Recent Advances in Molecular Theory

Kimihiko HIRAO
Special Advisor
Director, Planning Office for Institute for Computational Science
RIKEN
Meiji Seimei Kan
2-1-1 Marunouchi
Tokyo, Japan

Topics

1. Next-Generation Supercomputer Project
2. Recent Advances in Molecular Theory

Science in the 21st Century

The role of scientific research in solving the problems we face is more relevant than ever in the 21^{st} century.

Today we know that the world is dramatically being reshaped by scientific and technological innovations, global interdependence, cross-cultural encounters, and changes in the balance of economic and political power.

We strive to make significant contributions to solving the societal challenges of the 21st century by pursuing a course of leadership through technical excellence and innovation.

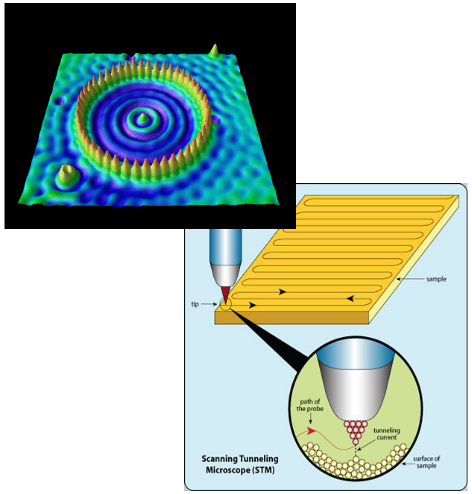
Linkage and Integration

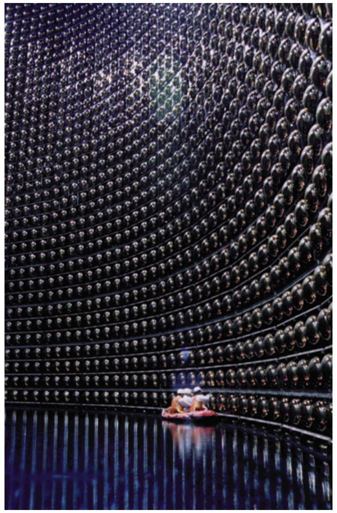
We live in an age characterized by two opposing trends – integration and fragmentation. The dominant trend of this age must be linkage and integration – not fragmentation.

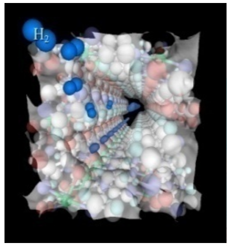
We are linked and integrated:

- because we inhabit a single planet. We share its thin and fragile environment. We share its ecosystems and its finite resources.
- by the common desire of people everywhere to live peacefully and in good health.
- by our common curiosity about the world, the universe, and each other.
- by the knowledge and understanding developed by many people in many places, in many periods of history.

Most Exciting Era of Science and Technology



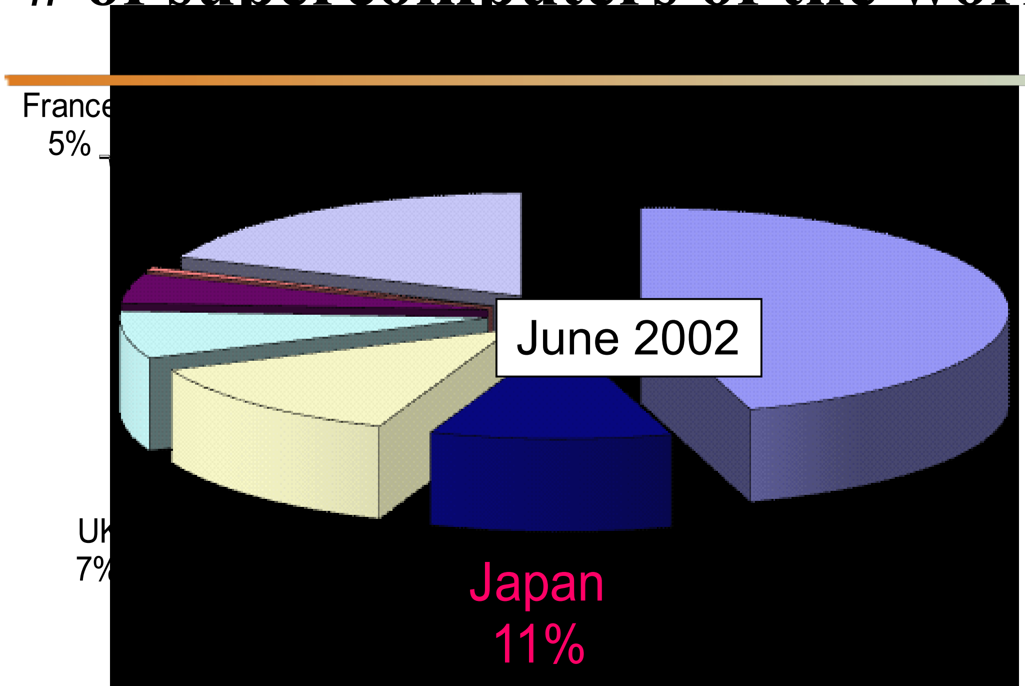




We live in the most exciting era of science and technology in human history. Recent decades have seen remarkable and rapid advances in the technological tools such as STM, Kamiokande, next generation GS, Subaru, Spring-8, etc. with which to do science. Science, in turn, has led to new technological products and processes.

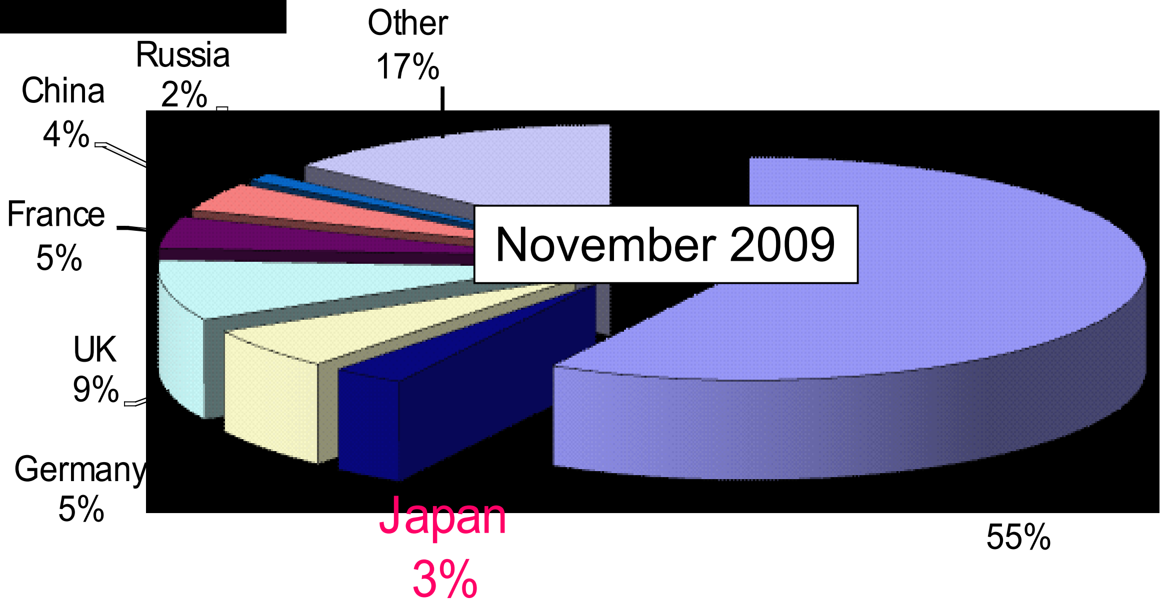
Supercomputer will undoubtedly accelerate this trend.

of supercomputers of the world's 500 most powerful ones



The number of supercomputers of Japan peaked in the early 1990s.

Japan, trying to cope with snowballing debt, has fallen far behind in supercomputer compared with US.



Objective of the Next-Generation Supercomputer Project

- Design, build, and set up the Next-Generation Supercomputer (NGSC), the world’s fastest and most advanced computer, with a speed of 10 Peta Flops

- Develop and distribute large-scale software applications (“Grand Challenge” software) that make full use of the supercomputer

- Set up a Center to run the supercomputer, to be the world’s COE in the field of supercomputing

System Configuration was Revised

- The original was a **vector-scalar** hybrid architecture. Vector unit is suitable for continuum physics simulation and scalar unit is for data base application, particle-based simulation.

- The project was thrown into confusion by the sudden departure of NEC and Hitachi from the project in May 2009.

- In the revised configuration, it keeps the original performance target of 10 PetaFlops in LINPACK benchmark with a **scalar** computing architecture manufactured by Fujitsu, eliminating the vector computing architecture.

Does Japan need to be No.1 in Science and Technology?

- The next crisis of the project came suddenly with a government panel's conclusion on November 13, 2009.

- The panel, a budget-cutting team concluded that the FY2010 budget for NGSC project should be virtually frozen.

- The panel also concluded on serious budget cuts for SPring-8, the world's largest synchrotron radiation facility as well as basic grant programs for post-doctoral researchers, etc.

Adverse Opinions Rapidly Exploded

- Following the broadcast decision, significant adverse opinions rapidly exploded in academia, industries and so on, including Nobel Prize laureates.

- The backlash was ignited by the televised discussions regarding the devitalizing of Japan’s Science and Technology capability - in particular, such questions as “Why is the project challenge to achieve No. 1 position? Why not No.2.”

Re-instated

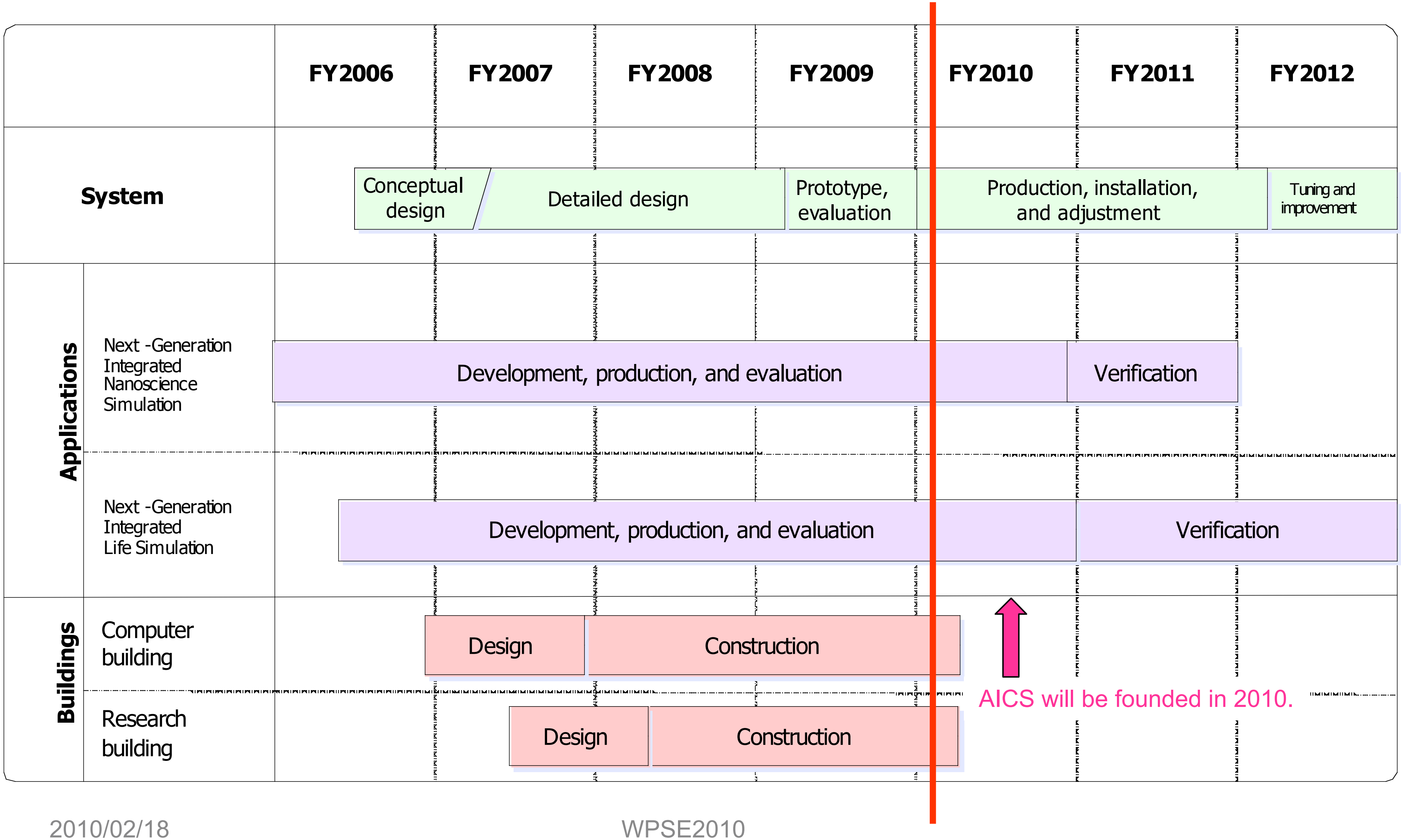
- On December 16, 2009, the Japanese government substantially decided to approve 22.7 billion yen (US$252 million) for the FY2010 budget of the NGSC project, which is 4.0 billion yen (US$44 million) smaller than the original proposal.
- The original performance target of 10 PetaFlops in the LINPACK benchmark is maintained.
- But the schedule was revised. The completion date of November 2011 was postponed to June 2012.

What happens twice will happen thrice.

Third time lucky!

Schedule Of The Project

We are here.



System
- Conceptual design
- Detailed design
- Prototype, evaluation
- Production, installation, and adjustment
- Tuning and improvement

Applications

Next -Generation Integrated Nanoscience Simulation
- Development, production, and evaluation
- Verification

Next -Generation Integrated Life Simulation
- Development, production, and evaluation
- Verification

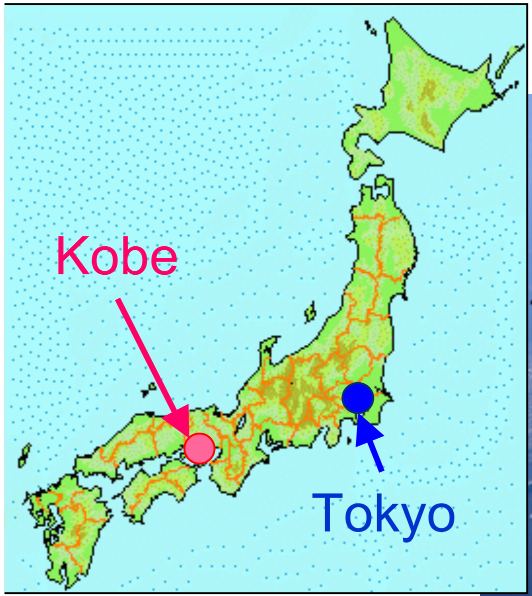
Buildings

Computer building
- Design
- Construction

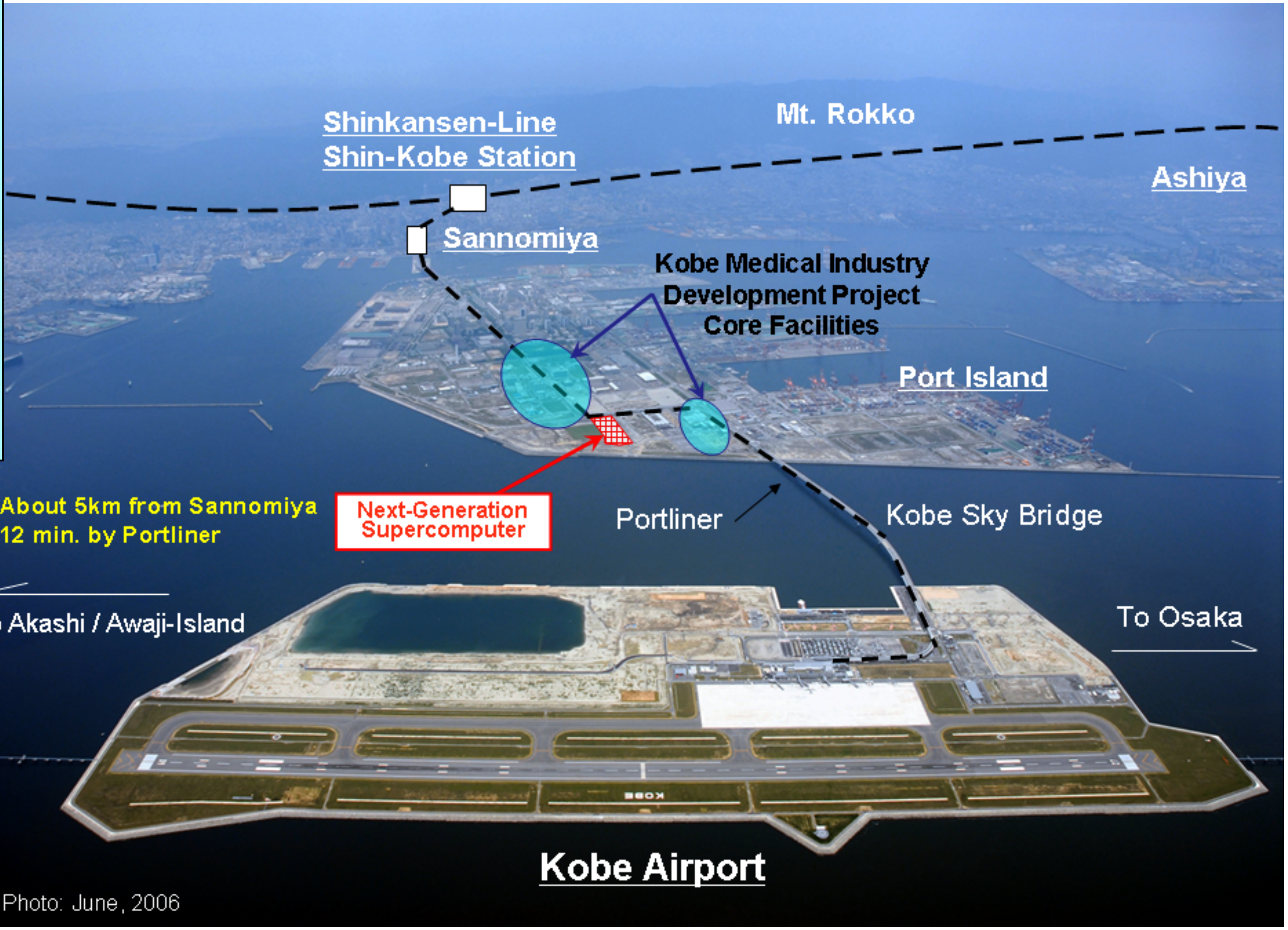
Research building
- Design
- Construction

AICS will be founded in 2010.

Location Of The Supercomputer, Kobe-City



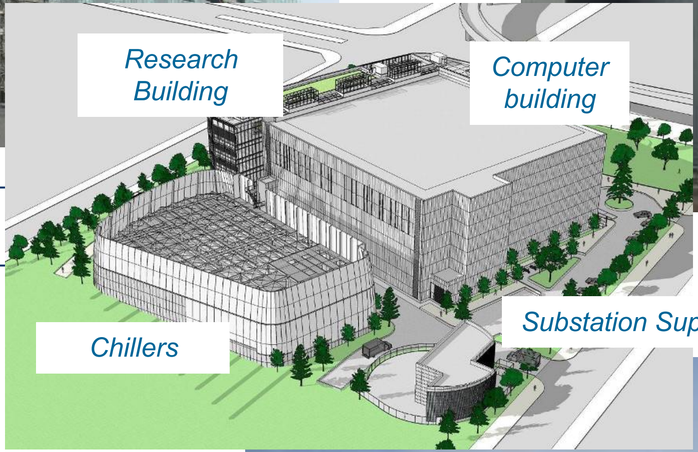
450km (280miles)
west from Tokyo

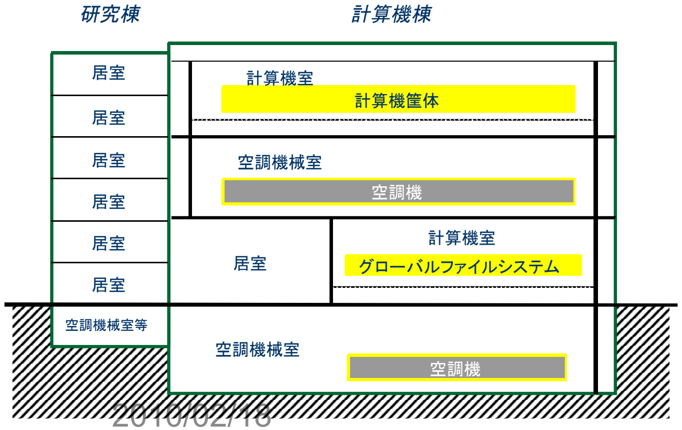


About 5km from Sannomiya
12 min. by Portliner

Photo: June, 2006

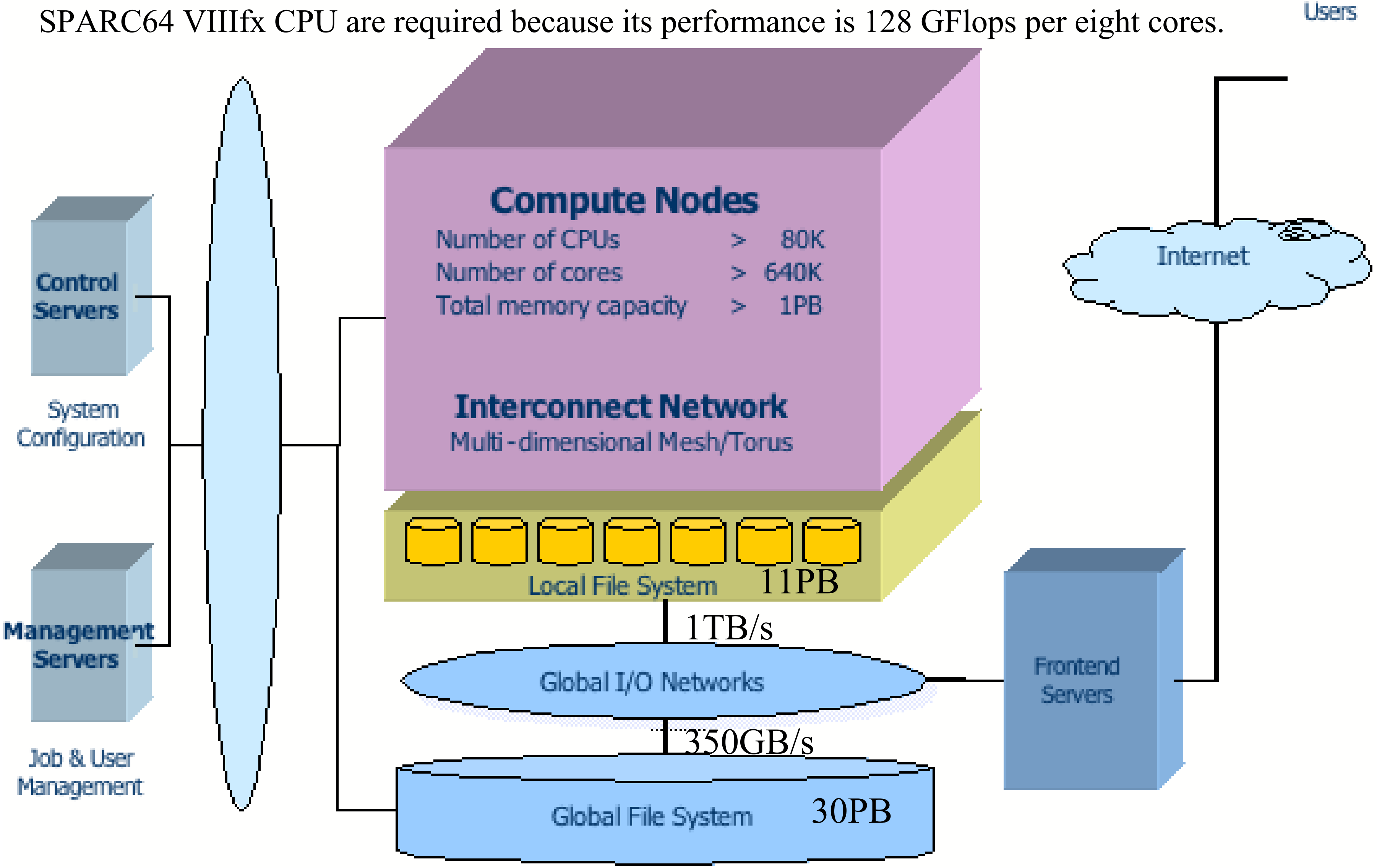
Photographs Of Facilities (January, 2010)





System Configuration

To achieve a 10 PFlops LINPACK performance, more than 640 thousand cores of SPARC64 VIIIfx CPU are required because its performance is 128 GFlops per eight cores.



- Petascale computing hardware is around the corner.
- Petascale computing chemistry should be science driven.
- Petascale resources will enable chemistry researchers to enter a new modeling.

- Parallel scaling to tens of thousands of processors
- Software not ready for petascale computing yet.
- Get insight into the computational chemistry science that will utilize petascale computing

- Operations are cheap, data movement is expensive.
- Off chip data bandwidth will become bottleneck for massively multicore processors.
- Need teams of systems software developers and application experts to make needed advances.

Advanced Institute for Computational Science

AICS

Image of Next-Generation Supercomputer facility

Missions of AICS

- It operates and enhances NGSC so to provide its resources effectively and efficiently to end-users.
- It leads computational science and technology so to form a world-class COE for computational science as a hub of the several strategic bases.
- It plots Japan's strategy for computational science and plays as a leader.

Good mutual collaboration between AICS and all the related communities is necessary to achieve the above missions.

Advanced Institute for Computational Science

● Division of Strategy and Management

- ・Planning of Strategy and Management
- ・Promotion of Collaboration and Cooperation with Academia and Industry throughout Japan
- ・Promotion of International Collaboration
- ・Outreach and Education

● Division of Computational Science (12 teams)･･･ Basic Research of Computational Science including Development of Theory, Modeling, Algorithm, etc.

- ・Quantum Systems : Nuclear Physics, Chemical Reactions, Catalyst, Functional Materials, Nanoscale Science, Device, Enzymes and Proteins
- ・Particle Systems : Big Bang Theory, Protein Dynamics, New Lightweight Materials, Drug Design, Virus
- ・Continuum Systems : Next-Generation Automobiles, Blood Stream, Global Warming, Earthquake Hazards, Extreme Weather
- ・Complex Systems : Solar Cell, Battery, Fuel Cell, People Mover, Disaster Mitigation, Nuclear Fusion, Cell Biology, Human Biology
- ・Discrete Systems : Genome, Social Phenomena and Economic Prediction, Security, Preventing Transmission of Infectious Diseases
- ・Research teams collaborating with the Strategic Organizations

● Division of Computer Science（8 teams）･･･Basic Research and Development of Computer Science

- ・Software Systems :System Software, Middleware
- ・Programming Systems : Compiler Systems, programming Environments
- ・Algorithms & mathematical Science : Algorithms, Performance Tuning
- ・Architectures : Computation Systems, Processor Architecture
- ・Data-Driven Computational Science and Visualization

● Division of Computer Operation and Maintenance

Your Cooperation Is Needed

- Computer simulation has significantly broadened its scope and is becoming more and more interdisciplinary !
- The NGSC is a highly sophisticated machine with massively parallel computer.

It is essential for AICS to promote interdisciplinary cooperation between computer scientists and computational scientists and also cooperation among computational scientists in various areas.

We are truly grateful for your interest, and eager for your cooperation, advice, and support

Research Activities

UTChem

Ab initio Theory

MRMP, MCQDPT
CASVB, QCAS, SPS-SCF&PT
Linear Scaling Method
(GFC)

DFT

OP Correlation
Parameter-Free Exchange
Long-range corrected functional
van der Waals
TDDFT
Dual-level DFT

Relativistic Theory

RESC, DK3
Dirac-Hartree-Fock
Dirac-Kohn-Sham
Ab initio Model Potential
Relativistic Basis Sets

Dynamics

Ab initio dynamics
(TD)DFT dynamics
Hybrid QM/MM
VSCF,VCI

Two Approaches To Molecular Theory

- **Wave Function Method**
  - Huge and successful efforts in the last 30 years
  - State-of-the-art methodology
  - Systematic, Converging to exact solution.
  - Accurate results for small systems
  - Steep N dependence of the computational effort

- **Density Method (Density Functional Theory)**
  - Not reached such a mature stage as wave function method
  - Not *ab initio*, but best semi-empirical
  - Simple and conceptual, Applicable to large systems
  - Accuracy depends on *xc* functionals

Present Status Of Quantum Chemistry

Owing to the theoretical developments and high-speed computers, quantum chemistry can now describe the properties of small to medium scale molecules with chemical accuracy (2 kcal/mol or 0.1eV) comparable to those of experiment.

Spectroscopic Constants of Diatomic Hydrides

Main-group elements across the second- through fifth-period of the periodic table

J.Chem.Phys. **120**, 3297 (2004)

BH	CH	NH	OH	FH
AlH	SiH	PH	SH	ClH
GaH	GeH	AsH	SeH	BrH
InH	SnH	SbH	TeH	IH

DK3-CCSD, DK3-CCSDT, DK3-CCSDTQ

Re-contracted relativistic cc-pVnZ (n=2-5)

Extrapolation

DK3 is our third-order Douglas-Kroll approximation

Spectroscopic Constants of CH and InH

	^{12}CH ($^2\Pi$)		^{115}InH($^1\Sigma^+$)	
	Theory	**Exp.**	**Theory**	**Exp.**
r_e / Å	1.120	1.120	1.841	1.838
r_0 / Å	1.131	1.130	1.855	1.851
B_e / cm^{-1}	14.449	14.457	4.977	4.995
B_0 / cm^{-1}	14.182	14.190	4.905	4.923
α_e / cm^{-1}	0.534	0.534	0.144	0.143
D_e / cm^{-1}	0.00148	0.00145	0.000229	0.000223
ω_e / cm^{-1}	2860	2859	1467	1476
$x\omega_e$ / cm^{-1}	66	63	26	26
ν_0 / cm^{-1}	1413	1413	727	732
ν_1 / cm^{-1}	4141	4146	2142	2157
ν_2 / cm^{-1}	6743	6752	3510	3535
D_0^0 / eV	3.47	3.47	2.46	2.48

Theory (DK3-CC) Can Predict Experimental

Bond lengths (r_e or r_0) within 0.002 Å
Rotational constants (B_e or B_0) within 0.02 cm^{-1}
Vibration-rotation constants (α_e) within 0.01 cm^{-1}
Centrifugal distortion constants (D_e) within 2 %
Harmonic vibrational constants (ω_e) within 9 cm^{-1}
Anharmonic vibrational constants ($\chi\omega_e$) within 2 cm^{-1}
Dissociation energies (D_0^0) within 0.02 eV (0.4 kcal/mol)

The theoretical best estimate will substitute for the missing experimental data for some fourth- and fifth-row hydrides.

Spectroscopic Constants Of The $^2\Pi$ State Of ^{130}TeH

	Theory	**Exptl.**
r_e / Å	1.656	1.656[a]
r_0 / Å	1.667	1.741[b]
B_e / cm^{-1}	6.149	
B_0 / cm^{-1}	6.067	5.56[b]
α_e / cm^{-1}	0.165	
D_e / cm^{-1}	0.000200	
ω_e / cm^{-1}	2144	(2137)[a]
$x\omega_e$ / cm^{-1}	39	
ν_0 / cm^{-1}	1062	
ν_1 / cm^{-1}	3128	
ν_2 / cm^{-1}	5119	
D_0^0 / eV	2.76	

Vibrational Theory

The Raman–Mizushima Lecture, IIT Bombay, 4th February 2006

The Present State and Future of Raman Spectroscopy

Prof. Mitsuo Tasumi

"Molecular spectra, including Raman spectra, are letters from the world of molecules.

…

it will become more important to develop methods to extract useful information from such letters. In this context, the role of quantum chemical calculations will expand…."

Prof. Mitsuo Tasumi continued

“The goal that I think is ultimate to vibrational spectroscopy is to calculate as accurately as possible the potential energy surfaces of a target molecule and derive anharmonic vibrational levels. The anharmonic interactions between vibrational levels (Fermi resonance) should be taken into account.

Then, it will be possible to understand completely the infrared and Raman spectra observed from the molecule”

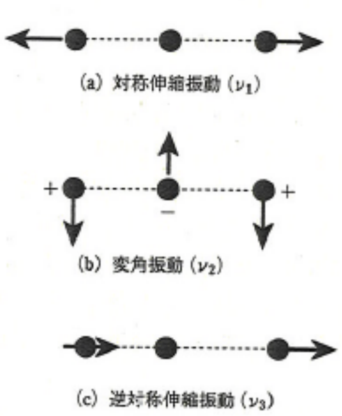
Fermi Resonance of CO_2

J.Chem.Phys. **126**, 124303 (2007)

A characterization of the spectroscopic properties of CO_2 is crucial for the understanding of chemistry of the seafloor, planetary atmospheres, and the greenhouse effect.

CO_2 has four normal modes of vibrations:

(a) a symmetric stretching mode (ν_1),
(b) degenerate bending modes (ν_2)
(c) an antisymmetric stretching mode (ν_3).



The accidental near degeneracy, $\nu_1 \approx 2\nu_2$, results in a significant anharmonic coupling between the two modes, which are pushed apart by the coupling. This is the well known ν_1-$2\nu_2$ Fermi resonance of CO_2.

Low-lying vibrational energy levels (in cm^{-1}) of CO_2 obtained with CCSD(T) and harmonic approximation

	Theory	Exptl.	Δ
ν_2 (01^10)	670.1	667.4	2.7
ν_1 (10^00)	1343.0	1285.4	57.6
ν_2^2 (02^00)	1344.6	1388.2	-43.6
ν_3 (00^01)	2354.5	2349.2	5.3

ν_1 is a symmetric stretching mode

ν_2 is degenerate bending modes

ν_3 is an antisymmetric stretching mode

ν_2^2 is the first overtone

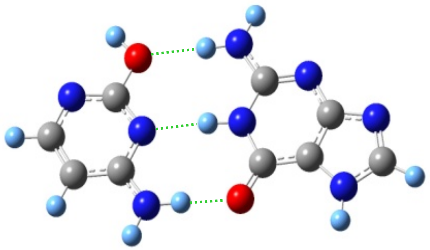
Low-lying vibrational energy levels (in cm^{-1}) of CO_2 obtained with CCSD(T) and vibrational CI

	Theory	**Exptl.**	
ν_2 (01^10)	668.5	667.4	
ν_1 (10^00)	1288.3	1285.4	Fermi doublet
ν_2^2 (02^20)	1338.2	1335.1	
ν_2^2 (02^00)	1388.7	1388.2	Fermi doublet
$\nu_1\nu_2$ (11^10)	1937.0	1932.5	Fermi doublet
ν_2^3 (03^30)	2009.6	2003.2	
ν_2^3 (03^10)	2078.6	2076.9	Fermi doublet
ν_3 (00^01)	2349.2	2349.2	

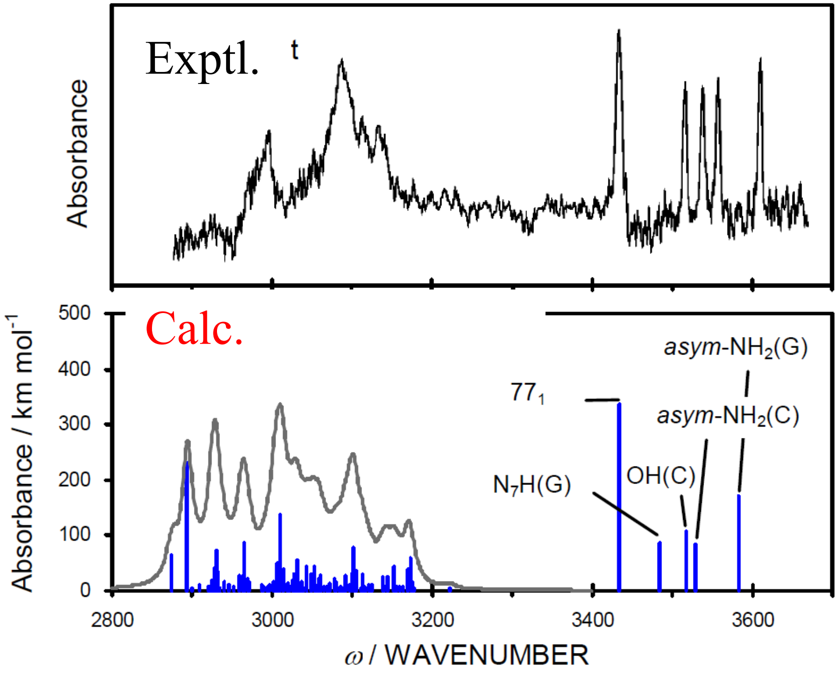
The accidental near degeneracy, $\nu_1 \approx 2\nu_2$, results in a significant anharmonic coupling between the two modes, ν_1-$2\nu_2$ Fermi resonance.

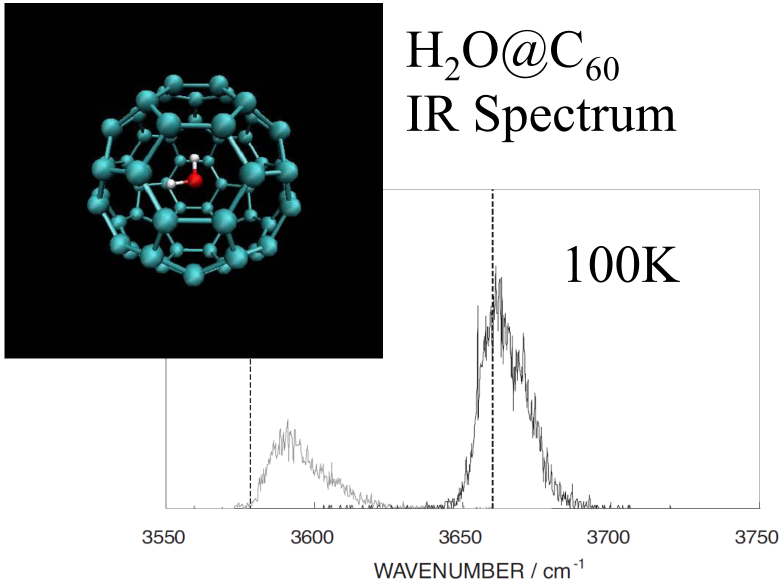
Vibrational Energies for Large Systems

Development of VSCF, VCI and VPT
Anharmonic vibrational state calculation



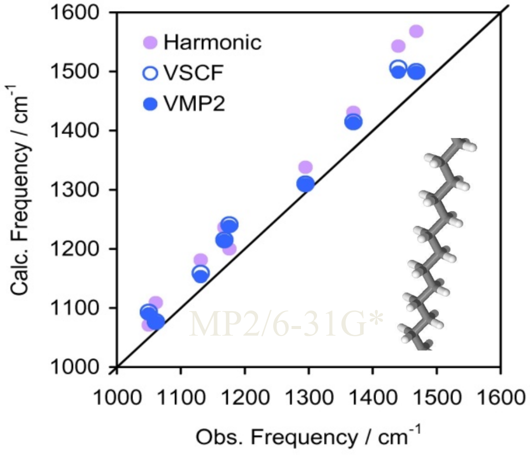
Vibrational Dynamics of
Guanine-Cytosine Base Pair



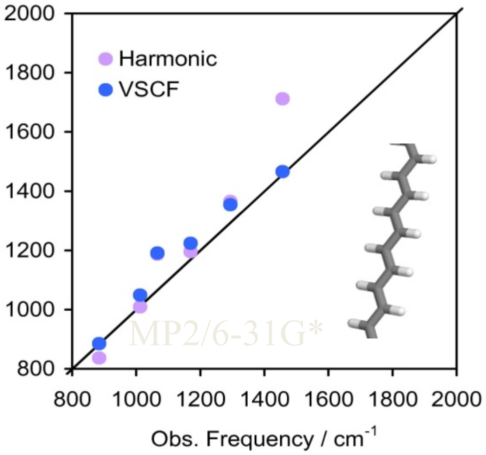


$H_2O@C_{60}$
IR Spectrum

Polymers



—Polyethylene—



—Polyacetylene—

Topics

1. **Hybrid GGA Functional Based on the Long-Range Correction**
   LC, LCgau, LC2gau functionals
2. **Systematic Errors in Computed Alkane Energies**
   DFT can be applied to large molecules?
3. **DFT Koopmans' Theorem**
   Kohn-Sham orbital energies and orbitals have physical meaning?
4. **Linear-Scaling Method For Computing Coulomb Potentials**
   Gaussian And Finite-Element Coulomb Method (GFC)

Acknowledgements

Collaborators

Dr T Nakajima Dr T Tsuneda Dr J-W Song Dr D Peng
Dr A Nakata Dr Y Nakatsuka Dr M Watson (Harvard)
Dr T Sato (Tokyo) Dr K Yagi (Yamanashi)

Dr A Savin (Paris) Dr H Sekino (Toyohashi)

Thanks to Ministry of Education, Japan Science and Technology Agency (JST), past and present group, and many others.

Density Functional Theory (DFT)

DFT is one of the most widely used quantum chemical methods, and it enables us to obtain reasonable chemical properties with an effectively short computational time. It is expected that DFT will be applied to real-size biochemical systems and nano-materials using parallelizing and linear-scaling techniques.

One-Electron Density Is Sufficient To Determine All Molecular Properties

The cusps of the density tell us where the nuclei are, the shape of the cusp tells us what they are

$$\frac{\partial}{\partial r_A} \rho(r_A) \Big|_{r_A = 0} = -2 Z_A \rho(0)$$

The integral of the density tells us the number of electrons

$$\int \rho(r) dr = N$$

Thus, Hamitonian is known from which everything is known!

Kohn-Sham DFT

For a given primary functional, the *xc* energy is written

$$E_{xc}[\rho] = \int \rho(r)\varepsilon_{xc}[\rho(r)]dr$$

KS operator associated with this functional is

$$F_{KS} = -\frac{1}{2}\nabla^2 - \sum \frac{Z_A}{|R_A - r|} + \int \frac{\rho(r')}{|r-r'|}dr' + V_{xc}[\rho], \qquad V_{xc}[\rho] = \varepsilon_{xc}[\rho] + \rho\frac{\delta\varepsilon_{xc}[\rho]}{\delta\rho}$$

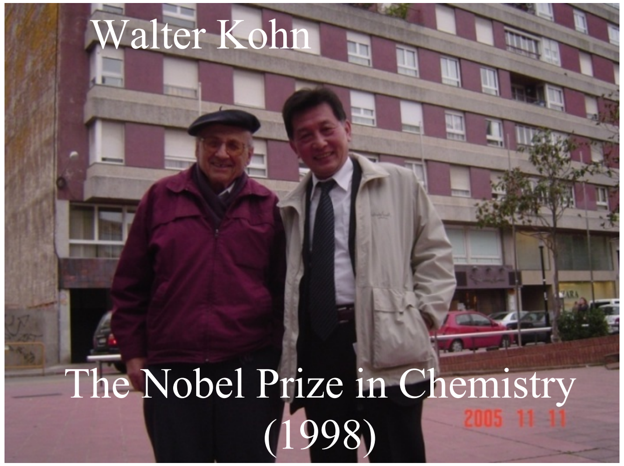
For a 2*n*-electron system, the KS orbitals satisfy

$$F_{KS}\varphi_i = \varepsilon_i \varphi$$

and related to the density by

$$\rho(r) = 2\sum_i^n \varphi_i^*(r)\varphi_i(r)$$

If V_{xc} were the *exact* exchange-correlation potential, the self-consistent Kohn-Sham solution would lead to the *exact* ground state charge density of the system.



Walter Kohn

The Nobel Prize in Chemistry (1998)

Topics

1. **Hybrid GGA Functional Based on the Long-Range Correction**
   LC, LCgau, LC2gau functionals
2. **Systematic Errors in Computed Alkane Energies**
   DFT can be applied to large molecules?
3. **DFT Koopmans' Theorem**
   Kohn-Sham orbital energies and orbitals have physical meaning?
4. **Linear-Scaling Method For Computing Coulomb Potentials**
   Gaussian And Finite-Element Coulomb Method (GFC)

Exchange Functionals

1. **LDA (SVWN)** - based upon exact exchange energy for a uniform electron gas (1980)

2. **GGA (BLYP)** - Becke's exchange functional (1988) is two-order more accurate than LDA

3. **Hybrid (B3LYP)** - The most popular functional in use today is Becke's hybrid functional (1993)

However, DFT has several problems to be solved even for ground state calculations.

Conventional GGA Has Problems

1. Barrier heights in chemical reactions underestimated.
2. Van der Waals interactions repulsive
3. Excitations using time-dependent DFT for Rydberg and CT states underestimated
4. Band gaps of insulators too small
5. Optical response function too large
6. And and and …

The failure arises from the wrong long-range behavior due to the local character of the approximate *xc* functional.

Long-Range Interactions

The conventional DFT models use local functional $E_{XC}(\rho)$ dependent only on ρ , or semilocal functions $E_{XC}(\rho,\nabla\rho)$ dependent on ρ and its gradient $\nabla\rho$. The most *xc* functionals do not satisfy the correct asymptotic behavior

$$\lim_{R\to\infty} V_{XC}(R) = -\frac{1}{R}, \quad V_{XC} = \frac{\delta E_{XC}}{\delta\rho}$$

Due to a local or semilocal character of these functionals, one cannot expect that the long range interaction will be described by current DFT models adequately.

Hybrid LC Functional

The electron repulsion operator, $1/r_{12}$, is divided into short-range (SR) and long-range (LR) parts using a standard error function:

$$\frac{1}{r_{12}} = \underbrace{\frac{1-erf(\mu r_{12})}{r_{12}}}_{\textbf{SR}} + \underbrace{\frac{erf(\mu r_{12})}{r_{12}}}_{\textbf{LR}}$$

where μ is the parameter which determines the proportion between the two ranges depending on the value of r_{12}.

DFT exchange functional is included through the first term, and the long range orbital–orbital exchange interaction is described using the HF exchange integral.

The Partitioning Of $1/r_{12}$ For Exchange Contribution

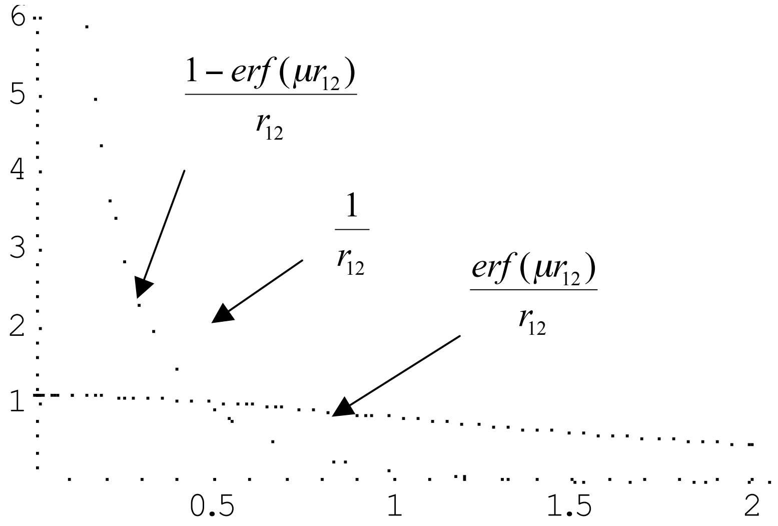
J.Chem.Phys., **120**, 8425 (2004)
J.Chem.Phys., **115**, 3540 (2001)

An Ewald partitioning

$$\frac{1}{r_{12}} = \frac{1-erf\left(\mu r_{12}\right)}{r_{12}} + \frac{erf\left(\mu r_{12}\right)}{r_{12}}$$

$$\Uparrow \qquad\qquad \Uparrow$$

DFT exchange *HF exchange*



$\mu = 0$ *DFT*

$\mu = \infty$ *WF*

$\mu \cong 0.5$ *LC*

LC Solves Many Of The Problems:

The underestimation of Rydberg excitation energies and corresponding oscillator strengths, the poor reproduction of CT excitations in TD-DFT calculations, and the overestimation of linear and nonlinear polarizabilities of long-chain molecules in CP-DFT calculations are improved. Moreover, the LC successfully provided a good description of vdW interactions as well as accurate reaction enthalpies and barrier heights.

Improved LC Functional

We modified the partitioning with a Gaussian correction such that the long-range part takes the form

$$L = \frac{erf(\mu r_{12})}{r_{12}} - k\frac{2\mu}{\sqrt{\pi}}e^{-(1/a)\mu^2 r_{12}^2} \qquad LCgau$$

$$L = \frac{erf(\mu r_{12})}{r_{12}} - k\frac{2\mu}{\sqrt{\pi}}e^{-(1/a)\mu^2 r_{12}^2} - k'\frac{2\mu}{\sqrt{\pi}}e^{-(1/a')\mu^2 r_{12}^2} \qquad LC2gau$$

LCgau preserves 100% exact exchange in the long range

$$r_{12} \rightarrow \infty \qquad HF$$

and allows a more flexible description at short-range.

In addition **LC2gau** removes all HF exchange contributions as the inter-electronic distance tends to zero, which agrees with our intuition.

$$r_{12} \rightarrow 0 \qquad DFT$$

Errors for Barrier Height and Reaction Enthalpy

Results from 148 atomization energies in the G2 test set.

Results from 72 barrier heights and 26 reaction enthalpies from Database/3. Root mean square errors are given for all calculations using the 6-311+ +G(2*d*,2*p*) basis set and optimized geometries

	Barrier Heights	Reaction Enthalpies	Atomization energies	
B3LYP	4.74	3.03	7.5	
LC-BOP	2.53	2.18	10.1	
CAM-BLYP	2.49	2.25	7.8	
LCgau-BOP	**2.43**	**2.00**	**5.3**	
LC2gau-BOP	**2.39**	**1.87**	**5.9**	*in kcal/mol*

Errors for Bond Length and Bond Angle

Results from 329 bond lengths and 271 bond angles in the G2 test set. Root mean square errors are given for all calculations using the 6-311++G(2*d*,2*p*) basis set and optimized geometries

	Bond Length (A)	**Bond Angle (Degree)**
B3LYP	0.0133	1.018
LC-BOP	0.0162	1.053
CAM-BLYP	0.0183	1.020
LCgau-BOP	**0.0127**	**0.936**
LC2gau-BOP	**0.0127**	**0.962**

Errors for Excitation Energies (eV)

Results from 27 valence excitation energies and 41 Rydberg excitation energies for N_2, CO, H_2CO, C_2H_4, C_6H_6. Root mean squarer errors are given for all calculations using Sadlej pVTZ basis set

	Valence Excitations	**Rydberg Excitations**
B3LYP	0.36	0.89
LC-BOP	0.36	0.36
CAM-BLYP	0.45	0.28
LCgau-BOP	**0.32**	**0.23**
LC2gau-BOP	**0.31**	**0.23**

Errors for Core-Excitation Energies (eV)

Results from 8 1s-valence (singlet), 3 1s-valence (triplet), excitation energies for N_2, CO, H_2CO, C_2H_4, C_6H_6. Root mean squarer errors are given for all calculations using Sadlej pVTZ basis set

	1s-valence (singlet)	**1s-valence (triplet)**
B3LYP	12.3	10.7
LC-BOP	18.8	15.6
CAM-BLYP	3.8	3.3
LCgau-BOP	**5.8**	**0.2**
LC2gau-BOP	**0.8**	**0.4**

LCgau and LC2gau

The LCgau and LC2gau partitioning preserves *100% exact exchange in the long range* (r_{12}>1.5 au), which is known to be important for the success of the LC scheme.

Although LC2gau includes an extra parameter (so 4 in total) it has *100% DFT at short range*, which agrees with our intuition.

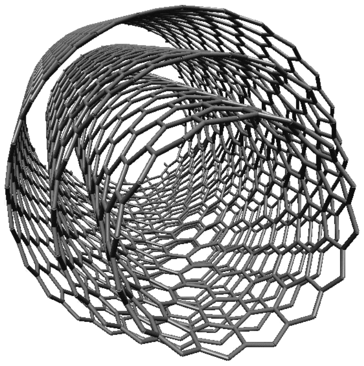
Accurate results can be obtained with such a partitioning in terms of the reproduction of atomization energies, barrier heights, reaction enthalpies, geometrical properties, and excitation energies (including CT and core excitations) over a wide range of molecular systems

Van Der Waals Interactions

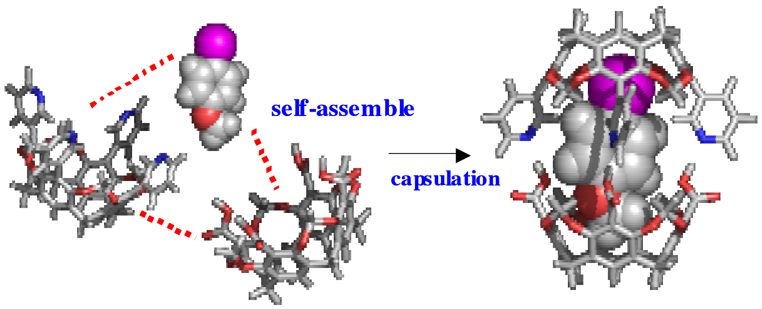
Van der Waals (dispersion) interactions play an important role in many chemical systems. They control the structure of DNA and proteins, the packing of crystals, the formation of aggregates, host-guest systems, the orientation of molecules on surfaces or in molecular films, etc.

Unfortunately almost all GGA DFTs are unable to describe dispersive interactions.

Biomolecules



Nanotubes



Host-Guest systems

Van Der Waals Interactions

$$\Delta E^{vdW} = E\ \text{(Pauli repulsion)} - E\ \text{(Dispersion attraction)}$$

LDA predicts the binding character of vdW interactions. However, LDA severely overestimates the binding energy and yields a too short vdW bond. The GGA predicts repulsive vdW interactions. Thus, none of the functionals account successfully for vdW interactions.

MP2 significantly overestimates the binding energies and MP2 results have strong basis set dependence.
Only CCSD(T) with a large basis set gives the accurate estimation.

Van Der Waals Interactions

$$E^{Total} = E^{LC-GGA} + E^{dispersion}$$

Van der Waals functional (Andersson et al, PRL 1996)

$$E^{dispersion} = -\frac{6}{4(4\pi)^{3/2}} \int_{V_1} d^3\mathbf{r}_1 \int_{V_2} d^3\mathbf{r}_2 \, \frac{\sqrt{\rho_1(\mathbf{r}_1)\rho_2(\mathbf{r}_2)}}{\sqrt{\rho_1(\mathbf{r}_1)} + \sqrt{\rho_2(\mathbf{r}_2)}} \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|^6}$$

The functional provides an accurate dispersion energy between well-separated electron distributions.

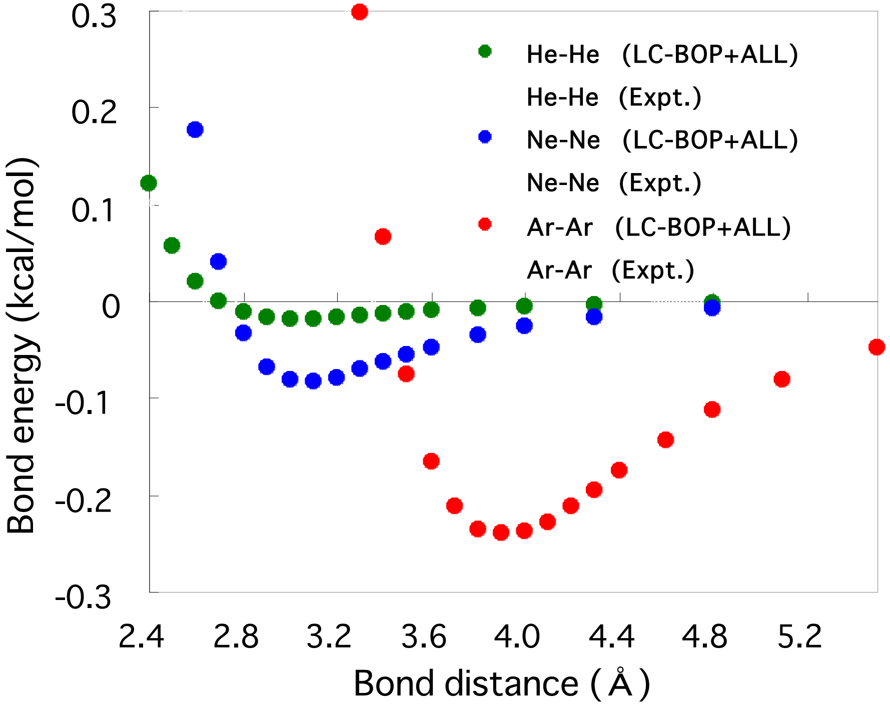
multiplied by a damping factor

$$f_{damp} = \exp\left[-\left(\frac{\alpha_{AB}}{r_{12}}\right)^6\right] \qquad (\alpha_{AB} = C_1 R_{AB} + C_2)$$

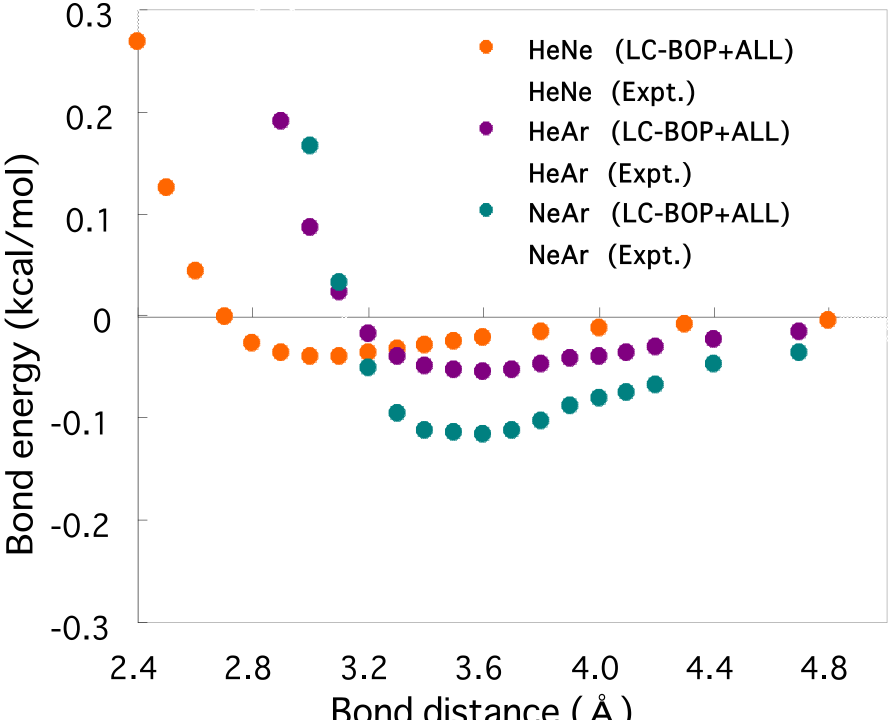
R_{AB} : van der Waals radius

Potential Energy Curves of Rare-gas Dimers

Basis set: aug-cc-pVTZ, BSSE-corrected



Homo-nuclear

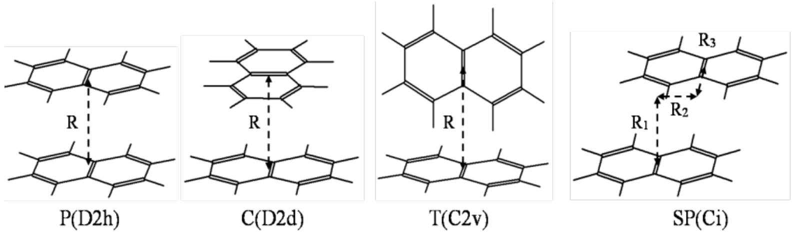


Hetero-nuclear

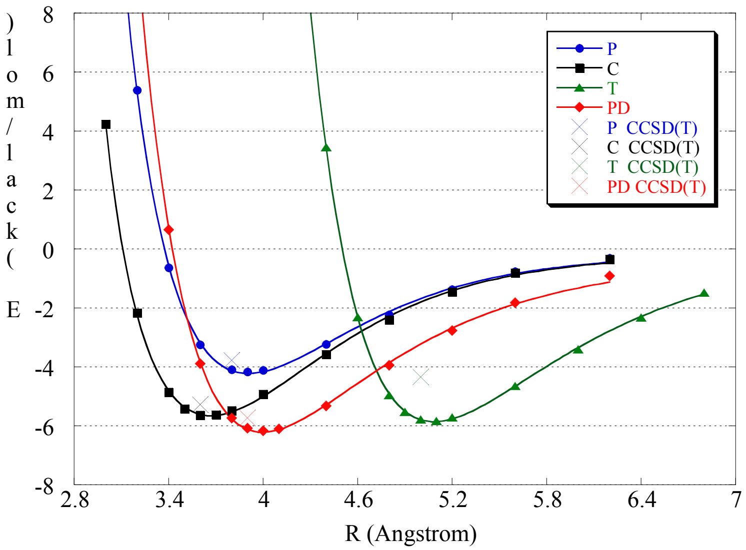
van der Waals interaction

Naphthalene Dimer

GGA predicts repulsive vdW interactions



Our DFT with 6-31+G**



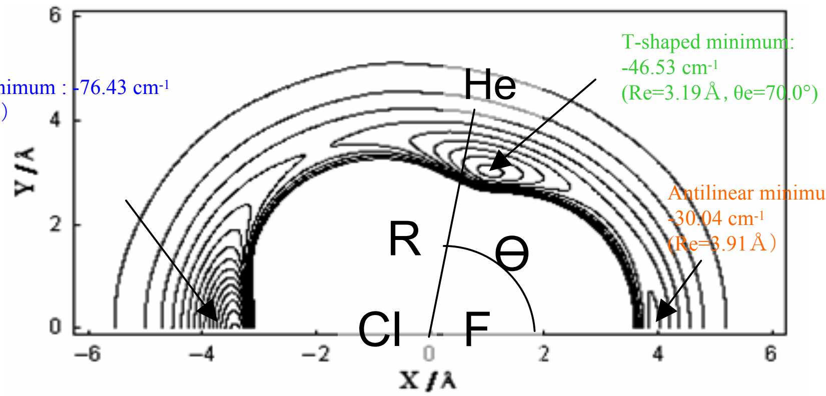
Our DFT 6.16 kcal/mol

CBS CCSD(T) 5.73 kcal/mol

ClF…He

GGA predicts only linear structures

Our DFT with aug-cc-pVQZ

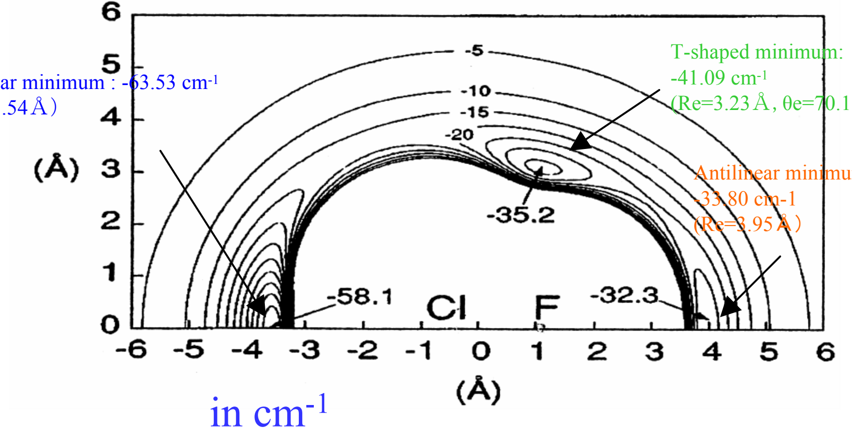


Collinear minimum : -76.43 cm^{-1}
(Re = 3.41 Å)

T-shaped minimum:
-46.53 cm^{-1}
(Re=3.19 Å, θe=70.0°)

Antilinear minimum:
-30.04 cm^{-1}
(Re=3.91 Å)

CCSD(T) with aug-cc-pVQZ+BF(3s3p2d2f1g)



Collinear minimum : -63.53 cm^{-1}
(Re = 3.54 Å)

T-shaped minimum:
-41.09 cm^{-1}
(Re=3.23 Å, θe=70.1°)

Antilinear minimum:
-33.80 cm-1
(Re=3.95 Å)

in cm^{-1}

Topics

1. **Hybrid GGA Functional Based on the Long-Range Correction**
   LC, LCgau, LC2gau functionals
2. **Systematic Errors in Computed Alkane Energies**
   DFT can be applied to large molecules?
3. **DFT Koopmans' Theorem**
   Kohn-Sham orbital energies and orbitals have physical meaning?
4. **Linear-Scaling Method For Computing Coulomb Potentials**
   Gaussian And Finite-Element Coulomb Method (GFC)

Systematic Errors In Computed Alkane Energies

Paul v R Schleyer et al, Org. Lett. (2006)

Energies computed by B3LYP and other popular DFT functionals are flawed by systematic errors, which can become considerable for larger molecules.

These errors, predominately due to inadequacies in assessing longer-range nonbonded attractive effects (dispersion), are illustrated by the isodesmic stabilization energies of *n*-alkanes

Newer functionals, designed to describe weak interactions, give somewhat better agreement with experiment, but are not fully satisfactory.

Isodesmic stabilization energies (*kcal/mol*) of the linear conformations of *n*-alkanes

$$n\text{-}CH_3(CH_2)_mCH_3 + mCH_3 \rightarrow (m+1)C_2H_6$$

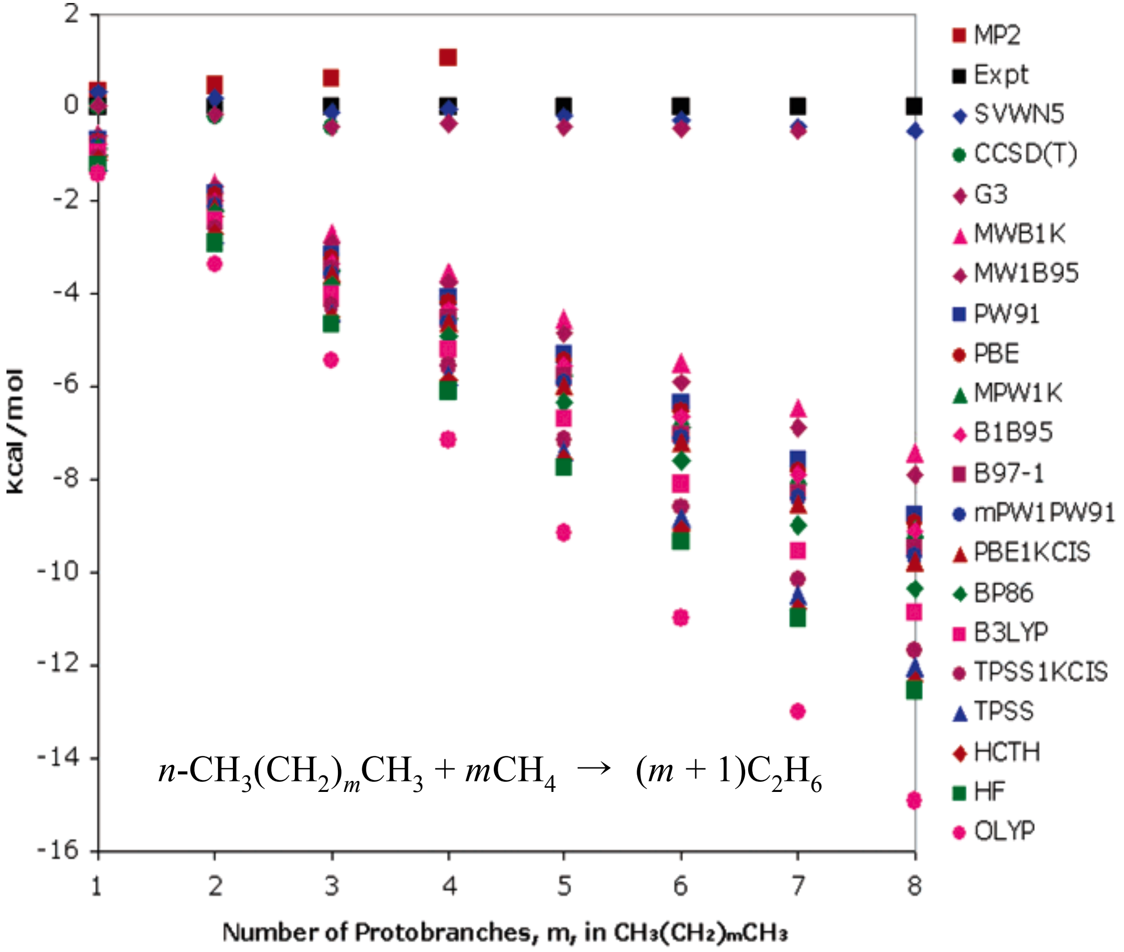
Deviations from experimental protobranching stabilization energies. Negative values denote underestimation

m	B3LYP	wB97X	M06	SVWN5
1	-0.97	-0.27	-0.54	0.29
2	-2.45	-1.07	-1.52	0.14
3	-4.01	-1.96	-2.60	-0.14
4	-5.22	-2.58	-3.61	-0.05
5	-6.69	-3.28	-4.31	-0.24
6	-8.04	-3.92	-5.53	-0.30
7	-9.51	-4.75	-6.16	-0.48
8	-10.84	-5.43	-7.47	-0.52

No available density functional is generally reliable for large molecules!

Systematic Errors

n-alkane reaction enthalpy calculations



$$n\text{-}CH_3(CH_2)_mCH_3 + mCH_4 \rightarrow (m+1)C_2H_6$$

MP2 : overestimation
HF : underestimation

DFT (GGA, meta-GGA, and hybrid GGA) : underestimation

Only SVWN5 (LDA) shows nearly no errors on these systems.

G3 and CCSD(T) provide accurate results.

Deviations of various DFT functionals from experimental (0 K) proto branching stabilization energies. Negative values denote underestimation. Stabilization energies are based on eq 1. CCSD(T) and MP2 refer to CCSD(T)/aug-cc-pVTZ//MP2/6-311+G(d,p) and MP2/aug-cc-pVTZ//MP2/6-311+G(d,p), respectively, and include MP2/6-311+G(d,p) zero-point corrections. All other computations employed the 6-311+G(d,p) basis set.

Isodesmic stabilization energies (*kcal/mol*) of the linear conformations of *n*-alkanes

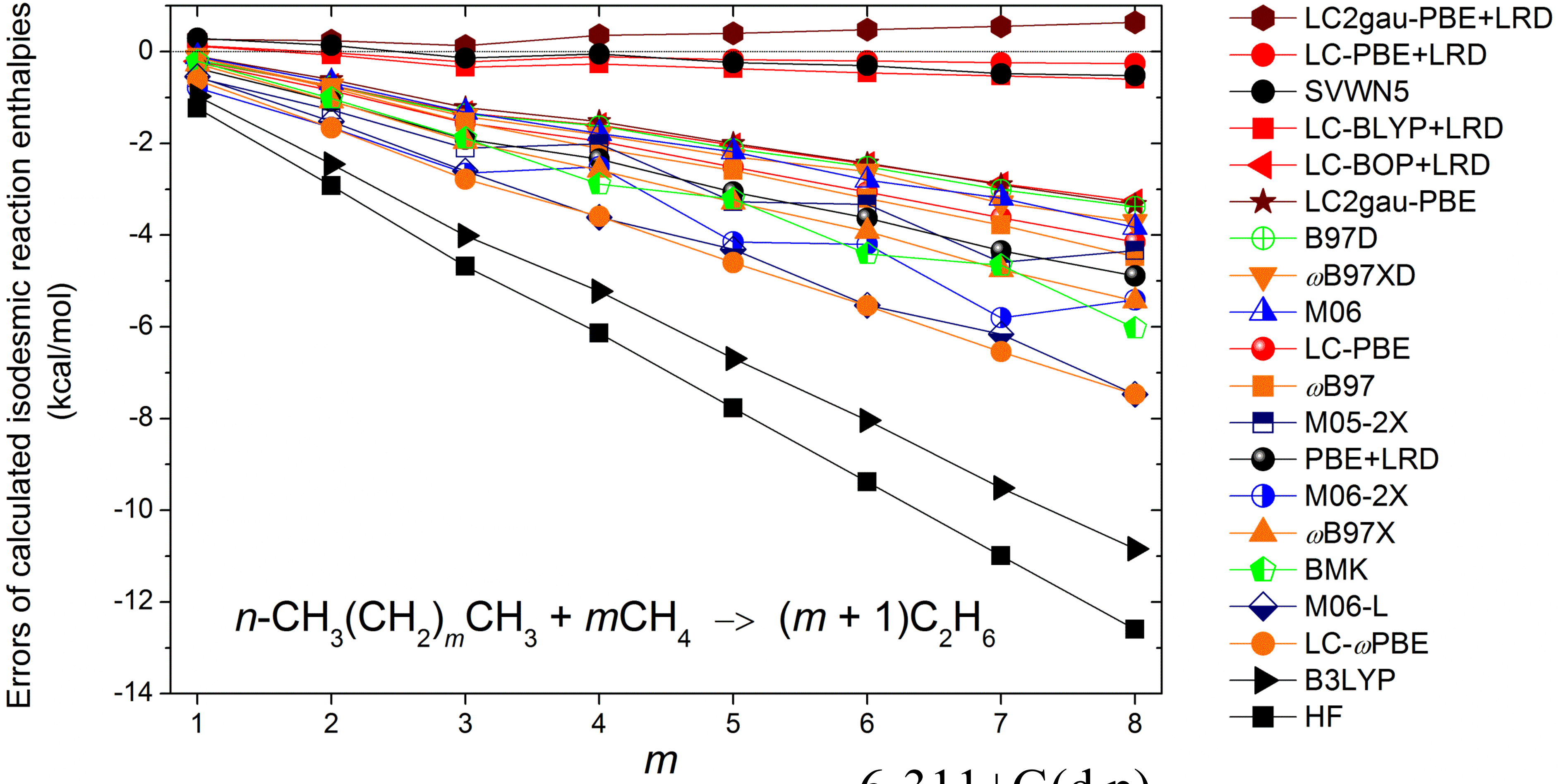
$$n\text{-}CH_3(CH_2)_mCH_3 + mCH_3 \rightarrow (m+1)C_2H_6$$

Organic Letters, 12, 1440-1443 (2010)

Deviations from experimental protobranching stabilization energies.
Negative values denote underestimation.

m	B3LYP	wB97X	M06	SVWN5	**LCgau-BLYP-LRD**	**LC-PBE-LRD**
1	-0.97	-0.27	-0.54	0.29	**0.28**	**0.21**
2	-2.45	-1.07	-1.52	0.14	**0.18**	**0.14**
3	-4.01	-1.96	-2.60	-0.14	**0.03**	**0.00**
4	-5.22	-2.58	-3.61	-0.05	**0.18**	**0.18**
5	-6.69	-3.28	-4.31	-0.24	**0.19**	**0.19**
6	-8.04	-3.92	-5.53	-0.30	**0.21**	**0.25**
7	-9.51	-4.75	-6.16	-0.48	**0.24**	**0.26**
8	-10.84	-5.43	-7.47	-0.52	**0.26**	**0.34**

Deviations of LC-BLYP+LRD from experimental protobranching stabilization energies



6-311+G(d,p)

Conclusions

The isodesmic reaction enthalpies of n-alkane, which have so far been known to give systematic errors in DFT calculations, are **successfully reproduced** using LC-DFT.

The failure of normal DFT calculations on isodesmic stabilization energies of *n*-alkanes exactly comes from the deficiencies of long-range exchange interactions and long-range dispersion energies.

Topics

1. **Hybrid GGA Functional Based on the Long-Range Correction**

   LC, LCgau, LC2gau functionals

2. **Systematic Errors in Computed Alkane Energies**

   DFT can be applied to large molecules?

3. **DFT Koopmans' Theorem**

   Kohn-Sham orbital energies and orbitals have physical meaning?

4. **Linear-Scaling Method For Computing Coulomb Potentials**

   Gaussian And Finite-Element Coulomb Method (GFC)

Hartree-Fock And Koopmans' Theorem

MO is the most fundamental quality in quantum chemistry.

Almost all of our understanding of "what the electrons are doing in molecule" is based on the MO concept. The success of Hartree-Fock method in describing the electronic structure has made it natural to analyze the wave function in terms of the MOs. The concept is simple and has a close relation to experiment through Koopmans' theorem.

$$\varepsilon_{HOMO} \cong -IP$$

(However, $\varepsilon_{LUMO} \neq -EA$)

DFT - Koopmans' Theorem

In DFT a similar theorem exists that relates the first ionization energy and electron affinity to the HOMO and LUMO energies.

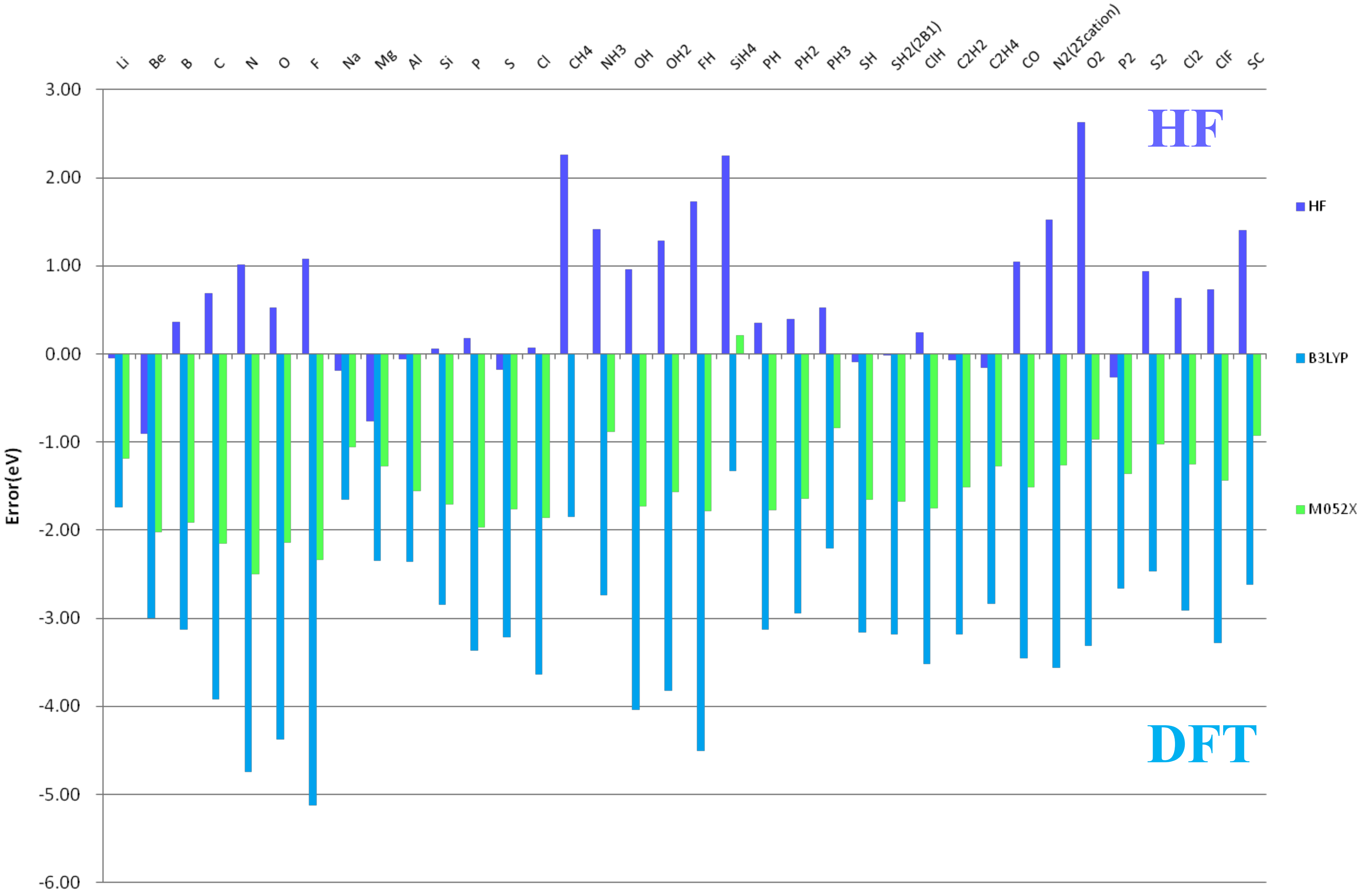
While these are *exact* statements in the formalism of DFT, the use of approximate *xc* potentials makes the calculated energies approximate.

It is believed that **Koopmans' theorem does not hold for DFT.** KS orbitals and orbital energies are thought to have no strict physical meaning.

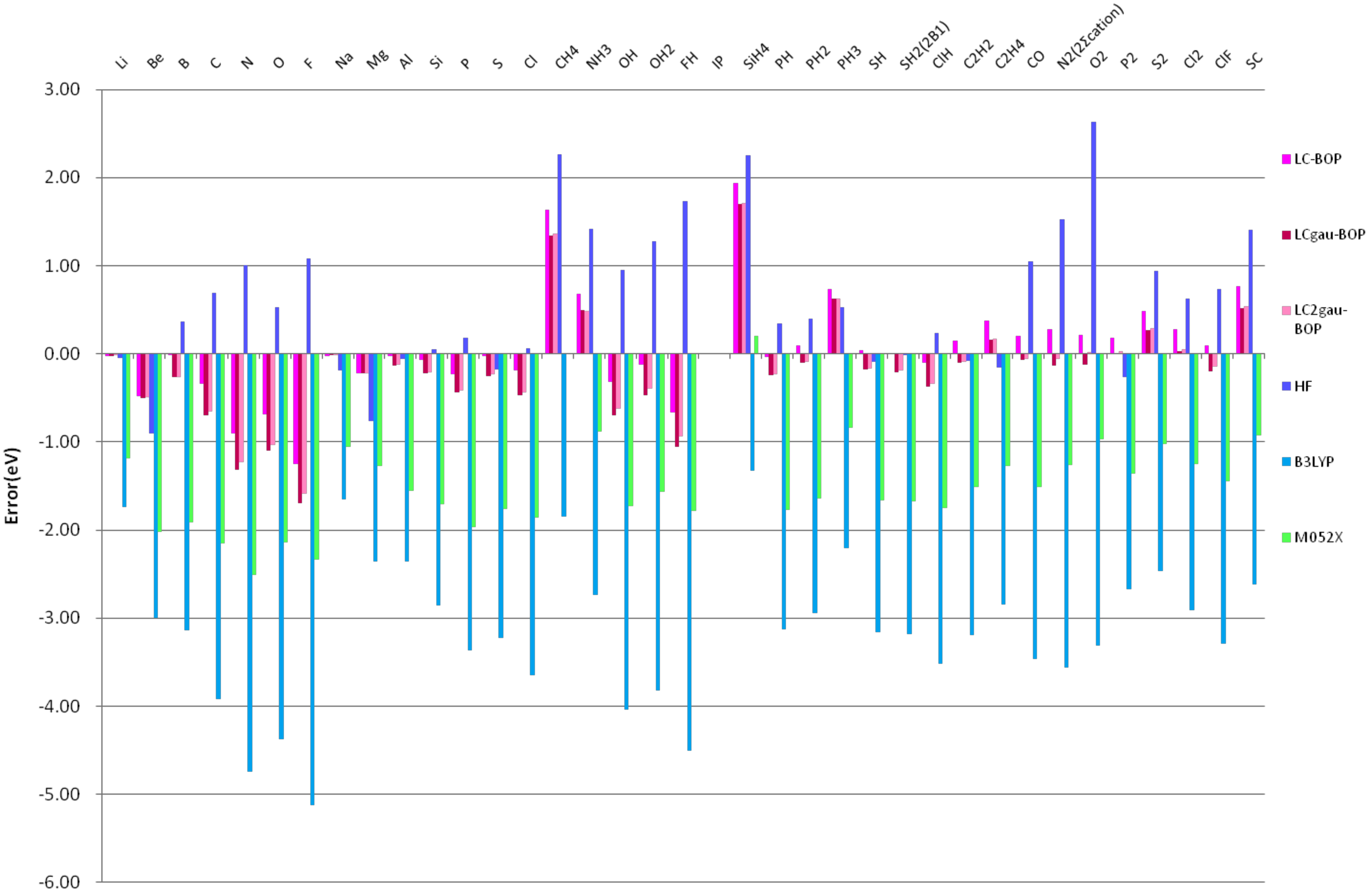
LC satisfies DFT-Koopmans' theorem

(T. Tsuneda et al)

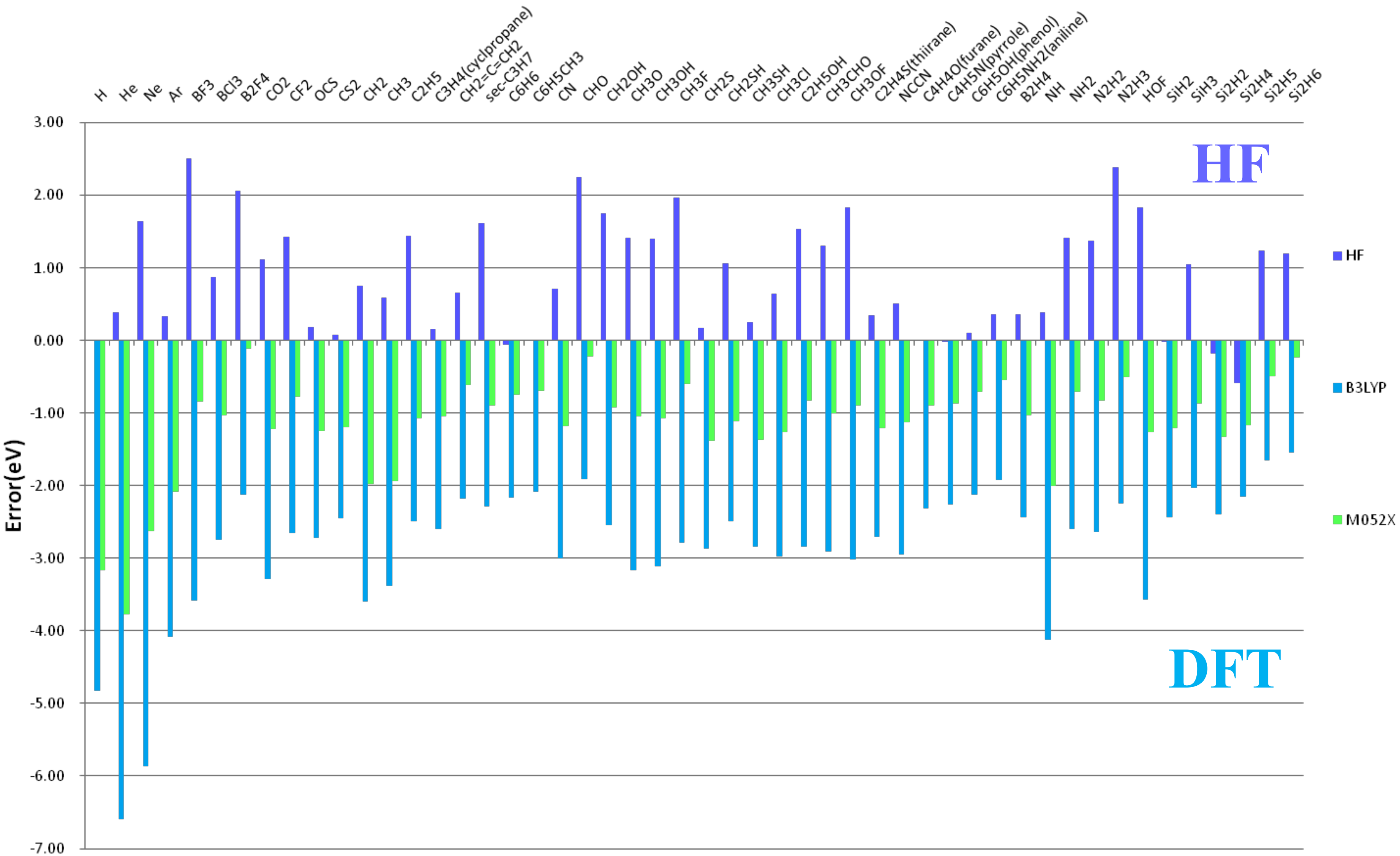
IP Koopmans' theorem - Deviation from experiment



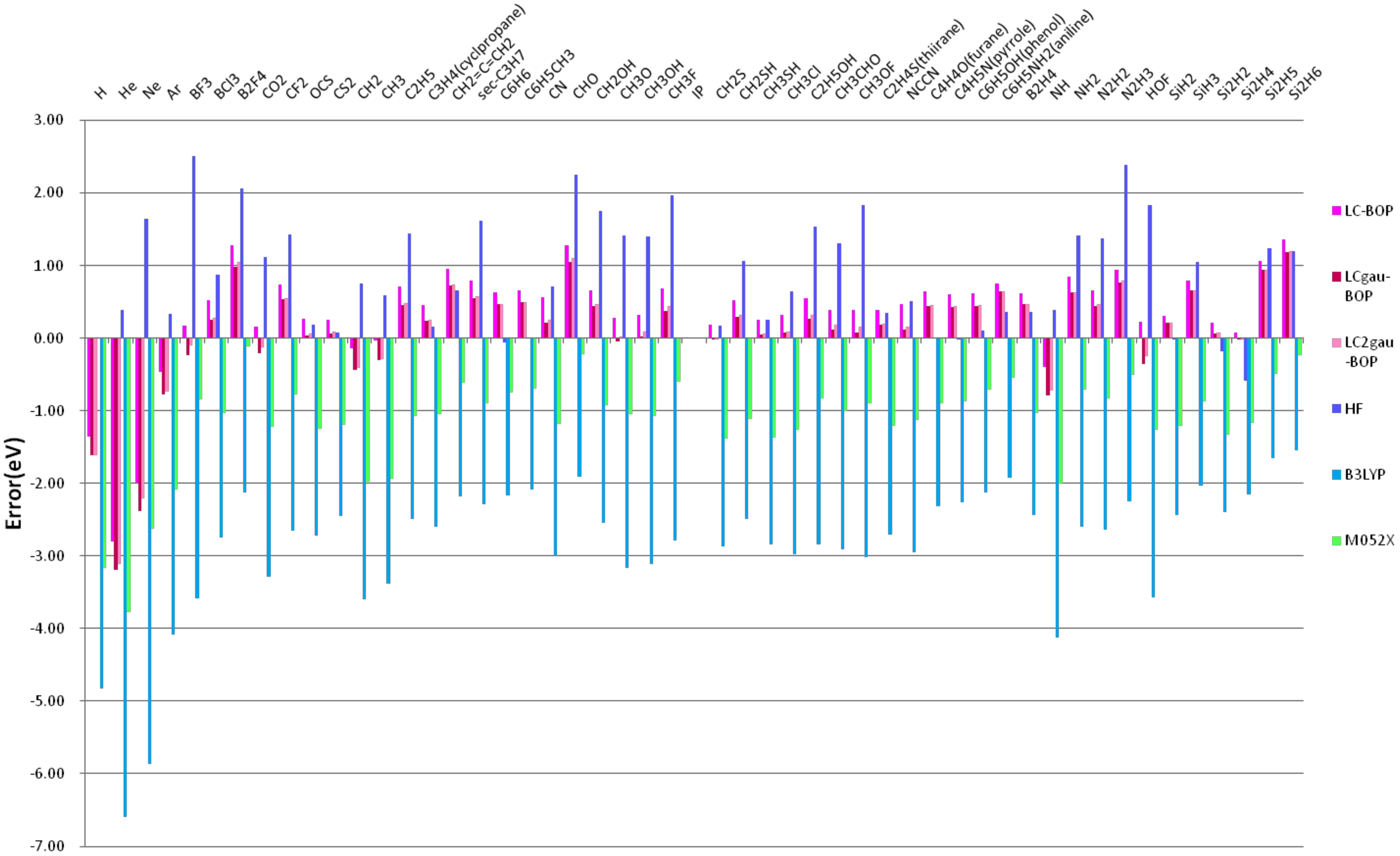
Ionization Potentials - Deviation from experiment



IP Koopmans’ theorem - Deviation from experiment



Ionization Potentials- Deviation from experiment

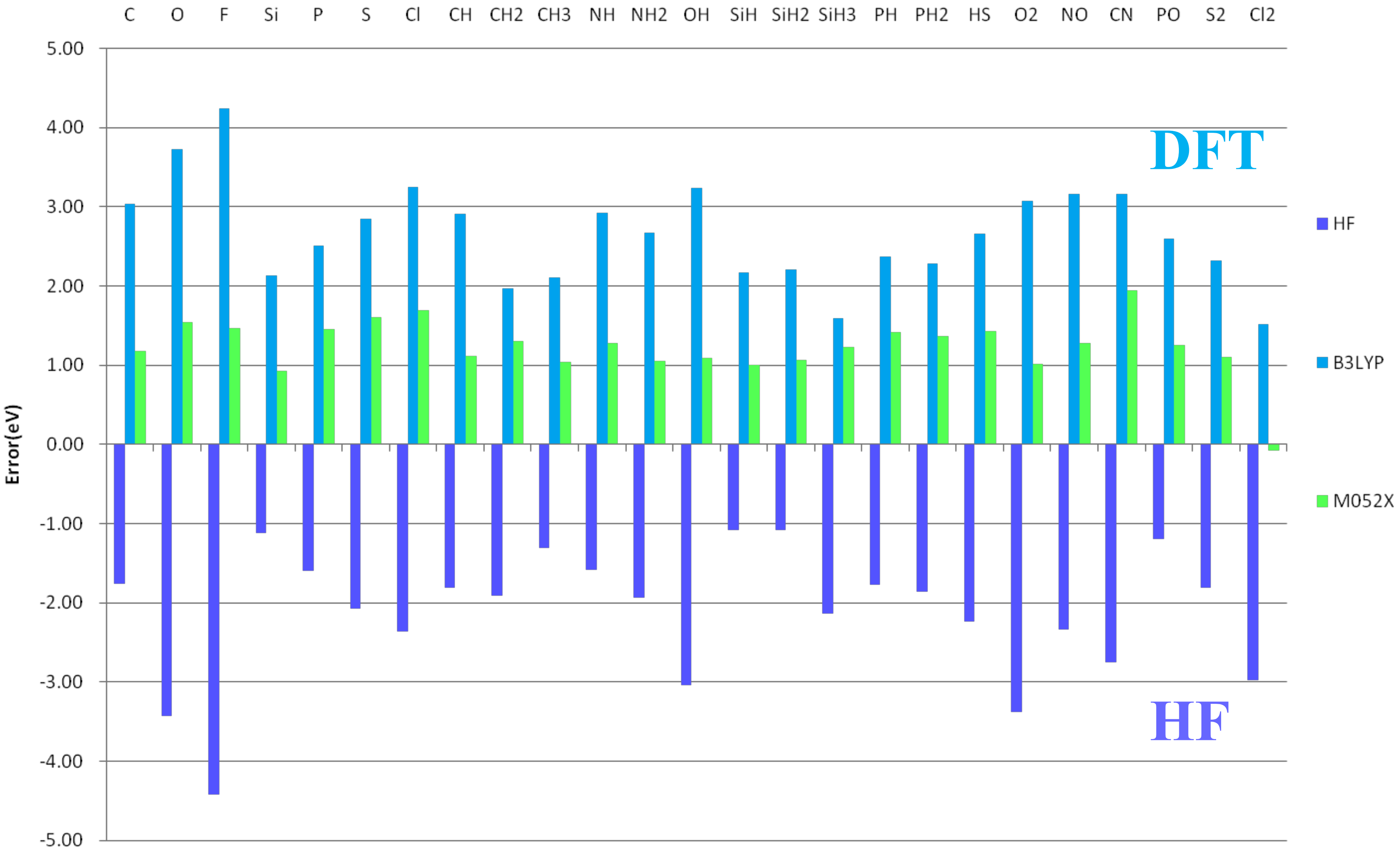


MAE for Ionization Potentials (eV)

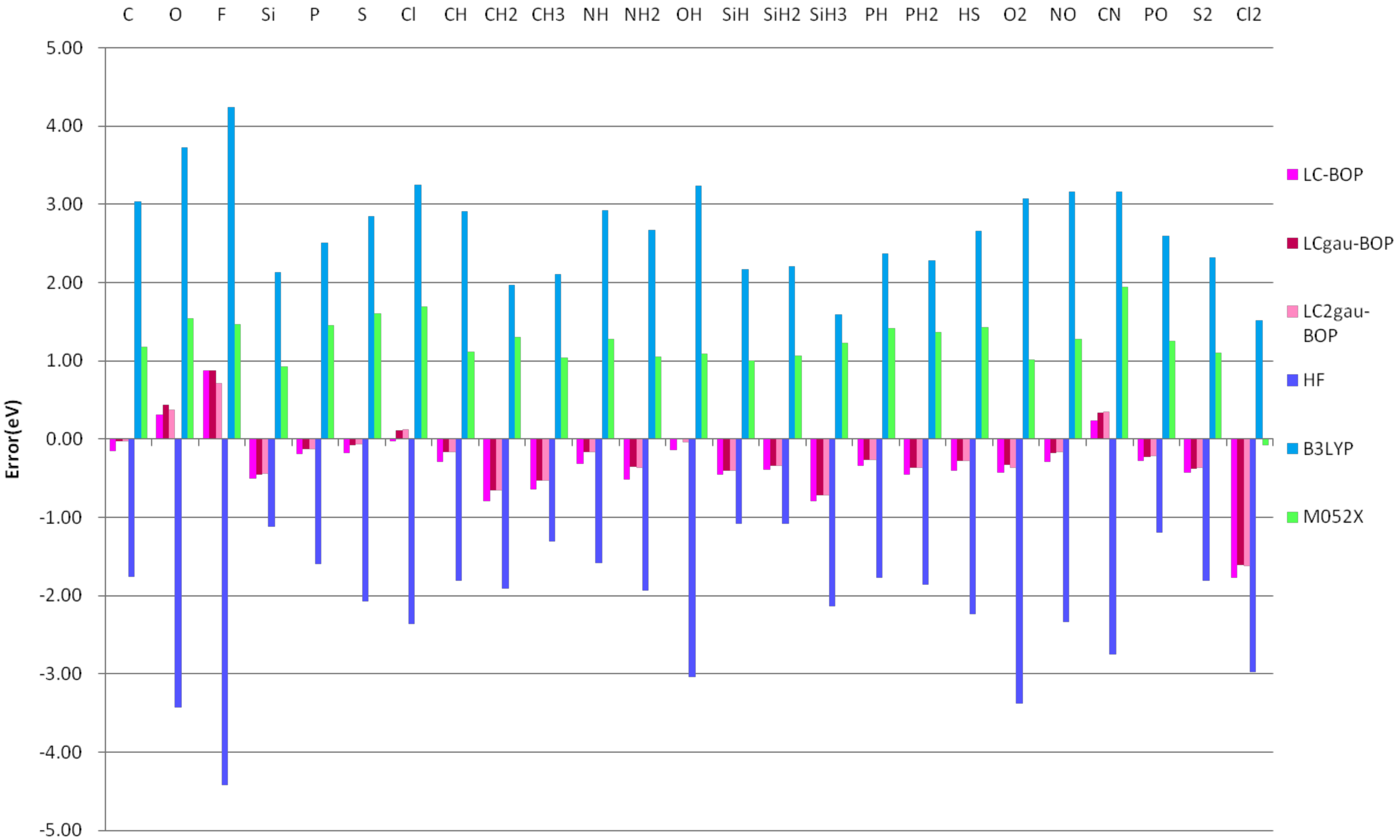
Results from molecules in the G2-1 and G2-2 sets. Mean absolute errors are given for all calculations using 6-311G++(2d,2p) basis set and optimized geometries.

	G2-1	**G2-2**
HF	0.78	0.86
B3LYP	3.22	3.01
M052X	1.57	1.33
LC-BOP	0.45	0.55
LCgau-BOP	0.53	0.52
LC2gau-BOP	0.51	0.51

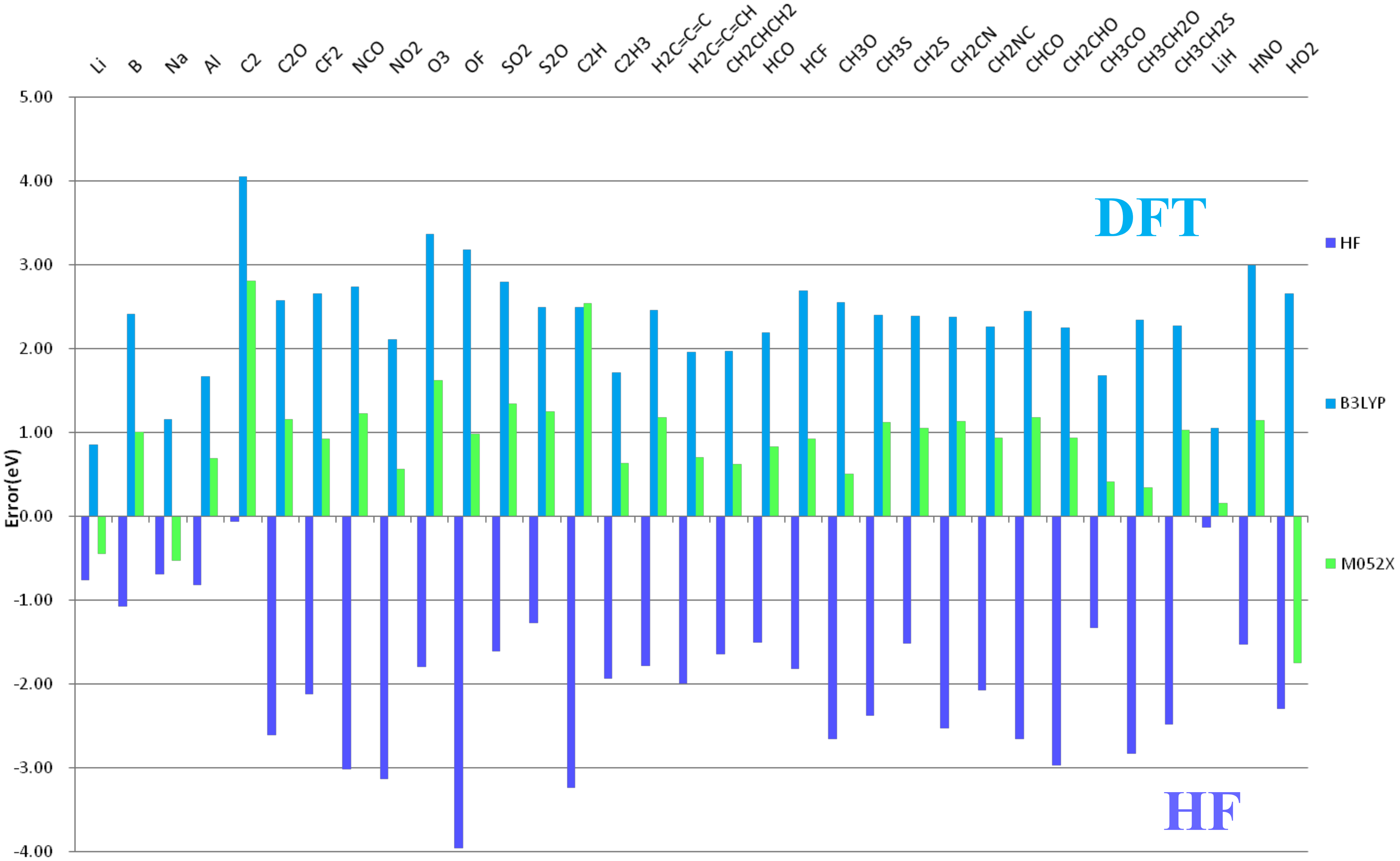
EA Koopmans' theorem - Deviation from experiment



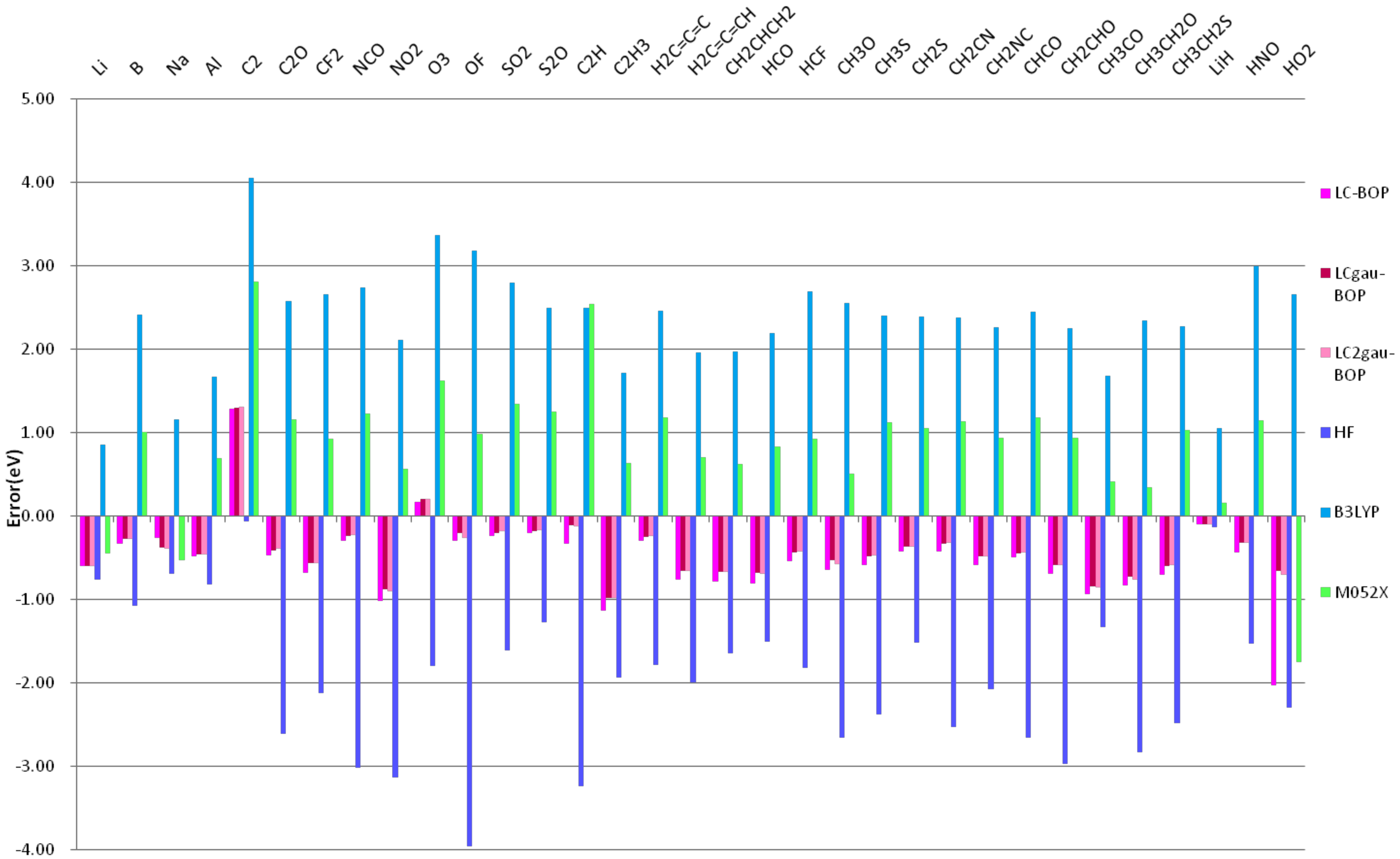
Electron Affinities - Deviation from experiment



EA Koopmans' theorem - Deviation from experiment



Electron Affinities - Deviation from experiment

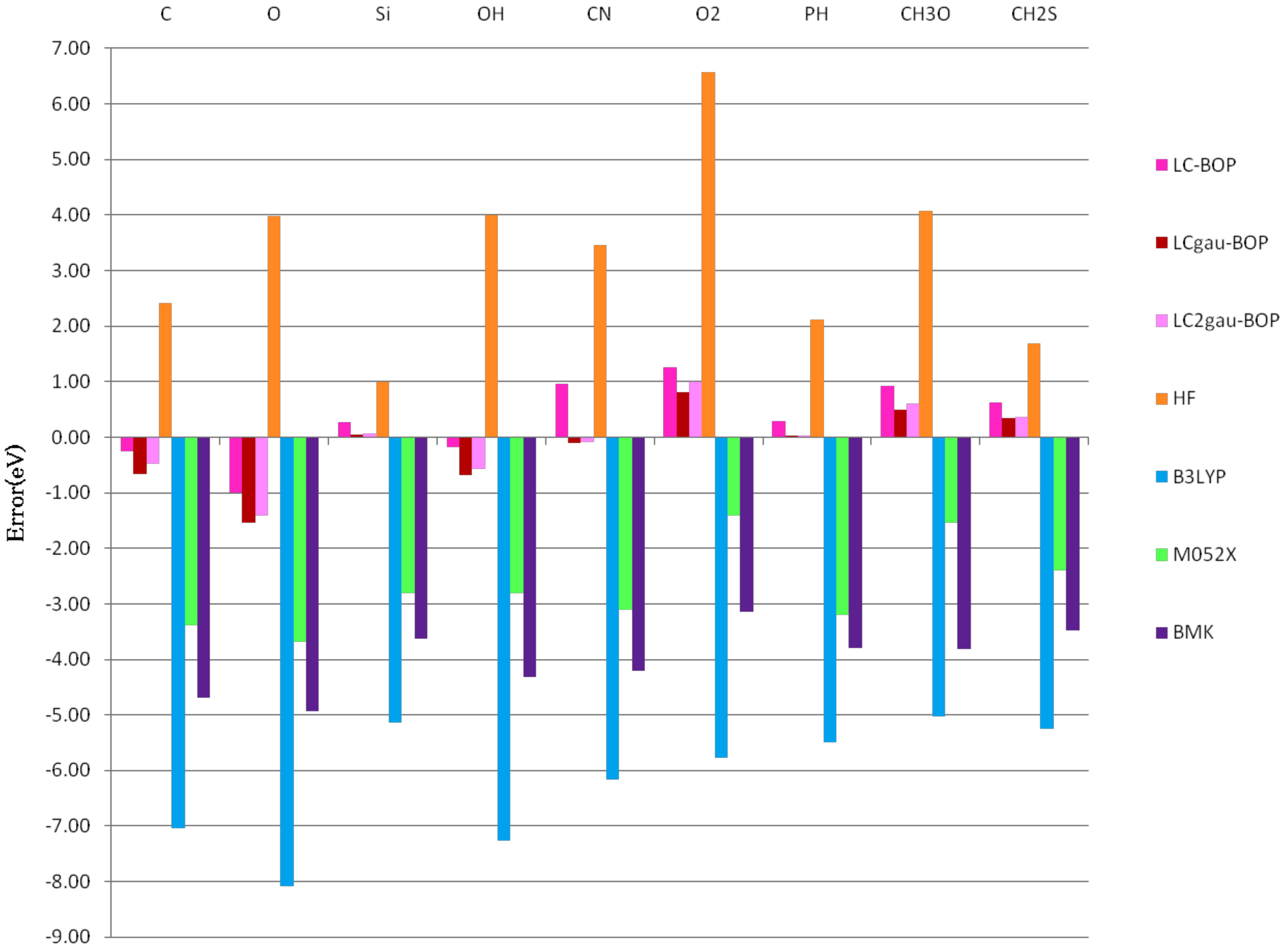


MAE for Ionization Potentials (eV)

Results from molecules in the G2-1 and G2-2 sets. Mean absolute errors are given for all calculations using 6-311G++(2d,2p) basis set and optimized geometries.

	G2-1	**G2-2**
HF	0.78	0.86
B3LYP	3.22	3.01
M052X	1.57	1.33
LC-BOP	0.45	0.55
LCgau-BOP	0.53	0.52
LC2gau-BOP	0.51	0.51

HOMO-LUMO Gap – Deviation from experiment



Janak’s theorem

The derivative of the total DFT energy, E, with respect to the occupation of a given orbital n_i is equal to the corresponding orbital energy

$$\frac{\partial E}{\partial n_i} = \varepsilon_i$$

This is true not only for the exact *xc* but for all commonly used approximate *xc* potentials.

Numerical illustration of Janak's theorem, $\frac{\partial E}{\partial n_i} = \varepsilon_i$

cc-pVQZ

Method	$C_2H_4^+$			C_2H_4			$C_2H_4^-$	
	ε_{LUMO}	dE/dN(+1)	dE/dN(0)	ε_{HOMO}	ε_{LUMO}	dE/dN(0)	dE/dN(-1)	ε_{HOMO}
LC-BOP	-10.22	-10.17	-10.71	-10.68	2.60	2.62	1.97	1.94
B3LYP	-13.48	-13.46	-7.59	-7.60	-0.22	-0.20	4.69	4.66
BOP	-14.40	-14.39	-6.43	-6.45	-0.78	-0.72	5.74	5.68
HF	-7.71	-7.61	-10.11	-10.22	3.56	3.95	2.23	2.11
M05-2X	-11.92	-11.67	-9.17	-9.16	1.12	1.15	3.49	3.47
LCgau	-10.37	-10.33	-10.53	-10.56	2.54	2.56	1.97	2.10

Janak' theorem is satisfied not only for the exact *xc* but for all commonly used approximate *xc* potentials.

DFT-Koopmans' theorem

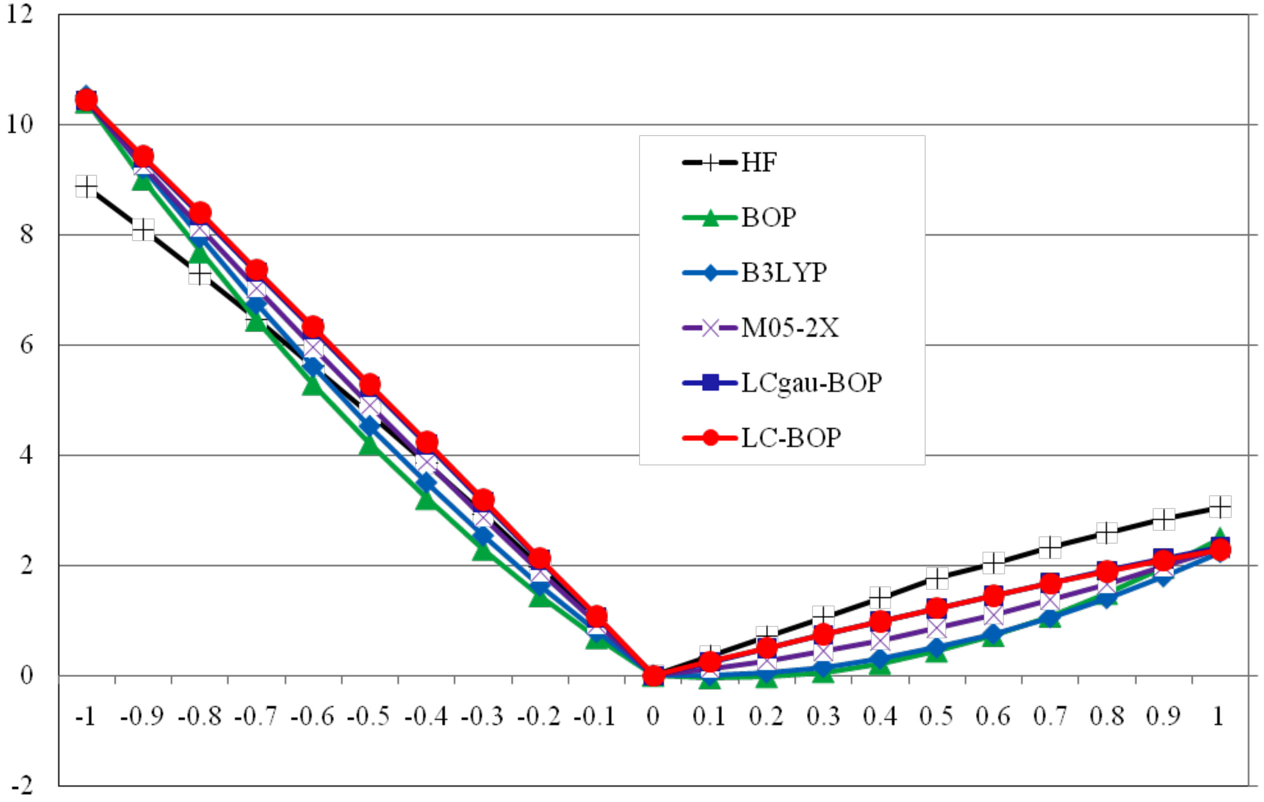
One can get IP by integrating ε_i over dn_i, between 1 and 0 (between 0 and 1 for EA).

$$IP = E(N-1,i) - E(N) = \int_1^0 \frac{\partial E}{\partial n_i} dn_i = \int_1^0 \varepsilon_i(n_i) dn_i$$

If the orbital energy does not depend on the occupation number

$$\frac{\partial \varepsilon_i}{\partial n_i} = 0 \quad \Longrightarrow \quad IP = -\varepsilon_i$$

$$\frac{\partial E}{\partial n_i} = \varepsilon_i$$



DFT-Koopmans' theorem

LC satisfies DFT-Koopmans’ theorem

$$\varepsilon_{HOMO}^{LC} \cong -IP$$

$$\varepsilon_{LUMO}^{LC} \cong -EA$$

The eigenvalues connected to the Kohn-Sham orbitals do have a strict physical meaning.

“What the electrons are doing in molecule”
can be understood in terms of Kohn-Sham orbitals.

Topics

1. **Hybrid GGA Functional Based on the Long-Range Correction**
   LC, LCgau, LC2gau functionals
2. **Systematic Errors in Computed Alkane Energies**
   DFT can be applied to large molecules?
3. **DFT Koopmans' Theorem**
   Kohn-Sham orbital energies and orbitals have physical meaning?
4. **Linear-Scaling Method For Computing Coulomb Potentials**
   Gaussian And Finite-Element Coulomb Method (GFC)

What is the Coulomb Problem?

Coulomb potential and Coulomb energy are given by

$$V(r) = \int_{-\infty}^{\infty} \frac{1}{r_{12}} \rho(r_2) dr_2 = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \int_{-\infty}^{\infty} e^{-t^2 (r_1 - r_2)^2} \rho(r_2) dr_2 dt$$

$$E_J = \int \rho(r) V(r) dr = \iint \frac{\rho(r_1)\rho(r_2)}{r_{12}} dr_1 dr_2$$

It is common to expand the charge density in a Gaussian basis set in order that the Coulomb potential may be evaluated by efficient analytical methods.

Linear-Scaling Gaussian and Finite-Element Coulomb (GFC) Method

T. Nakajima,   Y. Kurashige,   M. Watson

J.Chem.Phys., **129**, 184107 (2008)
J.Chem.Phys., **128**, 054105 (2008)
J.Chem.Phys., **126**, 144106 (2007)

Recently we have developed a new implementation of an approach to treating the Coulomb problem, which is rigorously linear scaling.

The Gaussian and Finite Element Coulomb (GFC) method

GFC evaluates the Coulomb potential by direct solution of the Poisson equation.

Coulomb potential in a molecular system is sharply spiked at the nuclei and has a long, diffuse tail far from the nuclei.

We use a mixture of finite-element and Gaussian functions to describe the Coulomb potential. This is the key to the efficiency of the GFC.

Gaussian and Finite-Element Expansion

Expand the potential in a mixed basis of Gaussians and finite-element functions,

$$V(r) = \sum_i c_i^{FE} \xi_i^{FE}(r) + \sum_i c_i^{Gauss} \xi_i^{Gauss}(r)$$

- Gaussians efficiently describes the core regions
- The finite-elements efficiently describe the residual smooth variations.

The expansion coefficients *c* are determined through Poisson Equation

$$\nabla^2 V(r) = -4\pi\rho(r)$$

Finite-element tensorial basis

Lagrange interpolating polynomials

$$L_k(x) = \prod_{j=0,\, j\leq k}^{n} \frac{x - jh\cdot n^{-1}}{(k-j)h\cdot n^{-1}}$$

$$(0 \leq k \leq n, \qquad 0 \leq x \leq h)$$

Finite-element basis functions

$$\xi_{klm}^{FE}(r) = \begin{cases} L_k(x - X_N)\cdot L_l(y - Y_N)\cdot L_m(z - Z_N) \\ 0 \ \ (i\text{-th grid-point}) \notin (N\text{-th finite-element}) \end{cases}$$

Determining *V*(r) via the Poisson equation

Solve the discretized Poisson equation algebraically

Galerkin Method

$$A \cdot c = b$$

where

$$A_{ij} = \int \nabla \xi_i(r) \nabla \xi_j(r) dr$$

$$b_i = 4\pi \int \xi_i(r) \rho(r) dr$$

A is large but extremely sparse.

Use the conjugate gradient method to solve the linear equations

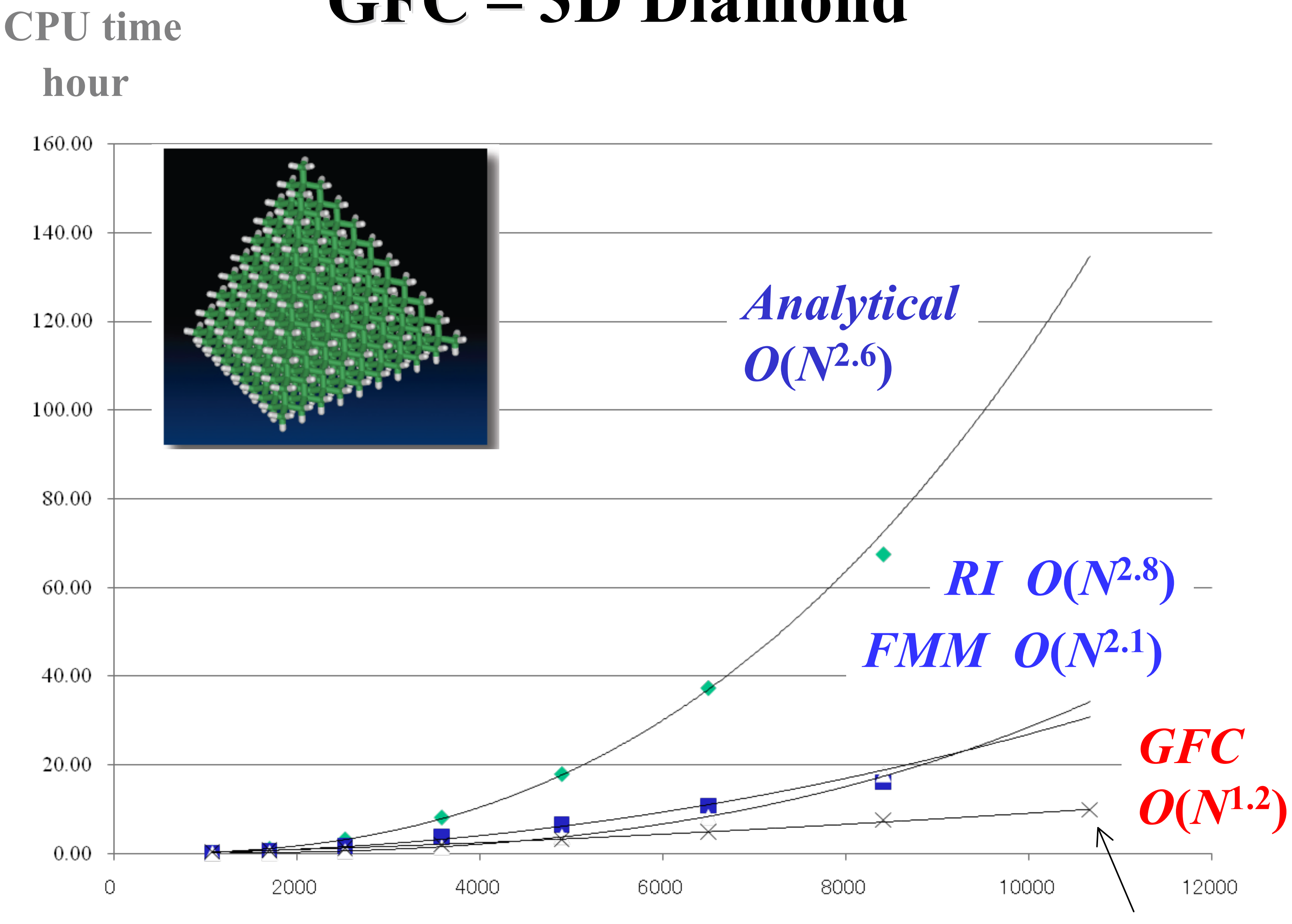
Construction of the Coulomb Matrix

- The Coulomb matrix

$$
\begin{aligned}
J_{pq} &= \int \chi_p(r)\chi_q(r)V(r)dr \\
&= \sum_i c_i^{Gauss} \int \chi_p(r)\chi_q(r)\xi_i^{Gauss}(r)dr + \\
&\quad \sum_i c_i^{FE} \int \chi_p(r)\chi_q(r)\xi_i^{FE}(r)dr
\end{aligned}
$$

- Only **overlap** and **kinetic-like integrals** are required
- But we must also consider the boundary condition…

GFC – 3D Diamond



962 atoms ($C_{650}H_{312}$)
10,660 basis functions

Energy Derivative

The Coulomb force can be evaluated efficiently by using the GFC method. The first derivative of the Coulomb energy E_J with respect to a nuclear coordinate X can be written as

$$\frac{\partial E_J}{\partial X} = \frac{\partial}{\partial X}\left\{\int dr_1 \int dr_2 \rho(r_1)\frac{1}{r_{12}}\rho(r_2)\right\} = 2\int dr_1 \left(\frac{\partial \rho(r_1)}{\partial X}\right) V(r_2)$$

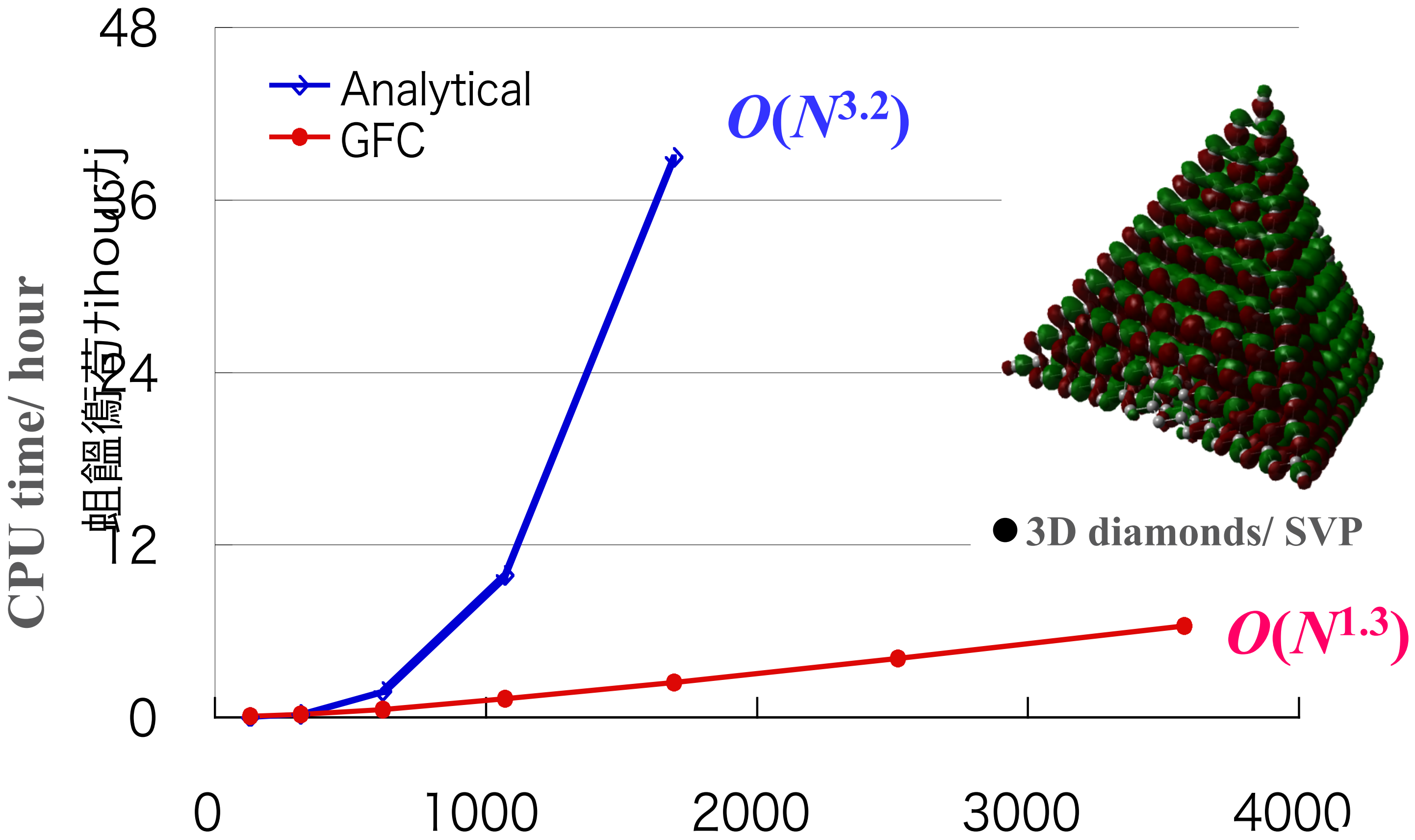
In the GFC method, the Coulomb potential $v(r)$ is expanded in the auxiliary functions. Thus the energy derivative can be evaluated approximately by

$$\frac{\partial E_J}{\partial X} = 2\sum_i c_i^{Gauss} \int dr \left(\frac{\partial \rho(r)}{\partial X}\right) \xi_i^{Gauss}(r) + 2\sum_i c_i^{FE} \int dr \left(\frac{\partial \rho(r)}{\partial X}\right) \xi_i^{FE}(r)$$

Computational cost for gradient evaluation is equivalent to that of one SCF iteration and scales with almost linear scaling.

GFC Energy Gradients

The GFC achieves linear scaling for gradient calculation



- The GFC is much faster than analytical gradient evaluation.
- The GFC scales with $O(N^{1.3})$, while the analytical method scales as $O(N^{3.2})$.

Conclusions

Gaussian and Finite-Element Coulomb (GFC) method offers the best performance for evaluating Coulomb integrals without loss of accuracy. The algorithm is found to scale linearly with system size.

GFC makes the electronic structure calculations affordable for truly large systems involving tens of thousands of basis functions.