New theoretical models for high-precision excited-state calculations:excited-state CC methods

XSCI Initiative: Development of Extreme-Scaling Algorithms for Molecular Modeling

Karol Kowalski

William R Wiley Environmental Molecular Sciences Laboratory and Chemical Sciences Division,

Pacific Northwest National Laboratory

Focus areas

- Development of scalable high-accuracy methods for molecular systems
  - Accurate coupled-cluster (CC) methods for excited state
    - Interaction of light with matter: solar energy conversion & energy storage processes
- Upscale time scales in molecular simulations
  - Development of multilevel parareal algorithms
    - from pico- to nano/micro-second molecular dynamics simulations
- Development of multiscale algorithms
  - Integration of high-level ab-initio methods with Multi-scale Multi-physics (MS/MP) approaches

Exact solution of Schrödinger equation

Weyl formula (dimensionality of full CI space) – exact solution of Schrödinger equation

$$f(n,N,S)=\frac{2S+1}{n+1}\binom{n+1}{N/2-S}\binom{n+1}{N/2+S+1}$$

$$H|\Psi\rangle = E|\Psi\rangle$$

n – total number of orbitals

N – total number of correlated electrons

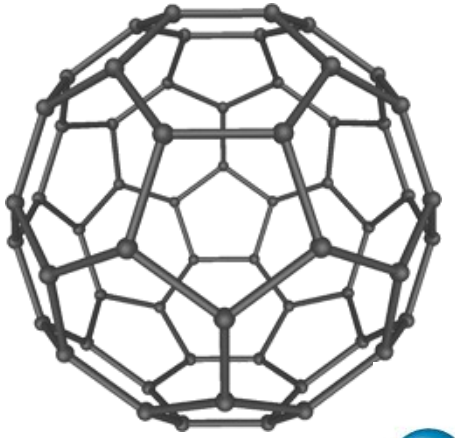
S – spin of a given electronic state

$$\begin{aligned}&C_2 \text{ molecule}:\\&12 \text{ electrons}, 100 \text{ orbitals}:\\&\#\text{FCI config.} \approx 10^{17}!!!\end{aligned}$$

Approximate WF methods

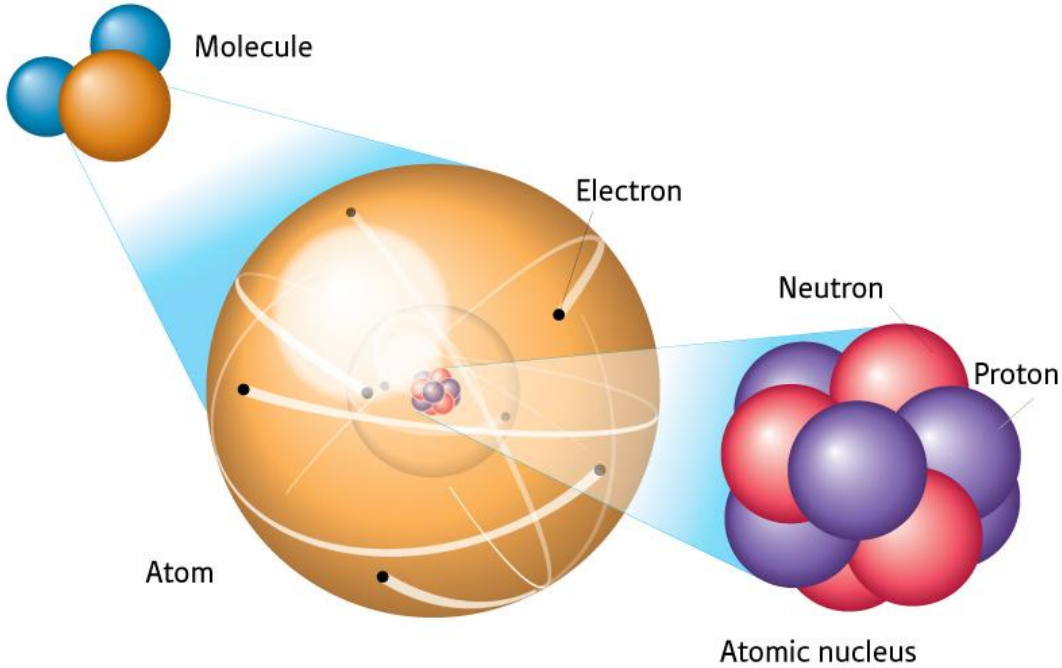
- Hartree-Fock method
- Truncated configuration interaction expansions
  - CI-n
- Perturbative methods (MBPT-n)
- Coupled Cluster formalisms
  - CCSD/EOMCCSD
  - CCSD(T)/CR-EOMCCSD(T)
- Multireference formulations
  - CASSCF
  - MRCI
  - CASPT2
  - MRMBPT-n
  - MRCC

CC methods



CC methods can be universally applied across energy and spatial scales!

Bartlett, Musial *Rev. Mod. Phys.* (2007)
Dean, Hjorth-Jensen, *Phys. Rev. B* (2004)
Kowalski, Dean, Hjorth-Jensen, Papenbrock, Piecuch, *Phys. Rev. Lett.* (2004)



CC implementations

- ACES III (parallel)
- DALTON (serial)
- GAMESS (CCSD/CCSD(T) – parallel)
- MOLPRO (parallel)
- NWCHEM (parallel)
- PQS (CCSD/CCSD(T) – parallel)

Coupled Cluster (CC) and Equation-of-Motion Coupled Cluster Methods

cluster operator

$$\left|\Psi_0\right\rangle = e^T \left|\Phi\right\rangle$$

reference function (HF determinant)

$$\left|\Psi_K\right\rangle = R_K e^T \left|\Phi\right\rangle$$

"excitation" operator

$$\overline{H} R_K \left|\Phi\right\rangle = E_K R_K \left|\Phi\right\rangle$$

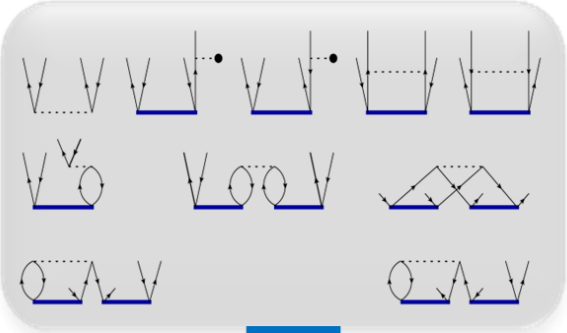
$$\overline{H} = e^{-T} H e^T$$

Method	**Numerical complexity**	**Size of the eigenvalue problem to be solved**
EOMCCSD (singles & doubles)	N^6	$\sim N^2+N^4$
CR-EOMCCSD(T) (perturbative triples)	N^7	$\sim N^2+N^4$
EOMCCSDT (singles & doubles & triples)	N^8	$\sim N^2+N^4+N^6$

Tensor Contraction Engine (TCE)

- Highly parallel codes are needed in order to apply the CC theories to larger molecular systems
- Symbolic algebra systems for coding complicated tensor expressions: Tensor Contraction Engine (TCE)

		Expression[a]
$D_i^a t_i^a$	=	$f_i^a + t_i^f I_f^a - t_n^a I'^n_i + t_{ni}^{fa} I_f^n + t_n^f v_{fi}^{na} - \frac{1}{2} t_{no}^{fa} v_{fi}^{no} + \frac{1}{2} t_{ni}^{fg} v_{fg}^{na} + \frac{1}{4} t_{ino}^{afg} v_{fg}^{no}$
$D_{ij}^{ab} t_{ij}^{ab}$	=	$v_{ij}^{ab} + P(a/b) I_f^a t_{ij}^{fb} - P(i/j) I_i^n t_{nj}^{ab} + \frac{1}{2} t_{ij}^{fg} I'^{ab}_{fg} + \frac{1}{2} t_{no}^{ab} I_{ij}^{no}$ $+ P(a/b) P(i/j) t_{in}^{af} I_{fj}^{nb} \frac{1}{2} P(a/b) I_{fg}^{na} t_{nij}^{fgb}$ $- \frac{1}{2} P(i/j) I_{fi}^{no} t_{noj}^{fab} + t_{nij}^{fab} I_f^n + P(i/j) t_i^f I'^{ab}_{fj} - P(a/b) t_n^a I'^{nb}_{ij} + \frac{1}{4} t_{ijno}^{abfg} v_{fg}^{no}$
$D_{ijk}^{abc} t_{ijk}^{abc}$	=	$P(a/bc) I_f^a t_{ijk}^{fbc} - P(i/jk) I_i^n t_{njk}^{abc} + \frac{1}{2} P(a/bc) t_{ijk}^{afg} I_{fg}^{bc} + \frac{1}{2} P(i/jk) t_{ino}^{abc} I_{jk}^{no}$ $+ P(ab/c) P(ij/k) t_{ijn}^{abf} I_{fk}^{nc} + P(a/bc) P(ij/k) t_{ij}^{af} I_{fk}^{bc} - P(ab/c) P(i/jk) t_{in}^{ab} I_{jk}^{nc}$ $+ t_{nijk}^{fabc} I_f^n + \frac{1}{2} P(a/bc) I_{fg}^{na} t_{nijk}^{fgbc} - P(i/jk) I_{fi}^{no} t_{nojk}^{fabc} + \frac{1}{4} t_{ijkno}^{abcfg} v_{fg}^{no}$
$D_{ijkl}^{abcd} t_{ijkl}^{abcd}$	=	$P(a/bcd) I_f^a t_{ijkl}^{fbcd} - P(i/jkl) I_i^n t_{njkl}^{abcd} + \frac{1}{2} P(ab/cd) t_{ijkl}^{abfg} I_{fg}^{cd} + \frac{1}{2} P(ij/kl) t_{ijno}^{abcd} I_{kl}^{no}$ $+ P(abc/d) P(ijk/l) t_{ijkn}^{abcf} I_{fl}^{nd} + P(ab/cd) P(ijk/l) t_{ijk}^{abf} I_{fl}^{cd} - P(abc/d) P(ij/kl) t_{ijn}^{abc} I_{kl}^{nd}$ $+ P(a/bcd) P(ij/kl) t_{ij}^{af} I'^{bcd}_{fkl} - P(ab/cd) P(ijk/l) t_{in}^{ab} I'^{ncd}_{jkl} + P(ab/cd) P(ij/kl) t_{ijn}^{abf} I'^{ncd}_{jk}$ $+ \frac{1}{2} P(abc/d) P(ij/kl) t_{ino}^{abc} I'^{nod}_{jkl} + t_{nijkl}^{fabcd} I_f^n + \frac{1}{2} P(a/bcd) I_{fg}^{na} t_{nijkl}^{fgbcd}$ $- \frac{1}{2} P(i/jkl) I_{fi}^{no} t_{nojkl}^{fabcd}$
$D_{ijklm}^{abcde} t_{ijklm}^{abcde}$	=	$P(a/bcde) I_f^a t_{ijklm}^{fbcde} - P(i/jklm) I_i^n t_{njklm}^{abcde} + \frac{1}{2} P(abc/de) t_{ijklm}^{abcfg} I_{fg}^{de} + \frac{1}{2} P(ijk/lm) t_{ijkno}^{abcde} I_{lm}^{no}$ $+ P(abcd/e) P(ijkl/m) t_{ijkln}^{abcdf} I_{fm}^{ne} + P(abc/de) P(ijk/lm) t_{ijkn}^{abcf} I_{flm}^{nde}$ $+ \frac{1}{2} P(abcd/e) P(ij/klm) t_{ijno}^{abcd} I'^{noe}_{klm} + \frac{1}{2} P(ab/cde) P(ijkl/m) t_{ijkl}^{abfg} I'^{cde}_{fgm}$ $+ P(abc/de) P(ijkl/m) t_{ijkl}^{abcf} I_{fm}^{de} - P(abcd/e) P(ijk/lm) t_{ijkn}^{abcd} I_{lm}^{ne}$ $+ P(ab/cde) P(ijk/lm) t_{ijk}^{abf} I'^{ncde}_{flm} - P(abc/de) P(ij/klm) t_{ijn}^{abc} I'^{nde}_{klm}$ $+ P(a/bcde) P(ij/klm) t_{ij}^{af} I'^{bcde}_{fklm} - P(ab/cde) P(i/jklm) t_{in}^{ab} I'^{ncde}_{jklm}$



OCE

$$+\frac{1}{4} v_{ef}^{mn} t_{ij}^{ef} t_{mn}^{ab} - \frac{1}{2} v_{ef}^{mn} t_{mi}^{ef} t_{nj}^{ab} +$$

TCE

```
next = NXTASK(nprocs, 1)
DO p3b = noab+1,noab+nvab
DO p4b = p3b,noab+nvab
DO h1b = 1,noab
DO h2b = h1b,noab
IF (next.eq.count) THEN
CALL GET_HASH_BLOCK(d_a,dbl_mb(k_a),dim
- 1 + (noab+nvab) * (h1b_1 - 1 + (noab+
+nvab) * (p3b_1 - 1)))))
CALL GET_HASH_BLOCK_I(d_a,dbl_mb(k_a),d
```

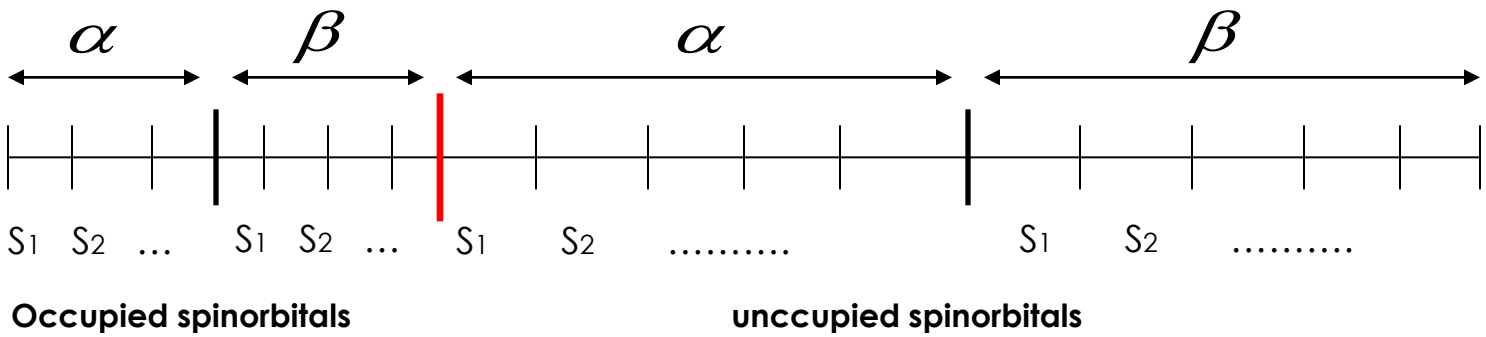
Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

Parallel performance

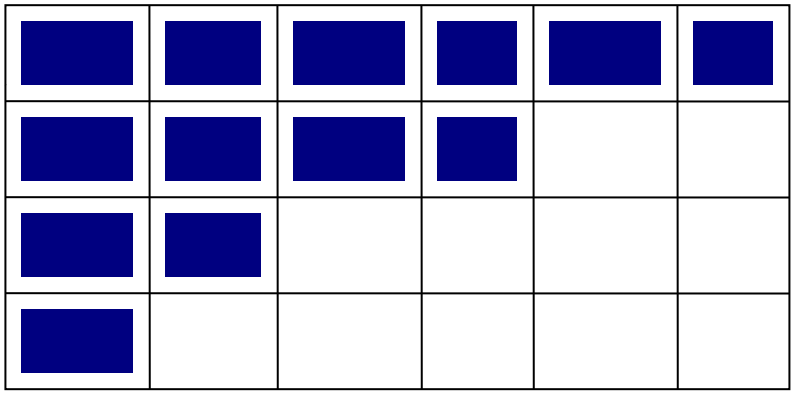
Parallel structure of the TCE CC codes

Tile structure:



Tensor structure:

$$T_a^i \Rightarrow T_{[p_n]}^{[h_m]}$$



Triples corrections to the EOMCCSD excitation energies

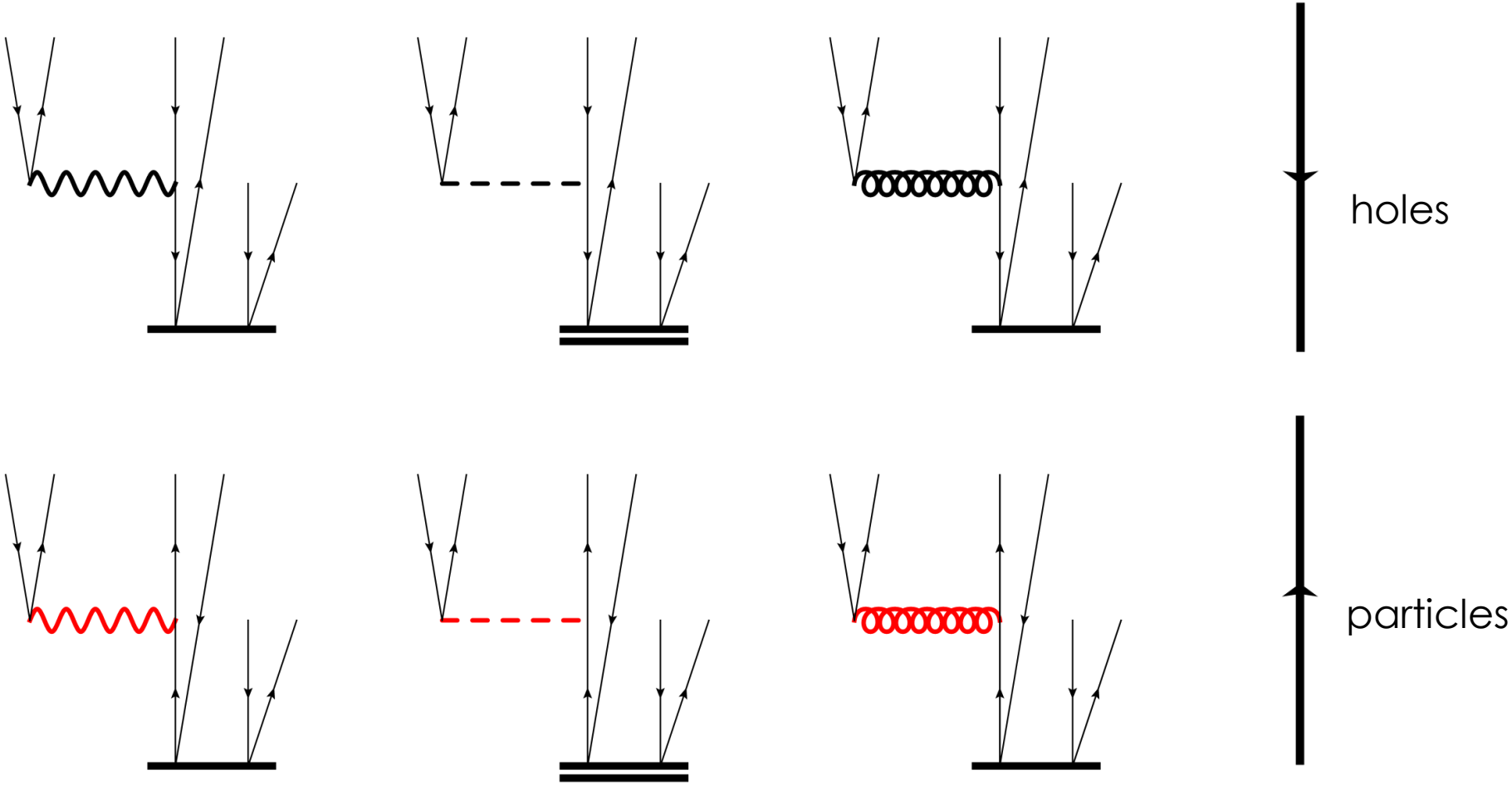
- Correlation effects due to the collective excitations of three electrons are very important in describing excited states
- Method of Moments of Coupled Cluster Equations (MMCC)
  - Non-iterative corrections can be expressed in terms non-zero components of the Schrödinger equation (moments)
  - Rigorous way of introducing higher order correlation effects

$$\delta_K(T) = \frac{\langle \Psi_K | Q_3 M_{K,3} | \Phi \rangle}{\langle \Psi_K | e^{T_1+T_2}(R_{K,0} + R_{K,1} + R_{K,2}) | \Phi \rangle}$$

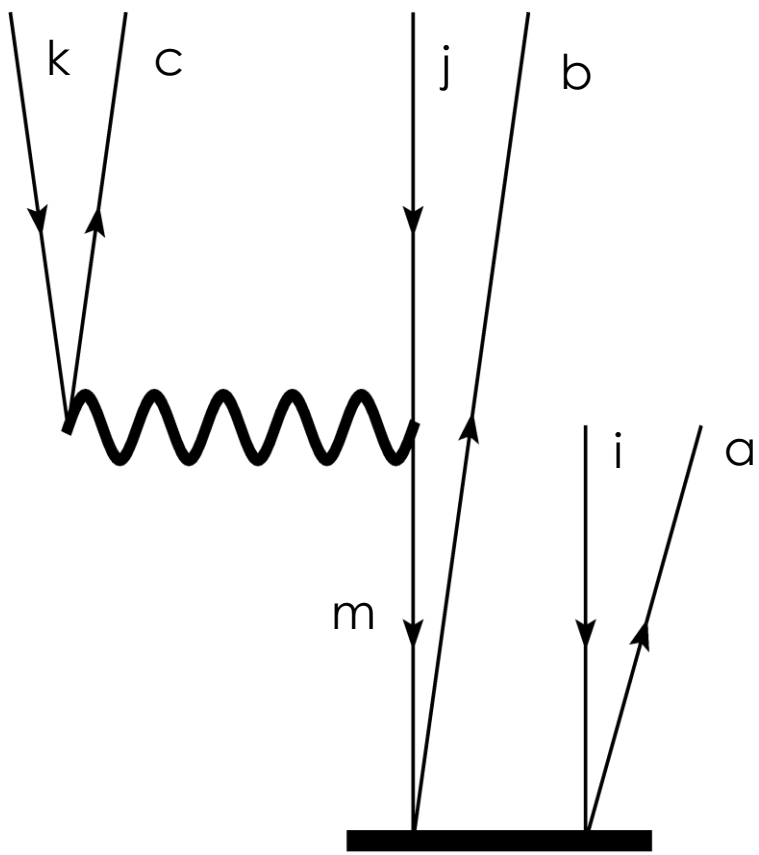
CR-EOMCCSD(T) correction
(Kowalski and Piecuch JCP 2004)

$$\omega_K^{CR-EOMCCSD(T)} = \omega_K^{EOMCCSD} + \delta_K^{CR-EOMCCSD(T)}$$

Diagrammatica



Diagrammatica



3.) Algebraic expression that represent the contribution of this process to the excited-state wavefunction can be derived:

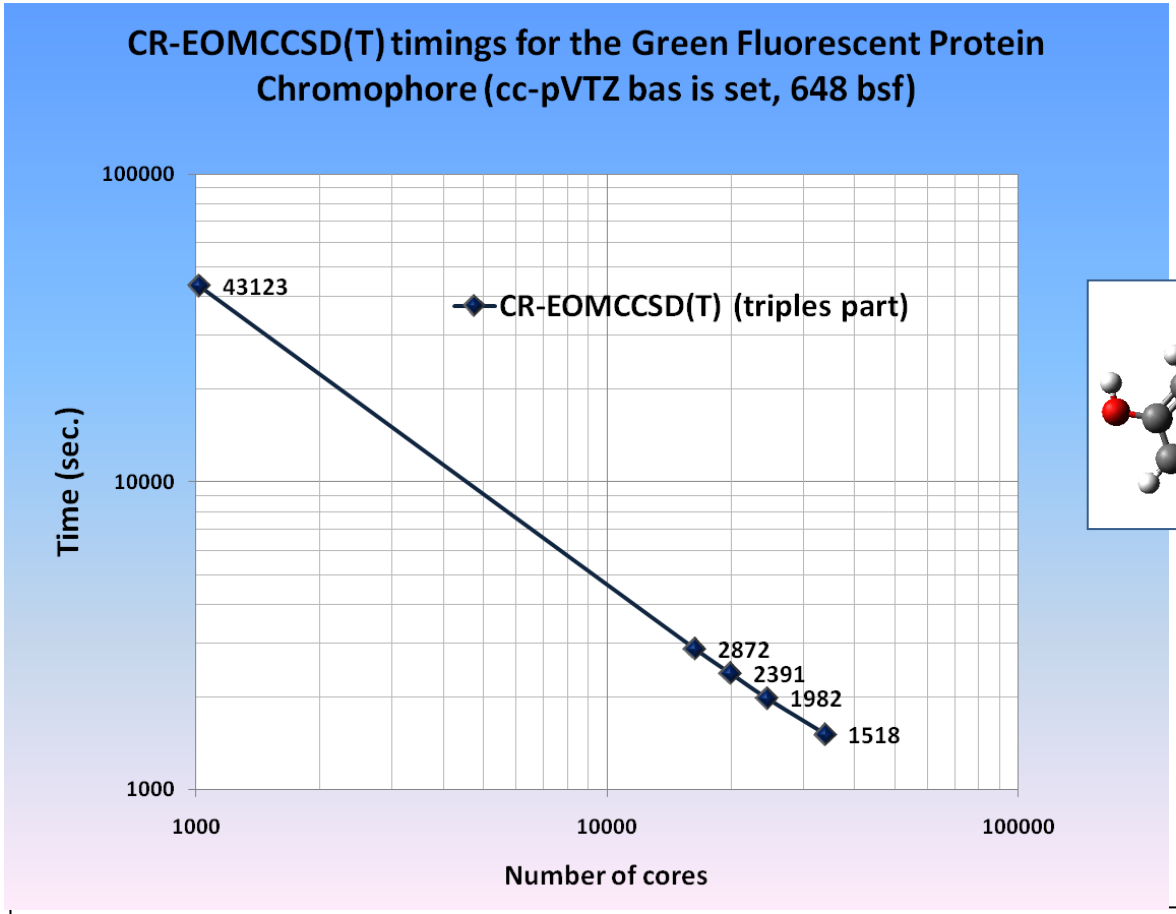
$$\frac{1}{4} A^{kj/i} A_{c/ba} V_{cm}^{kj}(R_K, T) T_{ba}^{mi}$$

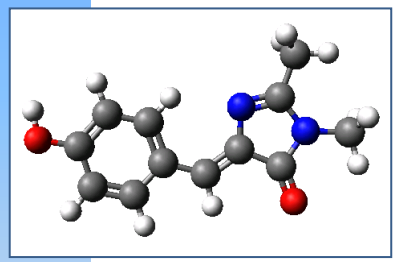
2.) hole "m" interacts with other particles and holes producing hole "k" and particle "c"; hole "m" can change the identity to "j"

1.) "correlation interaction" produces holes "i" & "m" and particles "a" & "b"

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

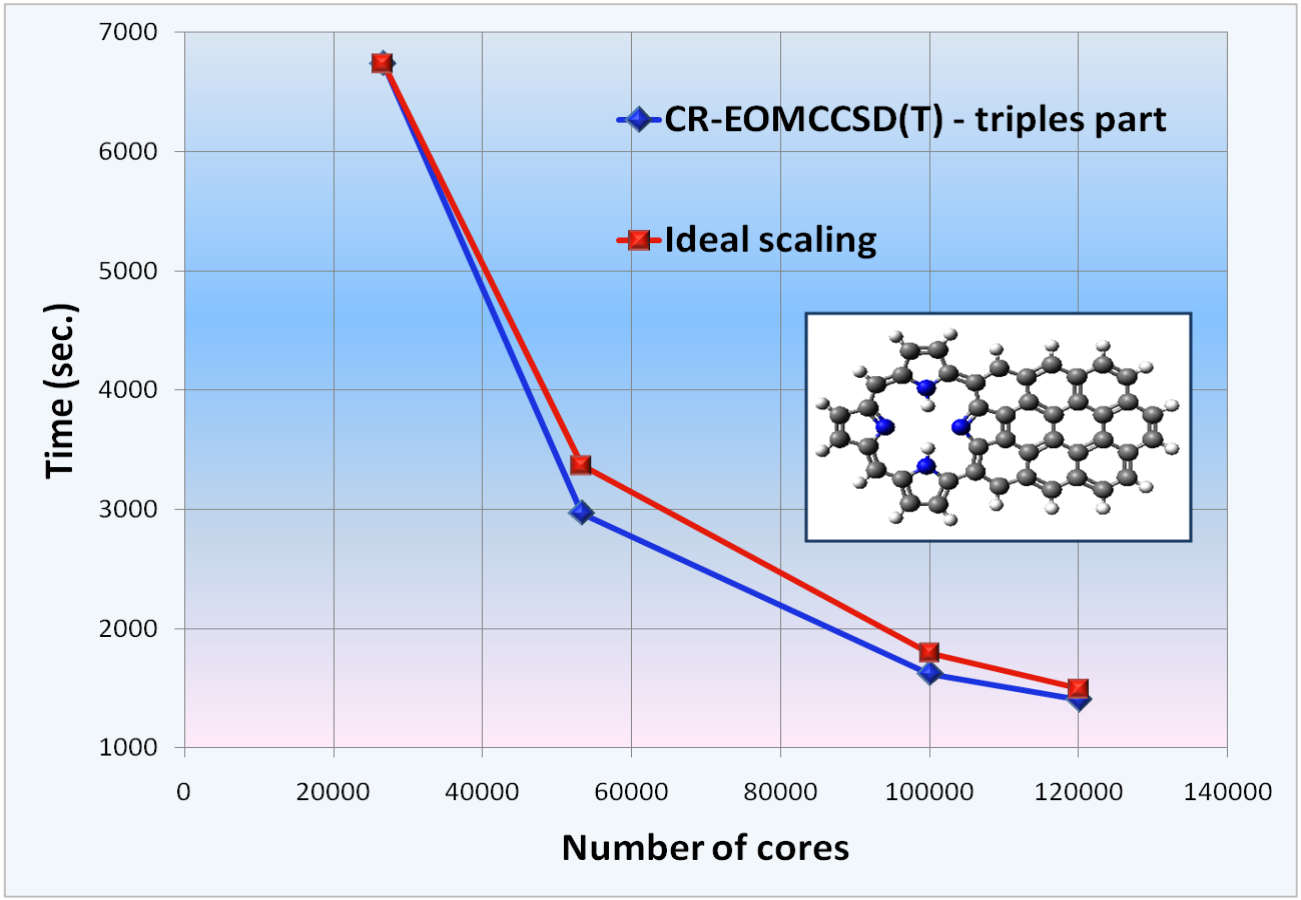
Parallel performance





An example of the scalability of the triples part of the CR-EOMCCSD(T) approach for GFPC described by the cc-pVTZ basis set (648 basis set functions). Timings were determined from calculations on the **Franklin Cray-XT4** computer at NERSC using 1024, 16384, 20000, 24572, and 34008 cores).

Parallel performance



An example of the scalability of the triples part of the CR-EOMCCSD(T) approach for the P1B1-f-coronene molecule described by the Ahlrichs-VTZ basis set (786 basis set functions). Timings were determined from calculations on the **Jaguar Cray-XT5** computer at ORNL.

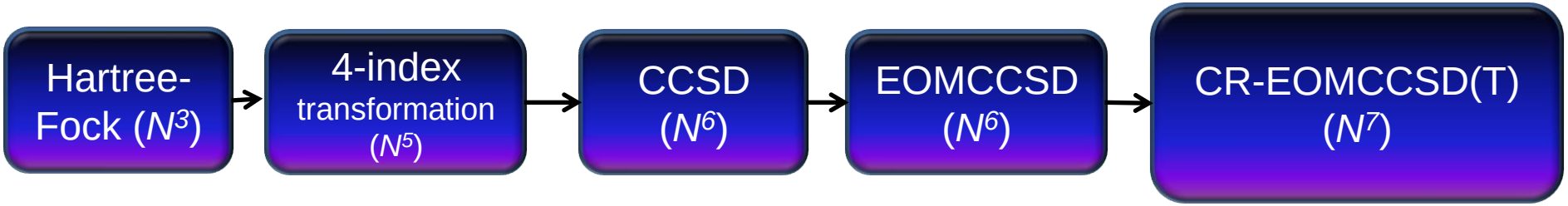
Parallel performance

- Major factors determining parallel performance of the (T) part
  - GA !!!
  - New antisymmetrization of 6-dimensional tensors
  - Improved local memory management
  - Possibility of controlling volume to surface ratio in the calculations (scalability can be achieved also for smaller systems)
  - Slicing 6-dimensional tensors (volume/surface ratio can be made arbitrarily large) in the CR-EOMCCSD(T) code

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

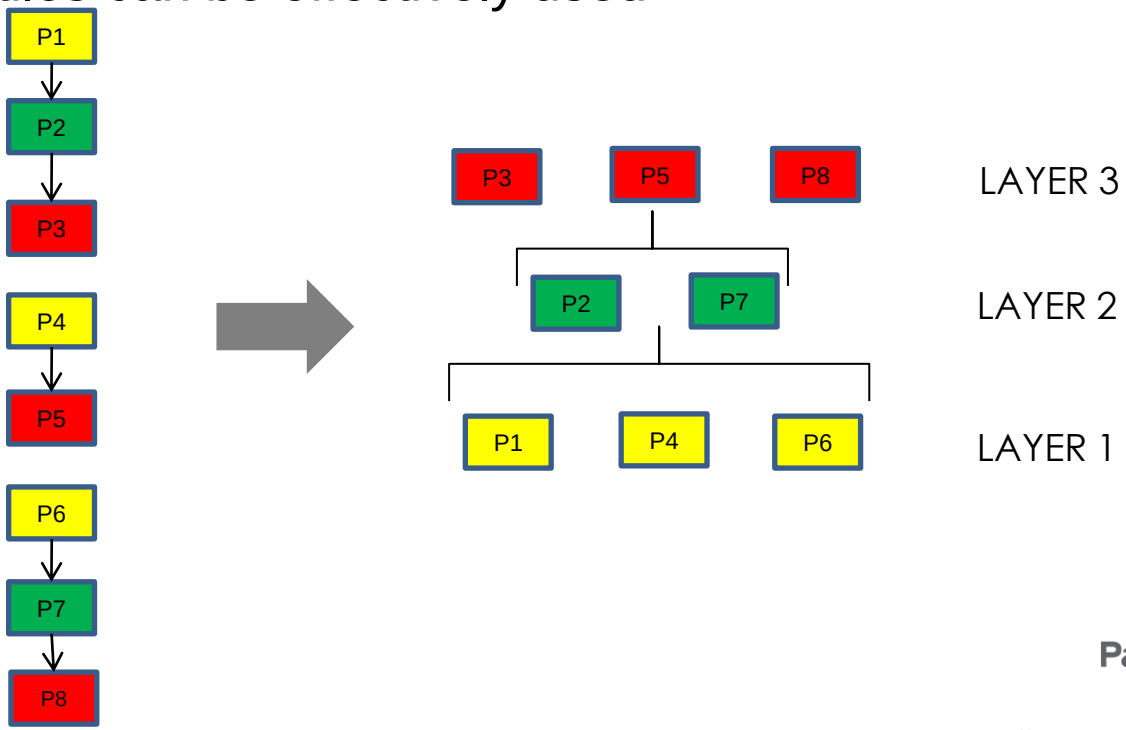
Parallel performance

- Iterative CCSD/EOMCCSD methods:
  - ga_get, ga_put, ga_acc, dgemm
  - very intensive communication pattern
  - possible problems with load balancing
- Non-iterative (T) methods
  - ga_get, dgemm

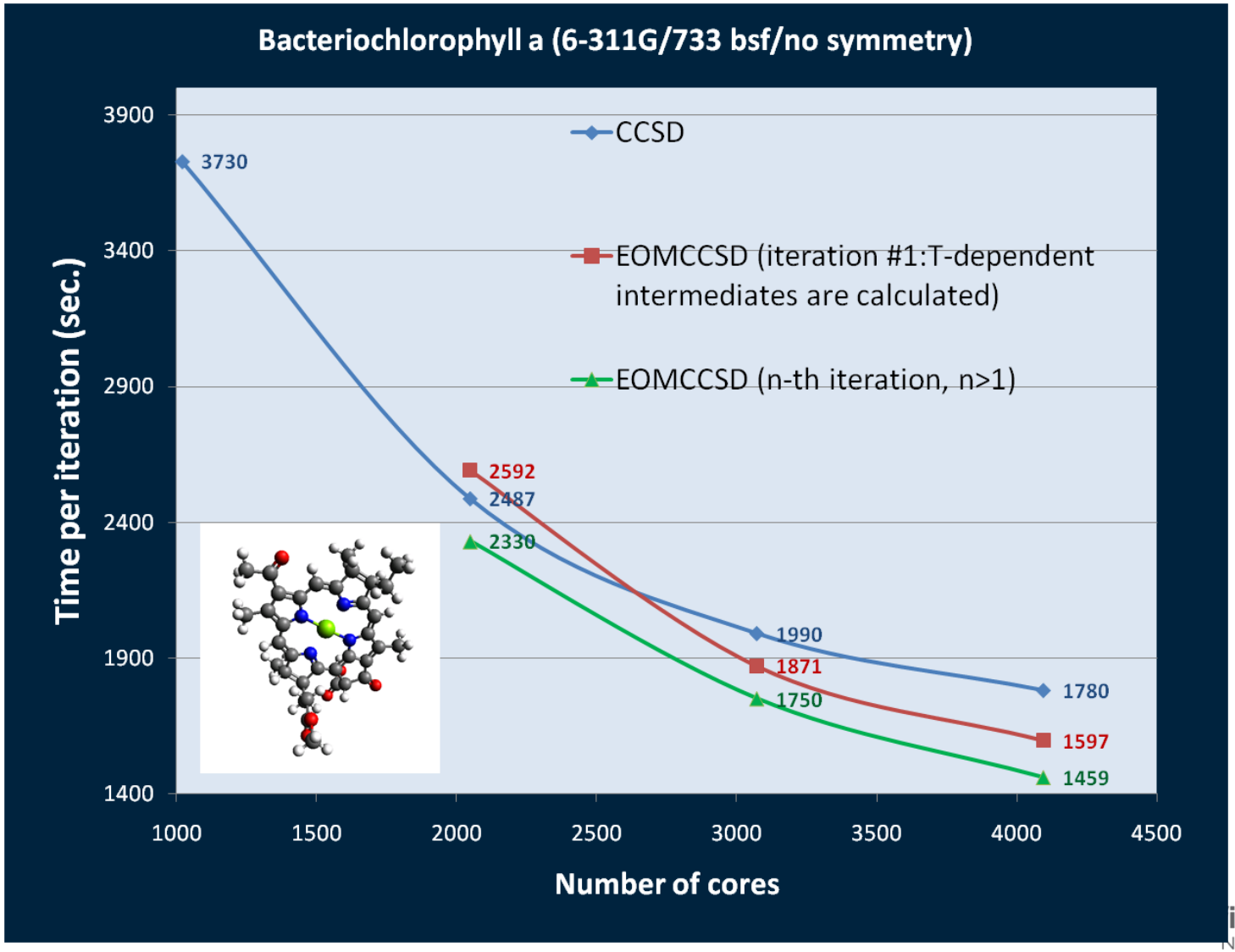


Parallel performance – iterative methods (Krishnamoorthy & Kowalski)

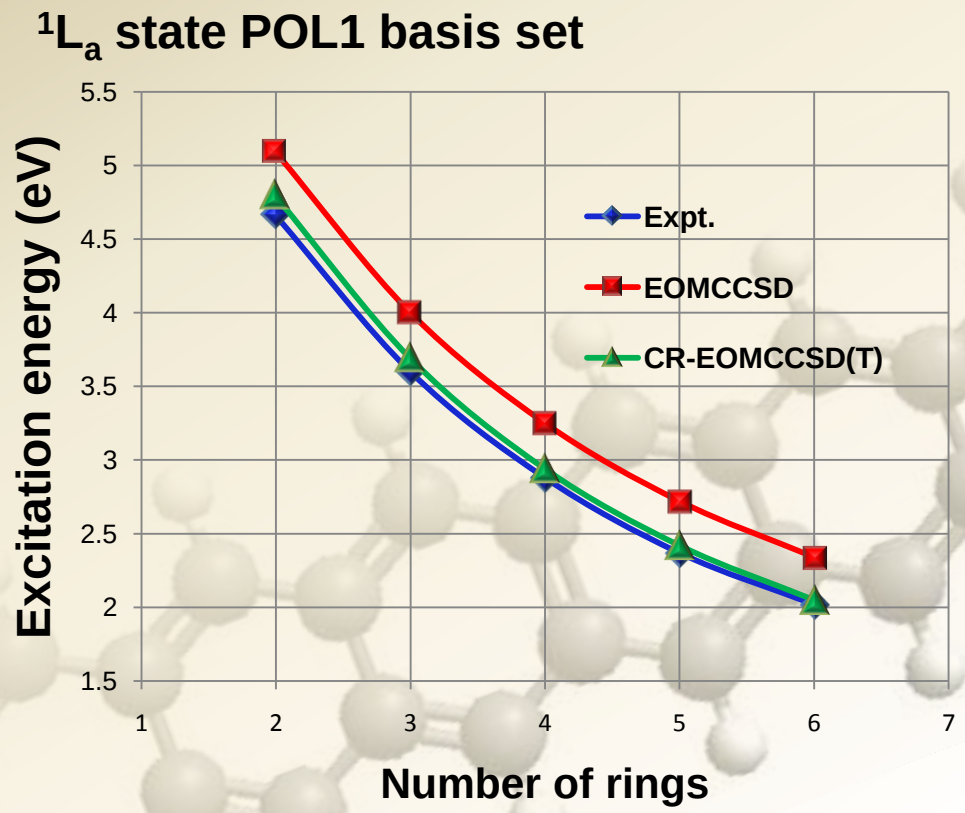
- Alternative task schedulers (independent counters)
  - use “global task pool”
  - improve load balancing
  - reduce the number of synchronization steps to absolute minimum (from 80 to 4 in the EOMCCSD case)
  - larger tiles can be effectively used



Parallel performance – iterative methods



Illustrative examples of large-scale calculations – components of light harvesting systems



Illustrative examples of large-scale calculations – components of light harvesting systems

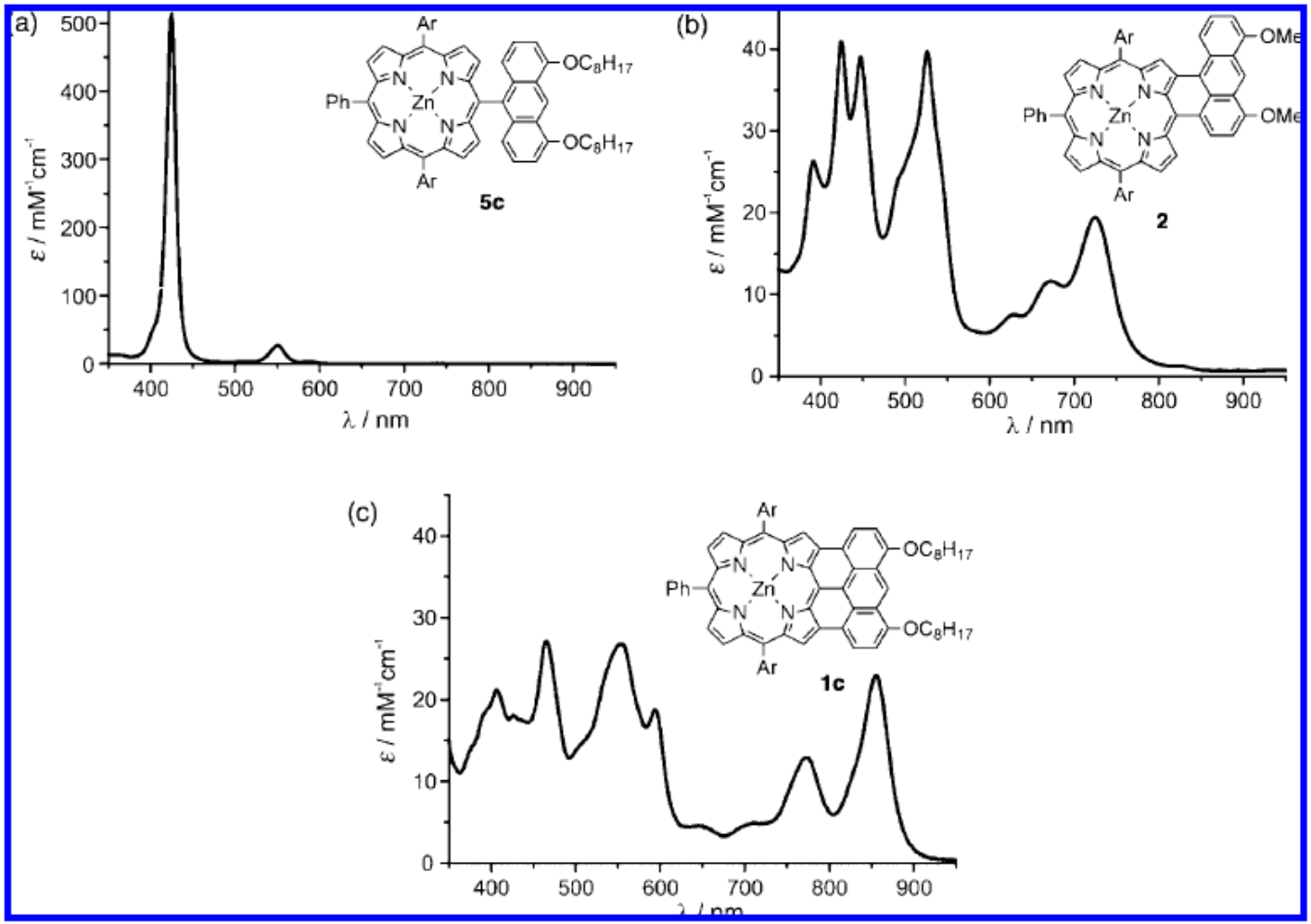
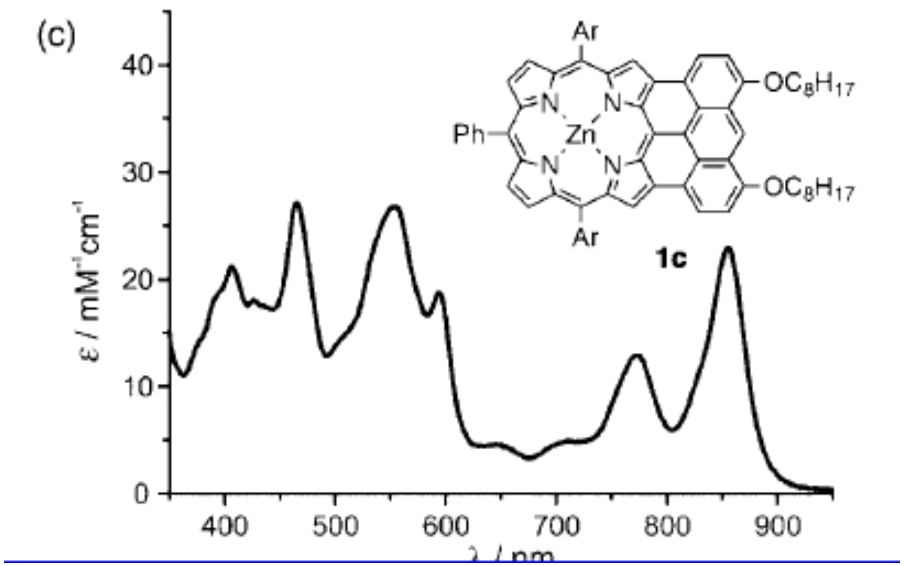
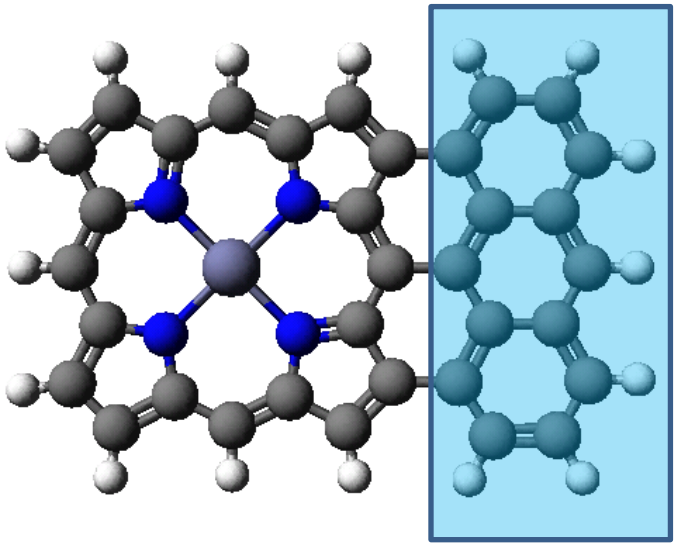


Figure 1. Absorption spectra in chloroform. *meso*-(9-anthryl)porphyrin **5c** (a) has a typical porphyrin spectrum, whereas the spectra of the *β,meso* doubly fused porphyrin **2** (b) and the *β,meso,β* triply fused porphyrin **1c** (c) are distorted and red-shifted.

Illustrative examples of large-scale calculations – components of light harvesting systems





ZnP

H➔L & H-1 ➔ L+1
H-1 ➔ L & H➔L+1

CR-EOMCCSD(T)	2.25 eV
Expt.	2.28 eV

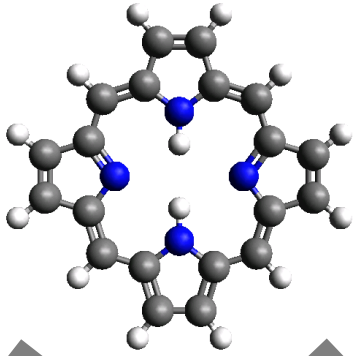
ZnP-f-anthracene

H➔L

CR-EOMCCSD(T)	1.79 eV
Expt.	1.71 eV

Illustrative examples of large-scale calculations – components of light harvesting systems

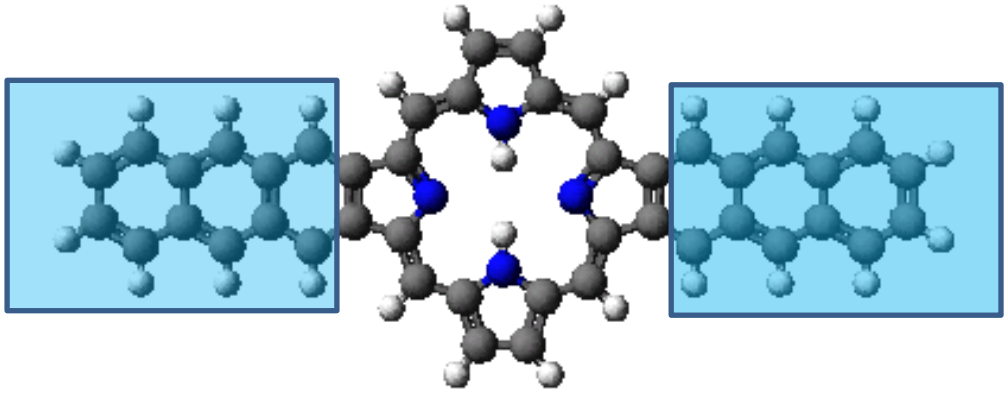
FBP



H→L CR-EOMCCSD(T) 2.32 eV

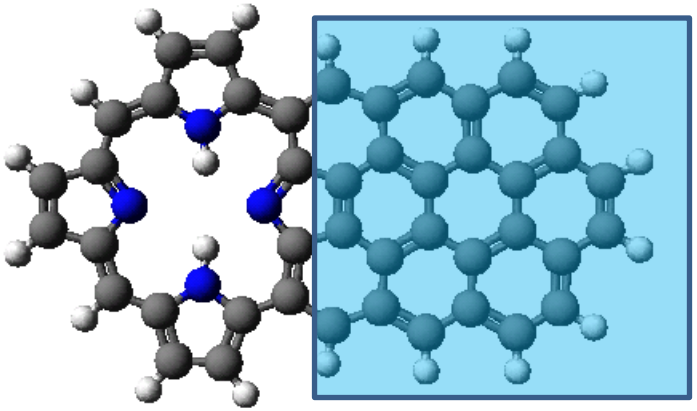
H-1→L
H→L+1 CR-EOMCCSD(T) 1.86 eV

P1B2-f-anthracene



H→L CR-EOMCCSD(T) 1.78 eV

P1B1-f-coronene



H→L CR-EOMCCSD(T) 1.36 eV

Illustrative examples of large-scale calculations – modeling charge transfer processes in open-shell systems

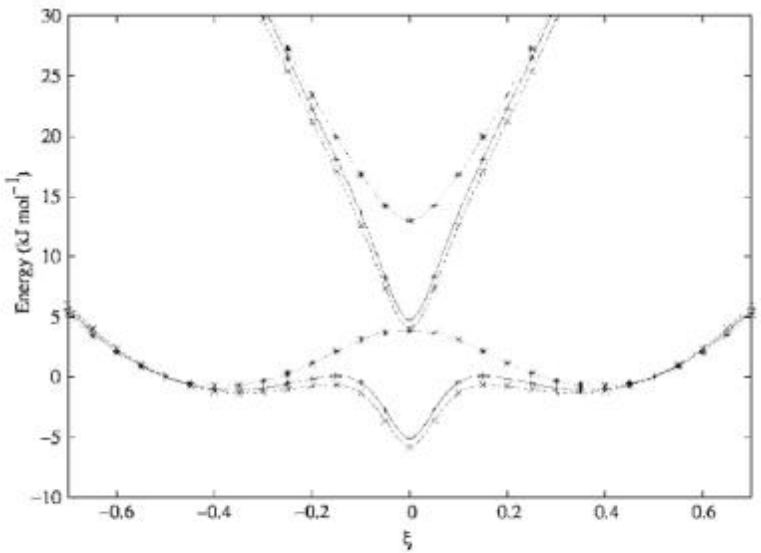
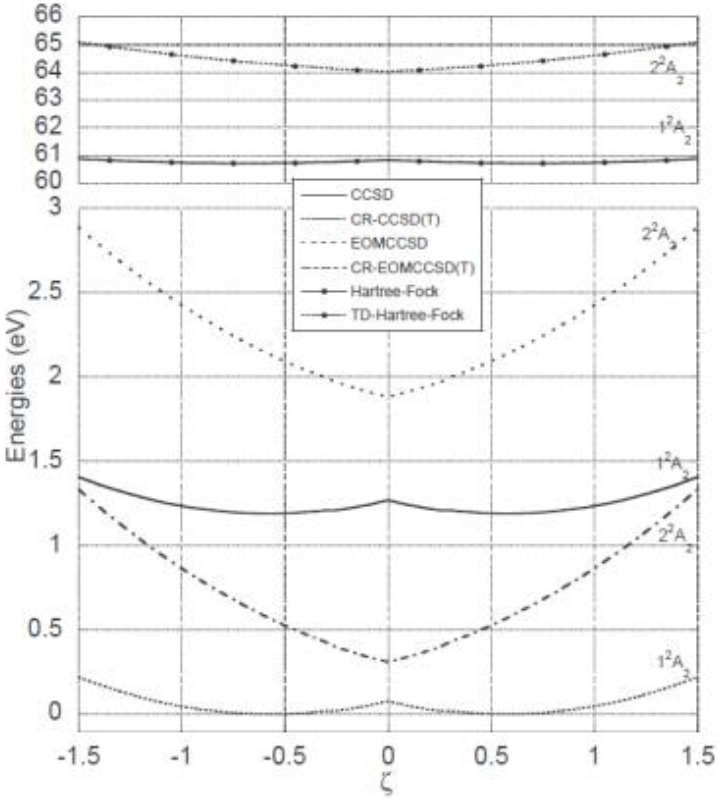


FIG. 2. SC-NEVPT2(can), SC-NEVPT3(can), and CASPT2(can) (no level shift) PES of the $^2A_2(1)$ and $^2A_2(2)$ states of the spiro cation. All the curves are shifted in order to have the two C_{2v} minima at zero energy. Full lines and "+" points, NEVPT2(can) energies; dashed lines and "×" points, CASPT2(can) energies; dotted lines and "*" points, NEVPT3(can) energies.



Reduced cost CR-EOMCCSD(T) methods

- Active-space CR-EOMCCSD(T) approach
  - Active space can be defined by orbitals with orbital energies falling into a predefined energy interval $[\alpha,\beta]$
  - Significant reduction of the numerical cost associated with calculating (T) correction: $N^7 \Rightarrow N_{act}^5 N^2$: Applicable to large molecular systems

Double excitations are allowed only within the active space

$$T_1,\ t_2$$

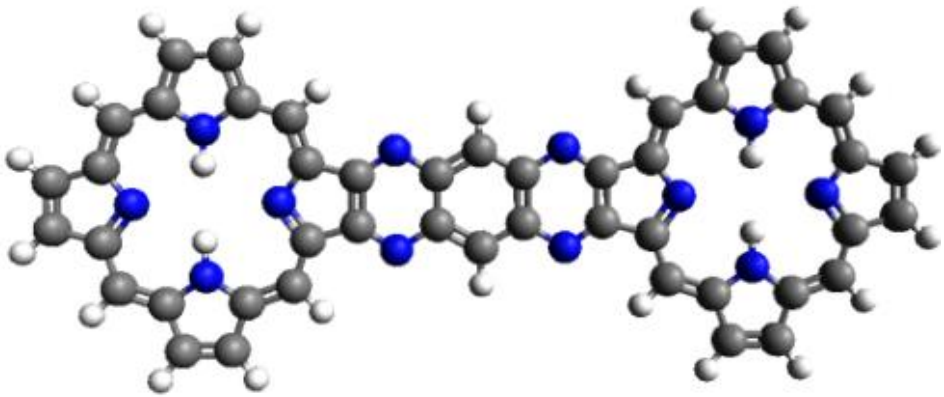
$$R_{K,1},\ r_{K,2}$$

Active-space CR-EOMCCSD(T) correction

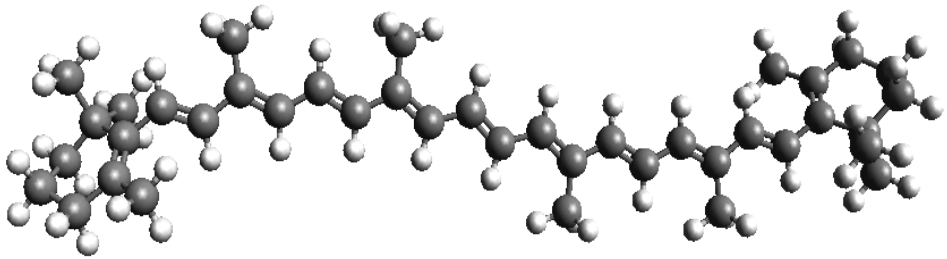
$$\delta_K(t) = \frac{\langle \Psi_K | q_3 M_{K,3} | \Phi \rangle}{\langle \Psi_K | e^{T_1+t_2}(R_{K,0} + R_{K,1} + r_{K,2}) | \Phi \rangle}$$

$$q_3 = \sum_{\substack{i<J<K \\ A<B<c}} | \Phi_{iJK}^{ABc} \rangle \langle \Phi_{iJK}^{ABc} |$$

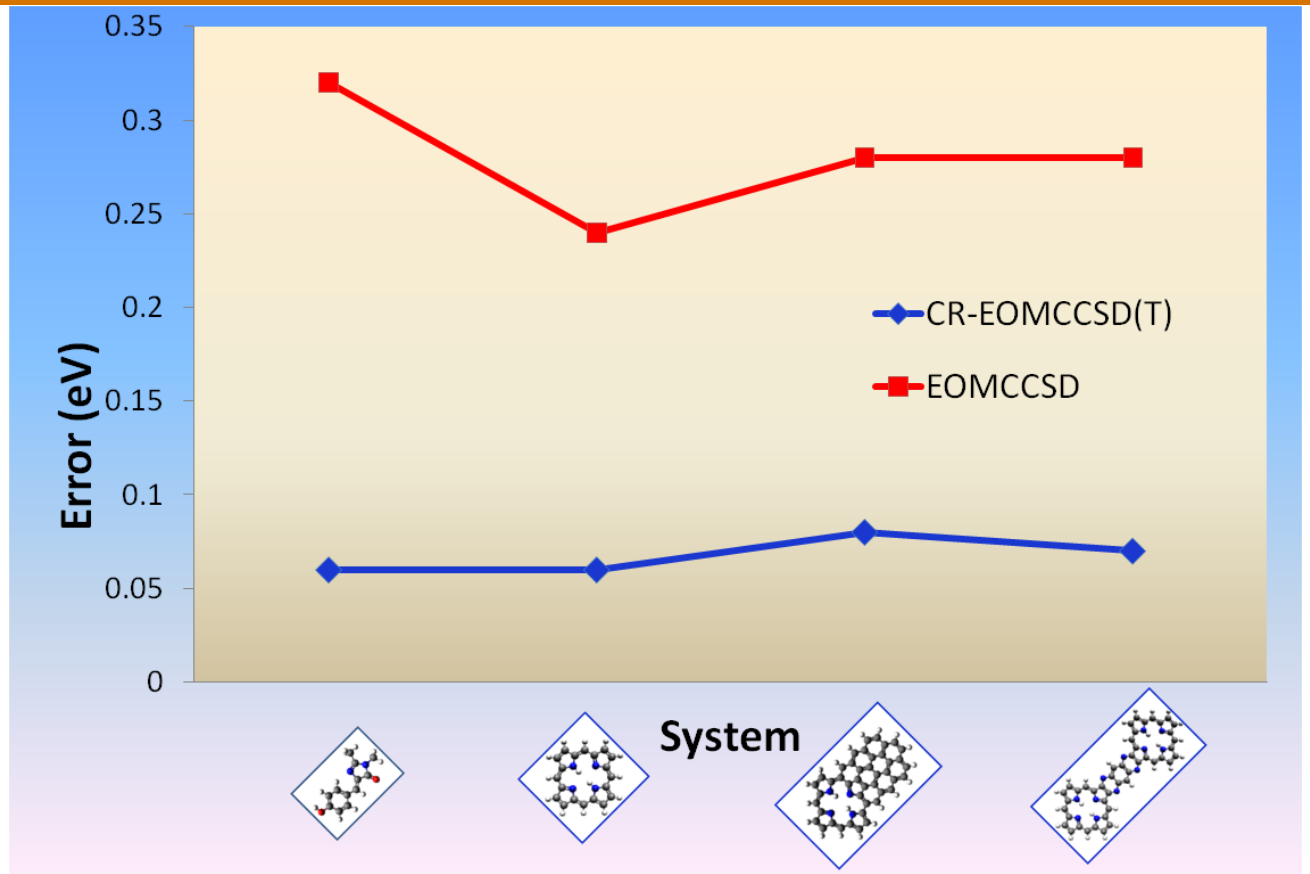
Reduced cost CR-EOMCCSD(T) methods



Basis/Method	1^1B_{3u}	1^1B_{1g}
CASPT2	**1.66**	**1.67**
cc-pVDZ CAM-B3LYP	**2.12**	**2.14**
cc-pVDZ (942 bsf)		
EOMCCSD	**2.13**	**2.14**
EOMCCSd[-1.0,1.5]	**2.19**	**2.20**
CR-EOMCCSd(t)[-1.0,1.5]	**1.97**	**1.98**
CR-EOMCCSD(T)-best estimate	**1.91**	**1.92**
Expt.	**1.84**	

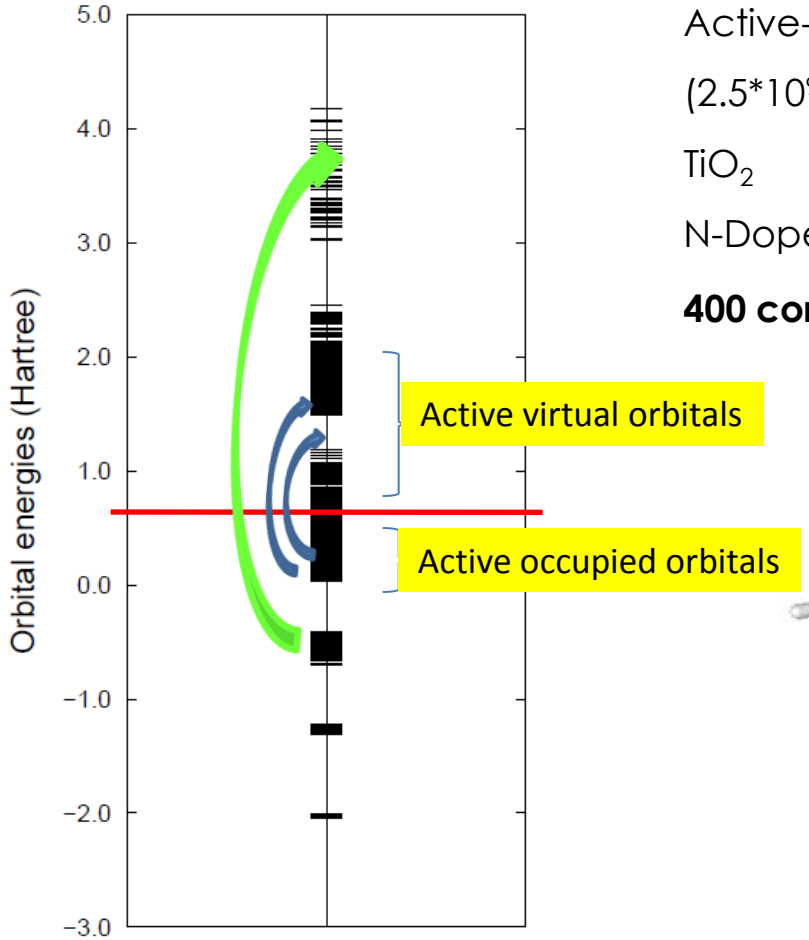


Correlation effects in the limit of large number of correlated particles



Errors of the EOMCCSD (blue) and the CR-EOMCCSD(T)/active-space CR-EOMCCSD(T) (green) vertical excitation energies with respect to the experimental data.

Multiscale CC approaches: modeling experiments



110: $Ti_{10}O_{40}(Z_O)_{60}$

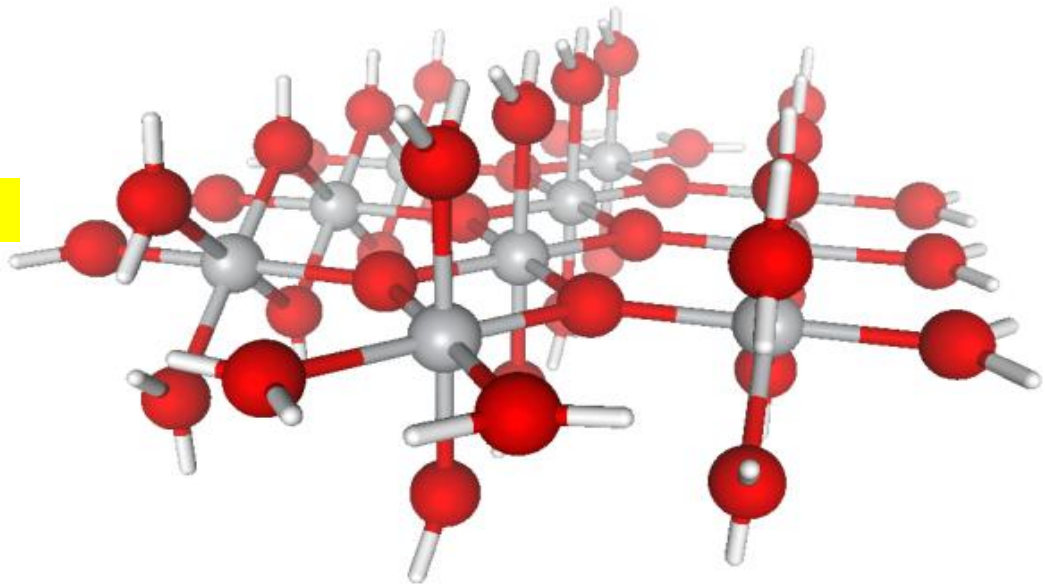
Active-space EOMCCSD method

($2.5*10^9$ doubly excited cluster amplitudes)

TiO_2 EOMCCSd: 3.69 eV

N-Doped TiO_2 EOMCCSd: 2.51 eV

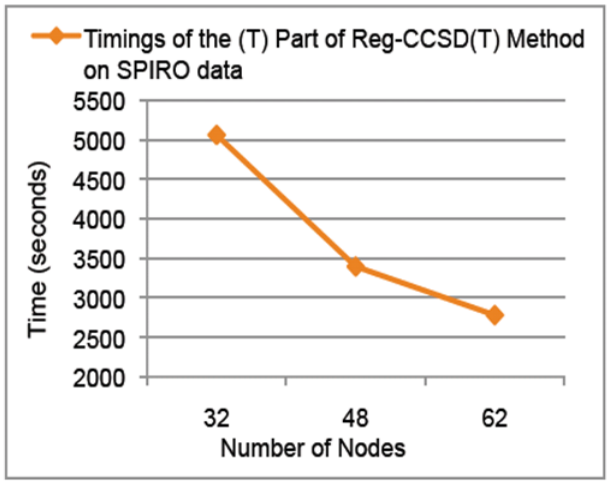
400 correlated electrons !!

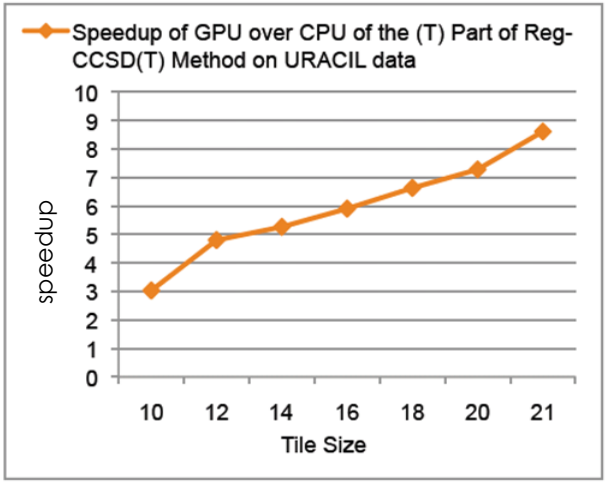


Govind, Sushko, Hess, Valiev, Kowalski, *Chem. Phys. Lett.* **470**, 353 (2009)
Govind, Rousseau, Andersen, Kowalski, *MRS Proceedings* (2010)

Towards future computer architectures (Villa,Krishnamoorthy, Kowalski)

- The CCSD(T)/Reg-CCSD(T) codes have been rewritten in order to take advantage of GPGPU accelerators
- Preliminary tests show very good scalability of the most expensive N^7 part of the CCSD(T) approach





Concluding remarks

- Significant progress in scalability of the most computationally intensive parts of the CC codes:
  - From 1K (before initiative) to 120K
- New parallel models for the iterative methods: CCSD & EOMCCSD
- GPU implementation of the CCSD(T) approach
- New theoretical models for excited-state processes:
  - New models for electron correlation effects in excited states
  - Active-space EOMCC methods – reduced numerical complexity
  - Effective Hamiltonian EOMCC formalisms – large number of excited states can be studied
- New research areas
  - Light harvesting systems – organic photovoltaic materials
  - Localized excited states in solids
  - Interaction of biological systems with radiation

Papers

- K. Kowalski, *J. Chem. Phys*. **130**, 194110 (2009).
- T.P. Straatsma, E.J. Bylaska, K. Kowalski, *SciDAC Review* **14**, 10 (2009).
- K. Kowalski, M. Valiev, *J. Chem. Phys.* **131**, 234107 (2009).
- N.Govind, P.V. Sushko, W.P. Hess, M. Valiev, K. Kowalski, *Chem. Phys. Lett.* **470**, (2009).
- Eric J. Bylaska, Kevin Glass, Doug Baxter, Scott B. Baden and John H. Weare (2009) *Hard scaling challenges for ab initio molecular dynamics capabilities in NWChem: Using 100,000 CPUs per second*, Journal of Physics: Conference Series, vol. 180, 012028, Proceedings of SciDAC 2009, San Diego, CA, 14-18 June 2009, Ed. Horst Simon.
- K. Kowalski, S. Krishnamoorthy, O. Villa, J.R. Hammond, N. Govind, *J. Chem. Phys.* 154103 (2010).
- E.J. Bylaska *et al.* *"Large Scale Plane Wave Density Functional Theory: Formalism, Parallelization, and Applications"* in *Computational Methods for Large Systems: Structure Approaches for Biotechnology and Nanotechnology,* (John Wiley Sons, Inc.) Ed. J.R. Reimers. (in press).
- K. Kowalski, *et al.* "*Coupled Cluster Methods for Large Molecular and Extended Systems,*" in "*Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*," (John Wiley & Sons, Inc.) Ed. J.R. Reimers (in press).
- W.A. de Jong, E.J. Bylaska, N. Govind, C.L. Janssen, K. Kowalski, T. Muller, I.M.B. Nielsen, H.J.J. van Dam, V. Veryazov, R. Lindh, *Phys. Chem. Chem. Phys.* **12**, 6896 (2010)
- K.R. Glaesemann, N. Govind, S. Krishnamoorthy, K. Kowalski, J. Phys. Chem. A **114**, 8764
- N. Govind, R. Rousseau,A. Andersen, K. Kowalski, 2010 Material Science Society, Meeting Proceedings **1263E** (2010).
- W. Ma, S. Krishnamoorthy, O. Villa, K. Kowalski, Cluster2010 proceedings.
- W. Ma, S. Krishnamoorthy, O. Villa, K. Kowalski, "GPU-based implementations of the regularized CCSD(T) method: applications to strongly correlated systems", *J. Chem. Theory and Computation* (submitted).

Acknowledgement

LDRD project – eXtreme Scale Computing Initiative

Marat Valiev, PNNL
(QM/MM formalilsms)

Eric Bylaska, PNNL
(Plane wave methods,parallel in time approaches)