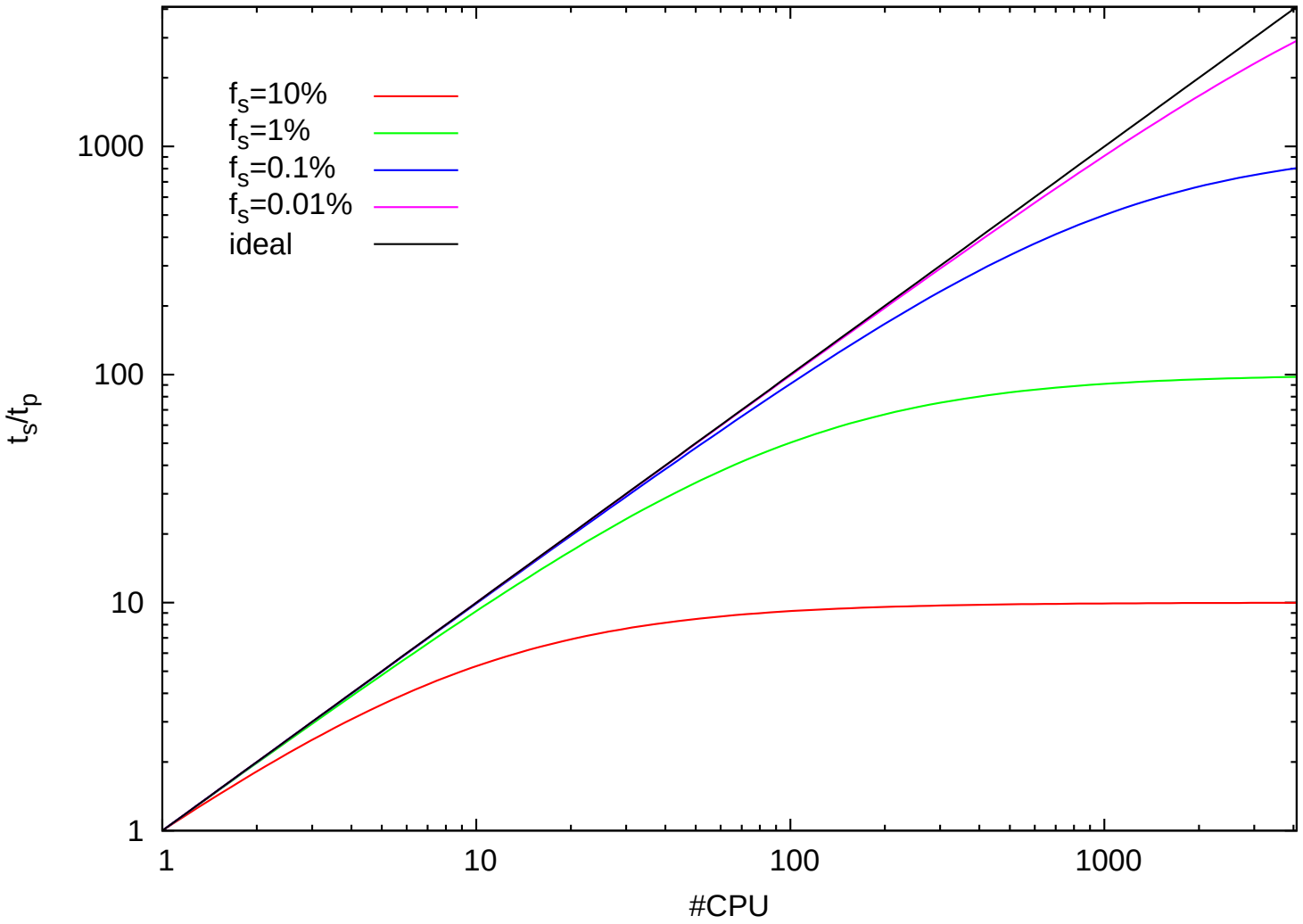
get ticket & retrieve task info

static assoc. task-proc

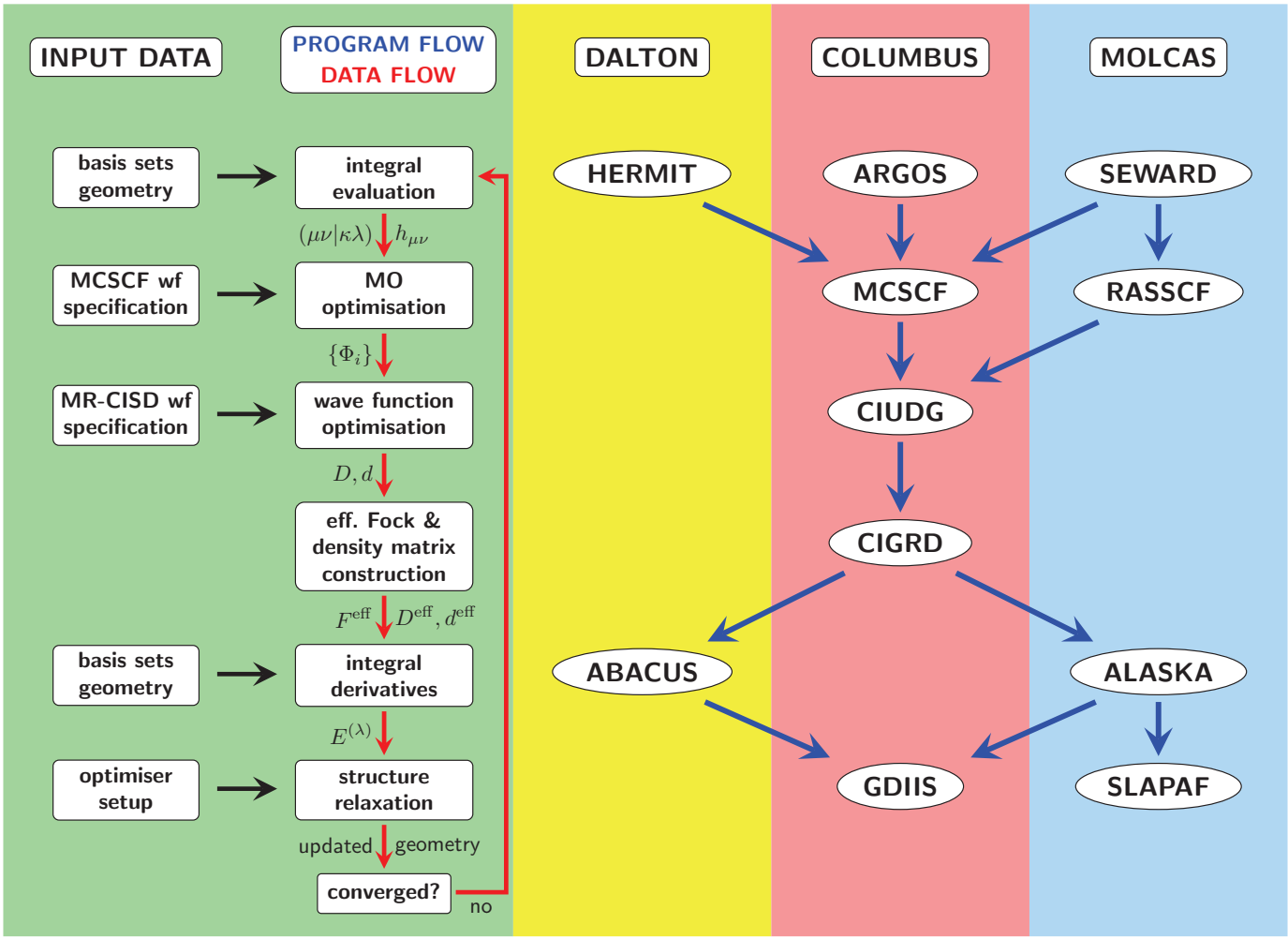
1 **2** **3** **4**

Process #1 **Process #2** **Process #3** **Process #4**

Amdahl's law



Interoperability of Quantum Chemistry Codes



Interoperability of Quantum Chemistry Codes

Classes of data interchange:

- small data
  - transferable: coordinates, gradients ...
  - less transferable: molecular orbitals
- large data
  - transferable: integrals, density matrices
  - less transferable: correlated wavefunctions

Molecular orbitals & integrals depend on
the (non-unique) definition of the basis functions
(e.g. phase, ordering, symmetry treatment, normalization)

Correlated wavefunctions are even less transferable
(determinants vs. CSFs, implicit ordering ...)

QM/MM: Electrostatic Embedding

- The interaction with the environment is taken into account through a modified Hamilton operator

$$
\begin{aligned}
E^{el}(I,O) &= E^{el}_{QM}(I,O) \\
\mathcal{H}^{el}_{IO} &= \sum_{A\in O}\sum_{B\in I}\frac{Z_A Z_B}{r_{AB}} - \sum_{A\in O}\sum_{i}\frac{Z_A}{r_{iA}} \\
&= V'_{nn} + V'_{ne}
\end{aligned}
$$

- The gradients evaluated in standard procedures with the modified Hamiltionian require an additional term

$$
E^{(x)}_{I,O} = Tr(\mathbf{D}V'^{(x)}_{ne}) + V'^{(x)}_{nn}
$$

- Non-adiabatic coupling vectors are corrected by

$$
h^{IJ}_{I,O} = Tr(\mathbf{D^{IJ}}V'^{(x)}_{ne})
$$

Electronic Structure Problem

N-particle expansion of the wavefunction

$$\Psi^{exact}(\tau_1, \tau_2 \ldots \tau_N) = \sum_{i=1}^{\infty} c_i \Psi_i(\tau_1, \tau_2 \ldots \tau_N)$$

Orthonormality:

$$\langle \Psi_i(\tau_1, \tau_2 \ldots \tau_N) | \Psi_j(\tau_1, \tau_2 \ldots \tau_N) \rangle = \delta_{ij}$$

Antisymmetry (fermions):

$$\Psi_i(\tau_1, \tau_2 \ldots \tau_N) = -\Psi_i(\tau_2, \tau_1 \ldots \tau_N)$$

Analytical Wavefunction Expansion

N-particle trial functions

Determinants: antisymmetrized product 1-particle functions

$$\Psi_i(\tau_1, \tau_2 \ldots \tau_N) = \frac{1}{\sqrt{N!}} \mathcal{A} \prod \phi_1(\tau_1), \phi_2(\tau_2), \ldots, \phi_n(\tau_N)$$

CSFs: linear combinations thereof to obtain eigenfunctions of $\mathbf{S}$, $\mathbf{S}_z$

1-particle trial functions

LCAO (linear combination of atomic orbitals) approximation

$$\phi_i(\mathbf{r}) = \sum_{\mu} C_{\mu,i} \chi_{\mu}(\mathbf{r})$$

Wavefunction degrees of freedom

Molecular orbitals

Given a set of n atomic orbitals (basis set), the 1-particle expansion (or molecular orbital) coefficients are optimized

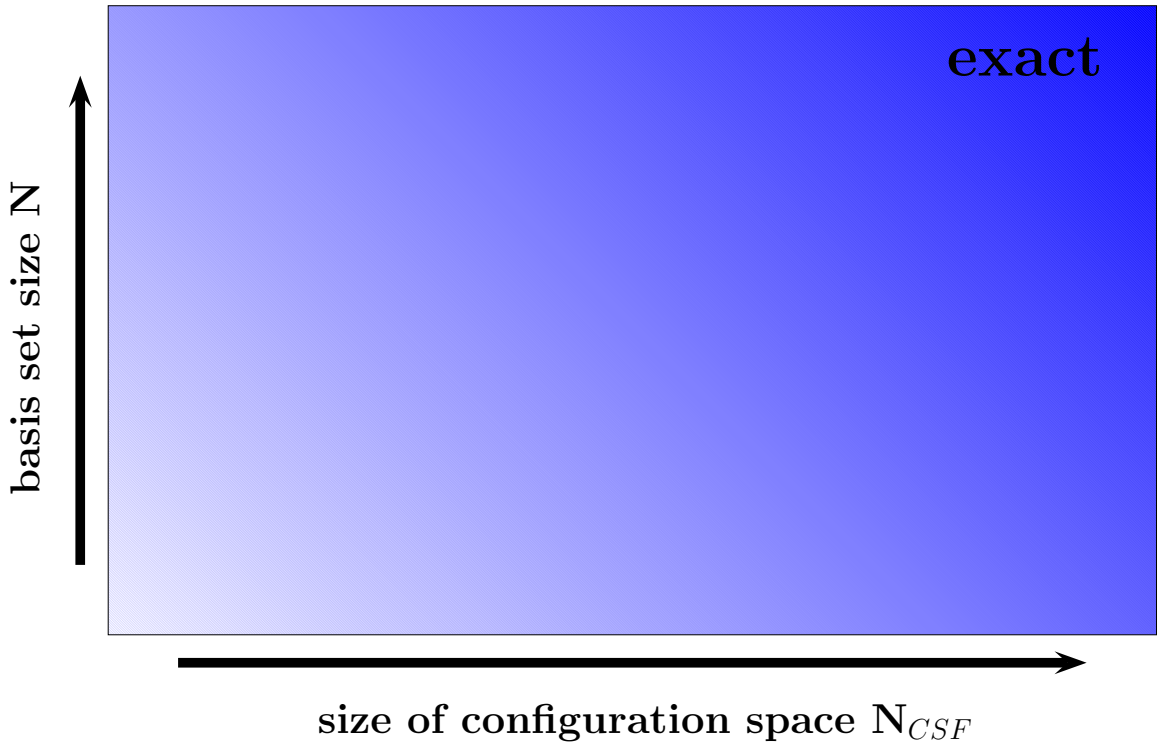
$$\phi_i(\mathbf{r}) = \sum_\mu C_{\mu,i}\chi_\mu(\mathbf{r}) \qquad \forall i$$

i.e. there are n^2 coefficients to be determined.

Independent particle approximation

Expanding $\Psi_i(\tau_1, \tau_2 \ldots \tau_N)$ in a *single determinant* and minimizing the energy subject to orthonormality yields the Hartree-Fock / Kohn-Sham equations.

N-particle expansion coefficients



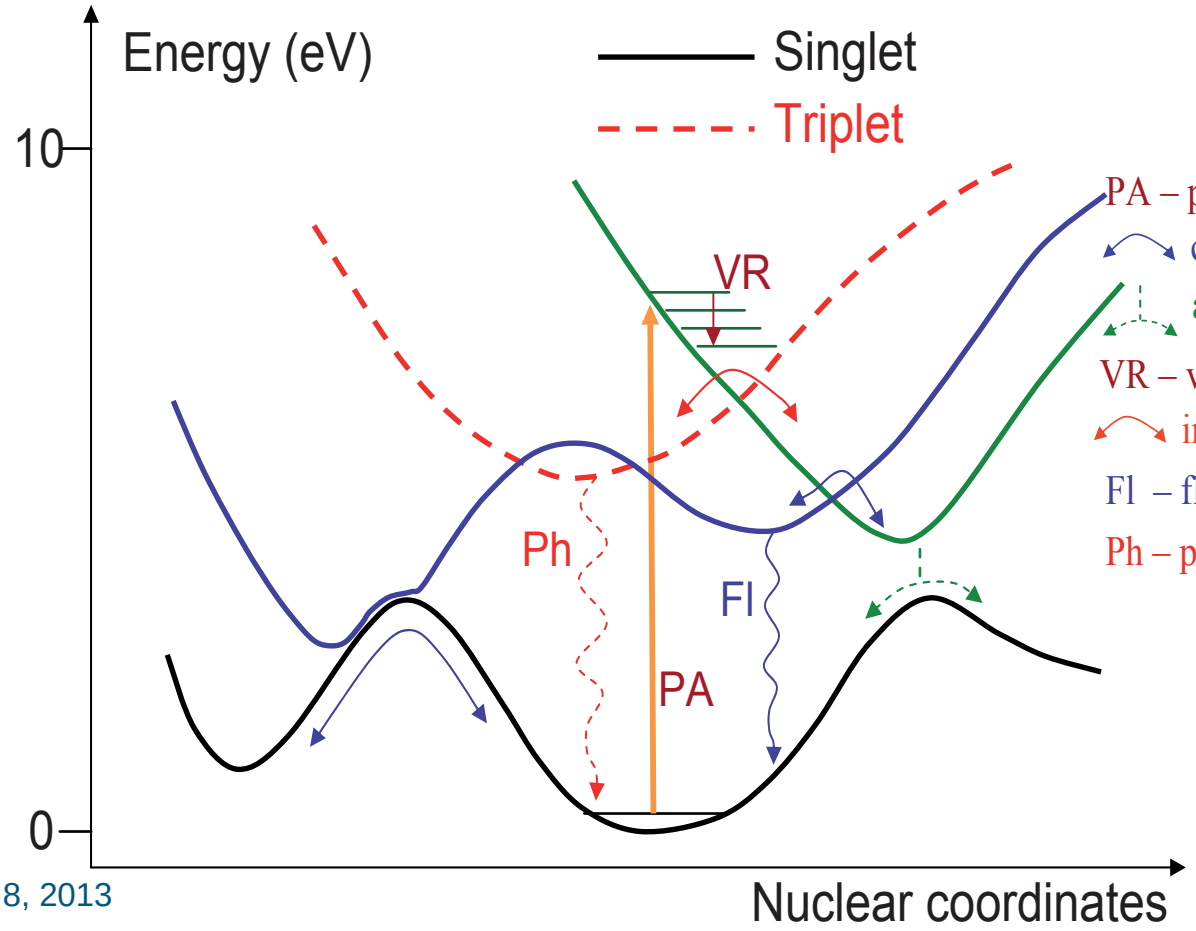
$$N_{CSF}^{FCI}(S, S_z = 0) = \frac{2S+1}{N+1}\binom{n+1}{\frac{1}{2}N - S}\binom{n+1}{\frac{1}{2}N + S + 1}$$

$$N_{Det}^{FCI} = \binom{n}{N_\alpha}\binom{n}{N_\beta}$$

Some illustrative numbers for Singlet states

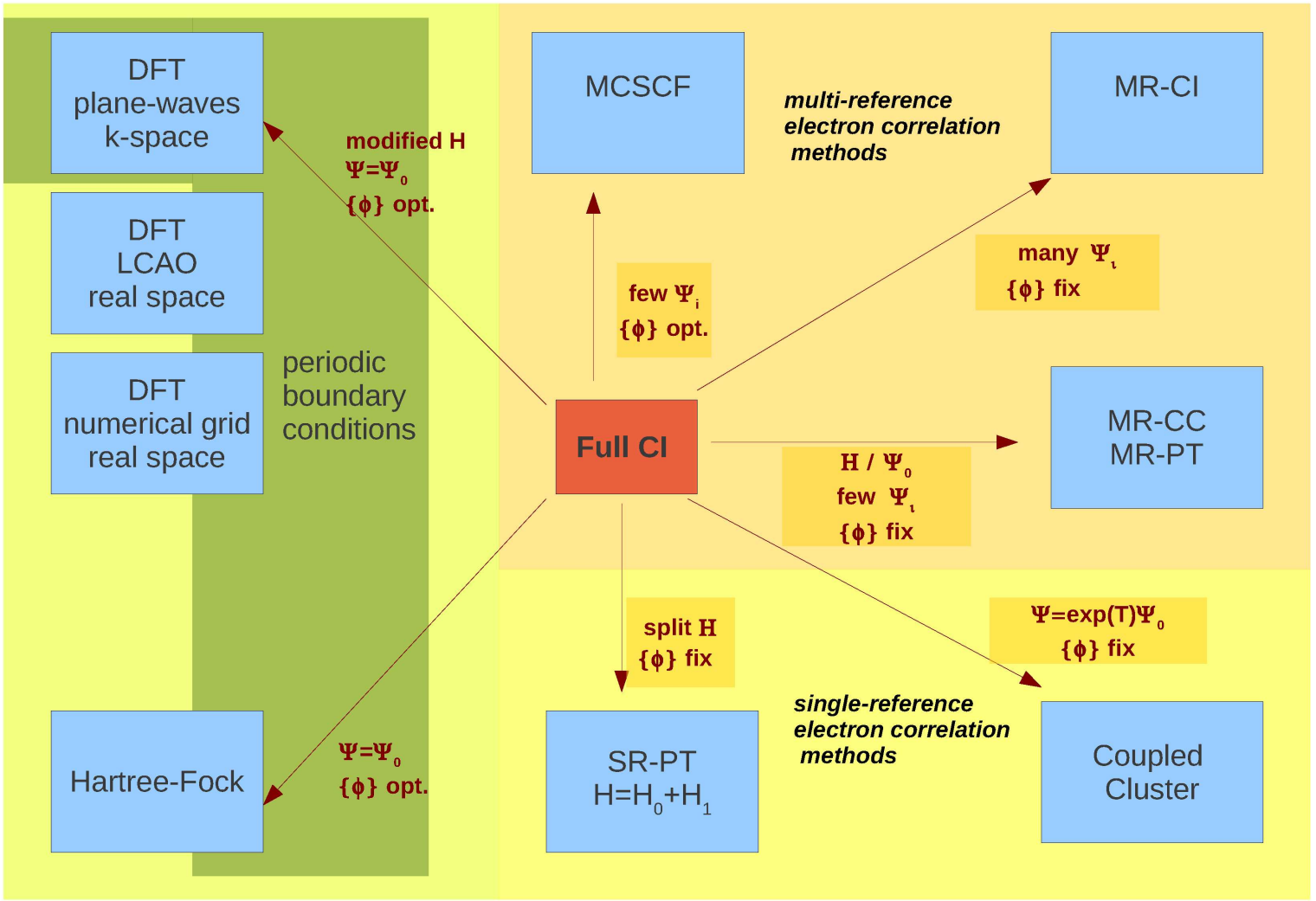
n	N	N_{CSF}^{FCI}	N_{Det}^{FCI}	Description
4	4	20	36	
8	8	1 764	4 900	
12	12	226 512	853 776	
16	16	34.8×10^{6}	165×10^{6}	
20	20	5.92×10^{9}	34.1×10^{9}	
56	8	29.0×10^{9}	134×10^{9}	C_2, cc-pVTZ
84	12	25.4×10^{15}	165×10^{15}	HCHO, cc-pVTZ
84	18	15.1×10^{21}	136×10^{21}	O_3, cc-pVTZ
238	30	3.14×10^{45}	47.1×10^{45}	pyridine, cc-pVTZ

Electronic Structure & Molecular Dynamics



PA – photoabsorption 1 fs
conical intersection $10\text{-}10^2$ fs
avoided crossing $10^2\text{-}10^4$ fs
VR – vibrational relaxation $10^2\text{-}10^5$ fs
intersystem crossing $10^5\text{-}10^7$ fs
Fl – fluorescence $10^6\text{-}10^8$ fs
Ph – phosphorescence $10^{12}\text{-}10^{17}$ fs

ab in


DFT
plane-waves
k-space
DFT
LCAO
real space
DFT
numerical grid
real space
periodic
boundary
conditions
Hartree-Fock
modified H
Ψ=Ψ0
{ϕ} opt.
Ψ=Ψ0
{ϕ} opt.
MCSCF
multi-reference
electron correlation
methods
MR-CI
few Ψi
{ϕ} opt.
many Ψt
{ϕ} fix
Full CI
H / Ψ0
few Ψt
{ϕ} fix
MR-CC
MR-PT
split H
{ϕ} fix
Ψ=exp(T)Ψ0
{ϕ} fix
single-reference
electron correlation
methods
SR-PT
H=H0+H1
Coupled
Cluster

Molecular DFT

$$F_{\mu\nu} = h_{\mu\nu} + \sum_{\kappa\lambda} D_{\kappa\lambda}(\mu\nu|\kappa\lambda) - \frac{\alpha}{2}\sum_{\kappa\lambda} D_{\kappa\lambda}(\mu\lambda|\kappa\nu) + \beta\epsilon^{X}_{\mu\nu}(D) + \gamma\epsilon^{C}_{\mu\nu}(D)$$

- ϵ^X and ϵ^C functionals evaluated on a 3D grid $\mathcal{O}(N^{1.3})$
- Coulomb part scales asymptotically $\mathcal{O}(N^2)$, formally $\mathcal{O}(N^4)$
  replace four-index by three-index integrals, formally $\mathcal{O}(N^3)$

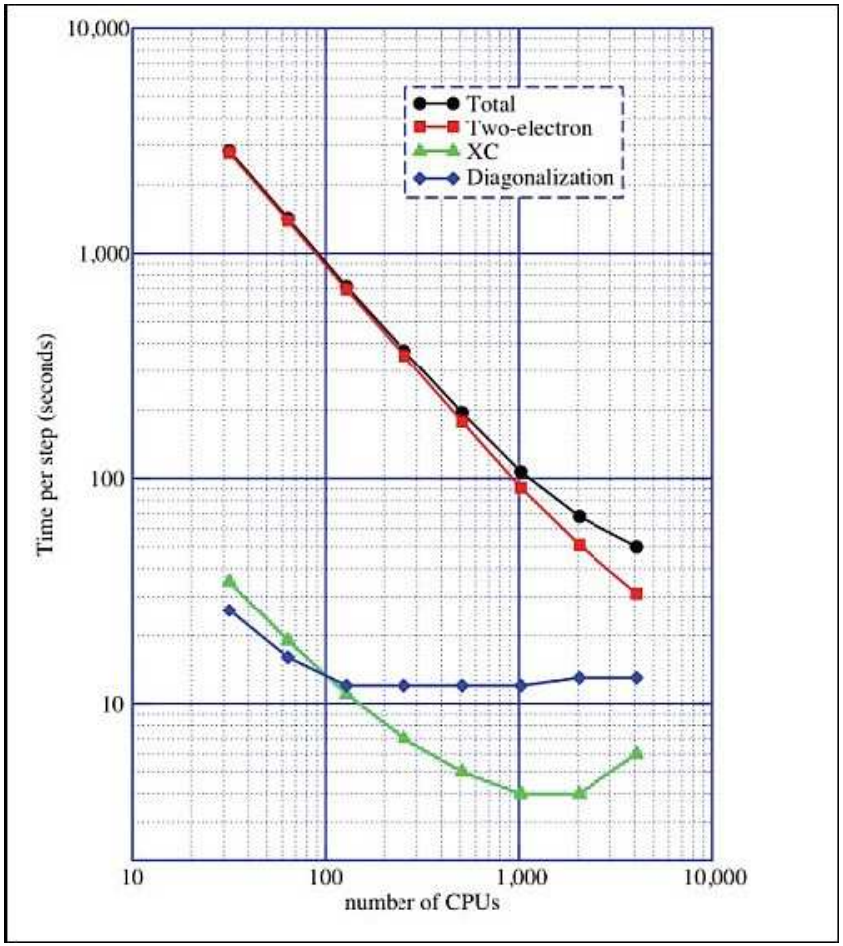
$$\chi_\alpha(\mathbf{r})\chi_\beta(\mathbf{r}) \approx \sum_Q c_{Q,\alpha\beta} Q(\mathbf{r})$$

$$(\alpha\beta|\gamma\delta) \approx \sum_{PQ} (\alpha\beta|Q)(Q|P)^{-1}(P|\gamma\delta)$$

spatial charge-density decomp. $\rightarrow$ multipole expansion $\mathcal{O}(NlogN)$

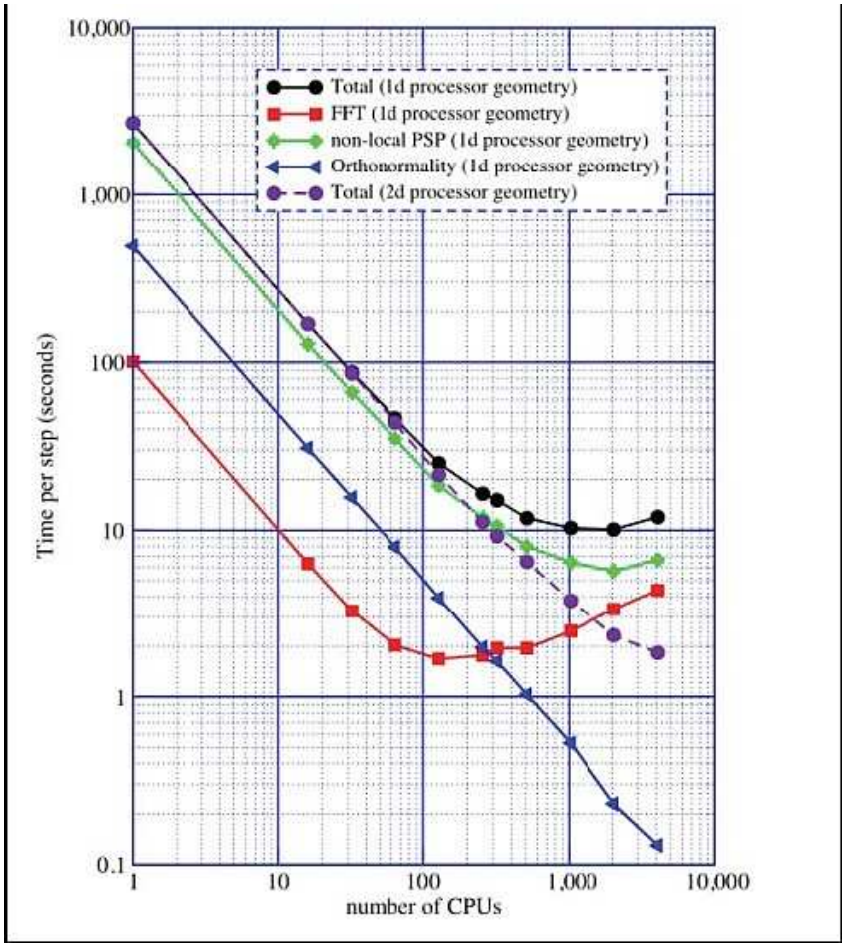
- conv. direct DFT scheme: integral evaluation dominates
- RI+multipole approximation: linear algebra dominates
- exact HF exchange is expensive

Parallel DFT Performance (NWCHEM)



C_{240}

Parallel Periodic DFT Performance (NWCHEM)



$UO_2^+ + 122\ H_2O$

Integral transformation

- 1-particle expansion coefficients are independently optimized
- hence, electron correlation methods require a basis change: AO basis $\longrightarrow$MO basis

$$(ij|kl) = \sum_{\mu} C_{\mu i} \sum_{\nu} C_{\nu j} \sum_{\kappa} C_{\kappa k} \sum_{\lambda} C_{\lambda l} (\mu\nu|\kappa\lambda)$$

- one-step evaluation: $\mathcal{O}(N^8)$ (cpu)
- four-step evaluation with intermediates: $\mathcal{O}(4N^5)$ (cpu)
- integral-direct transformation ($(\mu\nu|\kappa\lambda)$ evaluated on-the-fly) in particular useful, if $(ij|kl)$ require only simple processing

Multi-reference Configuration Interaction

$$\Psi = \sum_{I \in ref} c_I \Psi_I + \sum_{ia} c_i^a \Psi_i^a + \sum_{ijab} c_{ij}^{ab} \Psi_{ij}^{ab}$$

$$E = \sum_{ij} D_{ij} h_{ij} + \sum_{ijkl} d_{ijkl} (ij|kl)$$

$$\langle \Psi_I | \mathcal{H} | \Psi_J \rangle = H_{IJ} = \sum_{ij} a_{ij}^{IJ} h_{ij} + \sum_{ijkl} b_{ijkl}^{IJ} (ij|kl)$$

- minimization of the energy wrt the expansion coefficients yields an eigenvalue problem
- solving via subspace expansion technique
- quick evaluation of **H c** within a single pass through the integrals

Parallel Matrix-Vector Product

1,1	1,2	1,3	1,4
2,1	2,2	2,3	2,4
3,1	3,2	3,3	3,4
4,1	4,2	4,3	4,4

1
2
3
4

=

1
2
3
4

$$H_{1,1}v_1 + H_{1,2}v_2 + H_{1,3}v_3 + H_{1,4}v_4 = w_1$$
$$H_{2,1}v_1 + H_{2,2}v_2 + H_{2,3}v_3 + H_{2,4}v_4 = w_2$$
$$H_{3,1}v_1 + H_{3,2}v_2 + H_{3,3}v_3 + H_{3,4}v_4 = w_3$$
$$H_{4,1}v_1 + H_{4,2}v_2 + H_{4,3}v_3 + H_{4,4}v_4 = w_4$$

$$\sum_B \sum_{j \in B} H_{ij} v_j = w_i \qquad \forall i \in A$$

with $H_{ij} = H_{ji}$: $H_{AB}v_A \longrightarrow w_B$ and $H_{AB}v_B \longrightarrow_A$

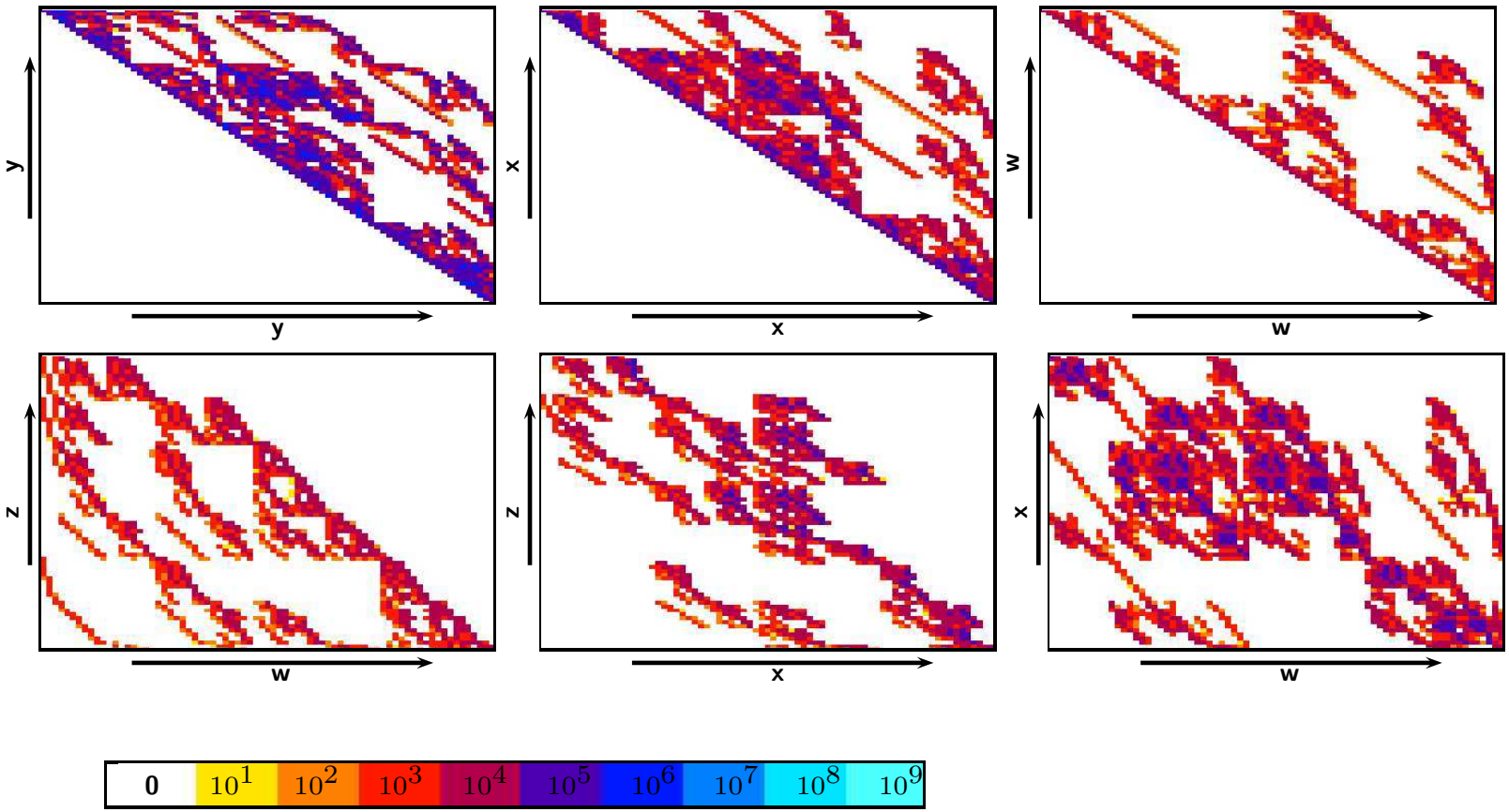
Strategy:

1. loop over all block pairs (A,B) and accumulate $w_A = w_A + H_{A,B}v_B$
2. loop over all block pairs ($A \geq B$)

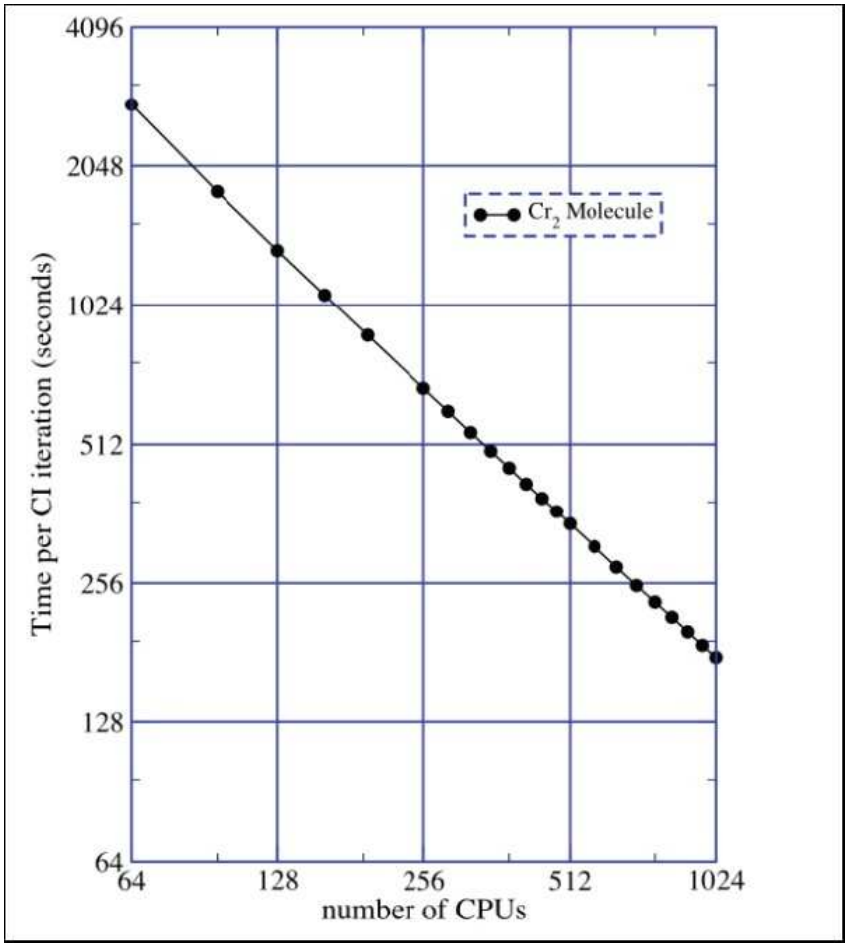
accumulate $w_A = w_A + H_{A,B}v_B$ and $w_B = w_B + H_{A,B}v_A$ if $A \neq B$

Load-balancing issues

Relative Cost (2-external integrals)



Parallel Multi-Reference CISD (COLUMBUS)



$^1\Sigma_g^+ Cr_2$
(1.7×10^9 CSFs)
IBM SP4 (32 cores, 128 GB/node)
Juropa: $\approx$4 times faster (8cores, 24GB/node)

Coupled Cluster Expansions

Wavefunction expansion

$$\Psi = exp(\mathcal{T})\Psi_0 = exp(\mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3 + \ldots)$$

$$\mathcal{T}_1 = \sum_{AI} t_I^A X_A^\dagger X_I$$

Coupled Cluster Equations

$$\langle \Psi_0 | exp(-\mathcal{T}) \mathcal{H} exp(\mathcal{T}) | \Psi_0 \rangle = E$$

$$\langle \Psi_{IJ}^{AB} | exp(-\mathcal{T}) \mathcal{H} exp(\mathcal{T}) | \Psi_0 \rangle = 0$$

Coupled Cluster Expansions

(finite) Hausdorff Expansion

$$exp(-\mathcal{T})\mathcal{H}exp(\mathcal{T}) = \mathcal{H} + [\mathcal{H},\mathcal{T}] + +\frac{1}{2}[[\mathcal{H},\mathcal{T}],\mathcal{T}]+$$

$$+\frac{1}{6}[[[\mathcal{H},\mathcal{T}],\mathcal{T}],\mathcal{T}] + \frac{1}{24}[[[[\mathcal{H},\mathcal{T}],\mathcal{T}],\mathcal{T}],\mathcal{T}]$$

- applying to Ψ_0 and projecting onto$\{\Psi_0, \Psi_I^A, \Psi_{IJ}^{AB} \ldots\}$ yields CC equations up to *forth order* in the amplitudes.
- terminating $\mathcal{T}$ at some excitation level and expanding the CC equations yields expressions in terms of amplitudes and numerous contractions of integrals.
- iterative solution with DIIS extrapolation.

Coupled Cluster Methods

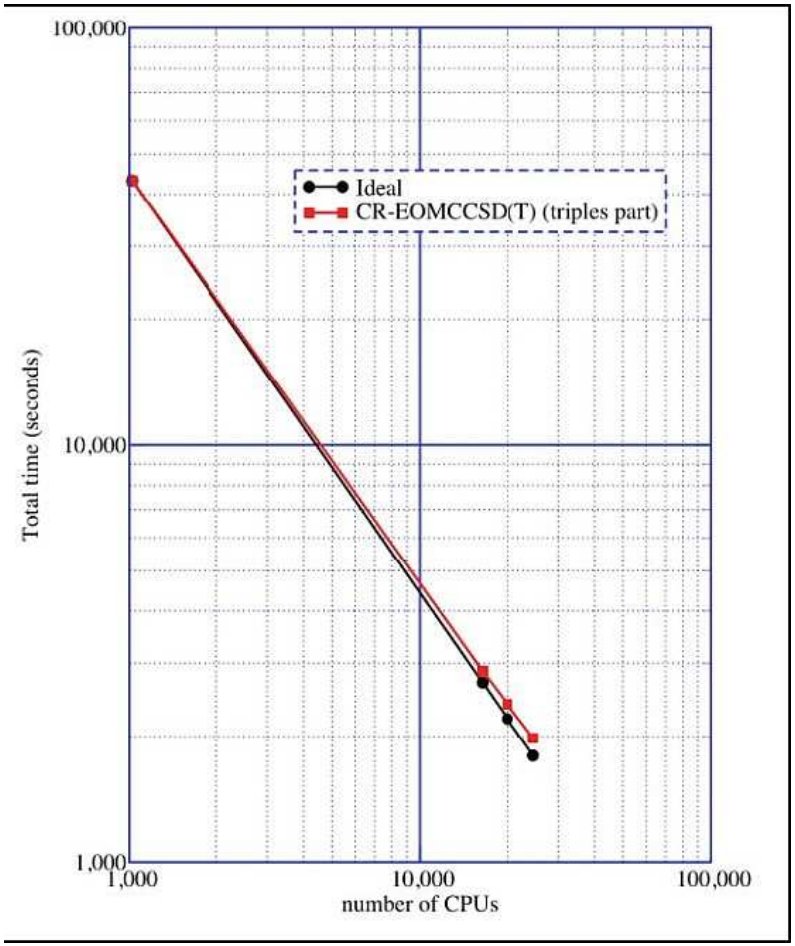
characteristics

- size-extensive
- asymptotically approaches FCI
- non-variational
- tensor-contraction engines

ressources

- computationally expensive:
  $\mathcal{O}(n^6)$ (CCSD), $\mathcal{O}(n^8)$ (CCSDT), $\mathcal{O}(N^{10})$ (CCSDTQ)
- memory-intense: $\mathcal{O}(n^4)$ (integrals) $\mathcal{O}(o^T(n-o)^T)$ (amplitudes)
  intermediates: balance memory, communication, cpu-time & I/O

Parallel EOM-CCSD(T) (NWCHEM)



some model protein
648 basis functions

Summary

- codes are ”alive” - method development and parallelization are concurrent tasks
- there is little interest in a specialized HPC code - rather a fast and universally applicable code with full compatibility within the entire package
- HPC systems tend to be specialized and respond well to little I/O, local or systolic communication patterns in combination with distributed memory
- traditionally, QC codes are a huge collection of continuously developed algorithms that do not directly rely on distributed memory
- our approach: using an interface for transparently supporting shared memory (per node replicated data) and fully distributed data along with both guarded and unguarded shared memory access and shared counters