

*Pacific Northwest National Laboratory*

*Jan 25, 2007*

**NATURAL LINEAR SCALING COUPLED-CLUSTER METHOD**  
**and some other advances for large calculations**



*National Champions, Football and Basketball!!!*

**RODNEY J. BARTLETT**

**NORBERT FLOCKE**

**TOM HUGHES**

*Quantum Theory Project*

*Departments of Chemistry and Physics*

*University of Florida*

*Gainesville, Florida USA*

**\$ NSF, AFOSR \$**

## OUTLINE

- **Our view on electronic structure theory, DFT and WFT.**
- **A few words about coupled-cluster theory.**
- **Some challenging serial applications to establish limits.**
- **Extending those limits through parallelization: ACES III.**
- **Further extension: Natural-linear scaling CC.**
- **Applications to n-alkanes, polyglycine, and others.**
- **Natural-linear scaling SCF problem.**

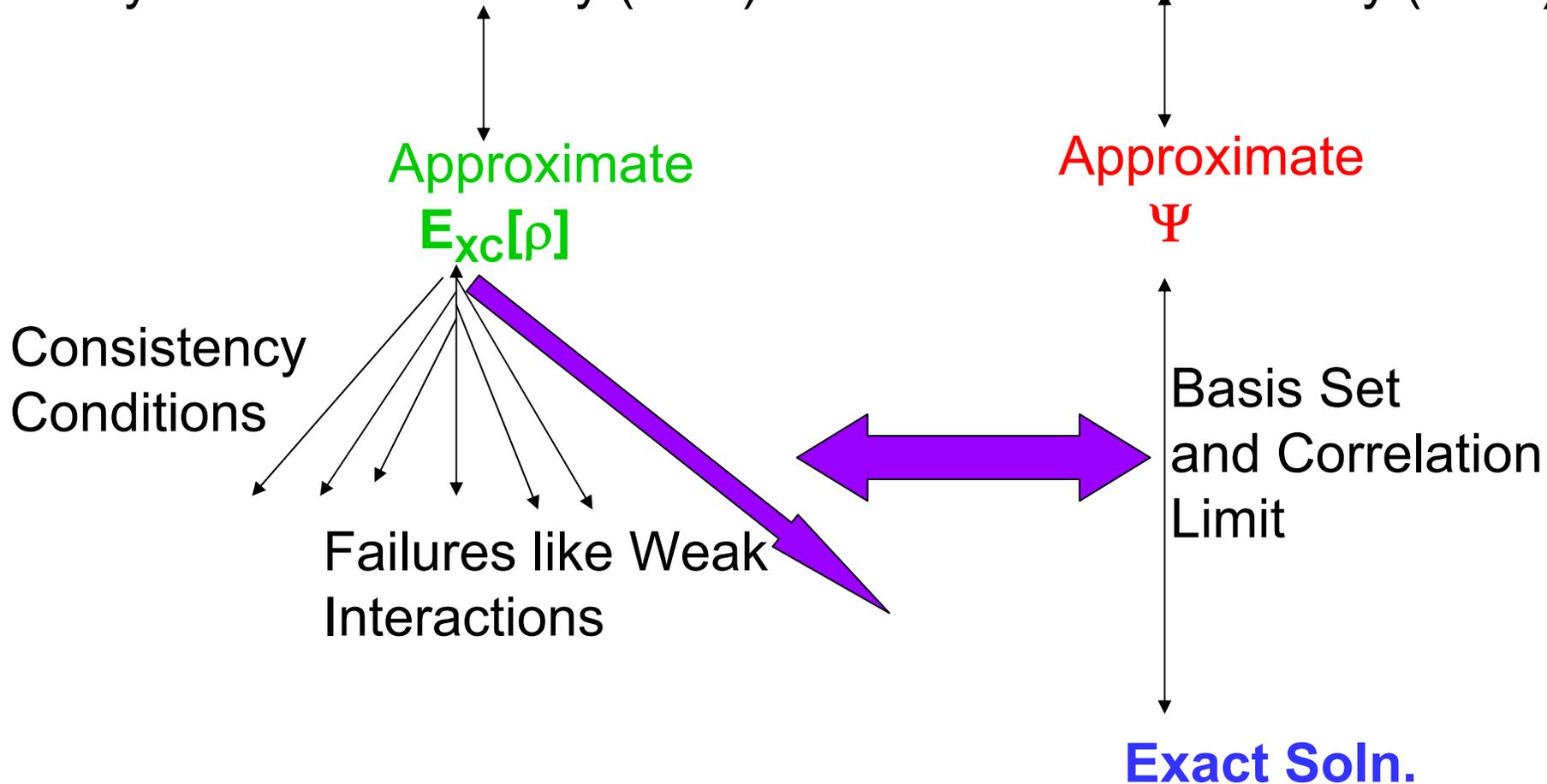
## **TWO ROUTES TOWARD ELECTRON CORRELATION IN QUANTUM CHEMISTRY**

- I. Effective one-particle theory:  
Density Functional Theory (and others like  
Dyson, etc.)**
  
- II. Explicit n-particle (2-particle theory):  
Coupled-cluster theory**

## Correlated one-particle orbital theories have....

- Significant computational advantages
- Applicable to polymers and crystalline solids
- Conceptual advantages (frontier MO theory; energy bands)
- Might expect to get principal ionization potentials and electron affinities as eigenvalues (band gaps).
- Electronic spectra (excitons) should require two-particle effects, but zeroth-order (one-particle)spectra can be improved.

# Density Functional Theory (DFT) vs. Wavefunction Theory (WFT)

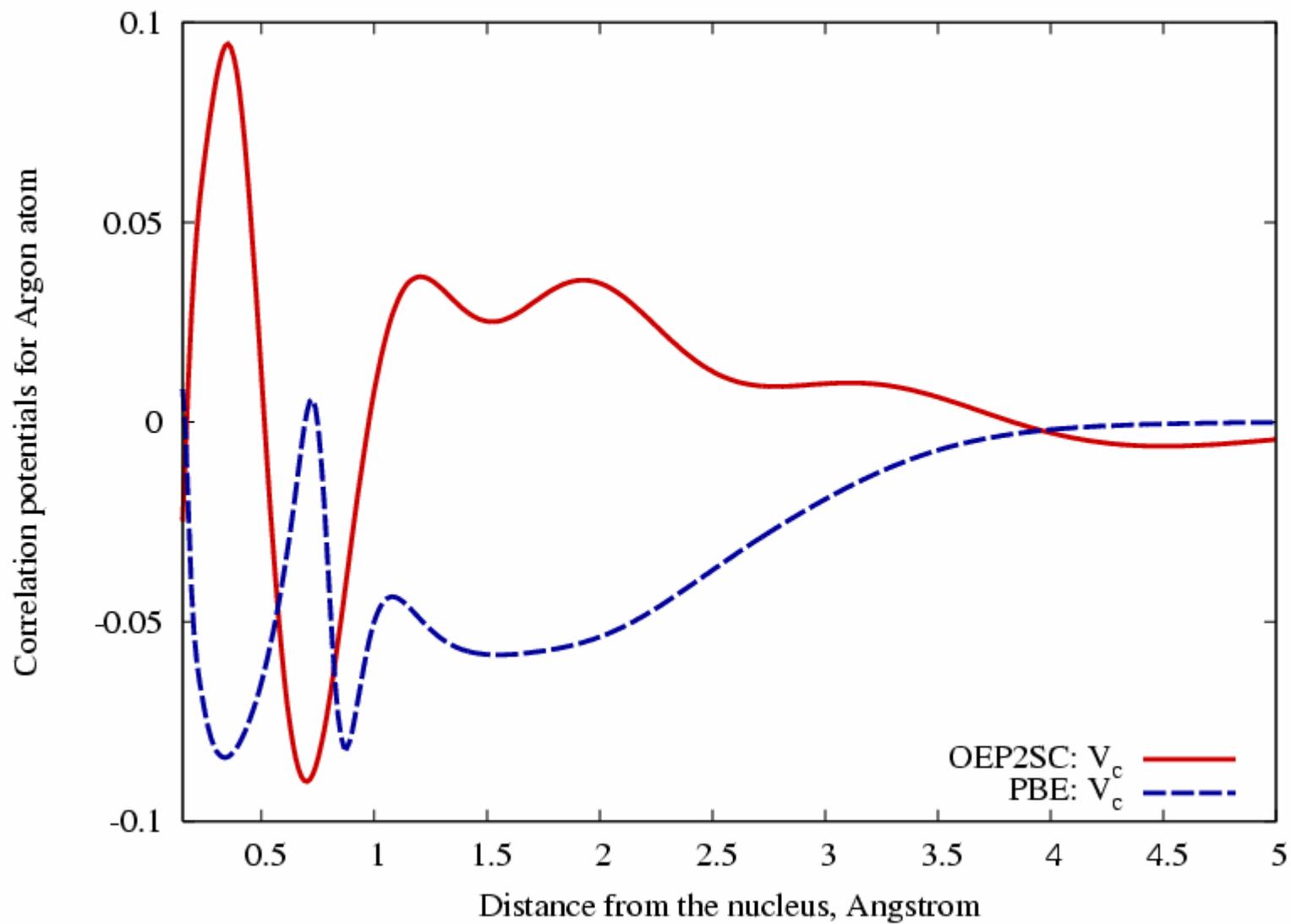


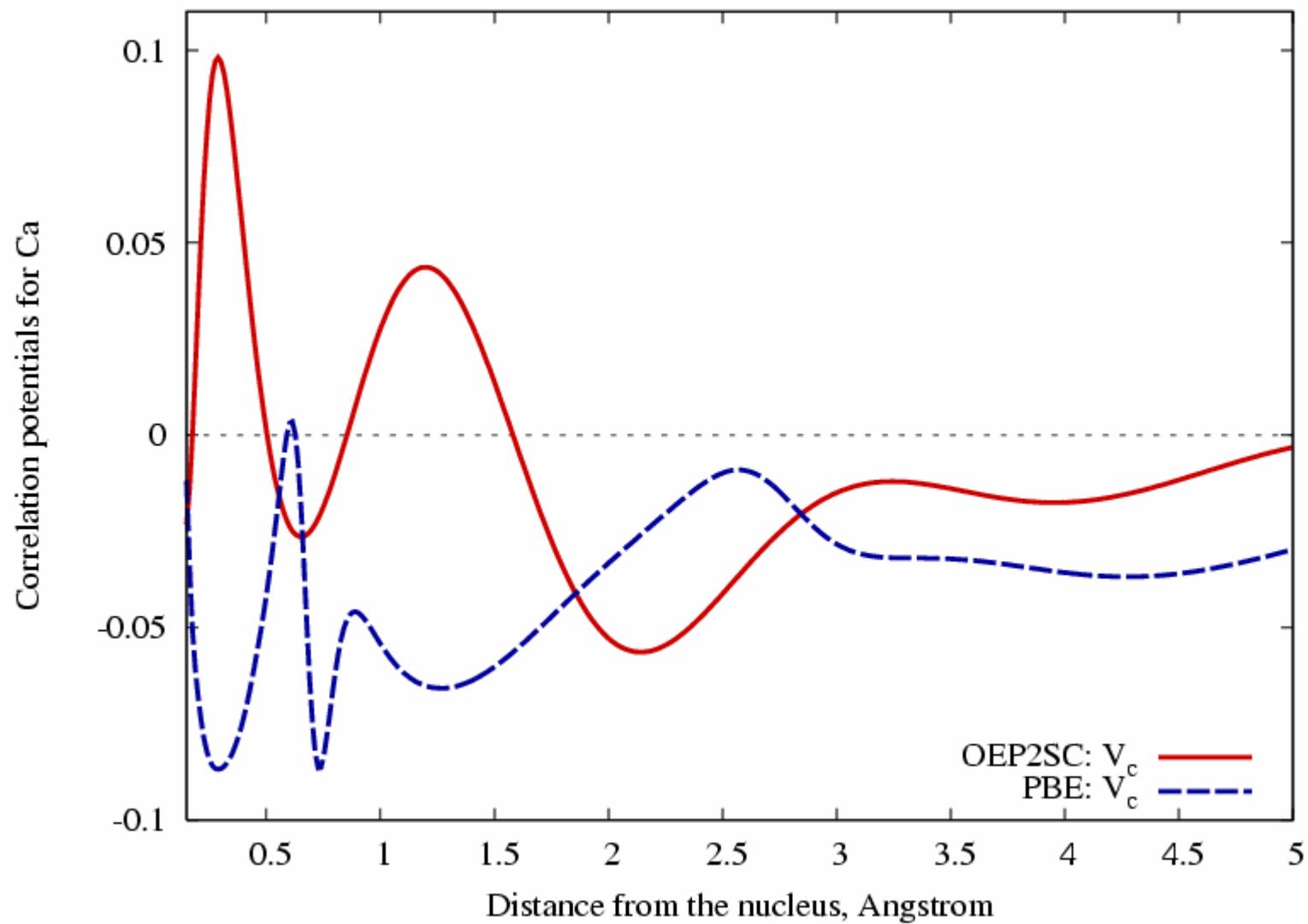
- To make the connection between WFT and DFT  $V_{xc}$  is defined by imposing the condition that the KS single determinant,  $\psi_{KS}$ , gives the exact  $\rho$ .

- THIS IS THE FUNDAMENTAL IDEA OF *ab initio dft*. Exc is an orbital dependent form taken from CC/MBPT.  $V_{xc}$  is unambiguously defined up to a constant.

It is analogous to other choices for a single determinant...

- The HF determinant,  $\psi_{HF}$ , gives the lowest SD energy
- The first natural determinant,  $\psi_N$ , gives the best SD approximation to the density matrix
- The Brueckner determinant,  $\psi_B$ , gives the best possible SD overlap with the exact wavefunction





If you do everything comparatively right in DFT (ie *ab initio dft*)---what have we learned so far? (RJB, et al,2005,2006)

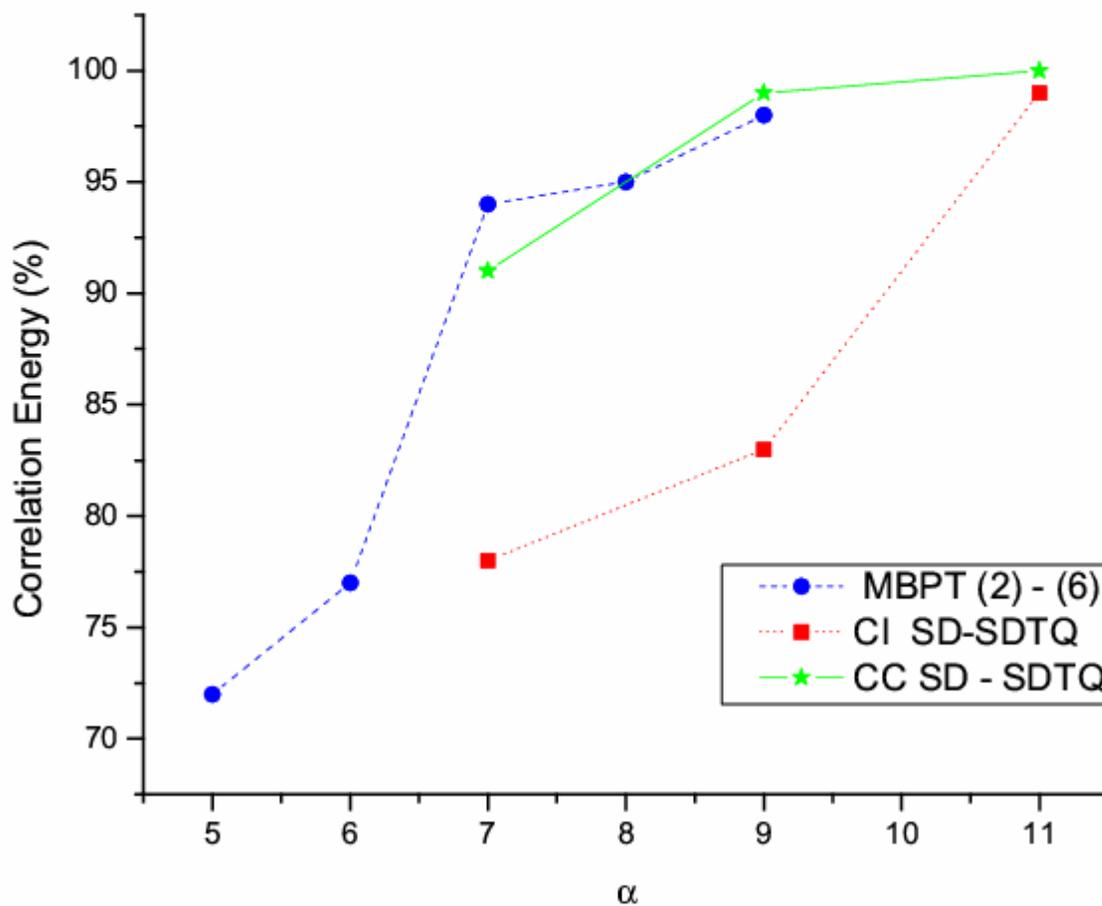
- Potentials have the correct shell structure and formally the correct asymptotic behavior.
- Self-interaction is properly handled.
- Dispersion interactions are correctly described.
- DFT orbital energies are given meaning by an analog of Koopmans' approximation for *all* ionization potentials.
- There is no integer discontinuity problem.
- DFT unoccupied orbital energies ( $\epsilon_a$ ) enable a reasonable zeroth-order approximation for excitation energies to be  $(\epsilon_a - \epsilon_i)$ , as in naïve Hückel theory.
- Results are usually better than that of standard DFT, and are in many cases competitive with coupled-cluster theory, even when using only a MBPT(2) functional.
- Using higher-order coupled-cluster functionals further improves the potentials, demonstrating convergence.

**Now let's return to CC theory...**

$$\star = \exp(T) \rightarrow 0$$

- SIZE-EXTENSIVE (No unlinked diagrams).
- RAPID SATURATION OF DYNAMIC CORRELATION
- CONNECTED EXPRESSIONS FOR AMPLITUDES (No CI evaluation.)

Performance of theories for the correlation energy in small molecules. To facilitate comparisons, the ordinate gives the size-scaling parameter of the approximation,  $\alpha = \alpha_n + \alpha_N + \alpha_{it}$  in the computational cost function  $n^{\alpha_n} N^{\alpha_N} N_{it}^{\alpha_{it}}$ .



# AB INITIO, CORRELATED, SIZE EXTENSIVE WAVEFUNCTION METHODS COMBINED WITH CONVERGING BASIS SETS HAVE ...

Established the now widely used paradigm for ground state

## *Ab Initio* Calculations

MBPT(2) < CCD < CCSD < CCSD[T] < CCSD(T) < CCSDT-1

1978

1982

1985

1989

1984

< CCSDT < CCSDT(Q<sub>f</sub>) < CCSDTQ < **CCSDTQP** < FULL CI

1987

1998

1992

2002

Automatic Code Generation by Hirata, Kallay, Olsen adds H

## OTHER REQUIRED EXTENSIONS FOR GENERALLY APPLICABLE THEORY

- Analytical gradients for CC Theory (1984-1991)
- CC Functional,  $E = \langle 0 | (1 + \text{thumbs up}) \exp(-T) H \exp(T) | 0 \rangle$
- $\partial E / X_\alpha = \langle 0 | (1 + \text{thumbs up}) \exp(-T) (\partial H / \partial X_\alpha) \exp(T) | 0 \rangle$ .
- CC generalization for response and relaxed density matrices and properties,  $\gamma_{pq} = \langle 0 | (1 + \text{thumbs up}) \exp(-T) p^\dagger q \exp(T) | 0 \rangle$
- EOM-CC Excited, Ionized, Electron Attached, ... States,  $[e^{-T} H e^T R_k]_C | 0 \rangle = \epsilon_k R_k | 0 \rangle$  (1993-1995)
- Excited state density matrices,  $\gamma_{pq}^k = \langle 0 | L_k \exp(-T) p^\dagger q \exp(T) R_k | 0 \rangle$  (1993)
- Second-order properties, (1993, 1994)
- $\text{smiley}(\mathbf{r}, \mathbf{r}) = \sum_k \langle 0 | (1 + \text{thumbs up}) e^{(-T)} (\mathbf{r} - \langle \mathbf{r} \rangle) e^{(T)} R_k | 0 \rangle \langle 0 | L_k e^{(-T)} (\mathbf{r} - \langle \mathbf{r} \rangle) e^{(T)} | 0 \rangle / \epsilon_k$

Evolving paradigm for CC/MBPT approaches for excited  
(ionized, electron attached) states -

EOM-CCSD <STEOM-CC< EOM-CCSDt<EOM-CCSDT-3

1984-1993

1997

1999

1996

<EOM-CCSDT<EOM-CCSDTQ<FULL CI

2001

NWChem (Hirata)

**ALL THE ABOVE GRADIENT, DENSITY MATRICES,  
PROPERTIES, EXCITED STATE, CAPABILITIES WERE  
WRITTEN INTO ACES II, 1991-1995.**

**If we were having this meeting ~20 years ago...**

**the emphasis would have been on CI, not CC.**

**And HF, not GGA/Hybrid DFT, for large molecules**

So when we're devoting so much effort to implementing 10~20 year-old theory for petascale computing, let's be aware that if it's not coupled to the development of NEW theory, the programs will soon become obsolete!

**So what's the next objective for CC?**

**Larger molecules, biomolecules, polymers, crystals.**

# Correlation Treatment of Extended Systems

## PROBLEM:

**Exponential Scaling Wall with # of basis functions  $M$**

$$\text{MP2} \rightarrow M^5$$

$$\text{CCSD} \rightarrow M^6$$

$$\text{CCSDT} \rightarrow M^8$$

...

## Challenge Class Calculations

- Excited States of Cyclopentadiene anion; 580 AO basis functions in  $C_{2v}$  symmetry. 36 electrons, 18 occupied orbitals. 5 dropped for correlation.
- ia64 (SGI, Altix), Linux, PG compilers
- At least 4.8 Gb of memory and 120 Gb Disk Space

	Wall clock time (hrs)
SCF	7
H formation (CC and Lambda)	94
Excited states	72

## CHALLENGING SERIAL APPLICATION

- Excited States of  $\text{Cr}(\text{CO})_6$ : 627 AO basis functions in  $D_{2h}$  symmetry. 108 electrons, 54 occupied.

(Villaume, Strich, Daniel, Strasbourg; Perera, RJB, Florida, NSF International grant)

- X86\_64 (OPTERON), Linux, Intel compilers
- At least 4 Gb of memory and 120 Gb Disk Space

	Wall clock time (hrs)
SCF	51
H formation (CC and Lambda)	351
Excited states	450

**So this is about the limit without parallelization...**

## **SOME NEW DEVELOPMENTS FOR ACES III...(Partly supported by CHSSI)**

- New, fast, integral and integral derivative program.
- Integral derivatives are open-ended
- Integral direct capabilities for HF, DFT, MBPT, CC to avoid storage bottlenecks
- Parallel implementation built upon new language, SIAL, (Super Instruction Assembler Language) that removes details of memory handling and message passing from the quantum chemical programmer.

**Dr. Victor Lotrich, QTP/ACES Q.C.**

**(Automatic CC Program Generation)**

**Dr. Norbert Flocke QTP (Integral Program)**

**Mr. Mark Ponton, ACES Q.C. (SIAL)**

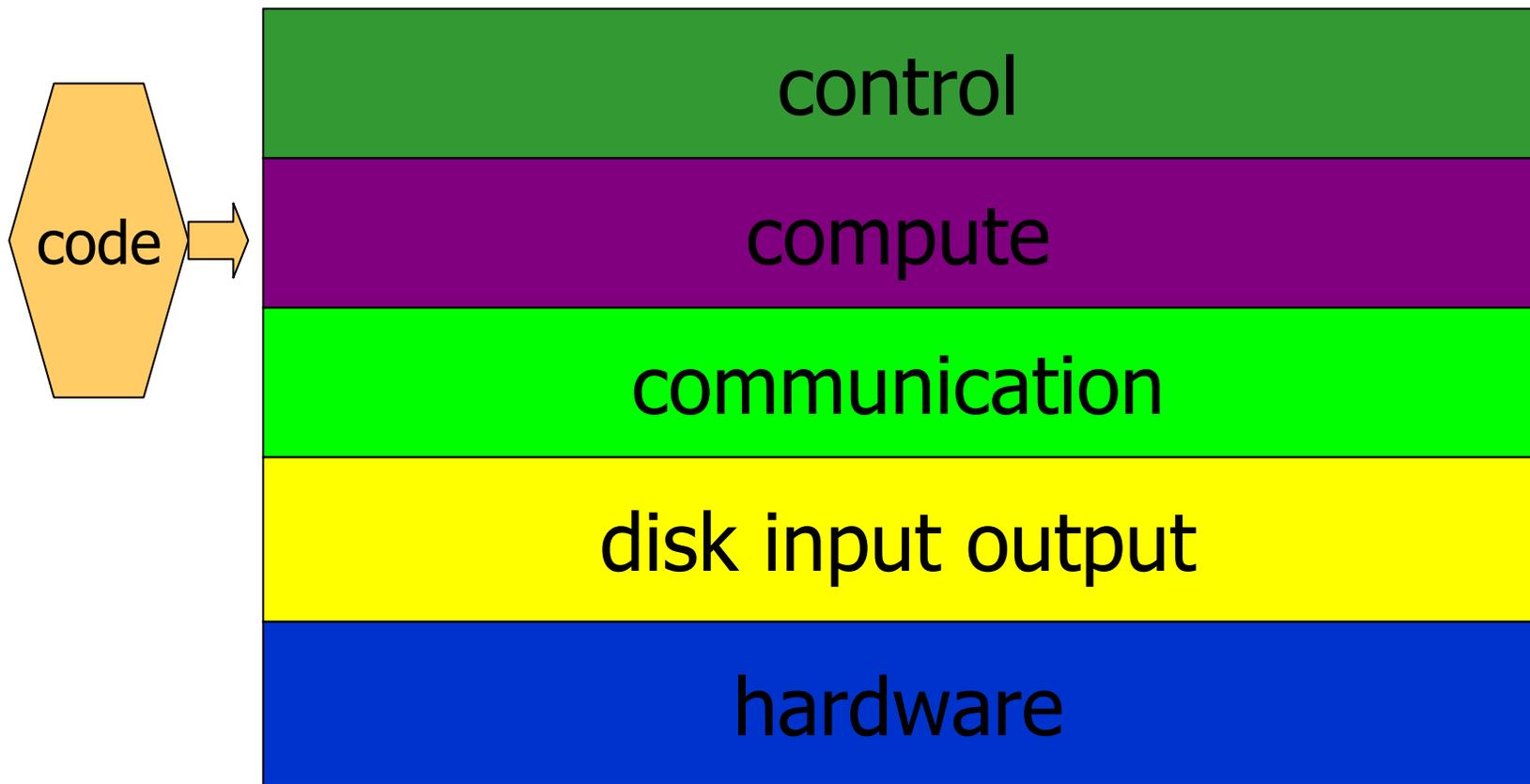
**Dr. Erik Deumens, QTP (Architect)**

**Dr. Ajith Perera, QTP**

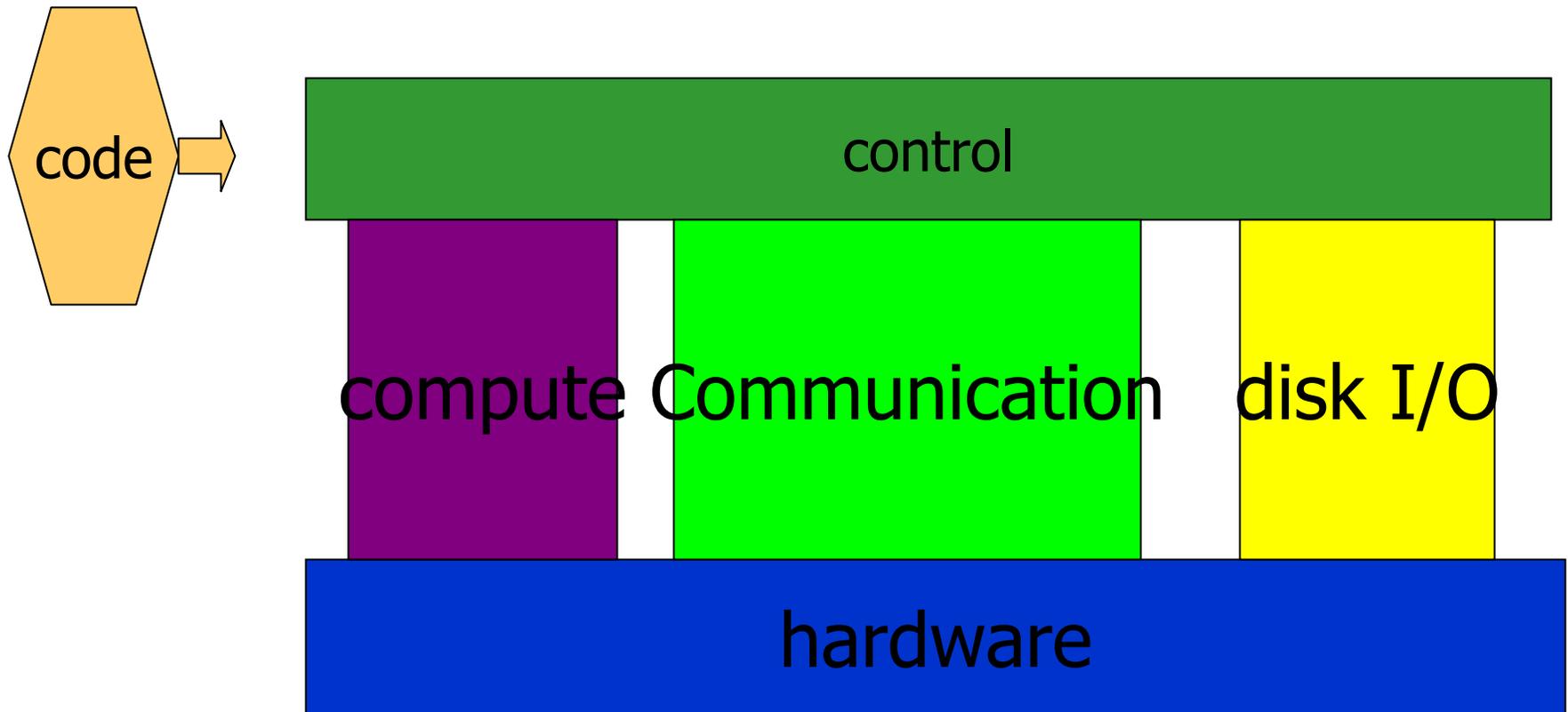
**Mr. H. Lei, ACES Q.C. (Compiler)**

**Dr. Anthony Yau, Aces Q.C. (HPTi)**

# *Traditional Design*



# *ACES III Design*



## *Aces III design*

- Extreme object oriented approach
- High level = problem domain specific
  - Concepts
  - Data structures
  - Algorithms
- Low level = focus on performance
  - Processor and memory speed
  - Communication latency and bandwidth

## *Super Instruction Assembly Language Compiler*

- The high level code is written in a simple language SIAL
  - Now it does linear algebra for ACES III
  - Some things like PARDO are general
  - It supports definition of “special instructions”
  - Domain specific structures and actions need to be defined
- The compiler produces “super assembly” code

# *Super Instruction Processor*

- The low level code is written in Fortran 77 and C
  - Executes “super instructions”
  - Read from the “super assembly” code binary
  - Is a MPMD (multiple program multiple data) parallel program
  - Uses MPI (message passing interface) with POSIX threads or shmem

*Ar<sub>4</sub> 36+164=200 bf 64 processors*

Machine	SCF	trans	CCSD 1 iteration
IBM P4 shelton	82 s	776 s	1,431 s .4 h
Compaq emerald	53 s	2,957 s	6,997 s 1.9 h

*Ar<sub>6</sub> 54+246=300 bf 64 processors*

Machine	SCF	trans	CCSD 1 iteration
IBM P4 shelton	313 s	4,242 s	16,363 s 4.5 h
Cray X1 diamond	582 s	6,452 s	19,601 s 5.4 h
Compaq emerald	132 s	4,180 s	29,188 s 8.1 h

## *Cray X1 on 64 processors*

Basis functions	SCF	trans	CCSD 1 iteration
Ar <sub>4</sub> 200 36+164	4,535 s <i>X1 busy</i>	26,871 s <i>X1 busy</i>	30 h <i>X1 busy</i>
Ar <sub>6</sub> 300 54+247	582 s	6,452 s	5.4 h
Ar <sub>10</sub> 500 90+410	2,810 s	32,855 s	77 h

# *IBM SP4 on 240 processors*

## *Real-world application*

- DEMP (Diethyl Methyl Phosphonate) + OH transition state
- 268 basis functions
- Transition state search converges in 5 iterations
- Timings in seconds
- CCSD Gradient in less than a day

## *IBM SP4 on 240 processors*

Code	1	2	3	4	5
SCF	808	762	728	791	795
trans	216	203	202	215	214
CCSD	19846	20085	20451	19558	20746
lambda	11036	10806	11834	10609	13667
1-grad	3362	3465	3415	3327	3471
2-grad	1156	1608	1535	1482	1577

## *IBM SP4 scaling*

- DMMP (Dimethyl Methyl Phosphonate) gradient
- The MBPT(2) gradient runs:
  - CC-PVTZ basis 397 functions
- The CCSD and LCCSD runs:
  - CC-PVDZ basis 208 functions

## *IBM SP4 scaling MBPT(2)*

proc	SCF	speedup	MP2 gradient	speedup
2	32281	<b>2.00</b>	-	-
4	13377	<b>4.83</b>	-	-
8	7508	<b>8.60</b>	41352	<b>8.00</b>
16	8555	<b>7.55</b>	22246	<b>14.87</b>
32	4658	<b>13.86</b>	12208	<b>27.10</b>
64	2971	<b>21.73</b>	5716	<b>57.88</b>
128	2017	<b>32.01</b>	2770	<b>119.43</b>

## *IBM SP4 scaling*

- SCF performance:
  - Notice the super linear scaling for SCF for 4 and 8 processors
  - Scaling trails off quickly, mostly for lack of work
  - Compare the timings for SCF in CCSD and LCCSD runs: the SCF is the same, timing difference show 5% to 40% variability on a normal system

## *IBM SP4 scaling CCSD*

proc	SCF	speedup	CCSD	speedup
2	5441	<b>2.00</b>	-	-
4	2066	<b>5.27</b>	-	-
8	1160	<b>9.38</b>	-	-
16	1129	<b>9.64</b>	64055	<b>16.00</b>
32	745	<b>14.61</b>	33139	<b>30.93</b>
64	649	<b>16.77</b>	16302	<b>62.87</b>
128	628	<b>17.33</b>	8808	<b>116.36</b>

## *IBM SP4 scaling LCCSD*

proc	SCF	speedup	LCCSD	speedup
2	5452	<b>2.00</b>	-	-
4	3066	<b>3.56</b>	-	-
8	2118	<b>5.15</b>	-	-
16	1687	<b>6.46</b>	37018	<b>16.00</b>
32	1014	<b>10.75</b>	12413	<b>47.72</b>
64	621	<b>17.56</b>	5267	<b>112.45</b>
128	624	<b>17.47</b>	2331	<b>254.09</b>

## *IBM SP4 scaling*

- CCSD performance:
  - Scaling is excellent (91%) up to 128 processors.
- LCCSD performance:
  - Scaling is super linear to the extreme with 198% for 128 processors. This is caused by the fact that less and less data needs to be stored using the I/O servers and eliminating the I/O delays causes extra speedup.

# Correlation Treatment of Extended Systems

## PROBLEM:

**Exponential Scaling Wall with # of basis functions  $N$**

$$\text{MP2} \rightarrow N^5$$

$$\text{CCSD} \rightarrow N^6$$

$$\text{CCSDT} \rightarrow N^8$$

...

**Some "Clever Approaches" needed !**

## CCSD calculations with periodic boundary conditions....

- Hirata, Grabowski, Tobita, and RJB, CPL 345, 475 (2001).
- Hirata, Podeszwa, Tobita, and RJB, JCP 120, 2581 (2004).

Applications to polyethylene, polyacetylene, and polyynes,  $(\text{HF})_\infty$ ,  $(\text{Be})_\infty$   
(STO-3G to 6-31G\*)

- Related DMRG studies, Hachmann, Cardoen, Chan, JCP 125, 144101 (2006)  
(STO-3G for polyenes, STO-5G,  $\text{H}_2$  chains)

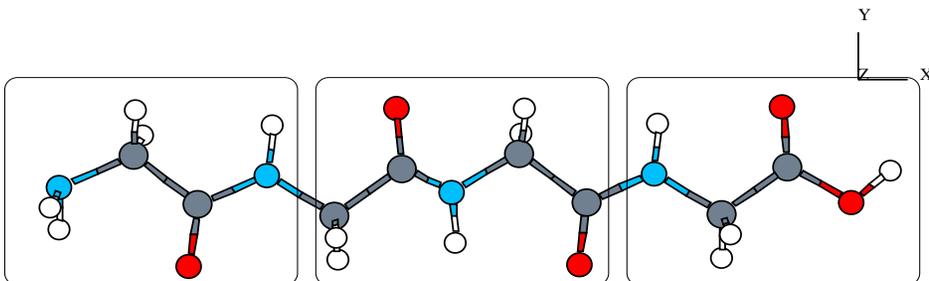
## Incremental methods for the energy based upon cluster calculations

- Stoll, Phys. Rev. B, 46, 6700 (1992).
- Paulus, Rosciszewski, Gaston, Schwerdtfeger, and Stoll, Phys. Rev. B 70, 165106 (2004).
- Flocke and RJB, JCP 118, 5326 (2003)
- Deev and Collins, JCP 122, 154102 (2005).

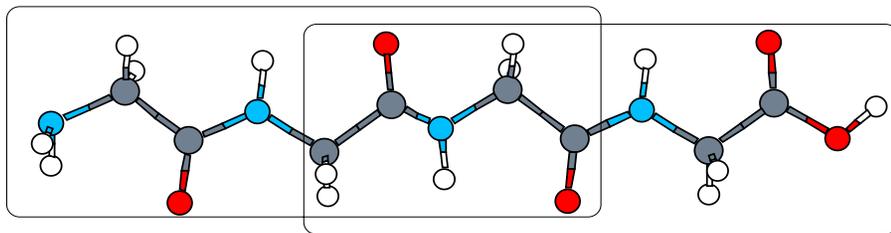
# Main Approaches for Correlation Treatment of Extended Systems

**FMOCC** : Fragment Molecular Orbital Coupled Cluster Method by *Kitaura et. al.*

Active



$$\Delta E_1^{corr} = \sum_I \Delta E_I^{corr}$$



$$\Delta E_2^{corr} = \sum_{I>J} (\Delta E_{IJ}^{corr} - \Delta E_I^{corr} - \Delta E_J^{corr})$$

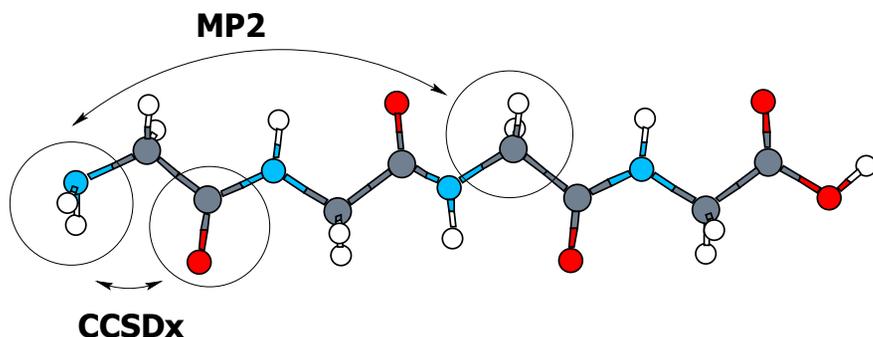
## Features

- No large scale SCF.
- Good accuracy > 99% using dimers.
- Biological size molecules possible.
- No correlated wavefunction.

**Reference:** D.G.Fedorov and K.Kitaura, JCP **123**, 134103 (2005).

# Main Approaches for Correlation Treatment of Extended Systems

**LCCSDx** : Localized Coupled Cluster Method by *H.J.Werner et. al.*



## Solve CCSDx equations:

$t$  and  $\int$  between distant domains = 0  
 $t$  and  $\int$  between medium domains: MP2  
 $t$  and  $\int$  between close domains: CCSDx

## Requirements:

localized  $t$  and  $\int$  in LMO basis

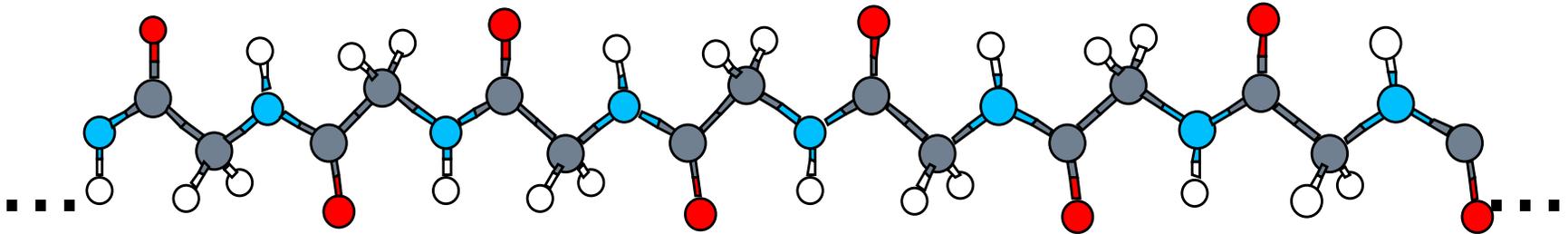
## Features

- Large scale, global SCF.
- Direct AO to LMO  $\int$  transform.
- Molecule size restricted by SCF.
- Non-orthogonal virtual space  $\rightarrow$  **S**
- Correlated wavefunction available.

**Reference:** M.Schütz and H.J.Werner,  
 JCP **114**, 661 (2000).

## Example : CC Correlation Energy of Polyglycine (GLY)<sub>x</sub>

Transferability that exists in chemistry suggests an ansatz...



$$\Psi(\text{polygly}) \cong \prod_P \psi_P(\text{electron pair or other unit})$$

$$E(\text{polyglycine}) \cong \sum_P E_P$$

$$\psi_P(\text{electron pair}) \cong \exp(T_P)|0\rangle$$

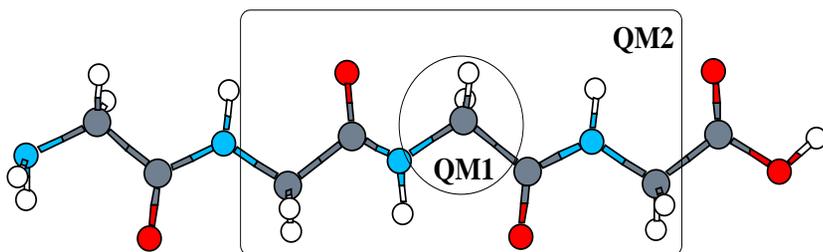
$$T_P = T_1(P) + T_2(P)$$

In general, these T's could be known in terms of

**localized orbitals** *after* the calculation of the whole system. So build an ansatz upon that expectation.

# New Approach for the Correlation Treatment of Extended Systems

**NLSCC** : Natural Linear Scaling Coupled Cluster Method



## Ansatz:

Target QM info  $\rightarrow$  QM1 region  
 QM2 is proper QM neighborhood for QM1  
 Solve QM2 sized CC equations only  
 Extract  $t$  and  $f$  info for QM1 only

## Requirements:

localized  $t$  and  $f$  in LMO basis  
 transferable  $t$  and  $f$  in LMO basis

## Features

- No large scale SCF.
- Biological size molecules accessible.
- Well suited to polymers.
- Correlated wavefunction /density matrices available.

**Reference:** N.Flocke and R.J.Bartlett,  
 JCP **121**, 10935 (2004).

## THE NLSCC ANSATZ

$$\Psi = \exp(T)|0\rangle$$

Introduce **localized orbitals** for occupied and virtual space....

$$\Psi \cong \prod_P \exp(T_P)|0\rangle$$

$$\Psi \cong \prod_P \prod_Q \exp(T_P)|0_Q\rangle$$

Where the molecule's reference function is

$$|0\rangle \cong \prod_Q |0_Q\rangle \quad P \subset Q$$

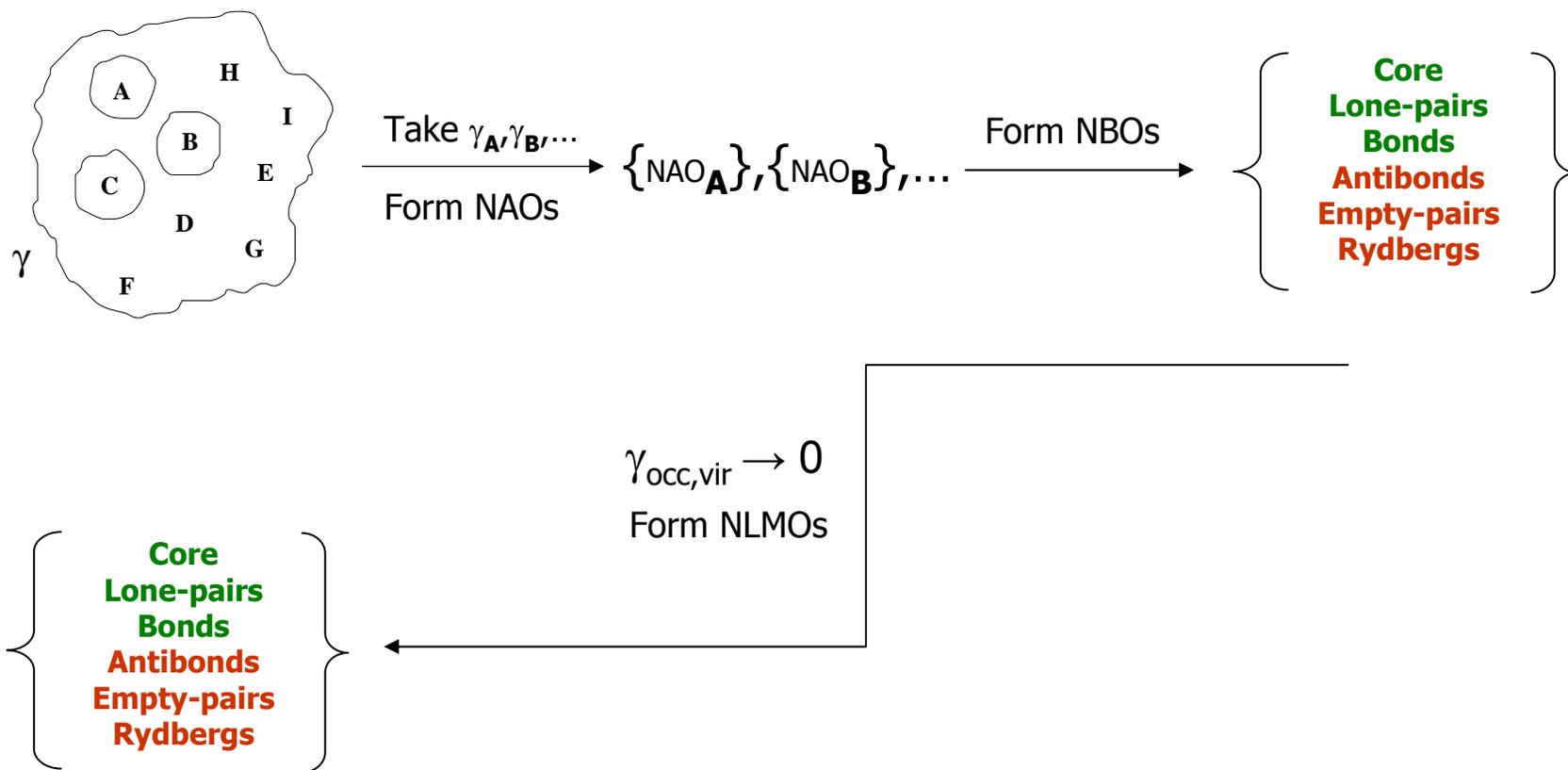
We obtain  $T_Q$  from  $\exp(T_Q)|0_Q\rangle$ , with atoms capped by H or a pseudopotential, and extract from  $T_Q, T_{P1}, T_{P2} \dots$  as 'local' transferable units, with 'local' correlation energies.

$$P=QM1, Q=QM2$$

## Requires...

- that correlation be 'short-sighted' making amplitudes and integrals fall off quickly ( $\sim 1/r^3$ )
- that units be largely transferable

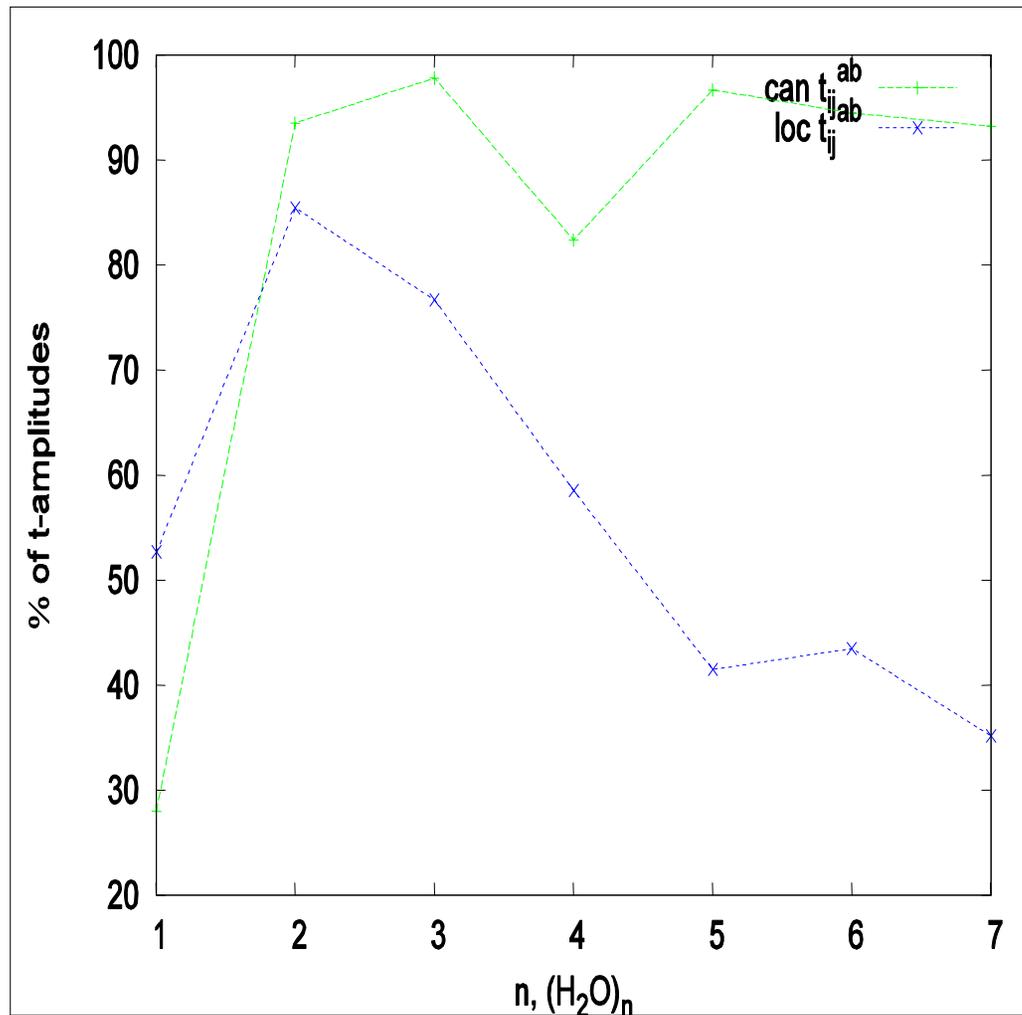
# Natural Localized MO Procedure



**Reference:** A.Reed, L.A.Curtiss and F.Weinhold,  
Chem. Rev. **88**, 889 (1988).

# Decay of second-order doubles and triples amplitudes in NLMOs for water clusters

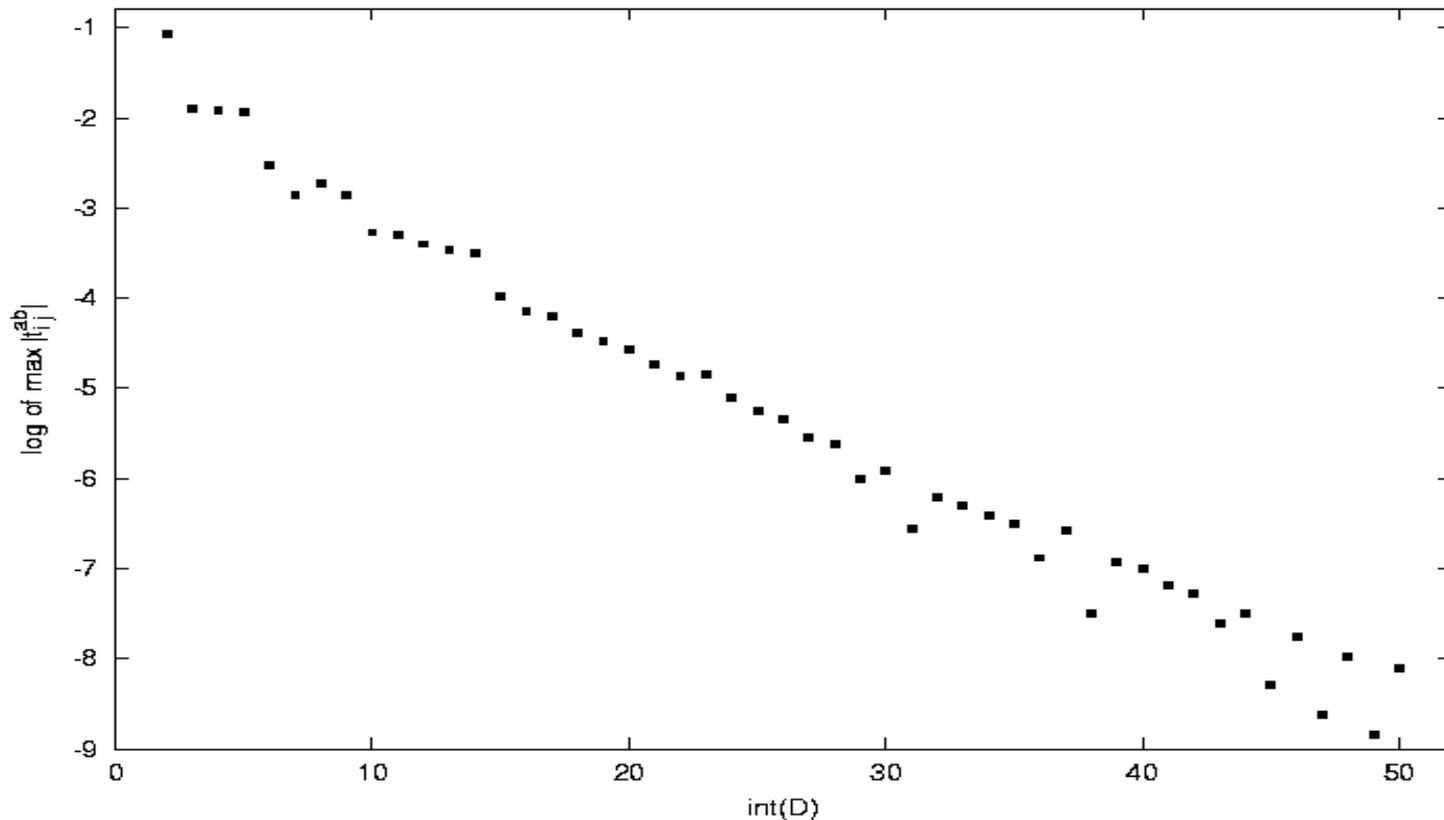
*For  $t \geq 5.0 \times 10^{-08}$*



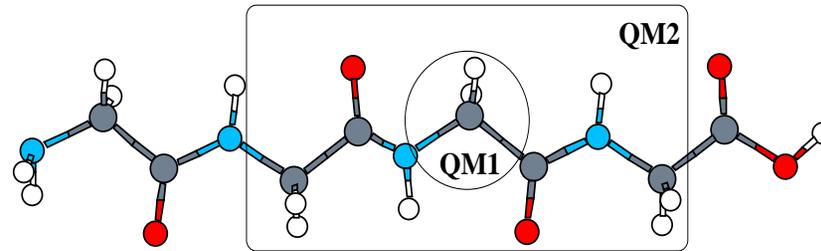
RHF, cc-pVDZ, bondsize = 2,  $C_1$  Symmetry

# Exponential decay of $T_2$ amplitudes

Example: Logarithmic plot of maximum of  $T_2$  amplitudes for each subset of  $T_2$  amplitudes defined in term of their integer parts  $\text{int}(D)$  of the corresponding sum  $D = d_{ia} + d_{jb}$  of their average NBO excitation distances  $i \rightarrow a$  and  $j \rightarrow b$  in the model paraffin  $C_{11}H_{24}$  using the 4-31G basis set.



# System $\phi_{ref}$ and $\psi$ in the NLSCC Method



## NLS Ansatz:

For each LMO  $i \in \text{QM1}$   $\left\{ \begin{array}{l} t_{ij..}^{ab..} \approx 0 \text{ if at least one of } j..ab.. \notin \text{QM2} \\ t_{ij..}^{ab..} = \text{transferable: } t_{ij..}^{ab..}(\text{QM2}) \approx t_{ij..}^{ab..}(\text{system}) \end{array} \right.$

$$T_n = \frac{1}{n!} \sum_{ij..,ab..}^{\text{system}} t_{ij..}^{ab..} a^\dagger b^\dagger \dots ji$$

$$\psi = e^{T_1 + T_2 + T_3 + \dots} \phi_{ref}$$

# NLS CC Correlation Energy

$$\Delta E = \sum_{ijab}^{system} f(t_i^a, t_{ij}^{ab}, \langle ij || ab \rangle) \quad \leftarrow \text{exact}$$

$$\Delta E = \sum_{QM1} \left[ \sum_{\substack{i \in QM1 \\ jab}}^{system} f(t_i^a, t_{ij}^{ab}, \langle ij || ab \rangle) \right] \quad \leftarrow \text{exact}$$

$$\Delta E \approx \sum_{QM1} \left[ \sum_{\substack{i \in QM1 \\ jab}}^{QM2} f(t_i^a, t_{ij}^{ab}, \langle ij || ab \rangle) \right] \quad \leftarrow \text{approx}$$

$$\Delta E \approx \sum_{QM1} \left[ \sum_{i \in QM1}^{QM2} \Delta E_i \right]$$

transferable  $t$  and  $\int \rightarrow$  transferable  $\Delta E_i$

**NLS Ansatz**

# QM1 vs QM2

We have a matrix of **QM1** choices and **QM2** regions.  
For example, for polyglycine, **QM1** could be

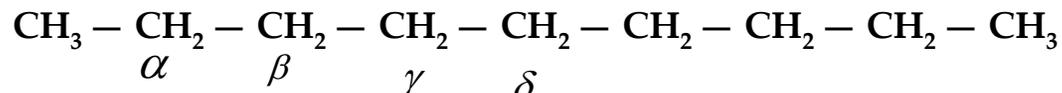
- (1) **C=O, N-H, N-C, C-H**. Electron pair bonds
- (2) **H-N-C=O, C-H**. Peptide Unit + Bond
- (3) **H<sub>2</sub>C-(NH)-C=O**. Combined Unit

**QM2** would be the complement consisting of 1,2,3,... $\infty$  units, plus **QM1**

**Self-consistency** would be achieved when the amplitudes obtained from **QM1** units are used to replace those units in **QM2**, until there is no change in the results for the **QM1** unit.

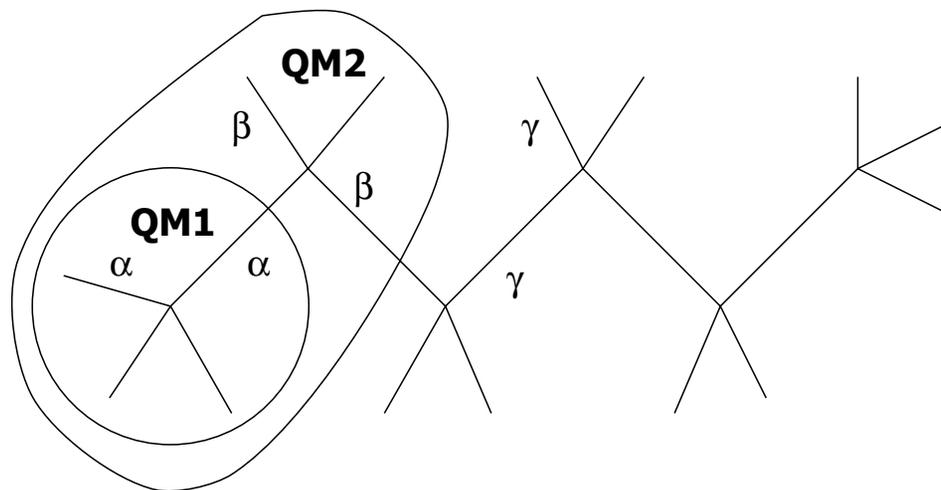
# Transferability of $T_2$ amplitudes

Example: carbon-carbon  $\sigma \rightarrow \sigma^*$  NBO excitation  $T_2$  amplitudes for model paraffins  $C_xH_{2x+2}$ . Greek letters indicate excitation location of bond/antibond pair inside each paraffin.



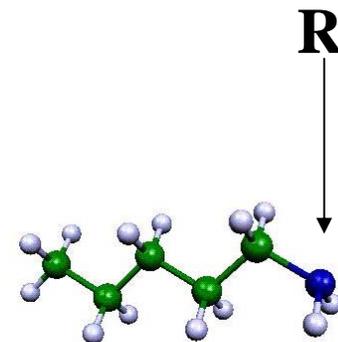
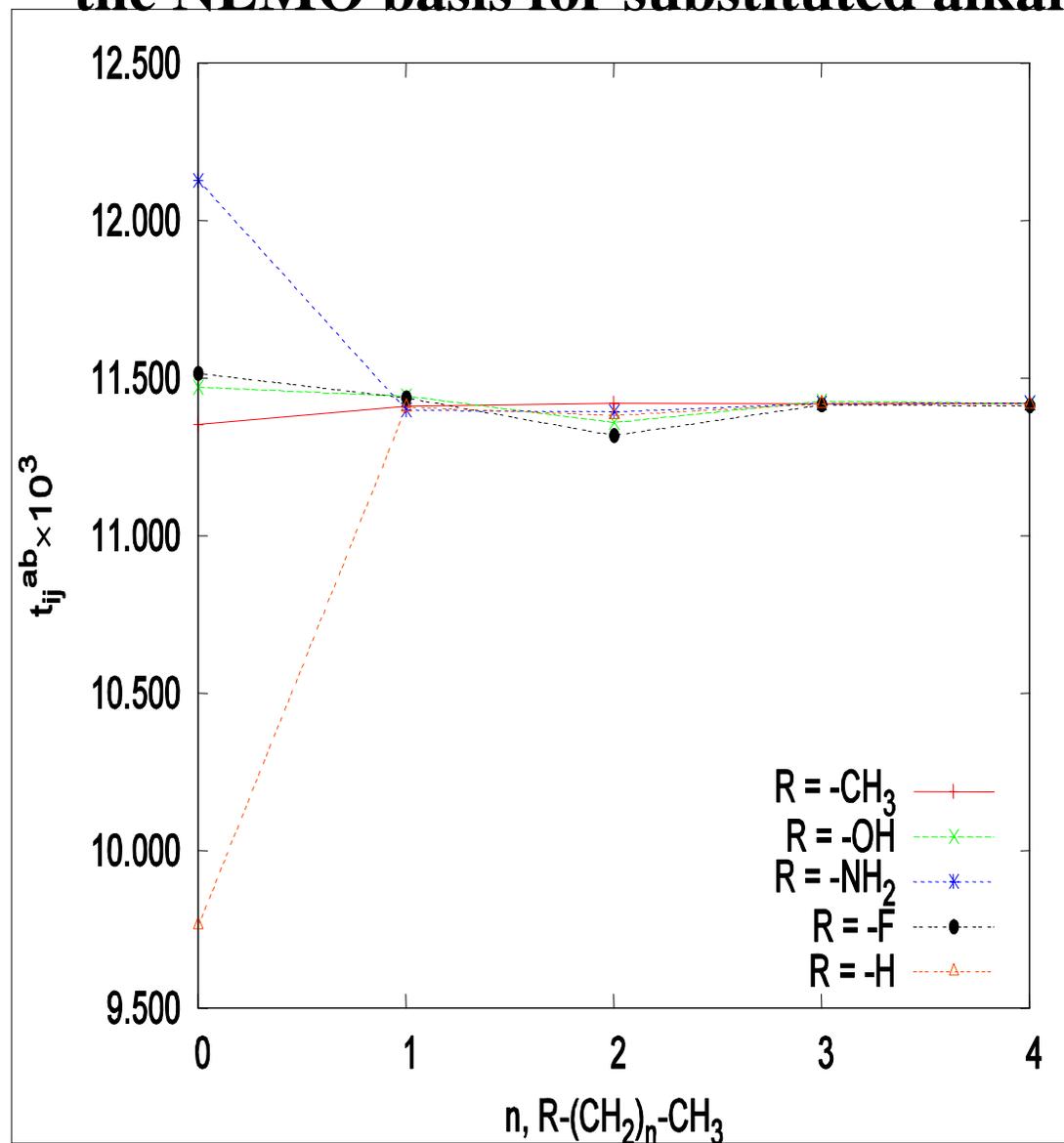
location	$C_3H_8$	$C_4H_{10}$	$C_5H_{12}$	$C_6H_{14}$	$C_7H_{16}$	$C_8H_{18}$
$\alpha\alpha$	-0.07548	-0.07544	-0.07537	-0.07535	-0.07534	<b>-0.07534</b>
$\beta\beta$		-0.07502	-0.07500	-0.07495	-0.07492	<b>-0.07492</b>
$\gamma\gamma$				-0.07500	-0.07493	<b>-0.07491</b>
$\delta\delta$						<b>-0.07487</b>
$\alpha\beta$	+0.01875	+0.01885	+0.01886	+0.01886	+0.01886	<b>+0.01886</b>
$\beta\gamma$			+0.01895	+0.01896	+0.01896	<b>+0.01896</b>
$\gamma\delta$					+0.01896	<b>+0.01897</b>

## Example I: NLS-CCSD Correlation Energy for n-Hexane



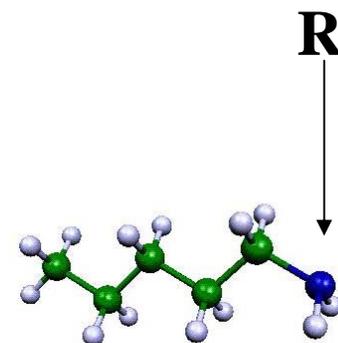
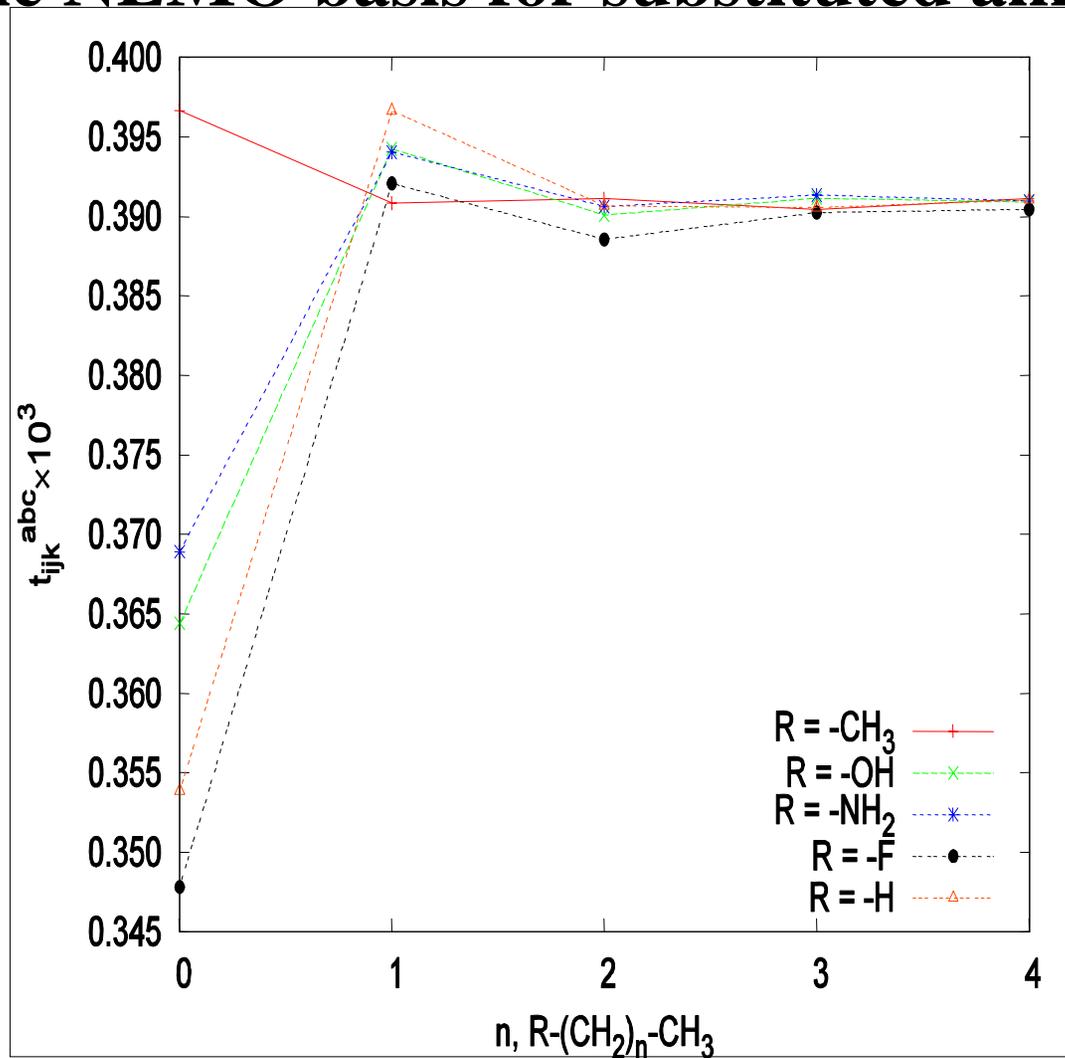
QM2 used	# of C-H				# of C-C			$\Delta E = \sum_i \Delta E_i$			
	$\alpha$	$\alpha'$	$\beta$	$\gamma$	$\alpha$	$\beta$	$\gamma$	6-31G	%	DZP	%
$C_2H_6$	14				5			-.58907	96.2	-.91617	97.5
$C_3H_8$	4	2	8		5			-.60249	98.4	-.92925	98.9
$C_4H_{10}$	4	2	8		2	3		-.60563	98.9	-.93561	99.5
$C_5H_{12}$	4	2	4	4	2	3		-.61138	99.9	-.93870	99.9
$C_6H_{14}$	4	2	4	4	2	2	1	-.61226	100	-.93987	100

# Transferability of CCSDT-3 doubles amplitudes in the NLMO basis for substituted alkanes



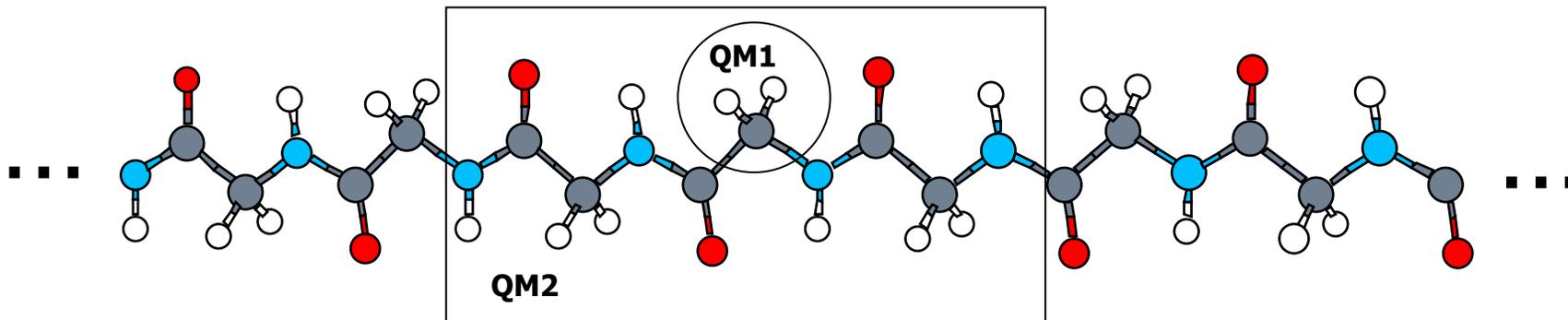
RHF, cc-pVDZ, bondsize = 2, High Symmetry

# Transferability of CCSDT-3 triples amplitudes in the NLMO basis for substituted alkanes

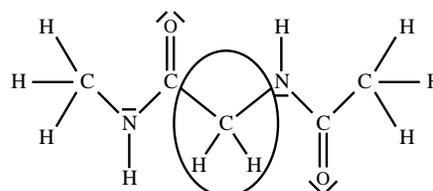


RHF, cc-pVDZ, bondsize = 2, High Symmetry

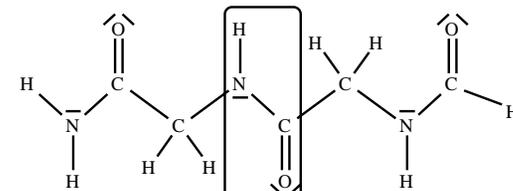
# Example : CCSD Correlation Energy of Polyglycine (GLY)<sub>x</sub>



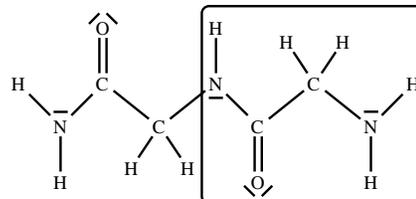
Dangling bonds  
of QM2 boxes  
terminated by  
H atoms



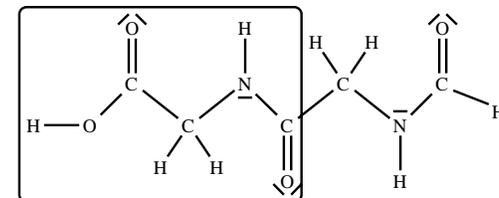
A



B



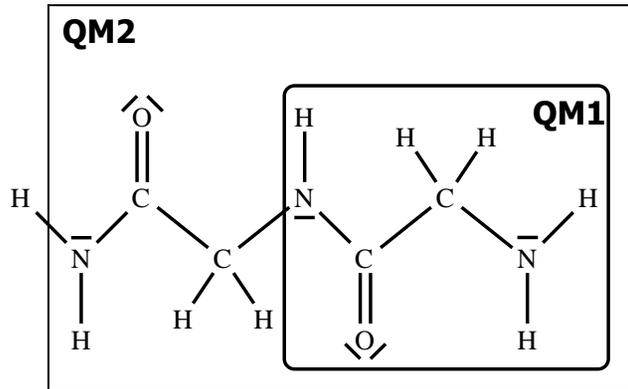
C



D

## Results

NLMO CCSD orbital correlation energies (cc-pVDZ) from QM2 box C:



C



$i$	$\Delta E_i$ (H)
C-H (methylene,2x)	-.0480
C-C (methylene)	-.0527
C-N (methylene)	-.0564
L.P. N (peptide)	-.0643
L.P. O (peptide)	-.0484
L.P. O (peptide)	-.0615
N-H (peptide)	-.0525
C-N (peptide)	-.0527
$\sigma$ C-O (peptide)	-.0495
$\pi$ C-O (peptide)	-.0746
L.P. N (amino)	-.0534
N-H (amino,2x)	-.0521

## Comparison between NLMO NLS CCSD and LCCSD Results

$(\text{GLY})_x$	NLMO <sup>[a]</sup>		LCCSD <sup>[b]</sup>	
	$\Delta E_{exact}$	$\Delta E_{approx}$	$\Delta E_{exact}$	$\Delta E_{approx}$
$x$				
2	-1.4305	-1.4307	-1.4574	-1.4427
4		-2.6542		-2.6783
6		-3.8777		-3.9141
8		-5.1012		-5.1499
10		-6.3247		-6.3858
12		-7.5482		-7.6216
14		-8.7716		-8.8575
-CH <sub>2</sub> -NHCO		-0.6117		-0.6179

Different core structure  $\approx$  -0.027

# Comparison of triples contribution to the correlation energy of poly-glycine

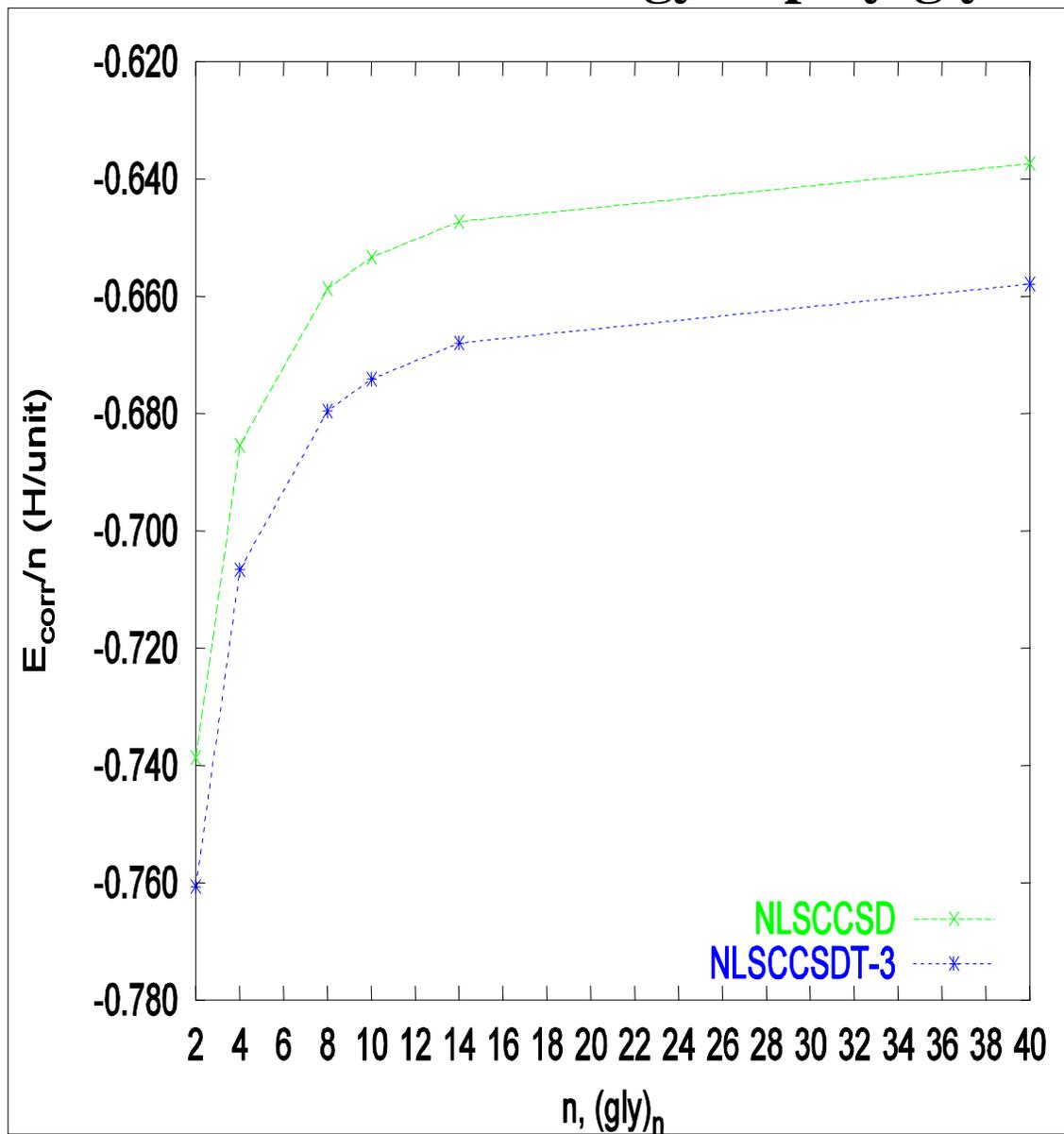
$n$	NLSCCSD	LCCSD	NLSCCSDT-3	LCCSDT-1b	LCCSD(T)
1	99.4%	99.1%	99.4%	98.9%	99.0%
2	-1.477300	-1.442654	-1.521269		
4	-2.741427	-2.678318	-2.826289	-2.755269	-2.757029
8	-5.269682	-5.149940	-5.436328	-5.300146	-5.303909
10	-6.533810	-6.385798	-6.741348		
14	-9.062064	-8.857490	-9.351388		
40	-25.495722		-26.316644		
⋮					
100					
⋮					

*M. Schütz, H.J. Werner, J. Chem. Phys. 114 (2001) 661.*

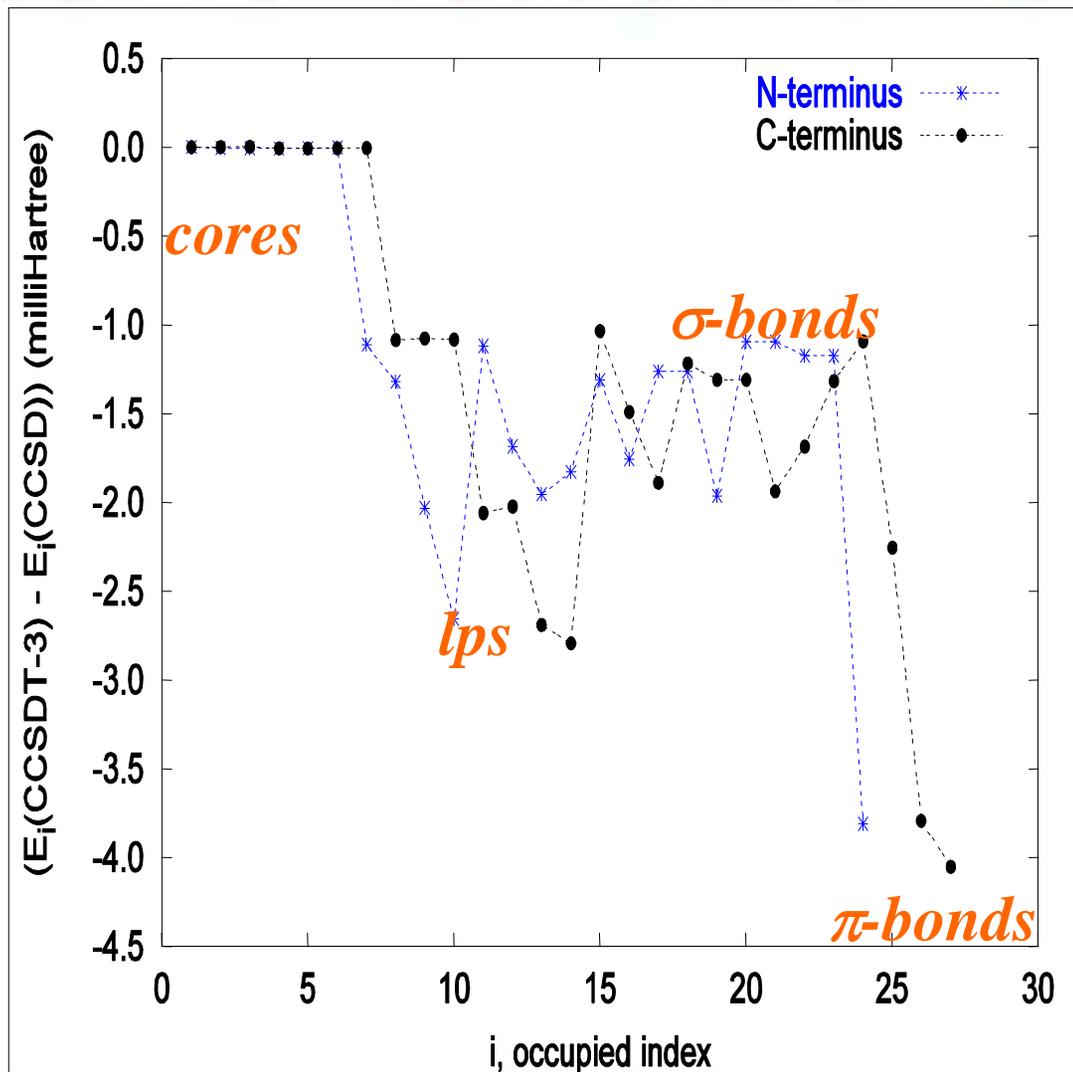
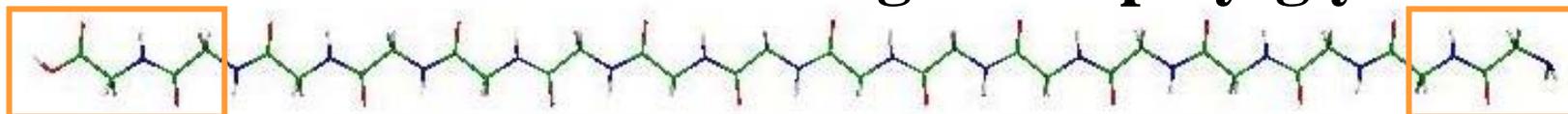
*M. Schütz, J. Chem. Phys. 116 (2002) 8772.*

*energies in Hartree*

# Comparison of triples contribution to the correlation energy of poly-glycine



# Difference between CCSDT-3 and CCSD orbital correlation energies for poly-glycine

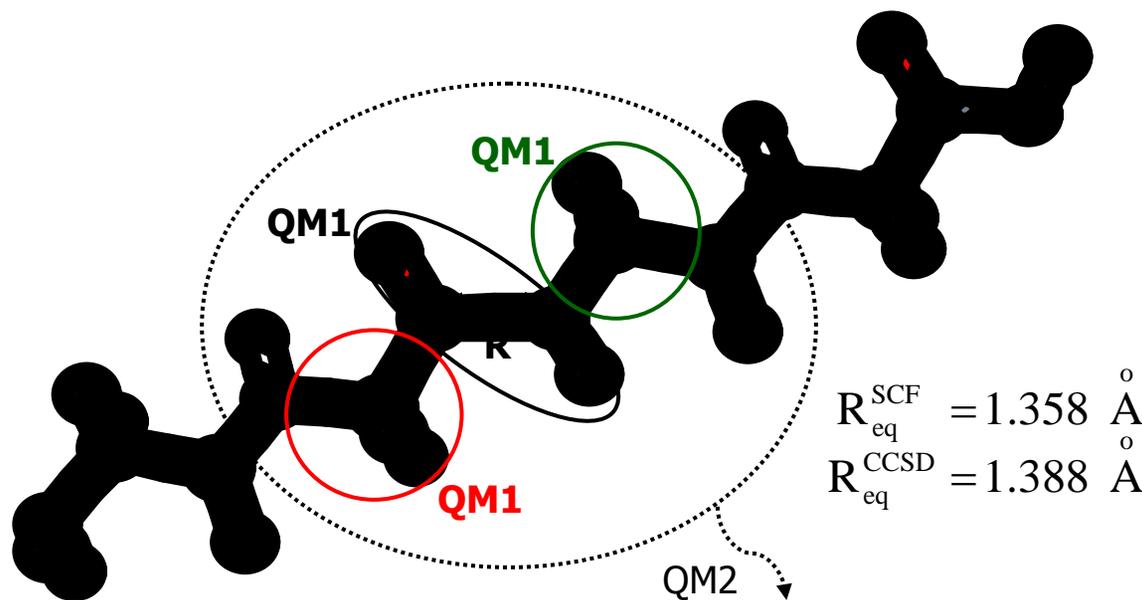


High  
Symmetry  
bondsize = 2  
cc-pVDZ  
RHF

# NLS CC Correlation Forces: (GLY)<sub>4</sub> Peptide Bond Stretch

$$\Delta F_P = \frac{\partial \Delta E}{\partial X_P} \approx \sum_{\text{QM1}} \left[ \sum_{i \in \text{QM1}}^{\text{QM2}} \frac{\partial \Delta E_i}{\partial X_P} \right]$$

Basis set : DZ  
 $\Delta F_P$  in  $10^{-2}$  H / Bohr



R	QM1	QM1	QM1	% error
1.1	- 3.108	+ 0.065	+ 0.153	2.10
1.3	- 2.875	+ 0.063	+ 0.025	2.19
1.5	- 2.732	+ 0.059	- 0.075	2.94
1.7	- 2.695	+ 0.105	- 0.064	1.55
1.9	- 2.555	+ 0.081	- 0.052	0.93

R		% error
1.1	- 2.854	0.85
1.3	- 2.763	1.34
1.5	- 2.713	1.63
1.7	- 2.652	1.49
1.9	- 2.517	0.58

# DENSITY MATRICES

- Density Matrices via the NLS approach

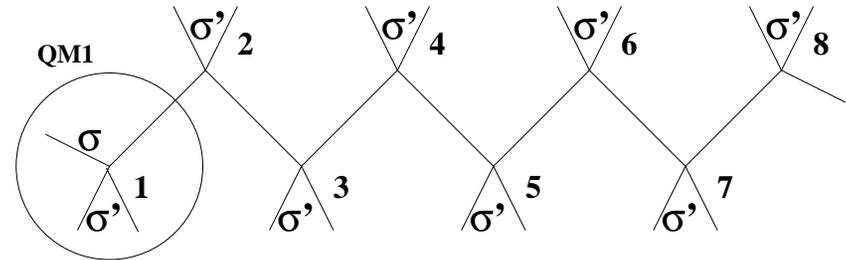
$$\Delta\gamma_{qp} = \left\langle \phi_{ref} \left| \begin{array}{c} e^{\hat{T}^\dagger} \{ \hat{p}^\dagger \hat{q} \} e^{\hat{T}} \\ (1 + \Lambda) \{ \hat{p}^\dagger \hat{q} \} e^{\hat{T}} \end{array} \right| \phi_{ref} \right\rangle_c$$

QM1 only  $\nearrow$   $\Delta\gamma_{qp}$   
 QM1 & QM2  $\nearrow$   $\Delta\gamma_{qp}$

$$\Delta\gamma^{system} = \left( \begin{array}{c} \text{QM1} \\ \downarrow \\ \text{QM1} \\ \& \\ \text{QM2} \\ \rightarrow \\ \text{[Black Box]} \\ \uparrow \\ \text{QM1} \\ \downarrow \\ 0 \\ \uparrow \\ 0 \end{array} \right)$$

# Transferability of Correlated Density Matrix Elements

**Example:** Hydrocarbon  $C_nH_{2n+1}$  Series



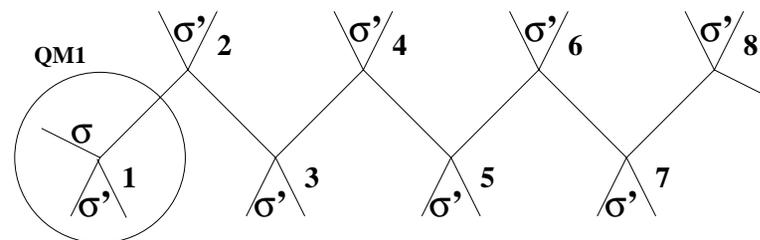
$$\Delta\gamma_{ji} = \left\langle 0 \left| e^{\hat{T}^\dagger} \{ \hat{i}^\dagger \hat{j} \} e^{\hat{T}} \right| 0 \right\rangle_C \quad e^{\hat{T}} \approx 1 + T_1 + \frac{T_1^2}{2} + T_2 \quad i,j \rightarrow \sigma, \sigma' \text{ CH bond}$$

dist $\rightarrow$	0 ( $\times 10^{-2}$ )	1 ( $\times 10^{-3}$ )	2 ( $\times 10^{-4}$ )	3 ( $\times 10^{-4}$ )	4 ( $\times 10^{-5}$ )	5 ( $\times 10^{-5}$ )	6 ( $\times 10^{-6}$ )	7 ( $\times 10^{-6}$ )
$C_3H_8$	-1.89143	1.88407	3.12581	-1.58426				
$C_4H_{10}$	-1.89521	1.88893	3.01042	-1.71922	2.48332			
$C_5H_{12}$	-1.89625	1.89127	3.00008	-1.68087	2.67793	-1.35472		
$C_6H_{14}$	-1.89691	1.89215	2.99313	-1.67711	2.49216	-1.43047	-3.01008	
$C_7H_{16}$	-1.89705	1.89295	2.99529	-1.67789	2.45620	-1.31002	3.68291	-1.24601

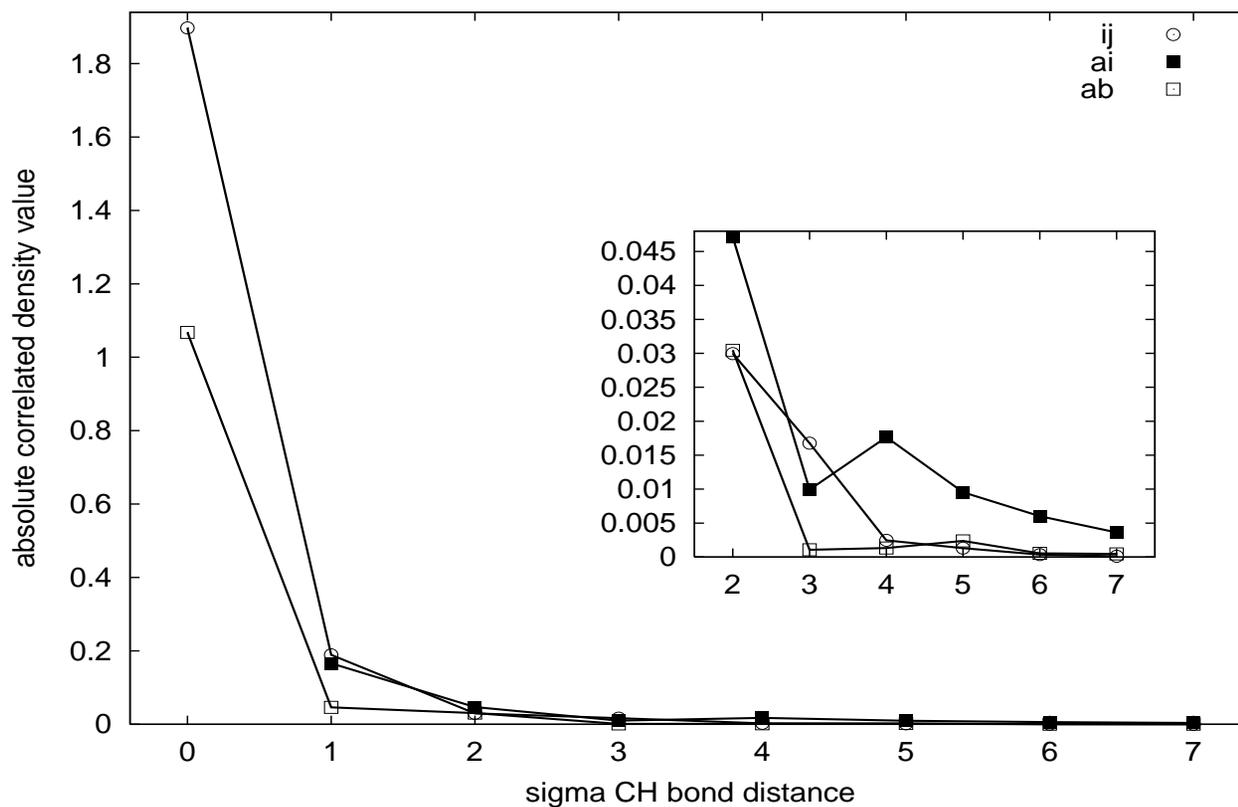
# Decay of Correlated Density Matrix Elements

**Example:** Heptane  $C_7H_{16}$

$i,j \oplus$   $\sigma$  CH bond ,  $\sigma'$  CH bond  
 $a,i \equiv$   $\sigma$  CH bond ,  $\sigma'^*$  CH antibond



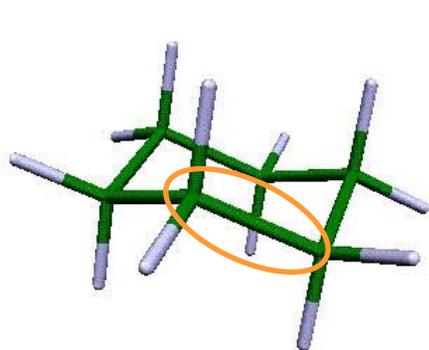
$a,b \otimes$   $\sigma^*$  CH antibond ,  $\sigma'^*$  CH antibond



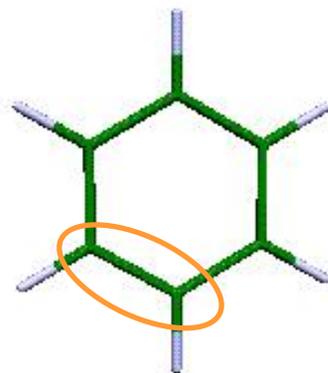
**What about delocalization?**

# Simple illustration of using bond sizes and NLMO occupations to determine locality

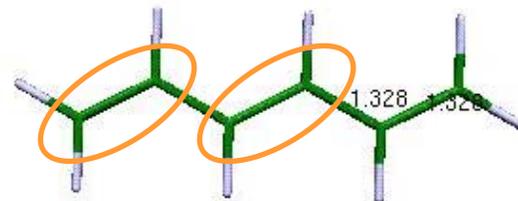
RHF, cc-pVDZ



$D_{3d}$



$D_{2h}$

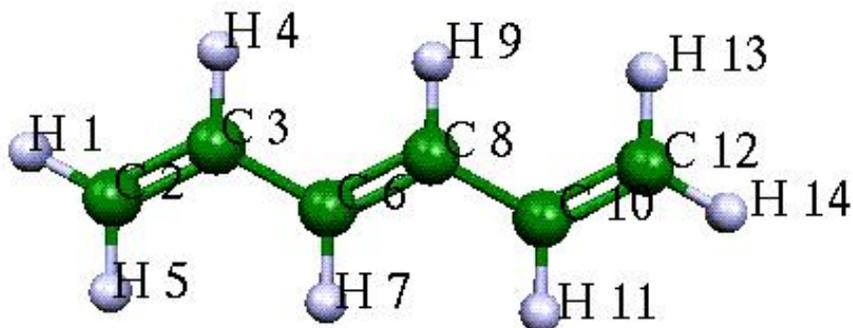


$C_s$

	NLMO occupation	NLMO occupation	NLMO occupation
bondsize = 2	1.9765	1.6604	1.9043    1.8290
bondsize = 6	1.9765	1.9899	1.9737

# Acetylene oligomers with different single:double carbon-carbon bond distance ratios, $u$

$$R_{\text{single}}=1.1948 \text{ \AA}, R_{\text{double}}=1.3276 \text{ \AA}, u=0.9$$



$$R_{\text{single}}=1.4714 \text{ \AA}, R_{\text{double}}=1.3276 \text{ \AA}, u=1.1$$

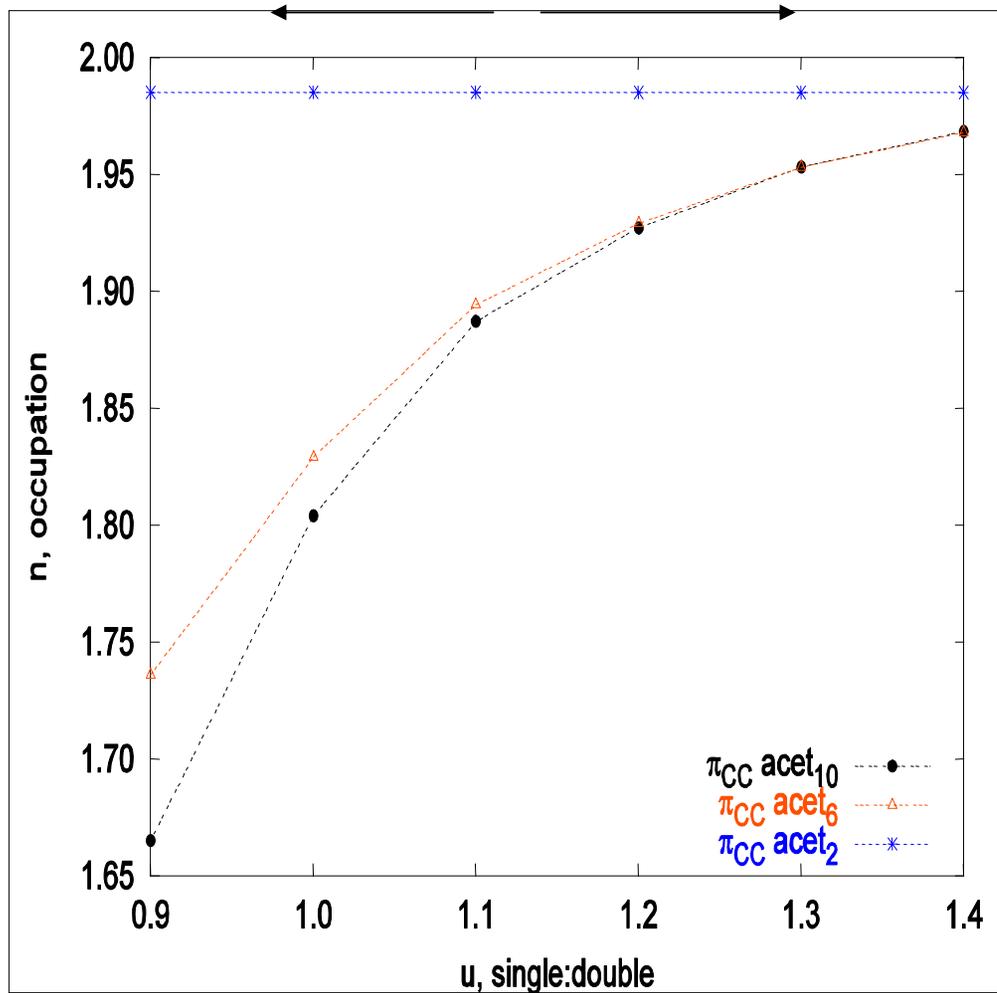
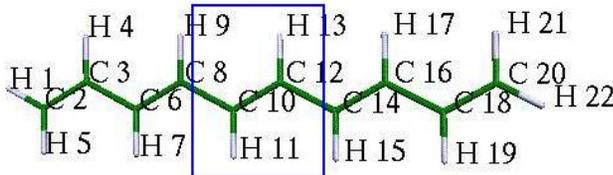
exp. est.

$$R_{\text{single}}=1.43 \text{ \AA}, R_{\text{double}}=1.38 \text{ \AA}, u \sim 1.0$$

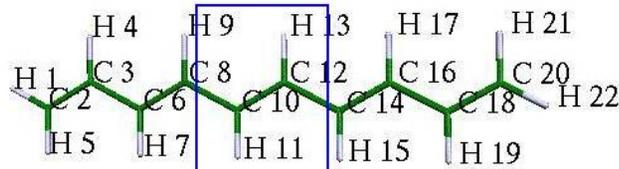
$$R_{\text{single}}=1.8586 \text{ \AA}, R_{\text{double}}=1.3276 \text{ \AA}, u=1.4$$

# Occupations of central $\pi$ NLMOs in 2-, 6-, and 10-carbon acetylene oligomers

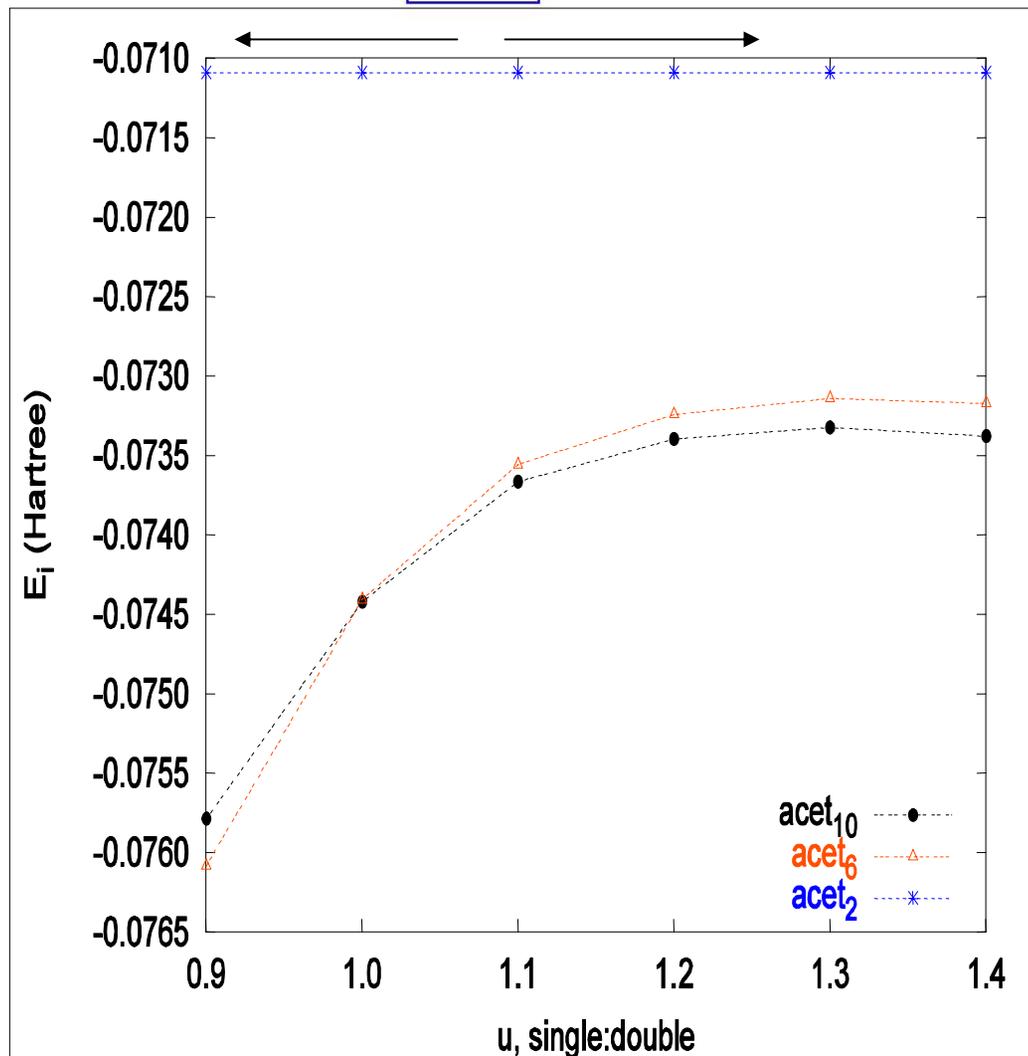
RHF  
 cc-pVDZ  
 bondsize = 2  
 $C_s$  Symmetry



# Orbital correlation energies of central $\pi$ NLMOs in 2-, 6-, and 10- carbon acetylene oligomers



RHF  
 CCSD  
 cc-pVDZ  
 bondsize = 2  
 $C_s$  Symmetry



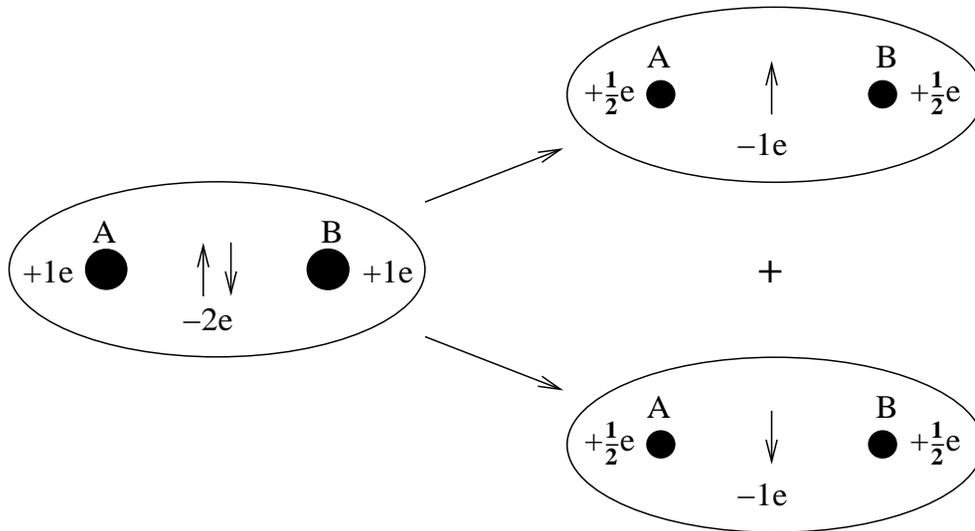
# NOW WHAT ABOUT THE SCF PROBLEM?

- Without adding that number to the NLSCC results obtained, we don't have the whole answer.
- Avoidance of the global SCF (or any global step) is very important in this approach.
- But we are faced with long-range Coulomb interactions that have to be included.
- And unlike the correlation problem, the SCF problem has to include the nuclear repulsion.

# Splitting of SCF charges into neutral units

$$\sum_A \frac{Z_A}{r_{kA}} = \sum_i \frac{1}{M_i} \sum_{Z_i}^{d_i} \frac{m_{Z_i}}{r_{kZ_i}}$$

Identity valid for any point  $k$



$M_i$  = number of electrons forming orbital  $i$

$d_i$  = number of nuclear centers in orbital  $i$

$m_{Z_i}$  = number of electrons from nuclear center  $Z_i$

$$\frac{1}{r_{kA}} + \frac{1}{r_{kB}} = \frac{1}{2} \left( \frac{1}{r_{kA}} + \frac{1}{r_{kB}} \right) + \frac{1}{2} \left( \frac{1}{r_{kA}} + \frac{1}{r_{kB}} \right)$$

## Natural Linear Scaling SCF

**Key idea:** Associate with each localized bond a +1 charge from nuclear framework

**Define:** Potential energy between neutral units:

$$p_{ij} = \langle ij | ij \rangle - \langle ij | ji \rangle - \langle i | \frac{1}{M_j} \sum_{Z_j} \frac{m_{Z_j}}{r_{Z_j}} | i \rangle - \langle j | \frac{1}{M_i} \sum_{Z_i} \frac{m_{Z_i}}{r_{Z_i}} | j \rangle + \frac{1}{M_i M_j} \sum_{Z_i \neq Z_j} \frac{m_{Z_i} m_{Z_j}}{r_{Z_i Z_j}}$$

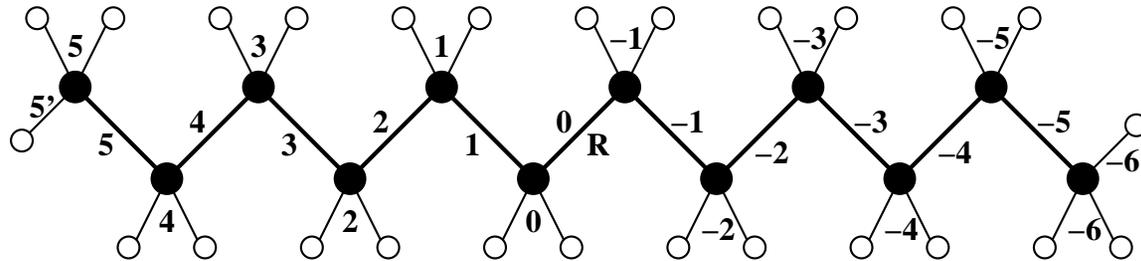
**Define:** Total interaction energy between one unit with all the rest:

$$e_i = k_i + \frac{1}{2} \sum_j p_{ij}$$

**Total SCF energy:**

$$E_{SCF} = \sum_i e_i$$

## Decay of $p_{ij}$ in n-dodecane and saturation of $e_j$



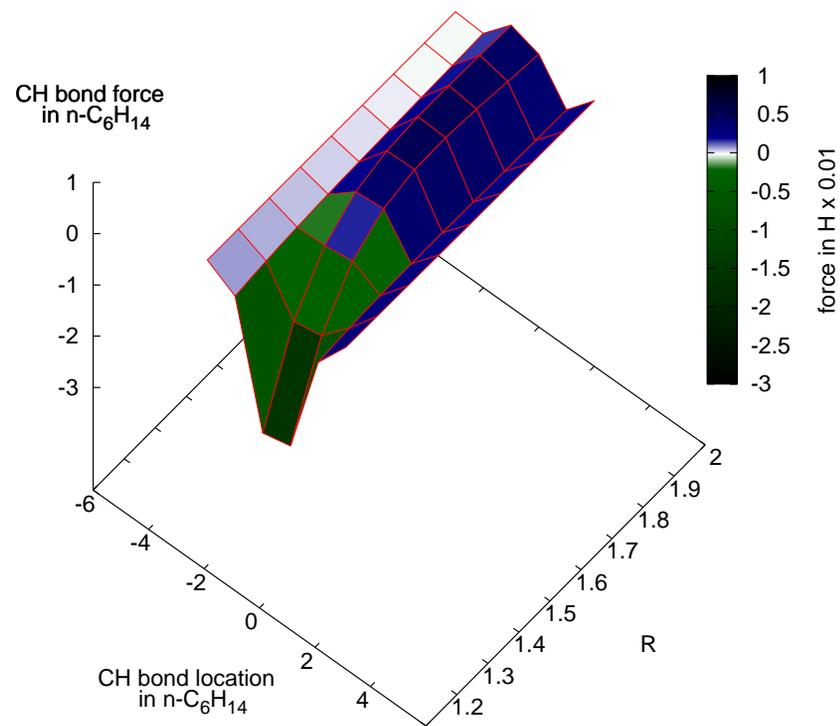
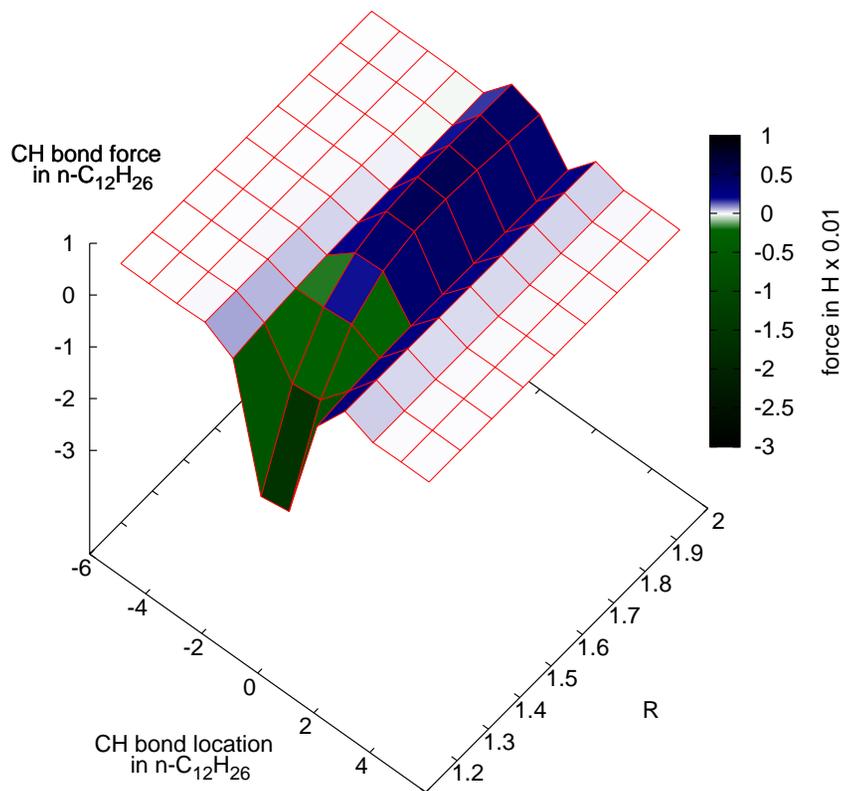
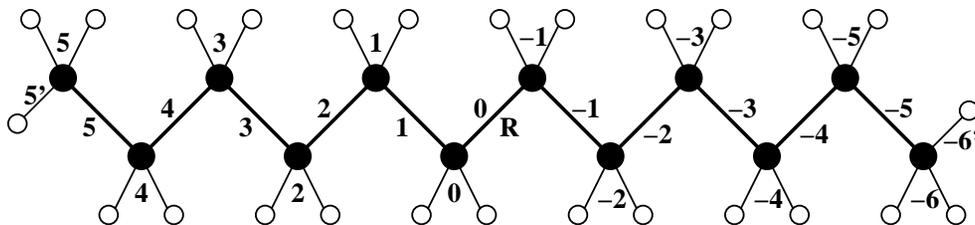
j	$p_{5j}$ CH bond	$p_{5j}$ CC bond
self	-1.3502286	-1.1212278
5	-.2888281	
4	.0031163	-.3097111
3	-.0004523	-.0032898
2	.0001119	.0003038
1	-.0000415	-.0001272
0	.0000369	.0000303
-1	-.0000113	-.0000097
-2	.0000139	.0000077
-3	-.0000050	-.0000012
-4	.0000064	.0000025
-5	-.0000028	.0000000
-6	.0000035	
-6'	.0000071	

	$e_5$ for CH bond 5	$e_5$ for CC bond 5
$C_2H_6$	-3.863031	-7.232628
$n-C_4H_{10}$	-3.863458	-7.225500
$n-C_6H_{14}$	-3.863602	-7.224779
$n-C_8H_{18}$	-3.863648	-7.224814
$n-C_{10}H_{22}$	-3.863667	-7.224965
$n-C_{12}H_{26}$	-3.863674	-7.224967
...	...	...
$n-C_xH_{2x+2} \quad x \rightarrow \infty$	-3.86378	-7.22497

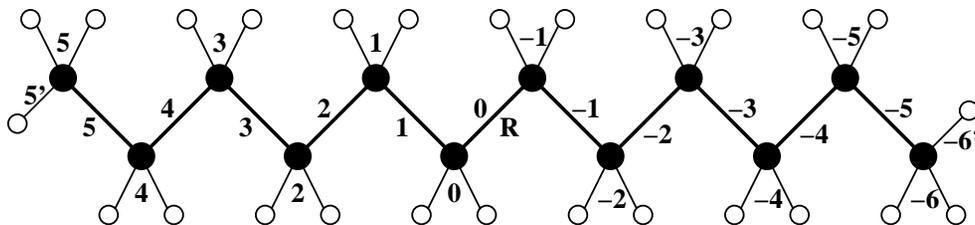
# Using $e_i$ transferability to assemble n-dodecane $E_{SCF}$

i	n-C <sub>12</sub> H <sub>26</sub>	C <sub>2</sub> H <sub>6</sub>	n-C <sub>4</sub> H <sub>10</sub>	n-C <sub>6</sub> H <sub>14</sub>
C Core 5	-4.5992 (x4)	-4.6033 (x24)	-4.5993 (x4)	-4.5993 (x4)
C Core 4	-4.5934 (x4)		-4.5939 (x20)	-4.5935 (x4)
C Core 3	-4.5901 (x4)			-4.5903 (x16)
C Core 2	-4.5898 (x4)			
C Core 1	-4.5897 (x4)			
C core 0	-4.5897 (x4)			
CH Bond 5	-3.8637 (x8)	-3.8630 (x48)	-3.8635 (x8)	-3.8636 (x8)
CH Bond 5'	-3.8614 (x4)	-3.8629 (x4)	-3.8614 (x4)	-3.8614 (x4)
CH Bond 4	-3.8554 (x8)		-3.8555 (x40)	-3.8553 (x8)
CH Bond 3	-3.8562 (x8)			-3.8561 (x32)
CH Bond 2	-3.8560 (x8)			
CH Bond 1	-3.8561 (x8)			
CH Bond 0	-3.8561 (x8)			
CC Bond 5	-7.2250 (x4)	-7.2326 (x22)	-7.2255 (x4)	-7.2248 (x4)
CC Bond 4	-7.2169 (x4)		-7.2175 (x18)	-7.2170 (x4)
CC Bond 3	-7.2169 (x4)			-7.2172 (x14)
CC Bond 2	-7.2160 (x4)			
CC Bond 1	-7.2163 (x4)			
CC Bond 0	-7.2159 (x2)			
$E_{SCF}$	-469.5970	-470.4729	-469.6663	-469.6134

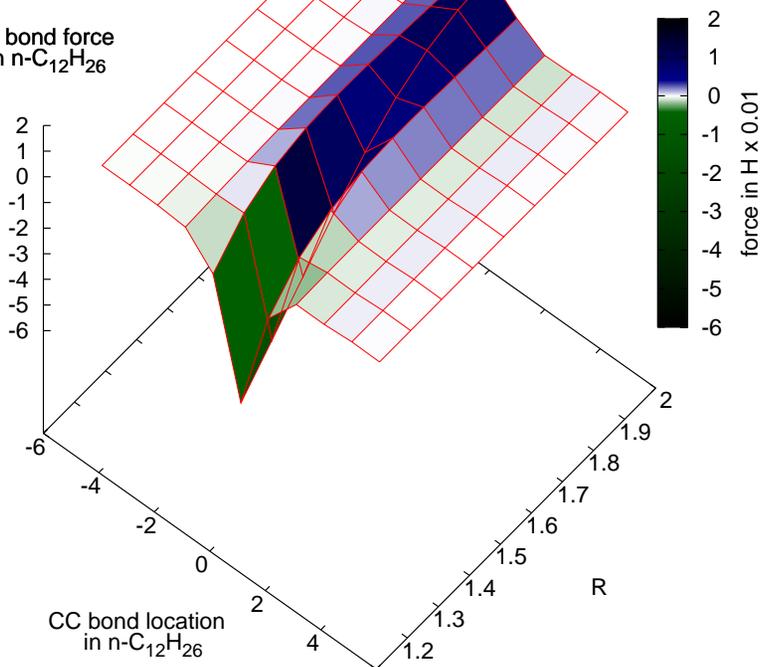
# Locality of neutral CH bond unit forces $f_i = \partial e_i / \partial R$



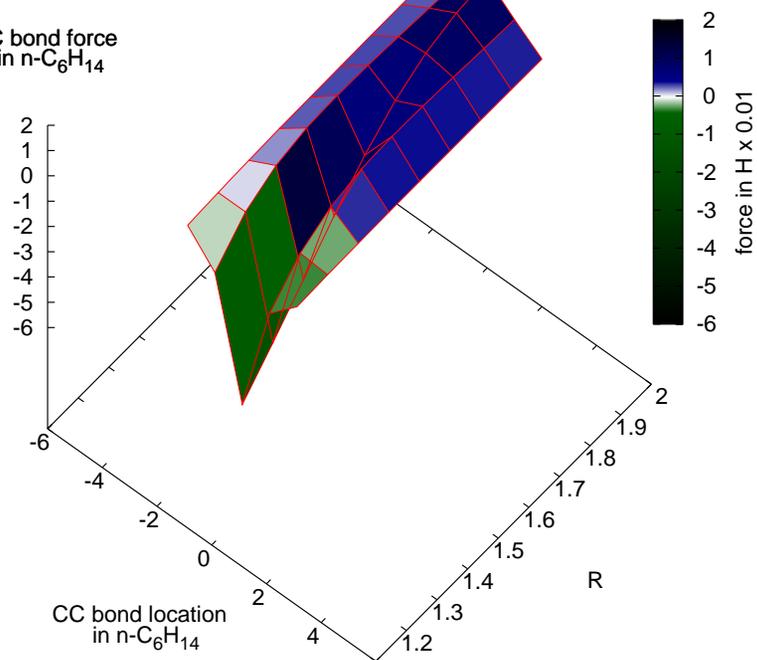
# Locality of neutral CC bond unit forces $f_i = \partial e_i / \partial R$



CC bond force  
in  $n\text{-C}_{12}\text{H}_{26}$



CC bond force  
in  $n\text{-C}_6\text{H}_{14}$



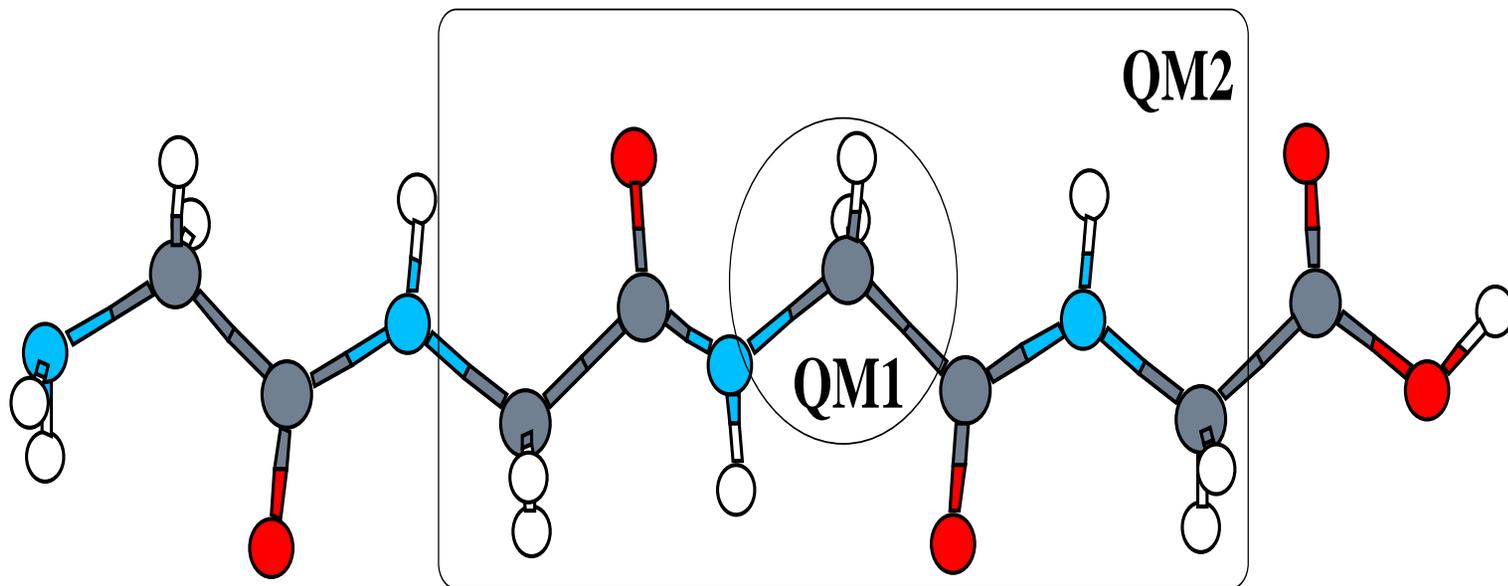
# CONCLUSIONS

- Introduced a manifestly, linear scaled simple localized orbital CC approach (NLSCC) for large systems built upon NLMO's that treat the occupied and virtual space equivalently.
- NLSCC avoids any use of non-orthogonal, projected atomic orbitals in its solution, reducing the computational dependence from  $\sim n^2(n+N)^4$  to  $\sim n^2N^4$ .
- NLSCC exploits the transferability of the correlation effects to replicate the same units throughout the system, and, thus, does not require that different parts of the correlation be treated by different approximations, like CCSD and MBPT2, as in other methods.
- NLSCC also applies to delocalized units like benzene rings, as long as such units constitute a QM1 region.
- NLSCC provides 'bond' or 'unit' correlation energies and densities for easy chemical interpretation.
- NLSCC can be applied to polymers using the highest levels of correlated theory. Intensive properties like band-gaps, exciton spectra, etc. are inherently different, however.

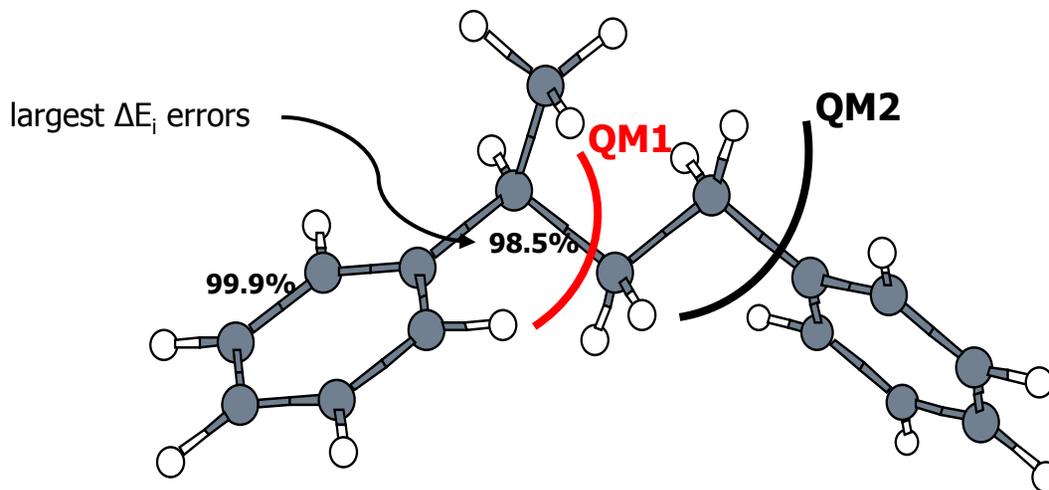
## CONCLUSIONS (cont)

- The underlying SCF (or KS) problem, can be done with a similar LS strategy, based upon local, neutral units.
- First report of ACES III, the parallel successor to ACES II
- Demonstrated 91% scaling for CCSD on up to 240 processors
- Adds new integral program and integral direct capabilities for very large scale, SCF, DFT, MBPT2, CCSD, CCSD gradients, and soon CCSD(T), EOM-CC, etc.
- Built upon a new language, SIAL, which offers a framework for writing quantum chemistry programs that really does take memory handling and message passing away from the quantum chemical programmer.

## NLSCC : Natural Linear Scaling Coupled Cluster Method



## Example II: NLS Correlation Energies for Distyrene



Compound	Avg $\Delta E_i$	CCSD (DZ)	$\Delta$ CCSDT-3 (DZ)	$\Delta$ CCSD (DZP)
Both QM2	Core	-0.01384	-0.00001	-0.00072
	$\sigma$ -Bonds	-0.03244	-0.00076	-0.01921
	$\pi$ -Bonds	-0.05451	-0.00296	-0.01898
Distyrene		-1.68406 (99.7%)	-0.04449	-0.79772
Polystyrene unit		-0.83041	-0.02235	-0.39283