Chemistry beyond the petascale

Why is it necessary?
How do we get there?

Robert J. Harrison

harrisonrj@ornl.gov

robert.harrison@utk.edu

The DOE funding

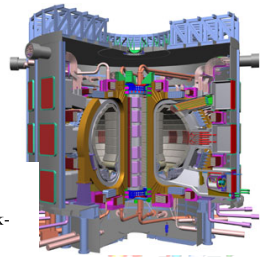
- This work is funded by the U.S. Department of Energy, the division of Basic Energy Science, Office of Science, under contract DE-AC05-00OR22725 with Oak Ridge National Laboratory. This research was performed in part using
  - resources of the National Energy Scientific Computing Center which is supported by the Office of Energy Research of the U.S. Department of Energy under contract DE-AC03-76SF0098,
  - and the Center for Computational Sciences at Oak Ridge National Laboratory under contract DE-AC05-00OR22725 .

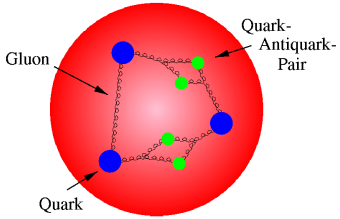
SciDAC

Scientific Discovery through Advanced Computing

Capability and Capacity Computing

- Unlike several disciplines, e.g.
  - Climate – one planet earth
  - Fusion – one ITER
  - HEP – one proton





- Chemistry is not defined by a few huge calculations on a few systems
  - There are numerous important molecules

Roles of Capability and Capacity Computing

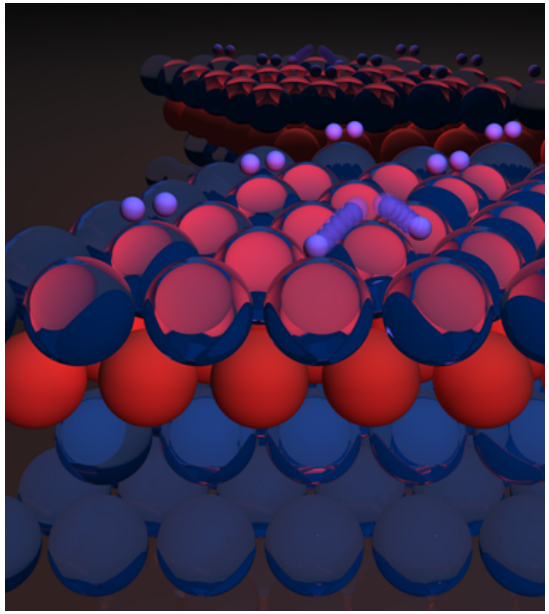
- Capability
  - Expand the frontier of feasible science
    - Size, complexity, accuracy, predictive ability
  - Calibrate widely applicable, cheaper methods
  - Today's capability is tomorrow's capacity
    - Must *constantly* advance leadership computing in part to reap benefits ASAP from future capacity
- Capacity
  - Satisfy the bulk of chemical needs (1:4 ?)
  - 2+ cost saving w.r.t. capability technology

Science Opportunities

- The chemical industry represents 10% of all U.S. manufacturing, employing more than one million Americans.
- These fields require petascale computing coordinated with experimental programs for significant innovation
- Clean energy innovations for automobiles and industry (catalysis, fuel cells, combustion).
  - Catalyst design is presently a misnomer … it is an Edisonian process
  - A partnership between experiment, theory, and advanced computing is required to enable rational design
  - Catalytic processes are directly involved in the synthesis of 20% of all industrial products.
- Molecular models for biological processes (protein and membrane function).
  - Modeling excited electronic states in photochemical systems and averaging over dynamics of the macromolecular structure potentially over very long timescales.
- Heavy element chemistry for advanced fuel cycles and environmental restoration.
  - These are special responsibilities of the DOE and petascale chemistry can replace many expensive experiments, and shorten timescales from decades to years.
- Dynamics of atoms and molecules interacting with electrons and powerful laser fields.
  - These are fundamental and defining challenges in physics and chemistry for the 21st century, and have been seeking for a solution for more than 50 years.
  - All major advances in this field are a result of new computational capabilities

Computational catalysis

- Currently allocated circa 16+M hours/year between NCCS, NERSC, EMSL
  - Need a sustained 100x increase
- Approach
  - Large systems accurately described with *modern* hybrid and meta DFT functionals
    - Chemistry codes have the advantage here
    - Full petascale simulation only for the largest runs
  - Higher-accuracy necessary for quantitative rates and for calibration on smaller systems
    - Many-body methods – demonstrated scaling to 10K processors and predicted to go to 100+K
    - 75% of peak speed on EMSL HP cluster 1700 cpu
- Outcomes
  - Rational design of novel catalyst(s)
  - Future savings of $B in various industries
  - Cleaner energy sources

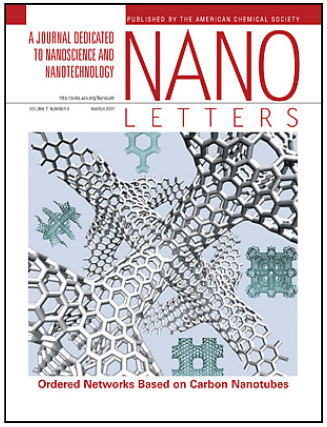


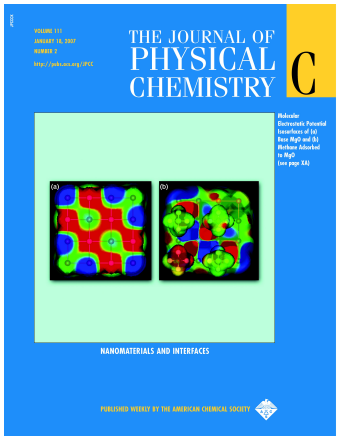
Mavrikakis, Wisconsin. H2 dissociation path on a bimetallic NSA surface.

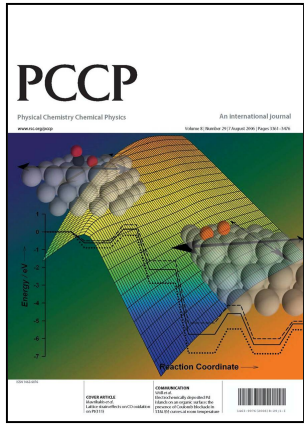
An Integrated Approach to the Rational Design of Chemical Catalysts

NCCS Incite Project

Robert J. Harrison: PI	**Oak Ridge National Laboratory, University of Tennessee**
Edoardo Aprà	**Oak Ridge National Laboratory**
Jerzy Bernholc	**North Carolina State University**
A.C. Buchanan III	**Oak Ridge National Laboratory**
Marco Buongiorno Nardelli	**North Carolina State University**
James M. Caruthers	**Purdue University**
W. Nicholas Delgass	**Purdue University**
David A. Dixon	**University of Alabama**
Sharon Hammes-Schiffer	**Pennsylvania State University**
Duane D. Johnson	**University of Illinois at Urbana Champaign**
Manos Mavrikakis	**University of Wisconsin at Madison**
Vincent Meunier	**Oak Ridge National Laboratory**
Mathew Neurock	**University of Virginia**
Steven H. Overbury	**Oak Ridge National Laboratory**
William F. Schneider	**University of Notre Dame**
William A. Shelton	**Oak Ridge National Laboratory**
David Sherrill	**Georgia Institute of Technology**
Bobby G. Sumpter	**Oak Ridge National Laboratory**
Kendall T. Thomson	**Purdue University**
Roberto Ansaloni	**Cray**
Carlo Cavazzoni	**CINECA, Italy**







INCITE request

- 2008 – 12M (got 10M)
- 2009 – 48M
- 2010 – 96M

Our need for leadership computing

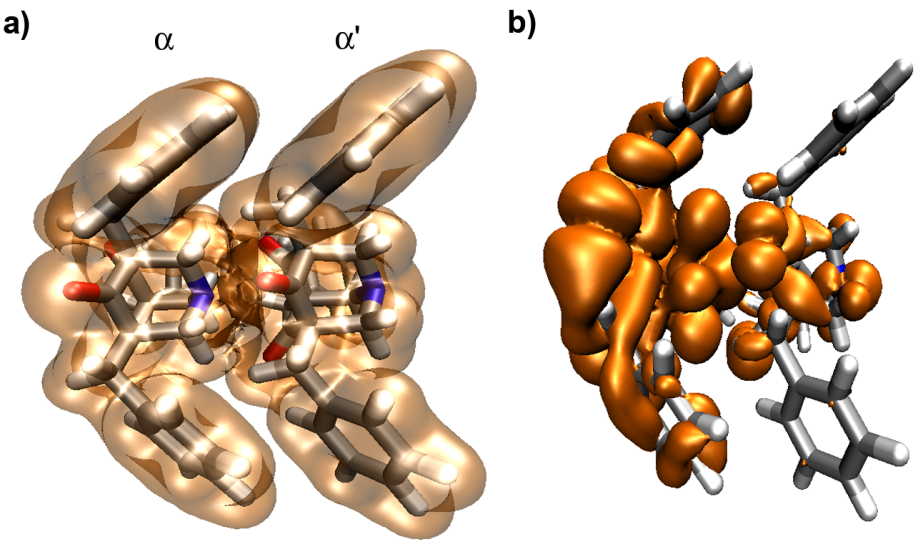
- Definitive, benchmark computations
  - The scope, scale and fidelity we expect from petascale computation will enable simulation to to answer truly hard questions about real systems. Fully quantitative computations are central to establishing fundamental understanding and enabling rational design.
- Integration of experiment and theory
  - Fast turnaround of reliable simulations is already enabling the intimate integration of theory and simulation into chemistry which is a predominantly experimental discipline.

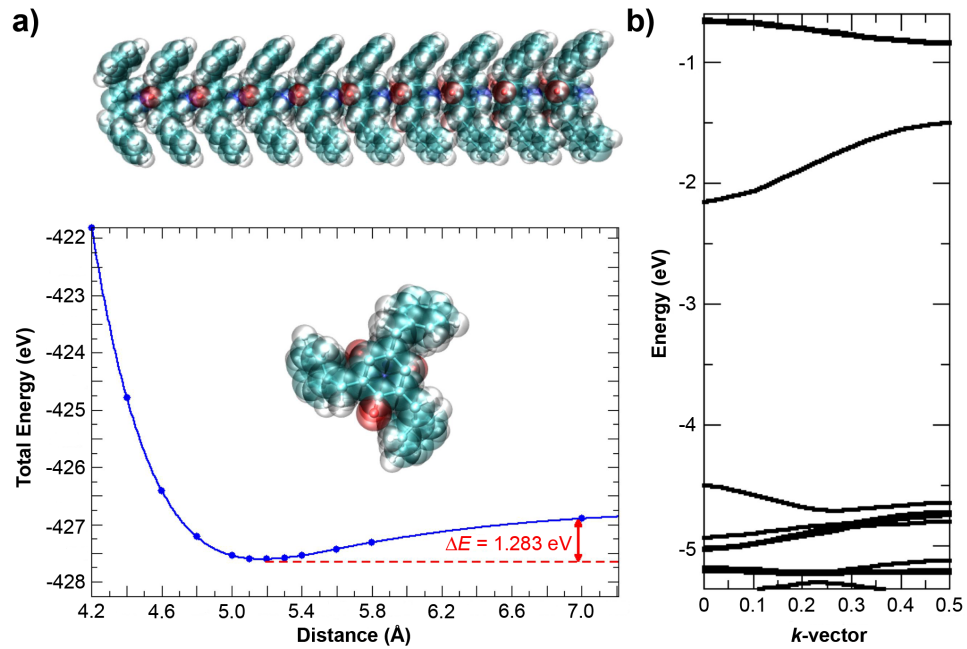
1.L. Grabow, Y. Xu and M. Mavrikakis, Lattice strain effects on CO oxidation on Pt(111), Physical Chemistry Chemical Physics 8, 3369 (2006).
2.Y. Xu, W. A. Shelton, and W. F. Schneider, Thermodynamic equilibrium compositions, structures, and reaction energies of PtxOy (x = 1 3) clusters predicted from first principles, J. Phys. Chem. B 110, 16591-16599, (2006).
3.Y. Xu, W. A. Shelton, and W. F. Schneider, Effect of Particle Size on the Oxidizability of Platinum Clusters, J. Phys. Chem. A, 110, 5839-5846, (2006).
4.Drummond ML, Sumpter BG, Shelton WA, Larese J., Density functional investigation of the adsorption of a methane monolayer on an MgO(100) surface , Phys. Rev. B 73, Art. No. 195313. (2006).
5.Drummond ML, Sumpter BG, Shelton WA, Larese J, Adsorption of Methane on the (100) Surface of MgO: Insight into Surface-Adsorbate and Adsorbate-Adsorbate Interations from First Principles Calculations , Mater. Res. Soc. Proc. 298, (2006).
6.E.E. Santiso, A.M. George, K.E. Gubbins and M. Buongiorno Nardelli, Effect of confinement by porous carnbons on the unimolecular decomposition of formaldehyde, J. Chem. Phys. 125, 084711 (2006).
7.M. Nunez and M. Buongiorno Nardelli, Tuning the Schottky Barrier Height in Metal-Alkaline Earth Oxide Interfaces, Physica Status Solidi, 243, 2081 (2006).
8.M. Nunez and M. Buongiorno Nardelli, First principles theory of metal-alkaline earth oxide interfaces, Phys. Rev. B, 73, 235422 (2006).
9.A. U. Nilekar, J. Greeley and M. Mavrikakis, Angewandte Chemie-International Edition, 45, 7046 (2006).
10.J. Zhou, S. Dag, S. D. Senanayake, B. C. Hathorn, S. V. Kalinin, V. Meunier, D. R. Mullins, S. H. Overbury, and A. P. Baddorf , Adsorption, desorption, and dissociation of benzene on TiO2 (110), and Pd/ TiO2 (110): Experimental characterization and first-principles calculations , Phys. Rev. B 74, 125318 (2006)
11.K. T. Park, M. H. Pan, V. Meunier, and E. W. Plummer, Surface Reconstructions of TiO2 (110) Driven by Suboxides, Phys. Rev. Lett. 226105 (2006)
12.J. Bernholc, W. Lu, S. M. Nakhmanson, P. H. Hahn , V. Meunier , M. Buongiorno Nardelli and W. G Schmidt, Atomic Scale Design of Nanostructures, Mol. Phys. 105, 147-156 (2007).
13.Drummond ML, Sumpter BG, Shelton WA, Larese J, An electronic structure and inelastic neutron scattering investigation of surface-adsorbate and adsorbate-adsorbate interactions in multilayers of CH4 on MgO(100) , J. Phys. Chem. B., 966-976 (2007).
14.R. B. Getman and W. F. Schneider, DFT-Based Characterization of the Multiple Adsorption Modes of Nitrogen Oxides on Pt(111), J. Phys. Chem. B, 2006, 111, 389-397 (2007).
15.J. Knudsen, A. U. Nilekar, R. T. Vang, J. Schnadt, E. L. Kunkes, J. A. Dumesic, M. Mavrikakis, F. Besenbacher, Journal of the American Chemical Society, 129, 6485-6490 (2007).
16.T. Mitsui, E. Fomin, D. F. Ogletree, M. Salmeron, A. U. Nilekar, M. Mavrikakis, Angewandte Chemie International Edition, 46, 5757-5761 (2007).
17.K. T. Park,1 M. Pan,2 V. Meunier,2,3 and E. W. Plummer, Reoxidation of TiO2 (110) via Ti interstitials and line defects, Phys. Rev. B 75, 245415 (2007)
18.A. A. Stekolnikov, K. Seino, and F. Bechstedt, S. Wippermann and W. G. Schmidt, A. Calzolari and M. Buongiorno Nardelli, Hexagon versus Trimer Formation in In Nanowires on Si(111): Energetics and Quantum Conductance, Phys. Rev. Lett. 98, 026105 (2007).
19.A. U. Nilekar, Y. Xu, J. L. Zhang, M. B. Vukmirovic, K. Sasaki, R. R. Adzic and M. Mavrikakis, Bimetallic and Ternary Alloys for Improved Oxygen Reduction Catalysis. Topics in Catalysis, (2007), in press.
20.R. R. Adzic, J. L. Zhang, K. Sasaki, M. B. Vukmirovic, M. Shao, J. X. Wang, A. U. Nilekar, M. Mavrikakis and F. Uribe, Platinum Monolayer Fuel Cell Electrocatalysts. Topics in Catalysis, (2007), in press.
21.A. Calzolari, A. Ferretti and M. Buongiorno Nardelli, Ab initio correlation effects on the electronic and transport properties of metal(II)-phthalocyanine based devices, Nanotechnology, in press (2007).
22.Ariana Beste, D. R. Mullins, S. H. Overbury, Robert J. Harrison, Adsorption and dissociation of methanol on the fully oxidized and partially reduced (111) cerium oxide surface: dependence on the configuration of the cerium 4f electrons, Surface Science (2007), accepted for publication.

Recent Science Progress

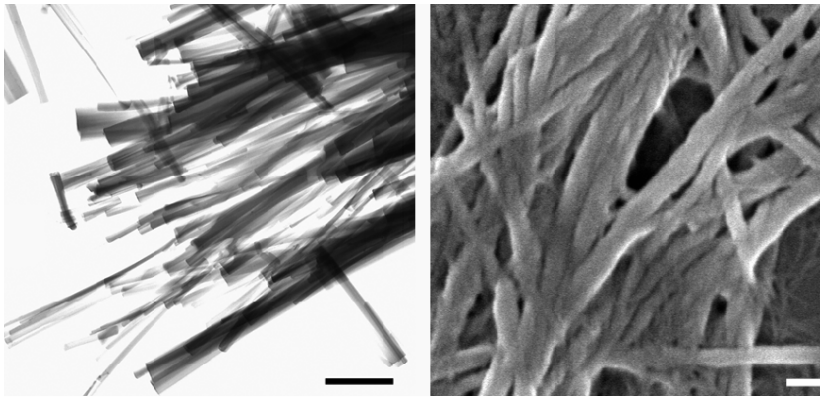
- Highlights from
  - Mavrikakis group at U. Wisconsin
  - Neurock's group at U. Virginia
  - ORNL Computational chemical sciences group

Nanoscale Self-Assembly of Donar-σ-Acceptor Molecules





parameter	1	$(1)_2$
		molecule α
HOMO (eV)	–5.80	–5.62
LUMO (eV)	–2.97	–3.20
E_{gap} (eV)	2.83	2.42
B.E. N^{1s} (eV)	373.55	373.4
B.E. N^{2s} (eV)	13.8	13.7
B.E. N^{2p} (eV)	6.6	6.4
partial charge (e)	7.47	7.47

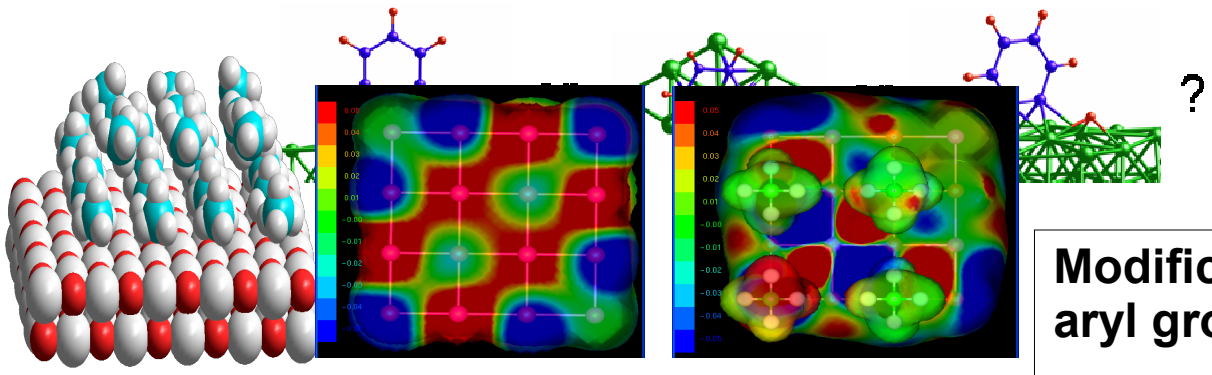


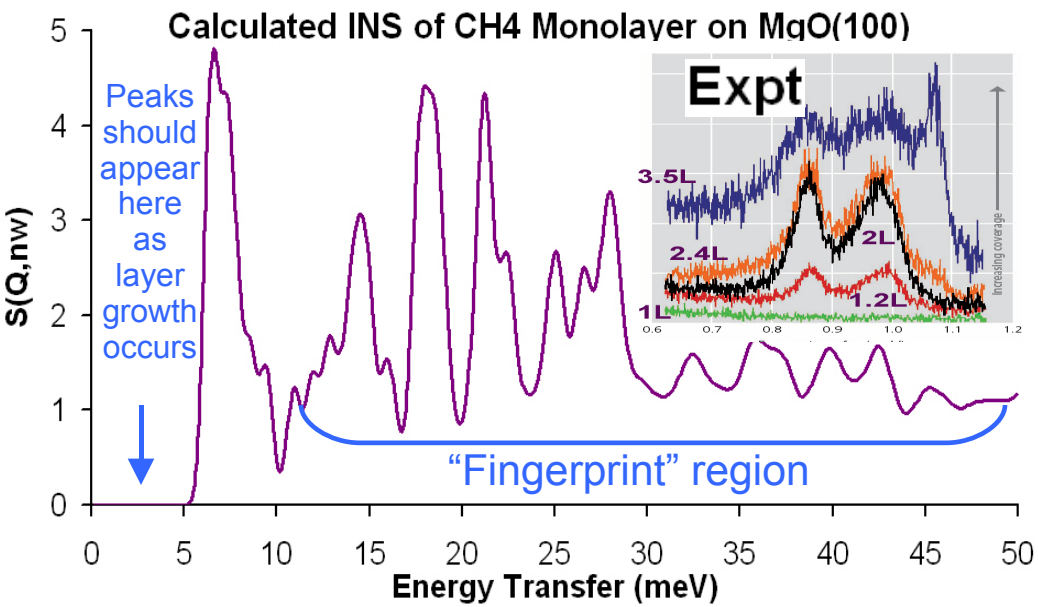
" Suitably functionalized 1-aza-admantanetrions self-assemble into well-defined solid-state architectures via strong dipolar interactions combined with weak dispersion forces.

" Peripheral substituents are found to interact directly with the core and tune the band gap into the semiconducting regime → tunable electronic structure

Castellano, Meunier, Sumpter, Valeev

Self-assembly on substrates





Modification of solid surfaces with aryl groups

Decreasing bond strength for transition metal elements from left to right.

Early transition metals favor the carbon-metal π-bond and therefore the flat-lying configuration, while late transition metals rich in d-electrons prefer the carbon-metal σ-bond and thus the upright fashion.

C6H5 is also found to undergo β-dehydrogentation on early transition metals.

First principles characterization of surface-adsorbate interactions and direct correlation to INS spectra

Dai, Jiang, Sumpter

Drummond, Larese, Shelton, Sumpter

Artificial DNA

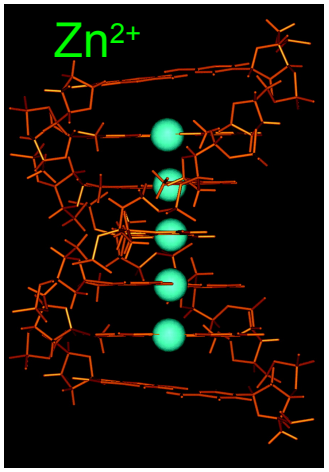
Uses recent advances in synthetic biology for:

- Fundamental investigations to enable the development of sensor techniques for detecting Single Nucleotide Polymorphs

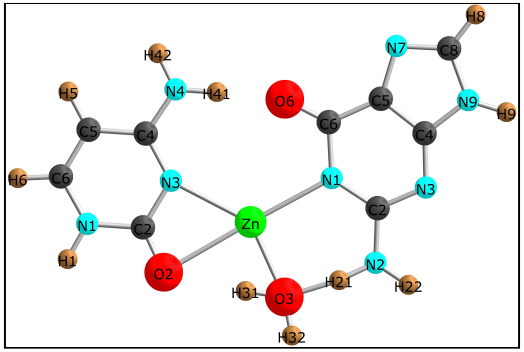
- Bionanowires and DNA by sequencing

- Probe the DNA replication process with unnatural DNA bases

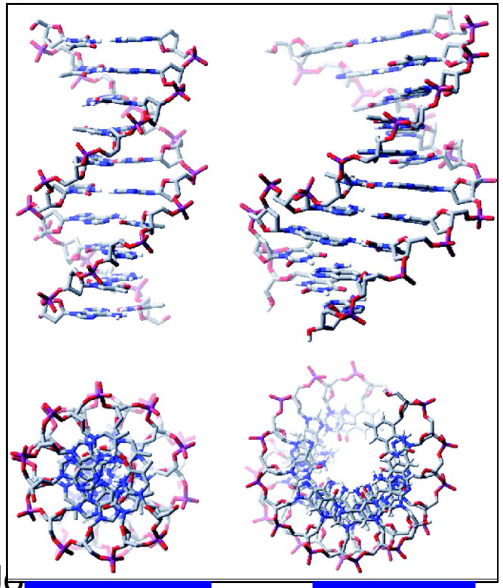
- Force-field generation for the simulation of novel synthetic biological systems



M-DNA



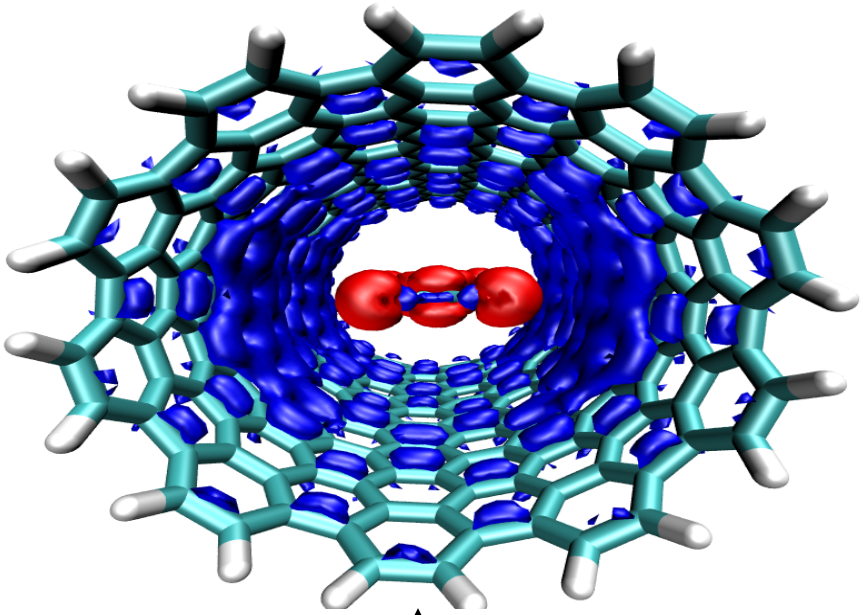
Metalated (M) and expanded (x) DNAs

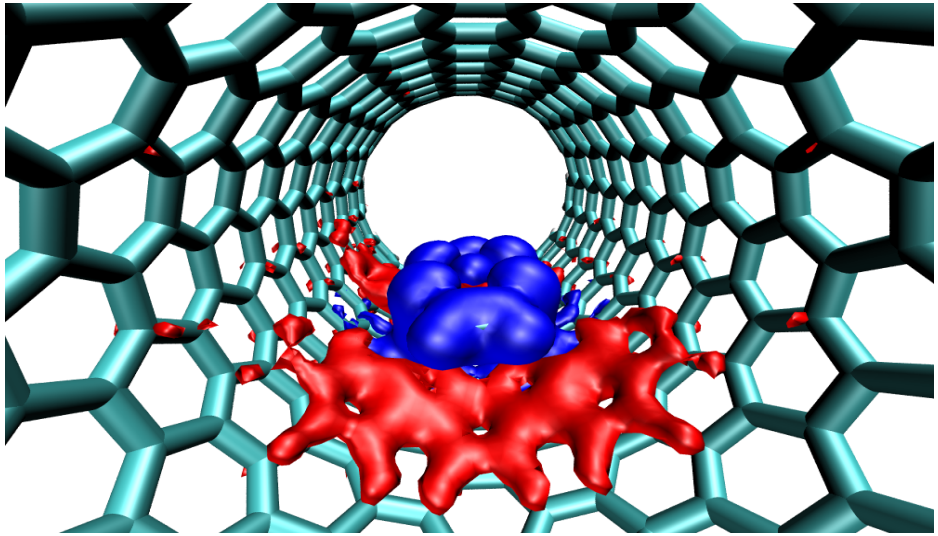


B-DNA xDNA

Fuentes, Šponer, Sumpter, Wells

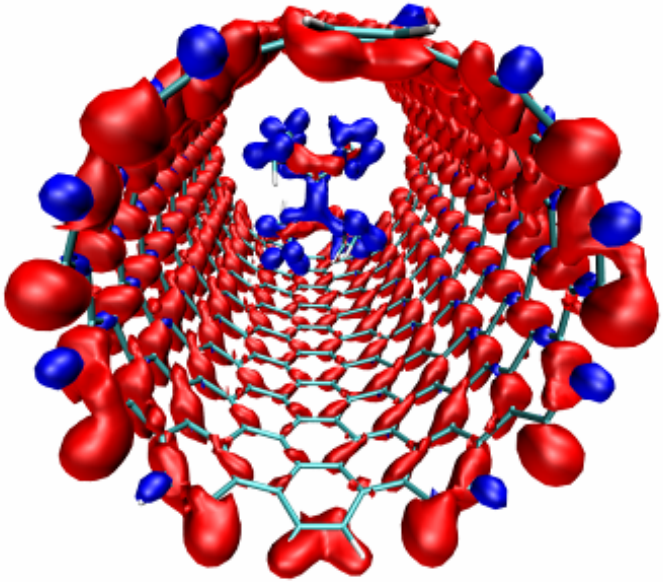
Amphoteric Doping of Carbon Nanotubes by Encapsulation of Organic Molecules: Electronic Transport and Quantum Conductance





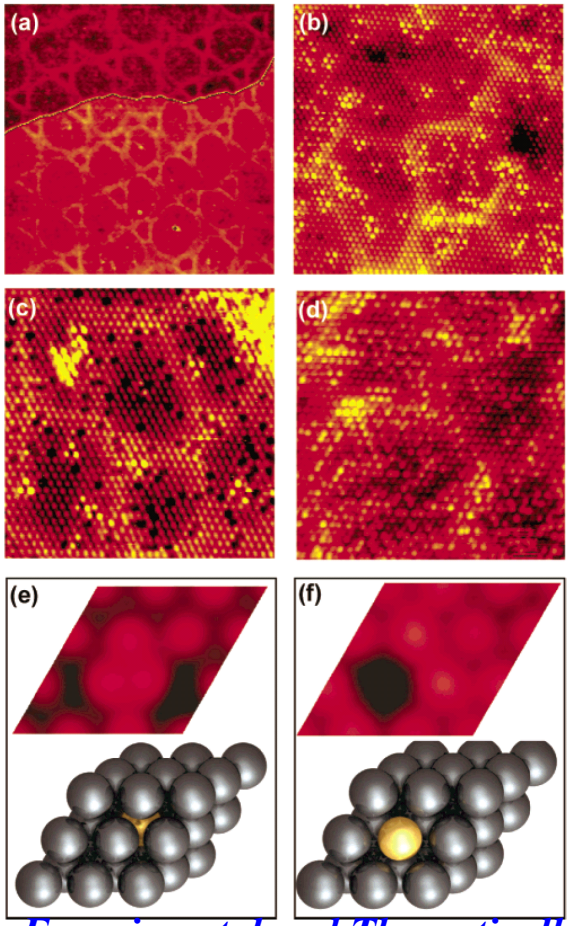
p-doped tube : holes are transferred from F_4-TCNQ to the nanotube

n-doped tube : holes are transferred from the tube to TTF and TDAE



Meunier, Sumpter

A Cu/Pt Near Surface Alloy for Low Temperature Water-Gas-Shift Reaction



Experimental and Theoretically simulated STM images.

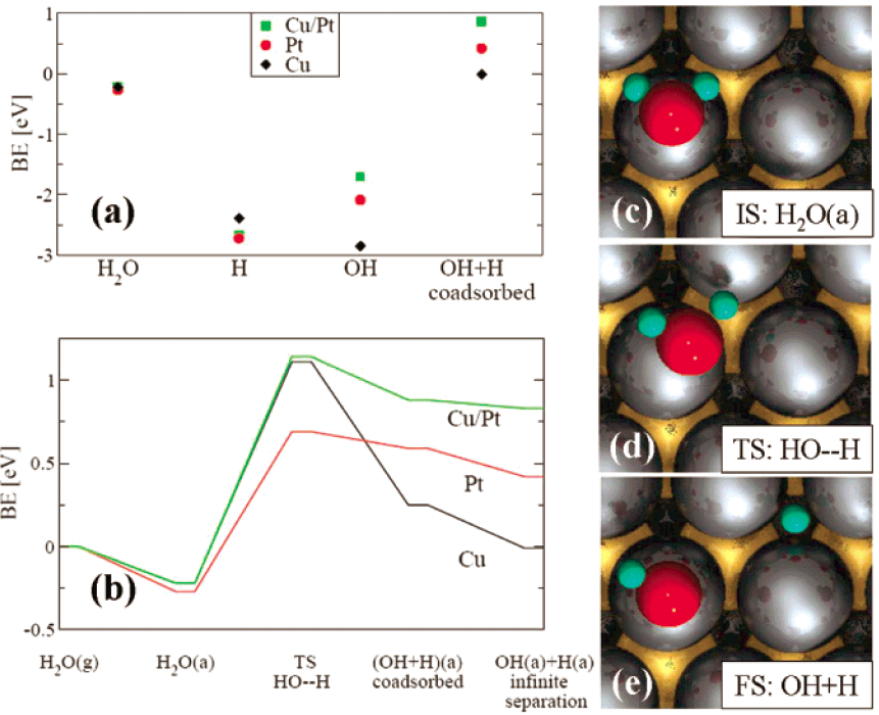
Water-Gas-Shift (WGS) reaction ($CO + H_2O \rightarrow CO_2 + H_2$) represents a crucial step in hydrogen production from fossil fuels. Because of the low efficiency and two-step nature of this process on current catalysts, there is a need for a better catalyst for small scale applications (e.g. on-board reformers in hydrogen fuel cell-powered vehicles).

We propose a Cu/Pt near surface alloy (NSA), which has Cu incorporated in the subsurface layer of Pt catalyst, as an improved WGS catalyst. Density Functional Theory (DFT) calculations and XPS results both show that Cu is most stable in the first subsurface layer of a Pt(111) surface.

A comparison between the experimental and theoretically simulated STM images reveals that post-annealing, the subsurface layer of the alloy is rich in Cu.

J. Knudsen, A. U. Nilekar, R. T. Vang, J. Schnadt, E. L. Kunkes, J. A. Dumesic, M. Mavrikakis, F. Besenbacher, Journal of the American Chemical Society, 129, 6485-6490 (2007).

A Cu/Pt Near Surface Alloy for Low Temperature Water-Gas-Shift Reaction



H_2O activation on a Cu/Pt NSA

Temperature Programmed Desorption (TPD) indicates that with increasing Cu coverage, CO-desorption T would decrease, a trend found with DFT, only if Cu existed in the subsurface layer. CO binds weaker to Cu/Pt NSA than to pure Pt surface. In addition, Cu/Pt binds formate (HCOO), a common spectator/poison of WGS catalysts, significantly weaker than Cu.

The Cu/Pt NSA activates H_2O – the Rate Limiting Step for WGS- with a barrier similar to Cu, and yet binds the H_2O dissociation products much weaker than Cu, which would facilitate subsequent elementary steps.

These properties make the Cu/Pt NSA a promising candidate for WGS catalysis, with increased activity and stability against poisoning by dominant reactive intermediates.

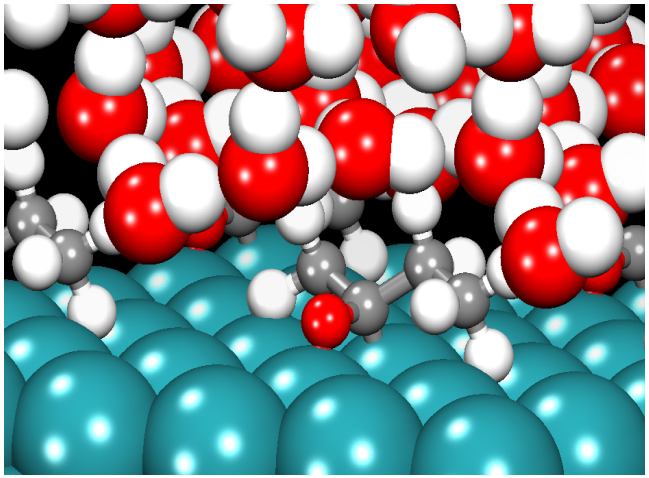
J. Knudsen, A. U. Nilekar, R. T. Vang, J. Schnadt, E. L. Kunkes, J. A. Dumesic, M. Mavrikakis, F. Besenbacher, Journal of the American Chemical Society, 129, 6485-6490 (2007).

Influence of Solvent and Metal on the Selective Hydrogenation of Oxygenates

Nishant Sinha and Matthew Neurock,
Department of Chemical Engineering, University of Virginia

Motivation:

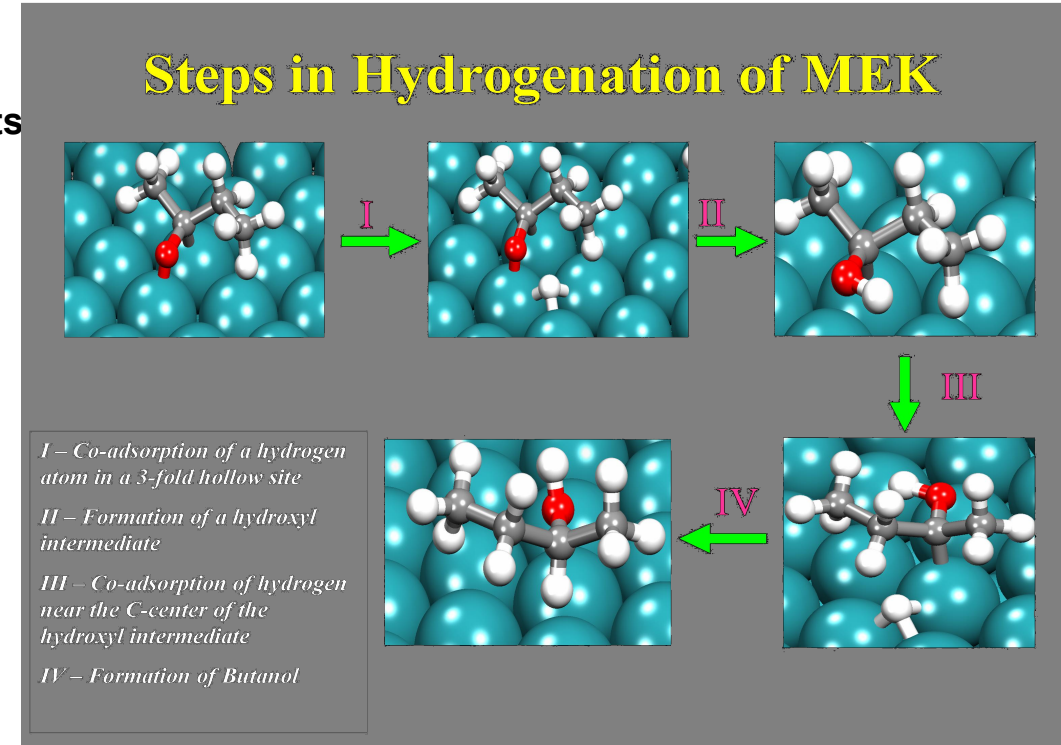
- Solvents enhance the activity and selectivity of industrially important hydrogenation reactions at metal-solution interface.
- kinetics and energetics of these reactions not understood by experiments.
- Ab initio DFT calculations used to elucidate reaction pathways and energetics.
- Examine in detail the periodic trends (of activation barrier etc.) to come up with a suitable catalyst and the influence of solvents



Methyl ethyl ketone (MEK) adsorbed on Ru(0001) in water

Methods:

- Large scale ab initio MD simulations are used to obtain the structure of solvent (water and methanol) on transition metal surfaces.
- Geometry optimization and climbing NEB in VASP are used to obtain ground state structures of reactants (or products) and transition states.



Results

Periodic Trends

L to R in period	$R_1R_2COH^* + H^* \rightarrow R_1R_2CHOH^*$		$R_1R_2CHOH^* + H^* \rightarrow R_1R_2CH2OH^*$	
	ΔE_{rxn}	E_a	ΔE_{rxn}	E_a
Co(0001)	0.11 eV	0.61 eV	0.00 eV	0.56 eV
Ni(111)	0.51 eV	1.31 eV	0.22 eV	0.77 eV
Cu(111)	0.20 eV	0.93 eV	-1.30 eV	0.17 eV
Ru(0001)	0.04 eV	0.21 eV	-0.27 eV	0.42 eV
Rh(111)	-0.50 eV	0.37 eV	-0.72 eV	0.48 eV
Pd(111)	-0.20 eV	0.47 eV	0.04 eV	0.63 eV
Ag(111)	0.02 eV	0.59 eV	-2.39 eV	0.06 eV
Os(0001)	-0.41 eV	0.52 eV	-0.60 eV	0.55 eV
Ir(111)	-0.79 eV	0.37 eV	-0.86 eV	0.37 eV
Pt(111)	**-1.12 eV**	**0.15 eV**	**0.81 eV**	**1.21 eV**
Au(111)	**-1.15 eV**	**0.22 eV**	**0.34 eV**	**0.86 eV**

	R1R2COH* + H* → R1R2CHOH*		R1R2CHOH* + H* → R1R2CH2OH*	
	ΔE_{rxn}	E_a	ΔE_{rxn}	E_a
Vapor	+0.15	0.38	-0.03	0.62
MeOH	-0.61	0.33	+0.05	0.60
H_2O	-0.57	0.21	+0.19	0.42

Energetics of MEK hydrogenation in methanol, water and vapor phase on Ru(0001)

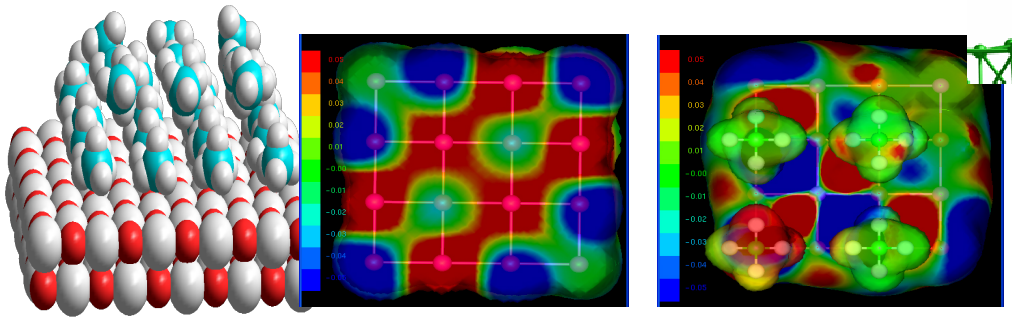
It is also observed that water provides an alternate solvent mediated mechanism which is 0.11 eV lower than metal-mediated mechanism.

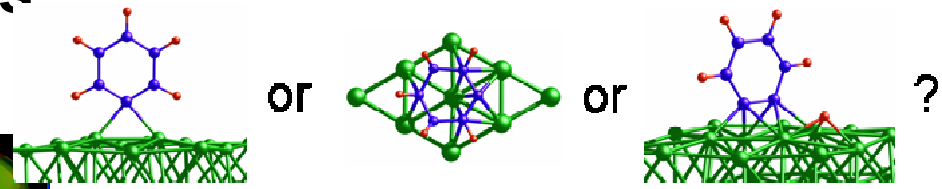
Conclusions:

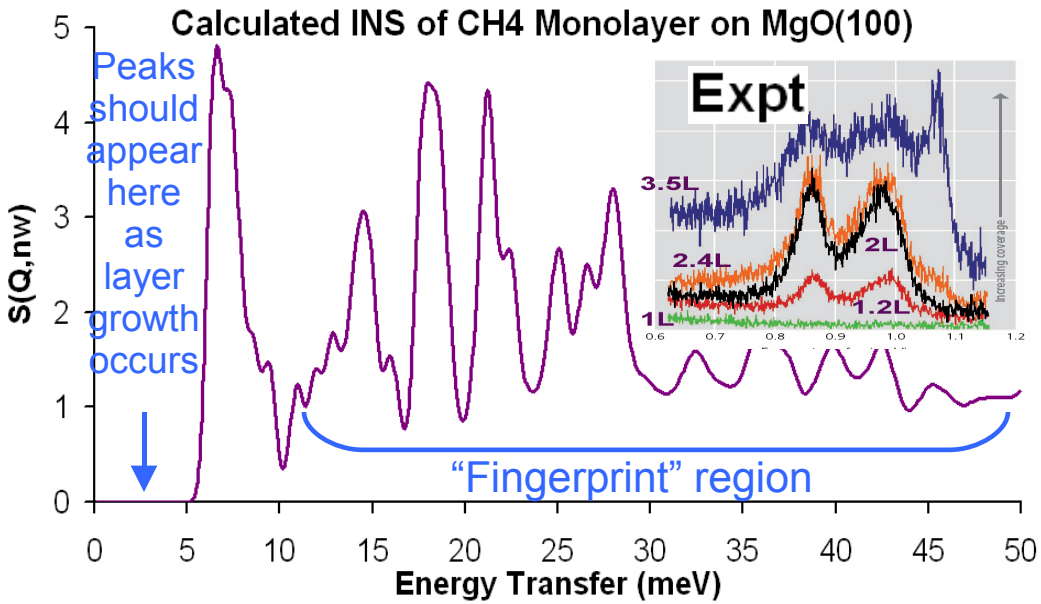
- **Systematic trends occur in the activation barriers occur over the periodic table that can be readily explained by simple bond order conservation principles. In the second hydrogenation step, we observe a maximum in activation barrier in Pd group metals.**
- **Protic solvents provides an alternate route for hydrogenation, with lower activation barrier.**
- **The more hydrophillic solvents provide ideal reaction environments.**

Self-assembly on substrates

Drummond, Sumpter, Shelton, Larese







Modification of solid surfaces with aryl groups

Decreasing bond strength for transition metal elements from left to right.

Early transition metals favor the carbon-metal π-bond and therefore the flat-lying configuration, while late transition metals rich in d-electrons prefer the carbon-metal σ-bond and thus the upright fashion.

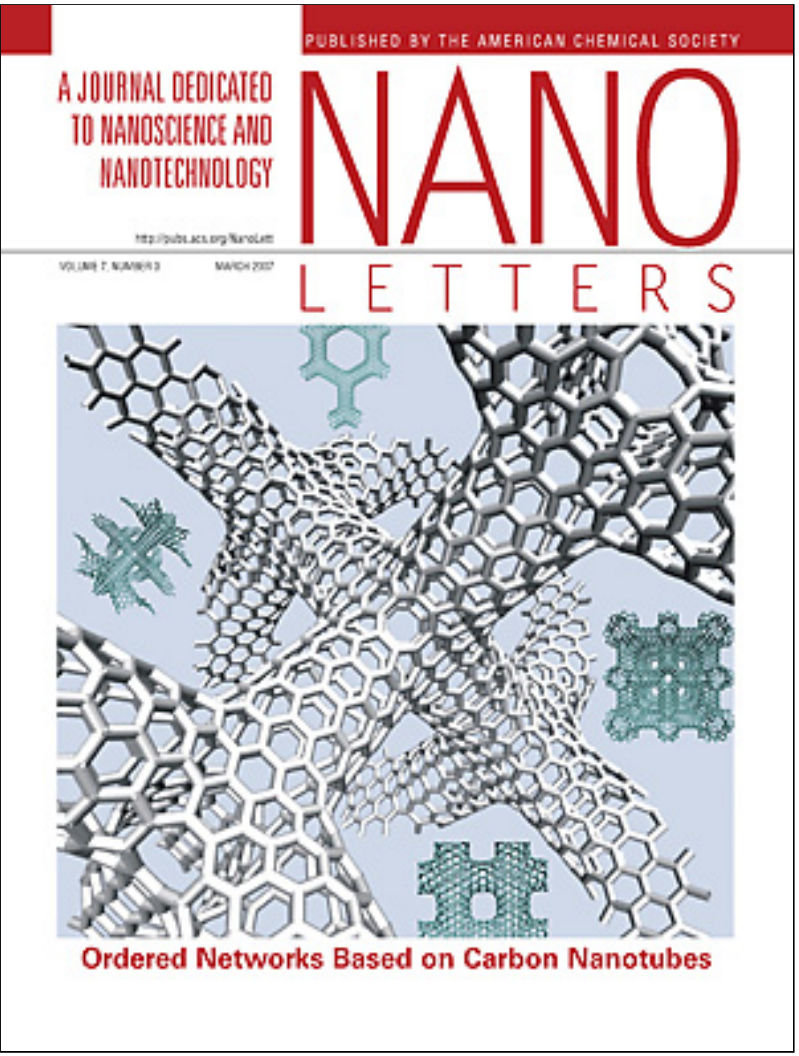
C6H5 is also found to undergo β-dehydrogentation on early transition metals.

First principles characterization of surface-adsorbate interactions and direct correlation to INS spectra

, J. Phys Chem C cover, 2007

Covalent 2D and 3D Networks from 1D Nanostructures: Designing New Material

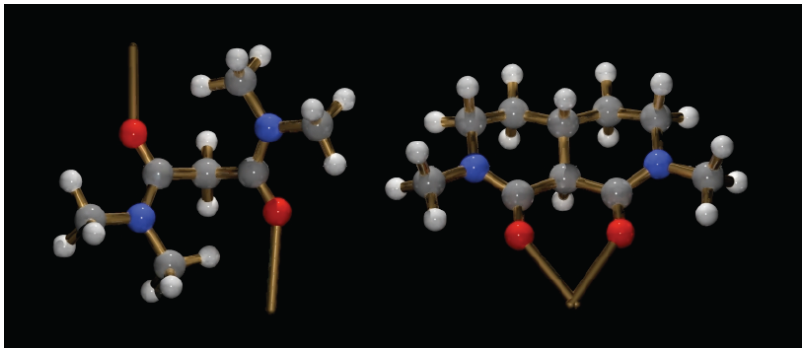
Romo-Herrera, Terrones, Dag, and Meunier



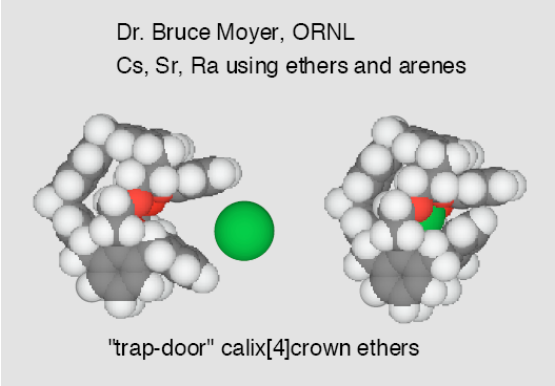
- characterization and property of carbon nanocomposites with different morphology or anomalous structures (defects, 2D, and 3D networks)
- Molecular models of various types of covalent carbon nanotube networks using nonhexagonal rings such as heptagons and octagons as linking units.
- The largest model (first plane) corresponds to a 3D nanotube network known as superdiamond (4 coordinated armchair nanotubes using octagons)
- the insets correspond to alternative 2D and 3D networks: supersquare, supercubic, and hexagonal; constructed from the Hierarchical Algorithm
- Electronic transport and the mechanical properties of these networks were also studied.

The role of simulation in heavy element chemistry for advanced fuel cycles

- Molecular-scale knowledge is vital to enable the rational design of new/enhanced agents
  - Reduced cost & risk with increased efficiency
  - Current experimental approach can generate only a fraction of required data over many years
    - The rest are guesstimated.
  - We can compute much of this
    - Need higher precision than currently feasible
  - Combinatorial methods use thermodynamics for screening, but this is not reliable enough
- Approach
  - Mixed MM/QM Gibbs-free energy computations of partition coefficients
  - Simulation of select liquid-liquid, gas-gas interfaces
  - Accurate thermo-chemistry and spectroscopy
    - Many-body methods incorporating relativistic effects
- Outcomes
  - Design of new separation chemistries on a timescale relevant to engineering requirements (months to years rather than decades)



B. Hay, EMSP project 73759



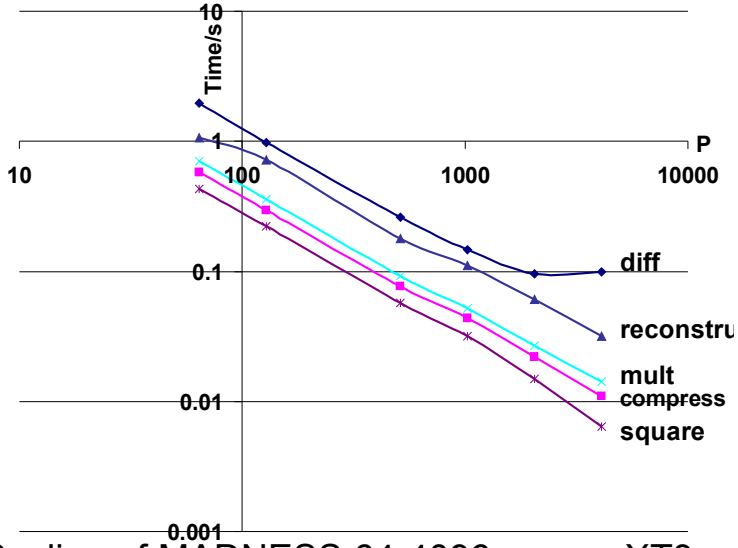
Computational Chemistry Endstation

International collaboration spanning 8 universities and 5 national labs

- Led out of UT/ORNL
- Focus
  - Actinides, Aerosols, Catalysis
- ORNL Cray XT3, ANL BG/L

Capabilties:

- Chemically accurate thermochemistry
  - Many-body methods required
- Mixed QM/QM/MM dynamics
  - Accurate free-energy integration
  - Simulation of extended interfaces
- Families of relativistic methods



Scaling of MADNESS 64-4096 cpu on XT3

NWChem: Largest CCSD(T) calculation

- Pollack, 2005.
- 1960 processor Itanium2 cluster
- 1468 basis functions (aug-cc-pVQZ)
- Perturbative triples (T)
  - 23 hours on 1400 processors
  - 75% of peak = 6.3 TFlops.

Issues

- Eliminate gulf between theoretical innovation in small groups and realization on high-end computers
- Eliminate the semantic gap so that efficient parallel code is no harder than doing the math
- Hardware/software architecture and programming models for 20xx
- Feasibility, support and maintenance

Major Molecular Electronic Structure Packages

- MOLPRO, MOLCAS, Turbomol, NWChem, GAMESS-US, GAMESS-UK, Gaussian, Jaguar, MPQC, ACES, QChem
- Wide range of functionality and algorithms
- Applicable to much of chemistry
- Vary greatly in speed, robustness, scalability, methods, …
- Several codes often used within any one project

- circa 20M lines of code …

Trends in Chemistry Codes – I

- *All* scalable codes use one-sided communication in various forms
  - Ease of programming; increased scalability
- QM/MM, multi-scale methods, direct dynamics, …
  - Multi-physics methods present scalability challenges
- Multi-level methods (FMM, multi-resolution, multi-grid, mixed-bases) in both space and time
  - A path to (near) linear or optimal scaling without sacrificing accuracy
  - Fully numerical real space methods
- Gaussian bases becoming attractive alternative to plane waves at low/medium precision? (Hütter)

Molecular Science Software Project

MS3
MOLECULAR SCIENCE SOFTWARE SUITE

ECCE
EXTENSIBLE COMPUTATIONAL CHEMISTRY ENVIRONMENT

NWCHEM
HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

GA TOOLS
PARALLEL COMPUTING LIBRARIES AND SOFTWARE TOOLS

<u>PNNL</u>
Yuri Alexeev,
Eric Bylaska,
Bert deJong,
Mahin Hackler,
Karol Kowalski,
Lisa Pollack,
Tjerk Straatsma,
Marat Valiev,
Theresa Windus

<u>ORNL</u>
Edo Apra,
Robert Harrison
Vincent Meunier
Ricky Kendall

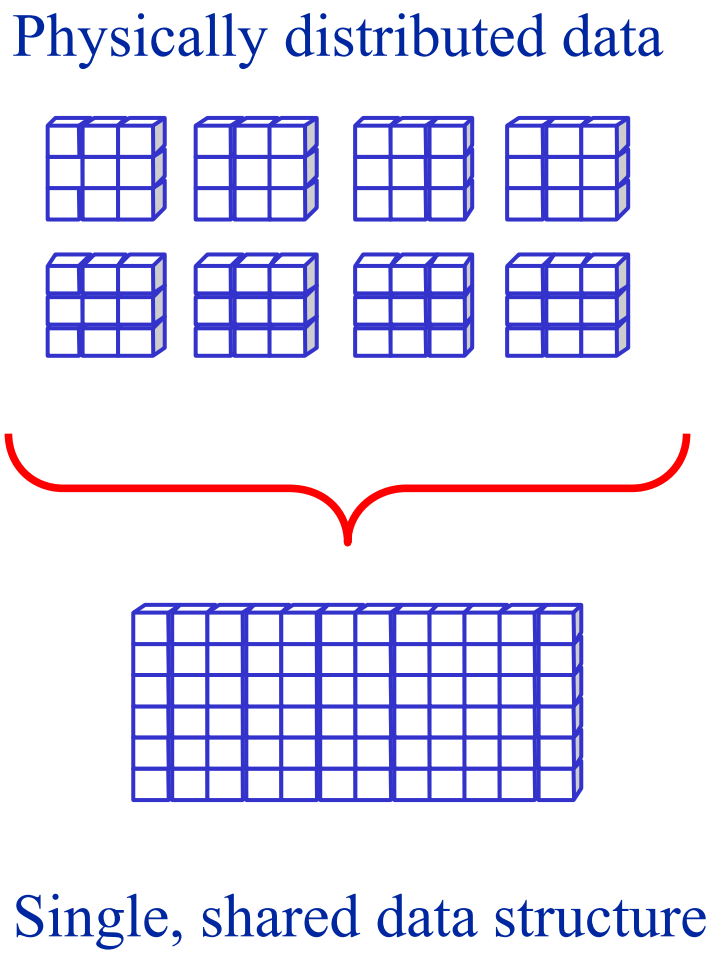
<u>Ames</u>
TL Windus

Gary Black,
Brett Didier,
Todd Elsenthagen,
Sue Havre,
Carina Lansing,
Bruce Palmer,
Karen Schuchardt,
Lisong Sun
Erich Vorpagel

Manoj Krishnan, Jarek Nieplocha,
Bruce Palmer, Vinod Tipparaju

http://www.emsl.pnl.gov/docs/nwchem/nwchem.html

Global Arrays

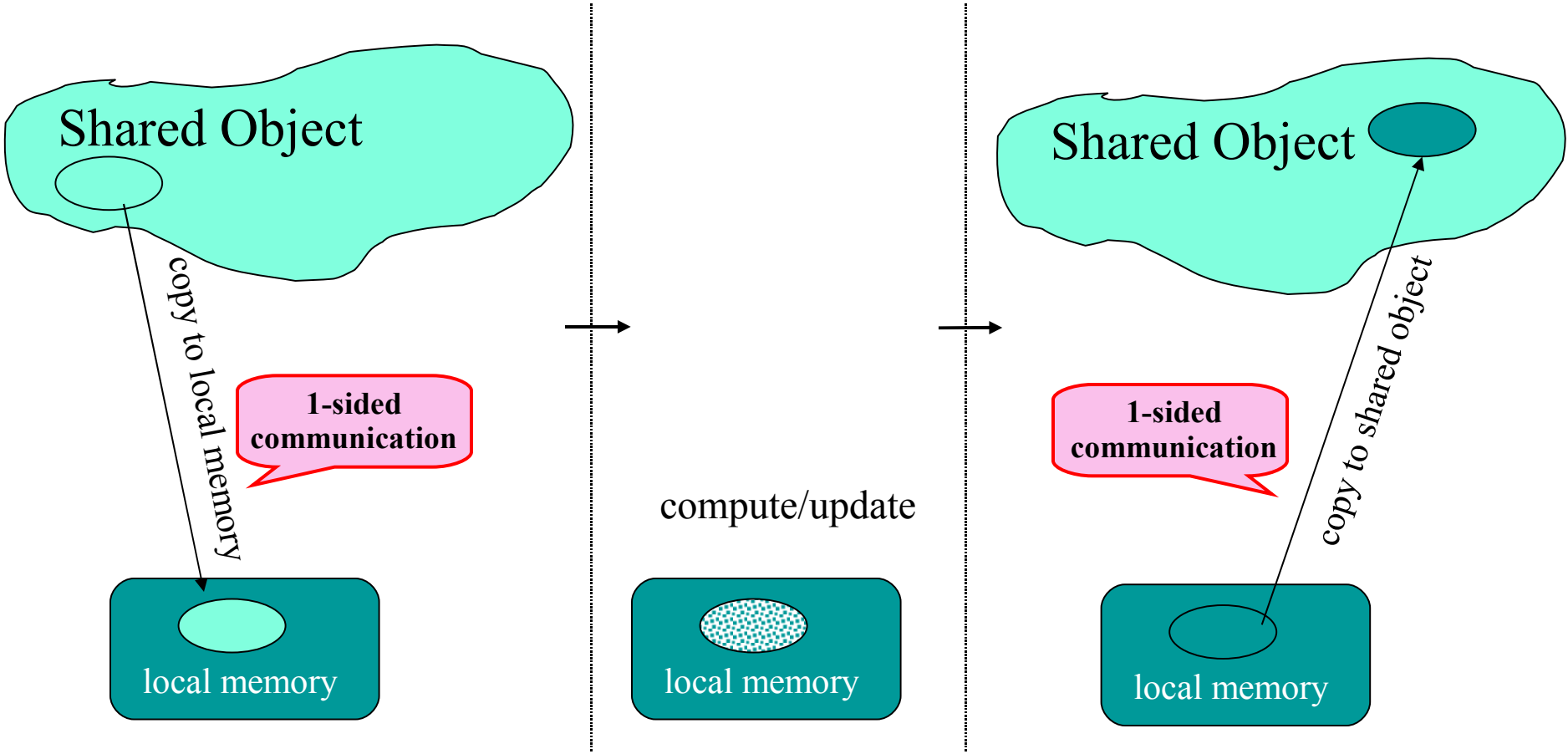


- Shared-memory-like model
  - Fast local access
  - NUMA aware and easy to use
  - MIMD and data-parallel modes
  - Inter-operates with MPI, …
- BLAS and linear algebra interface
- Ported to major parallel machines
  - IBM, Cray, SGI, clusters,...
- Originated in an HPCC project
- Used by most major chemistry codes, financial futures forecasting, astrophysics, computer graphics
- Supported by DOE

- Jarek Nieplocha, PNNL

http://www.emsl.pnl.gov/docs/global/

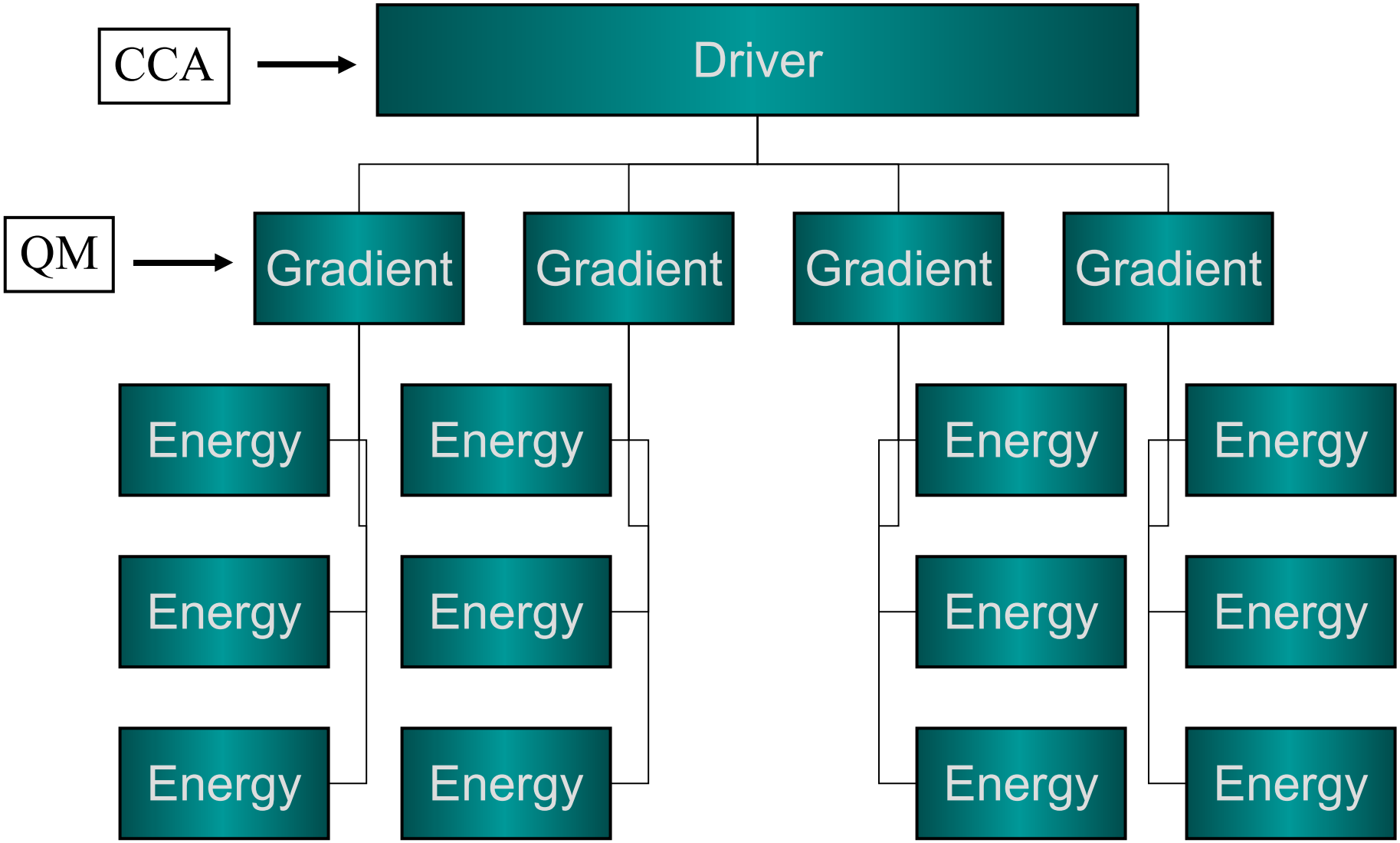
Non-uniform memory access model of computation



Trends in Chemistry Codes - II

- Multi-level parallelism
  - An effective path for most applications to scale to 100K processors without vast effort
    - Coarse grain over vibrational degrees of freedom in numerical hessian, or geometries in a surface scan or parameter study
    - Conventional distributed memory within each subtask
    - Fine grain parallelism within a few processor SMP (multi-threads, OpenMP, parallel BLAS, …)
  - Efficient exploitation of fine grain parallelism is a major concern on future architectures
- MADNESS side steps some of these issues
- Community effort to increase interoperability and leverage new capabilities between codes
  - CCA being adopted

Numerical Hessians



TL Windus, formerly PNNL

Synthesis of High Performance Algorithms for Electronic Structure Calculations

http://www.cis.ohio-state.edu/~gb/TCE

- Collaboration between DOE/SciDAC, NSF/ITR and ORNL/LDRD
- Objective: develop a high level programming tool that translates many-body quantum theory into efficient massively parallel codes. This is anticipated to revolutionize the rate of progress in this field by eliminating man-years of programming effort.
- NSF Project:
- Sadayappan (PI), Baumgartner, Cociorva, Pitzer (OSU)
  Bernholdt, Harrison (unfunded) (ORNL)
  Ramanujam (LSU)
  Nooijen (Waterloo)
- DOE SciDAC: Harrison (PI), Hirata (PNNL)
- DOE ORNL/LDRD: Bernholdt (PI, 2002-3)
- Other SciDAC projects adopting this tool: Piecuch, Gordon
- Also being applied to nuclear physics (Bernholdt and Dean)

CCSD Doubles Equation

```
hbar[a,b,i,j] == sum[f[b,c]*t[i,j,a,c],{c}] -sum[f[k,c]*t[k,b]*t[i,j,a,c],{k,c}] +sum[f[a,c]*t[i,j,c,b],{c}] -sum[f[k,c]*t[k,a]*t[i,j,c,b],{k,c}]
    -sum[f[k,j]*t[i,k,a,b],{k}] -sum[f[k,c]*t[j,c]*t[i,k,a,b],{k,c}] -sum[f[k,i]*t[j,k,b,a],{k}] -sum[f[k,c]*t[i,c]*t[j,k,b,a],{k,c}]
    +sum[t[i,c]*t[j,d]*v[a,b,c,d],{c,d}] +sum[t[i,j,c,d]*v[a,b,c,d],{c,d}] +sum[t[j,c]*v[a,b,i,c],{c}] -sum[t[k,b]*v[a,k,i,j],{k}]
    +sum[t[i,c]*v[b,a,j,c],{c}] -sum[t[k,a]*v[b,k,j,i],{k}] -sum[t[k,d]*t[i,j,c,b]*v[k,a,c,d],{k,c,d}] -sum[t[i,c]*t[j,k,b,d]*v[k,a,c,d],
    {k,c,d}] -sum[t[j,c]*t[k,b]*v[k,a,c,i],{k,c}] +2*sum[t[j,k,b,c]*v[k,a,c,i],{k,c}] -sum[t[j,k,c,b]*v[k,a,c,i],{k,c}]
    -sum[t[i,c]*t[j,d]*t[k,b]*v[k,a,d,c],{k,c,d}] +2*sum[t[k,d]*t[i,j,c,b]*v[k,a,d,c],{k,c,d}] -sum[t[k,b]*t[i,j,c,d]*v[k,a,d,c],{k,c,d}]
    -sum[t[j,d]*t[i,k,c,b]*v[k,a,d,c],{k,c,d}] +2*sum[t[i,c]*t[j,k,b,d]*v[k,a,d,c],{k,c,d}] -sum[t[i,c]*t[j,k,d,b]*v[k,a,d,c],{k,c,d}]
    -sum[t[j,k,b,c]*v[k,a,i,c],{k,c}] -sum[t[i,c]*t[k,b]*v[k,a,j,c],{k,c}] -sum[t[i,k,c,b]*v[k,a,j,c],{k,c}]
    -sum[t[i,c]*t[j,d]*t[k,a]*v[k,b,c,d],{k,c,d}] -sum[t[k,d]*t[i,j,a,c]*v[k,b,c,d],{k,c,d}] -sum[t[k,a]*t[i,j,c,d]*v[k,b,c,d],{k,c,d}]
    +2*sum[t[j,d]*t[i,k,a,c]*v[k,b,c,d],{k,c,d}] -sum[t[j,d]*t[i,k,c,a]*v[k,b,c,d],{k,c,d}] -sum[t[i,c]*t[j,k,d,a]*v[k,b,c,d],{k,c,d}]
    -sum[t[i,c]*t[k,a]*v[k,b,c,j],{k,c}] +2*sum[t[i,k,a,c]*v[k,b,c,j],{k,c}] -sum[t[i,k,c,a]*v[k,b,c,j],{k,c}]
    +2*sum[t[k,d]*t[i,j,a,c]*v[k,b,d,c],{k,c,d}] -sum[t[j,d]*t[i,k,a,c]*v[k,b,d,c],{k,c,d}] -sum[t[j,c]*t[k,a]*v[k,b,i,c],{k,c}]
    -sum[t[j,k,c,a]*v[k,b,i,c],{k,c}] -sum[t[i,k,a,c]*v[k,b,j,c],{k,c}] +sum[t[i,c]*t[j,d]*t[k,a]*t[l,b]*v[k,l,c,d],{k,l,c,d}]
    -2*sum[t[k,b]*t[l,d]*t[i,j,a,c]*v[k,l,c,d],{k,l,c,d}] -2*sum[t[k,a]*t[l,d]*t[i,j,c,b]*v[k,l,c,d],{k,l,c,d}]
    +sum[t[k,a]*t[l,b]*t[i,j,c,d]*v[k,l,c,d],{k,l,c,d}] -2*sum[t[j,c]*t[l,d]*t[i,k,a,b]*v[k,l,c,d],{k,l,c,d}]
    -2*sum[t[j,d]*t[l,b]*t[i,k,a,c]*v[k,l,c,d],{k,l,c,d}] +sum[t[j,d]*t[l,b]*t[i,k,c,a]*v[k,l,c,d],{k,l,c,d}]
    -2*sum[t[i,c]*t[l,d]*t[j,k,b,a]*v[k,l,c,d],{k,l,c,d}] +sum[t[i,c]*t[l,a]*t[j,k,b,d]*v[k,l,c,d],{k,l,c,d}]
    +sum[t[i,c]*t[l,b]*t[j,k,d,a]*v[k,l,c,d],{k,l,c,d}] +sum[t[i,k,c,d]*t[j,l,b,a]*v[k,l,c,d],{k,l,c,d}]
    +4*sum[t[i,k,a,c]*t[j,l,b,d]*v[k,l,c,d],{k,l,c,d}] -2*sum[t[i,k,c,a]*t[j,l,b,d]*v[k,l,c,d],{k,l,c,d}]
    -2*sum[t[i,k,a,b]*t[j,l,c,d]*v[k,l,c,d],{k,l,c,d}] -2*sum[t[i,k,a,c]*t[j,l,d,b]*v[k,l,c,d],{k,l,c,d}] +sum[t[i,k,c,a]*t[j,l,d,b]*v[k,l,c,d],
    {k,l,c,d}] +sum[t[i,c]*t[j,d]*t[k,l,a,b]*v[k,l,c,d],{k,l,c,d}] +sum[t[i,j,c,d]*t[k,l,a,b]*v[k,l,c,d],{k,l,c,d}]
    -2*sum[t[i,j,c,b]*t[k,l,a,d]*v[k,l,c,d],{k,l,c,d}] -2*sum[t[i,j,a,c]*t[k,l,b,d]*v[k,l,c,d],{k,l,c,d}] +sum[t[j,c]*t[k,b]*t[l,a]*v[k,l,c,i],
    {k,l,c}] +sum[t[l,c]*t[j,k,b,a]*v[k,l,c,i],{k,l,c}] -2*sum[t[l,a]*t[j,k,b,c]*v[k,l,c,i],{k,l,c}] +sum[t[l,a]*t[j,k,c,b]*v[k,l,c,i],{k,l,c}]
    -2*sum[t[k,c]*t[j,l,b,a]*v[k,l,c,i],{k,l,c}] +sum[t[k,a]*t[j,l,b,c]*v[k,l,c,i],{k,l,c}] +sum[t[k,b]*t[j,l,c,a]*v[k,l,c,i],{k,l,c}]
    +sum[t[j,c]*t[l,k,a,b]*v[k,l,c,i],{k,l,c}] +sum[t[i,c]*t[k,a]*t[l,b]*v[k,l,c,j],{k,l,c}] +sum[t[l,c]*t[i,k,a,b]*v[k,l,c,j],{k,l,c}]
    -2*sum[t[l,b]*t[i,k,a,c]*v[k,l,c,j],{k,l,c}] +sum[t[l,b]*t[i,k,c,a]*v[k,l,c,j],{k,l,c}] +sum[t[i,c]*t[k,l,a,b]*v[k,l,c,j],{k,l,c}]
    +sum[t[j,c]*t[l,d]*t[i,k,a,b]*v[k,l,d,c],{k,l,c,d}] +sum[t[j,d]*t[l,b]*t[i,k,a,c]*v[k,l,d,c],{k,l,c,d}]
    +sum[t[j,d]*t[l,a]*t[i,k,c,b]*v[k,l,d,c],{k,l,c,d}] -2*sum[t[i,k,c,d]*t[j,l,b,a]*v[k,l,d,c],{k,l,c,d}]
    -2*sum[t[i,k,a,c]*t[j,l,b,d]*v[k,l,d,c],{k,l,c,d}] +sum[t[i,k,c,a]*t[j,l,b,d]*v[k,l,d,c],{k,l,c,d}] +sum[t[i,k,a,b]*t[j,l,c,d]*v[k,l,d,c],
    {k,l,c,d}] +sum[t[i,k,c,b]*t[j,l,d,a]*v[k,l,d,c],{k,l,c,d}] +sum[t[i,k,a,c]*t[j,l,d,b]*v[k,l,d,c],{k,l,c,d}] +sum[t[k,a]*t[l,b]*v[k,l,i,j],
    {k,l}] +sum[t[k,l,a,b]*v[k,l,i,j],{k,l}] +sum[t[k,b]*t[l,d]*t[i,j,a,c]*v[l,k,c,d],{k,l,c,d}] +sum[t[k,a]*t[l,d]*t[i,j,c,b]*v[l,k,c,d],
    {k,l,c,d}] +sum[t[i,c]*t[l,d]*t[j,k,b,a]*v[l,k,c,d],{k,l,c,d}] -2*sum[t[i,c]*t[l,a]*t[j,k,b,d]*v[l,k,c,d],{k,l,c,d}]
    +sum[t[i,c]*t[l,a]*t[j,k,d,b]*v[l,k,c,d],{k,l,c,d}] +sum[t[i,j,c,b]*t[k,l,a,d]*v[l,k,c,d],{k,l,c,d}] +sum[t[i,j,a,c]*t[k,l,b,d]*v[l,k,c,d],
    {k,l,c,d}] -2*sum[t[l,c]*t[i,k,a,b]*v[l,k,c,j],{k,l,c}] +sum[t[l,b]*t[i,k,a,c]*v[l,k,c,j],{k,l,c}] +sum[t[l,a]*t[i,k,c,b]*v[l,k,c,j],{k,l,c}]
    +v[a,b,i,j]
```

Tensor Contraction Engine (TCE)

- High-level domain-specific language for a class of problems in quantum chemistry/physics based on contraction of large multi-dimensional tensors
- Specialized optimizing compiler
  - Produces F77+GA code, linked to runtime libs

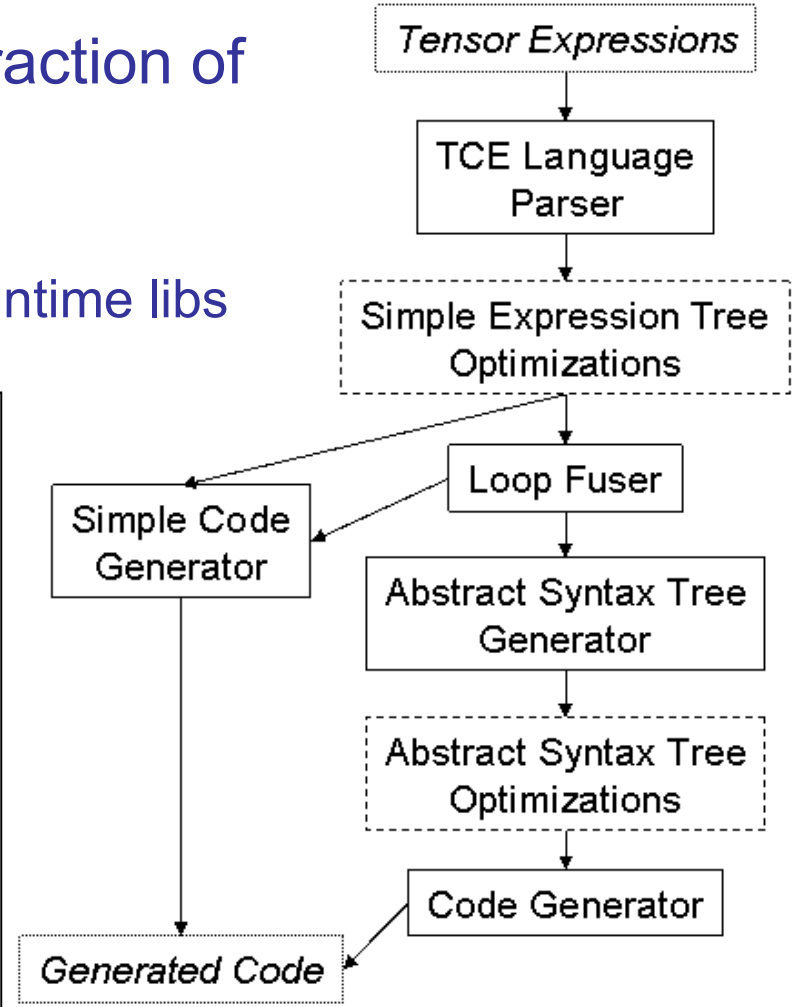
$$S_{abij} = \sum_{cefkl} A_{acik} B_{befl} C_{dfjk} D_{cdel}$$

```
range V = 3000;
range O = 100;

index a,b,c,d,e,f : V;
index i,j,k,l : O;

mlimit = 100GB;

procedure P(in A[V,V,O,O], in B[V,V,V,O],
            in C[V,V,O,O], in D[V,V,V,O],
            out S[V,V,O,O])=
begin
  S[a,b,i,j] == sum[ A[a,c,i,k] * B[b,e,f,l]
                   * C[d,f,j,k] * D[c,d,e,l],
                     {c,e,f,k,l}];
end
```



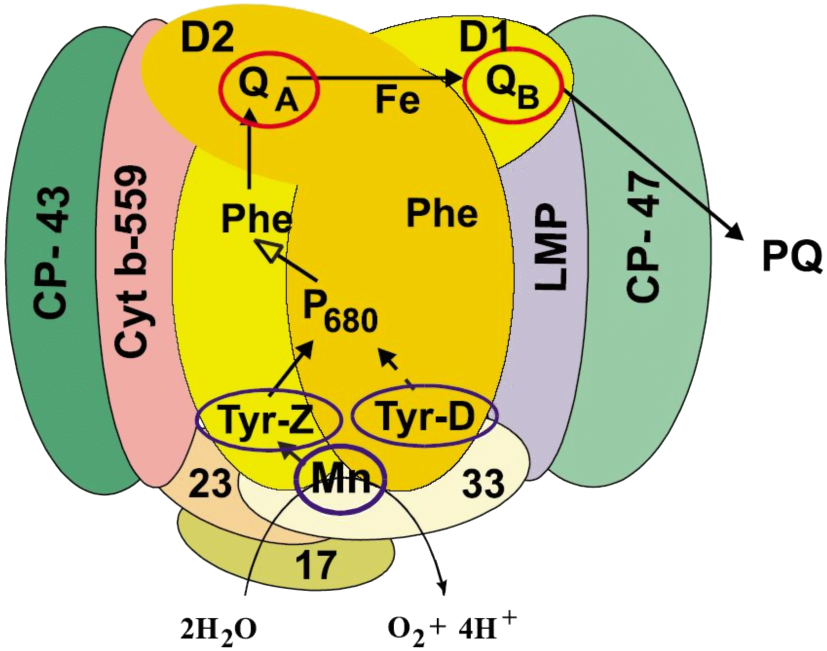
Chemical Computations on Future High-end Computers Award #CHE 0626354, NSF Cyber Chemistry

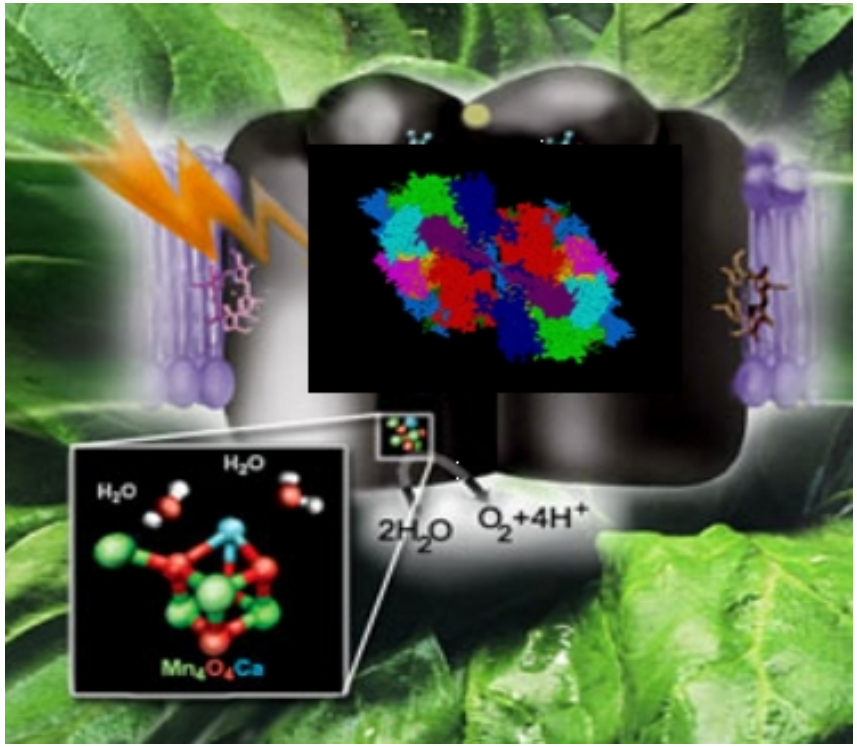
- NCSA (T. Dunning, PI)
- U. Tennessee (Chemistry, EE/CS)
- U. Illinois UC (Chemistry)
- U. Pennsylvania (Chemistry)
- Conventional single processor performance no longer increasing exponentially
  - Multi-core
- Many “corners” of chemistry will increasingly be best served by non-conventional technologies
  - GP-GPU, FPGA, MD Grape/Wine, Clearspeed

An Example Petascale Problem

- Important application area are in chemical and biological processes
  - A large number are often characterized by events that take place in a relatively local region of an otherwise large molecular aggregate.
    - A solute molecule in solution, the active region of an enzyme, adsorption on substrates, catalytic processes, and solute/solvent effects for reactions in condensed media.
- A particularly important example is the structure-function relationship for the water-splitting reactions in Photosystem II (PSII)
  - The fundamental chemical problem that has been solved by the PSII water splitting organisms is the ability to control the two-electron reduction of oxygen by creating stable intermediates while avoiding high-energy losses through relaxations.
  - Overall, four electrons have to be removed to allow the formation of molecular oxygen, and this has to be achieved while keeping activation energies low.
    - Small activation energies ensure sufficiently rapid reaction to avoid competitive losses and convert energy at the required rate.
    - It is important to remember that the key action of the water splitting site is as a catalyst, operating on reactants and products that do not differ much in free energy.
  - The four-electron electrochemistry, which oxygen-evolving organisms successfully achieve, remains poorly understood and yet these reactions are the foundation of the modern biosphere.
- Both biomass and hydrogen production are high on the agenda of the world s energy research community

Petascale Problem





" Lattice constant approximately 531.01 a.u.
" Approximately 17406 atoms (C,N,O, Mn)
" Need to get to 100,000 atoms to accurately model system properties
" Recent structural advances that have been reported on the PSII reaction center, which has crossed an important threshold by providing more details about the structure of the manganese/calcium complex and the arrangement of the amino acids surrounding it.

K.N. Ferreira, T.M. Inverson, K. Maghlaoui, J. Barber, and S. Iwata, Science 303, 1831 (2004)

Benchmark CCSD(T) Energies

- 1-PFLOP/s week = $6*10^{20}$ FLOPs
- 20 basis functions per electron
- Cost = $O^3V^4 = O^7 * 20^4$
- Hence, O=150 electrons = (CH2)*25
- But this is just one energy
  - Real chemistry needs
    - Multiple geometries including transition states
    - Harmonic frequencies with anharmonic corrections
    - 1000s of species