EMSL

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

U.S. DEPARTMENT OF ENERGY

NWChem approaching its 23rd birthday

Edoardo Aprà

Acknowledgements

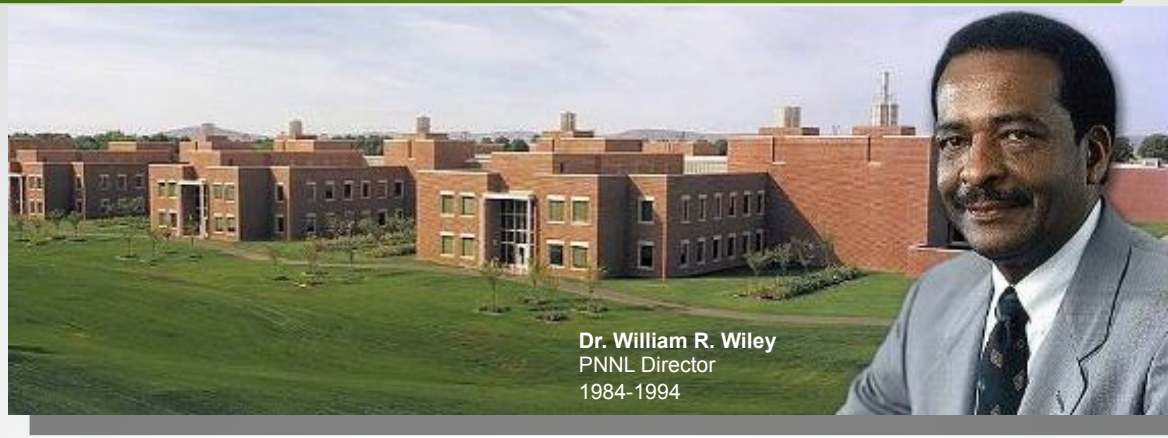
- Eric Bylaska, Sean Fisher, Niri Govind, Karol Kowalski and Marat Valiev
- PNNL HPC Group
- The rest of the NWChem developers

Acknowledgements ($$$ and Computer time)

- Dept. of Energy Office of Biological and Environmental Research
- Intel Parallel Computing Center
- Computing resources
  - EMSL Molecular Science Computing facility
  - PNNL Institutional Computing
  - NERSC

EMSL is a national scientific user facility

William R. Wiley's Vision:
An innovative multipurpose user facility providing *"synergism between the physical, mathematical, and life sciences."*



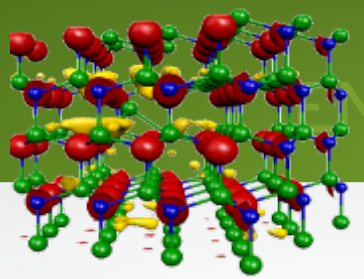
Mission

EMSL, a national scientific user facility at Pacific Northwest National Laboratory, provides ***integrated experimental and computational resources*** for ***discovery and technological innovation*** in the environmental molecular sciences to ***support the needs of DOE and the nation***.

- Funded by DOE **Office of Science's Office of Biological and Environmental Research (BER)**
- Located in Richland, WA - **Eastern Washington state**

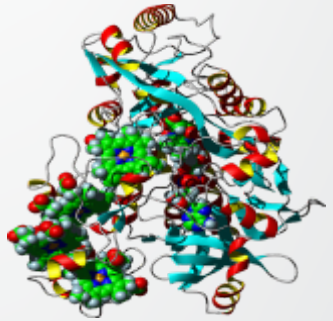
NWCHEM

HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE



- Developed at EMSL/PNNL since 1993
  - Large %age of funding: DOE-SC BER
- Version 3.0 distributed in 1997
- Provides major modeling and simulation capability for molecular science
  - Broad range of molecules, including catalysts, biomolecules, and heavy elements
  - Solid state capabilities
- Software development supported by funding for EMSL user facility
- Open Source → active user community
- A new version every year
- NWChem 6.6 released in October 2015

http://www.nwchem-sw.org



NWChem Developers/Collaborators Community

EMSL

NWCHEM

HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

J. Heyrovský Institute of Physical Chemistry
Academy of Sciences of the Czech Republic

BERKELEY LAB
Lawrence Berkeley National Laboratory

Argonne NATIONAL LABORATORY

NORTHWESTERN UNIVERSITY

Universitat de Girona

UBC THE UNIVERSITY OF BRITISH COLUMBIA

UCSD

PENNSTATE 1855

THE UNIVERSITY of WISCONSIN MADISON

UCLA

UCI University of California, Irvine

Massachusetts Institute of Technology

The Hebrew University of Jerusalem
האוניברסיטה העברית בירושלים

University at Buffalo
The State University of New York

UNIVERSITY OF CAMBRIDGE

UNT UNIVERSITY OF NORTH TEXAS
Discover the power of ideas.

IOWA STATE UNIVERSITY

OAK RIDGE NATIONAL LABORATORY
Managed by UT-Battelle for the Department of Energy

UNIVERSITY OF MINNESOTA
Driven to Discover℠

THE UNIVERSITY of TENNESSEE KNOXVILLE

TEXAS TECH UNIVERSITY

Duke UNIVERSITY

BROOKHAVEN NATIONAL LABORATORY

UNIVERSITY of WASHINGTON

UCL

Cal

vrije Universiteit amsterdam

PRINCETON UNIVERSITY

UNIVERSITÄT DES SAARLANDES

Science & Technology Facilities Council
Daresbury Laboratory

LSU

Weizmann Institute of Science

FOUNDED 1901 DOMINICAN UNIVERSITY

OHSU

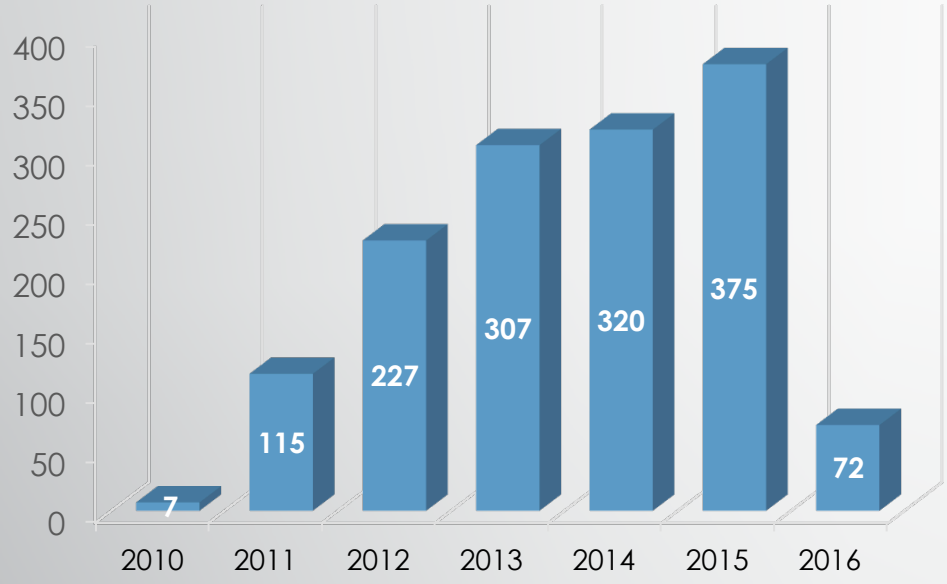
AKADEMIE VĚD ČESKÉ REPUBLIKY

Kitware

NWChem impact

- NWChem release papers are more than 3,000 times (Google)
- Last 6.6 release - October 2015

Number of citations (2010 release paper)



NWChem User Community: some highlights

EMSL

PNAS

Defusing redox bombs?

Nicholas F. Polizzi[a], Agostino Migliore[b], Michael J. Therien[b], and David N. Beratan[a,b,c,1]

[a]Department of Biochemistry, Duke University, Durham, NC 27708; [b]Department of Chemistry, Duke University, Durham, NC 27708; and [c]Department of Physics, Duke University, Durham, NC 27708

nature chemistry

ARTICLES

PUBLISHED ONLINE: 13 JULY 2014 | DOI: 10.1038/NCHEM.1999

Observation of an all-boron fullerene

Hua-Jin Zhai[1], Ya-Fan Zhao[2], Wei-Li Li[3], Qiang Chen[1], Hui Bai[1], Han-Shi Hu[2], Zachary A. Piazza[3], Wen-Juan Tian[1], Hai-Gang Lu[1], Yan-Bo Wu[1], Yue-Wen Mu[1], Guang-Feng Wei[4], Zhi-Pan Liu[4], Jun Li[2]*, Si-Dian Li[1]* and Lai-Sheng Wang[3]*

J|A|C|S

JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

Photoexcited Surface Frustrated Lewis Pairs for Heterogeneous Photocatalytic CO_2 Reduction

Kulbir Kaur Ghuman,[†] Laura B. Hoch,[‡] Paul Szymanski,[§] Joel Y. Y. Loh,[⊥] Nazir P. Kherani,[⊥,†] Mostafa A. El-Sayed,[§] Geoffrey A. Ozin,[*,‡] and Chandra Veer Singh[*,†]

PRL 113, 175502 (2014) PHYSICAL REVIEW LETTERS week ending 24 OCTOBER 2014

Stability, Energetics, and Magnetic States of Cobalt Adatoms on Graphene

Yudistira Virgus,* Wirawan Purwanto, Henry Krakauer, and Shiwei Zhang

Department of Physics, College of William and Mary, Williamsburg, Virginia 23187-8795, USA

Materials Views www.MaterialsViews.com

ADVANCED FUNCTIONAL MATERIALS www.afm-journal.de

A–D–A-Type Oligothiophenes for Small Molecule Organic Solar Cells: Extending the π-System by Introduction of Ring-Locked Double Bonds

*Roland Fitzner, Elena Mena-Osteritz, Karsten Walzer, Martin Pfeiffer, and Peter Bäuerle**

Angewandte International Edition Chemie

Photochemistry

International Edition: DOI: 10.1002/anie.201505183
German Edition: DOI: 10.1002/ange.201505183

Precise Design of Phosphorescent Molecular Butterflies with Tunable Photoinduced Structural Change and Dual Emission

Chenkun Zhou, Yu Tian, Zhao Yuan, Mingu Han, Jamie Wang, Lei Zhu, Maliheh Shaban Tameh, Chen Huang, and Biwu Ma**

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Molecular Rectifiers: A New Design Based on Asymmetric Anchoring Moieties

Colin Van Dyck* and Mark A. Ratner*

NWCHEM

HIGH-PERFORMANCE COMPUTATIONAL CHEMISTRY SOFTWARE

Dalton Transactions

ROYAL SOCIETY OF CHEMISTRY

PAPER View Article Online View Journal

CrossMark

Cite this: DOI: 10.1039/c5dt03508b

Molecular dynamics simulation study of various zeolitic imidazolate framework structures

Min Gao, Alston J. Misquitta, Leila H. N. Rimmer and Martin T. Dove*

PCCP

ROYAL SOCIETY OF CHEMISTRY

PAPER View Article Online View Journal | View Issue

CrossMark

Cite this: Phys. Chem. Chem. Phys., 2015, 17, 20093

Computational simulation and interpretation of the low-lying excited electronic states and electronic spectrum of thioanisole†

Shaohong L. Li, Xuefei Xu and Donald G. Truhlar*

JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes

Rebecca Sure and Stefan Grimme*

REPORTS

Tunable Electrical Conductivity in Metal-Organic Framework Thin-Film Devices

A. Alec Talin,[1,2]* Andrea Centrone,[2,3] Alexandra C. Ford,[1] Michael E. Foster,[1] Vitalie Stavila,[1] Paul Haney,[2] R. Adam Kinney,[2,3] Veronika Szalai,[2] Farid El Gabaly,[1] Heayoung P. Yoon,[2,3] François Léonard,[1] Mark D. Allendorf[1]*

nature COMMUNICATIONS

ARTICLE

Received 5 May 2014 | Accepted 23 Oct 2014 | Published 3 Dec 2014 DOI: 10.1038/ncomms6642

Freezing-in orientational disorder induces crossover from thermally-activated to temperature-independent transport in organic semiconductors

K.P. Goetz[1], A. Fonari[2], D. Vermeulen[3], P. Hu[4], H. Jiang[4], P.J. Diemer[1], J.W. Ward[1], M.E. Payne[1], C.S. Day[5], C. Kloc[4], V. Coropceanu[2], L.E. McNeil[3] & O.D. Jurchescu[1]

Inorganic Chemistry

Article

pubs.acs.org/IC

Predicting Stability Constants for Uranyl Complexes Using Density Functional Theory

Sinisa Vukovic,[†] Benjamin P. Hay,[‡] and Vyacheslav S. Bryantsev*

CHEMICAL REVIEWS

Review

pubs.acs.org/CR

Recent Advances in the Theory and Molecular Simulation of Biological Electron Transfer Reactions

Jochen Blumberger*

Angewandte International Edition Chemie

Chemical Bonding Hot Paper

International Edition: DOI: 10.1002/anie.201503845
German Edition: DOI: 10.1002/ange.201503845

The Nature of Bonding between Argon and Mixed Gold–Silver Trimers**

Armin Shayeghi, Roy L. Johnston, David M. Rayner, Rolf Schäfer, and André Fielicke**

NANO LETTERS

Letter

pubs.acs.org/NanoLett

Following the Transient Reactions in Lithium–Sulfur Batteries Using an In Situ Nuclear Magnetic Resonance Technique

Jie Xiao,[†] Jian Zhi Hu,*[,†] Honghao Chen,[†] M. Vijayakumar,[†] Jianming Zheng,[†] Huilin Pan,[†] Eric D. Walter,[†] Mary Hu,[†] Xuchu Deng,[†] Ju Feng,[†] Bor Yann Liaw,[‡] Meng Gu,[†] Zhiqun Daniel Deng,[†] Dongping Lu,[†] Suochang Xu,[†] Chongmin Wang,[†] and Jun Liu*[,†]

ARTICLES

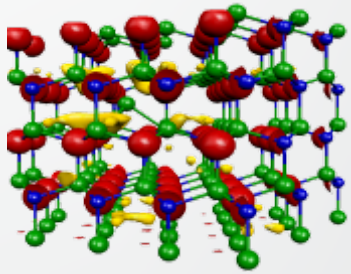
PUBLISHED ONLINE: 26 OCTOBER 2014 | DOI: 10.1038/NNANO.2014.246

nature nanotechnology

Long-range charge transport in single G-quadruplex DNA molecules

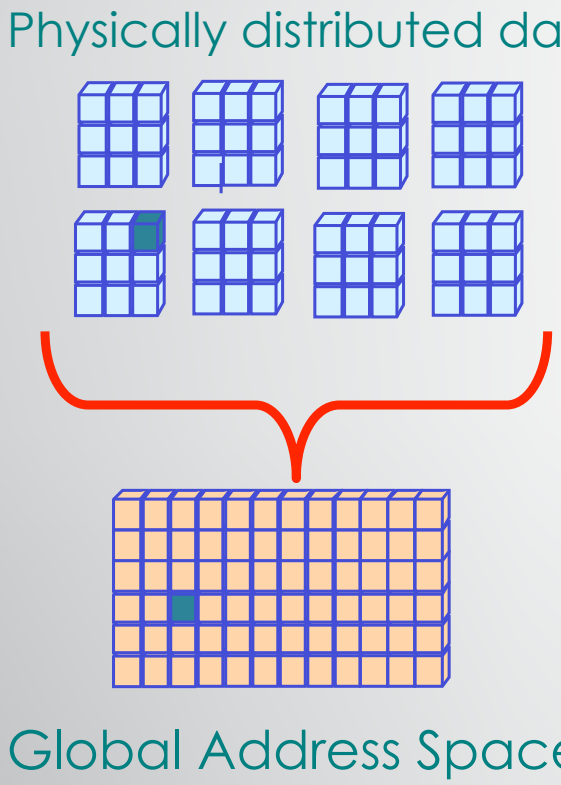
Gideon I. Livshits[1], Avigail Stern[1], Dvir Rotem[1], Natalia Borovok[2], Gennady Eidelshtein[2], Agostino Migliore[3,4], Erika Penzo[5], Shalom J. Wind[5], Rosa Di Felice[6,7], Spiros S. Skourtis[8], Juan Carlos Cuevas[9], Leonid Gurevich[10], Alexander B. Kotlyar[2]* and Danny Porath[1]*

- Modular structure
- Most of the code FORTRAN77 (bits o F90 and C)
- Performance characteristics – designed to run efficiently on parallel computers
- Emphasis on portability
  - Runs on a wide range of computers
- Uses Global Arrays/ARMCI for parallelization
- Some use of OpenMP directives
- Repository (SVN)
- Quality Assurance suite



Global Arrays

- **Distributed dense arrays** that can be accessed through a shared memory-like style
- **High level abstraction** layer for the application developer
- **One-sided** model = no need to worry and send/receive



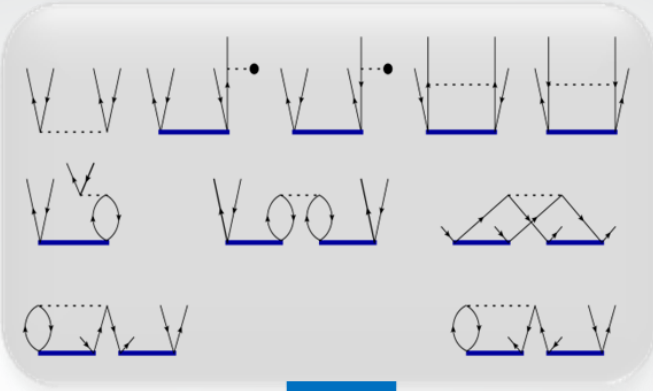
single, shared data structure/ global indexing

e.g., access A(4,3) rather than buf(7) on task 2

Tensor Contraction Engine (TCE)

K.Kowalski

- Symbolic algebra systems for coding complicated tensor expressions: Tensor Contraction Engine (TCE)
  - Hirata, *J. Phys. Chem. A* **107**, 9887 (2003)
  - Sadayappan, Krishnamoorthy, et al. *Proceedings of the IEEEE*, 93, 276 (2005).
  - Lai, Zhang, Rajbhandari, Valeev, Kowalski, Sadayappan, *Procedia Computer Science* (2012)
- New implementation of CC methods (since 2003)
  - more effective for implementing new methods
  - Easier tuning and porting



OCE

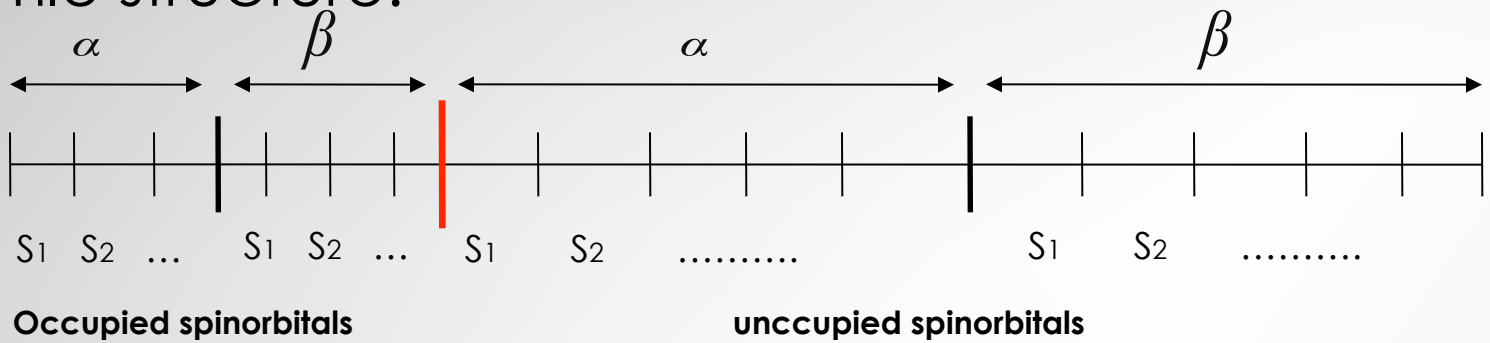
$$+\frac{1}{4}v_{ef}^{mn}t_{ij}^{ef}t_{mn}^{ab}-\frac{1}{2}v_{ef}^{mn}t_{mi}^{ef}t_{nj}^{ab}+$$

TCE

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

Tensor Contraction Engine (TCE)

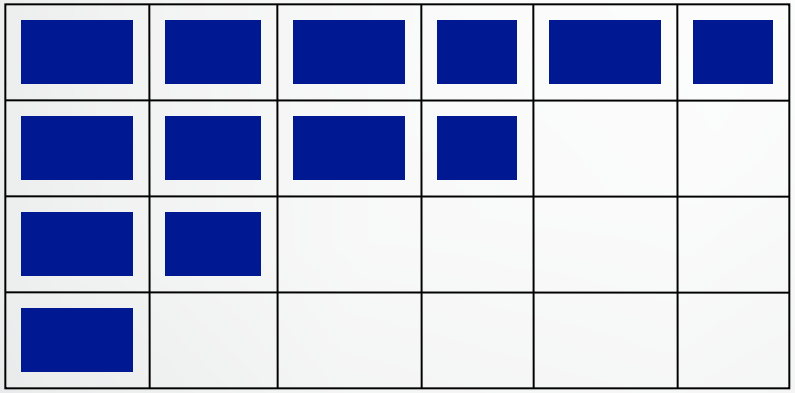
Tile structure:



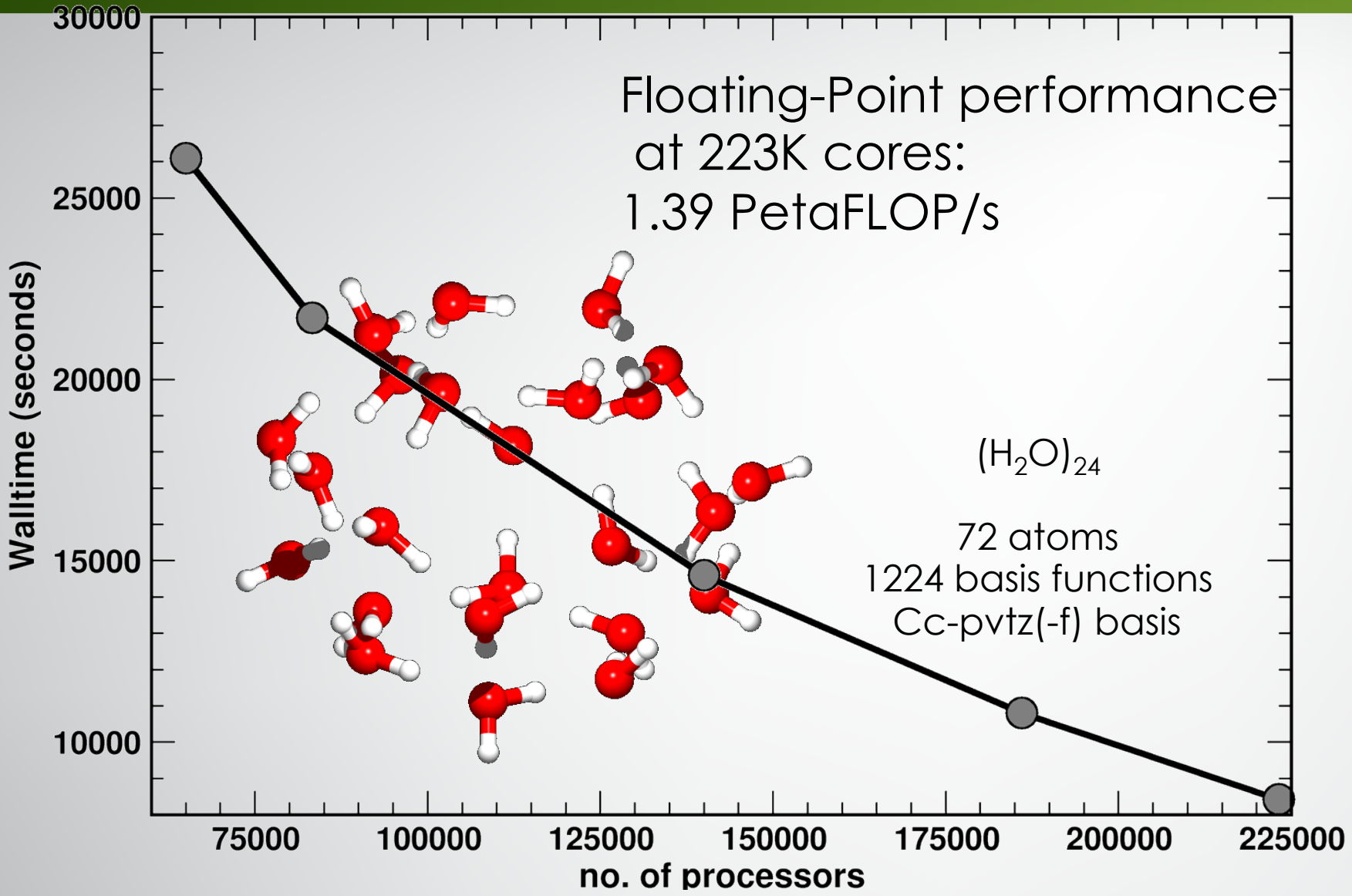
Occupied spinorbitals **unccupied spinorbitals**

Tensor structure:

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

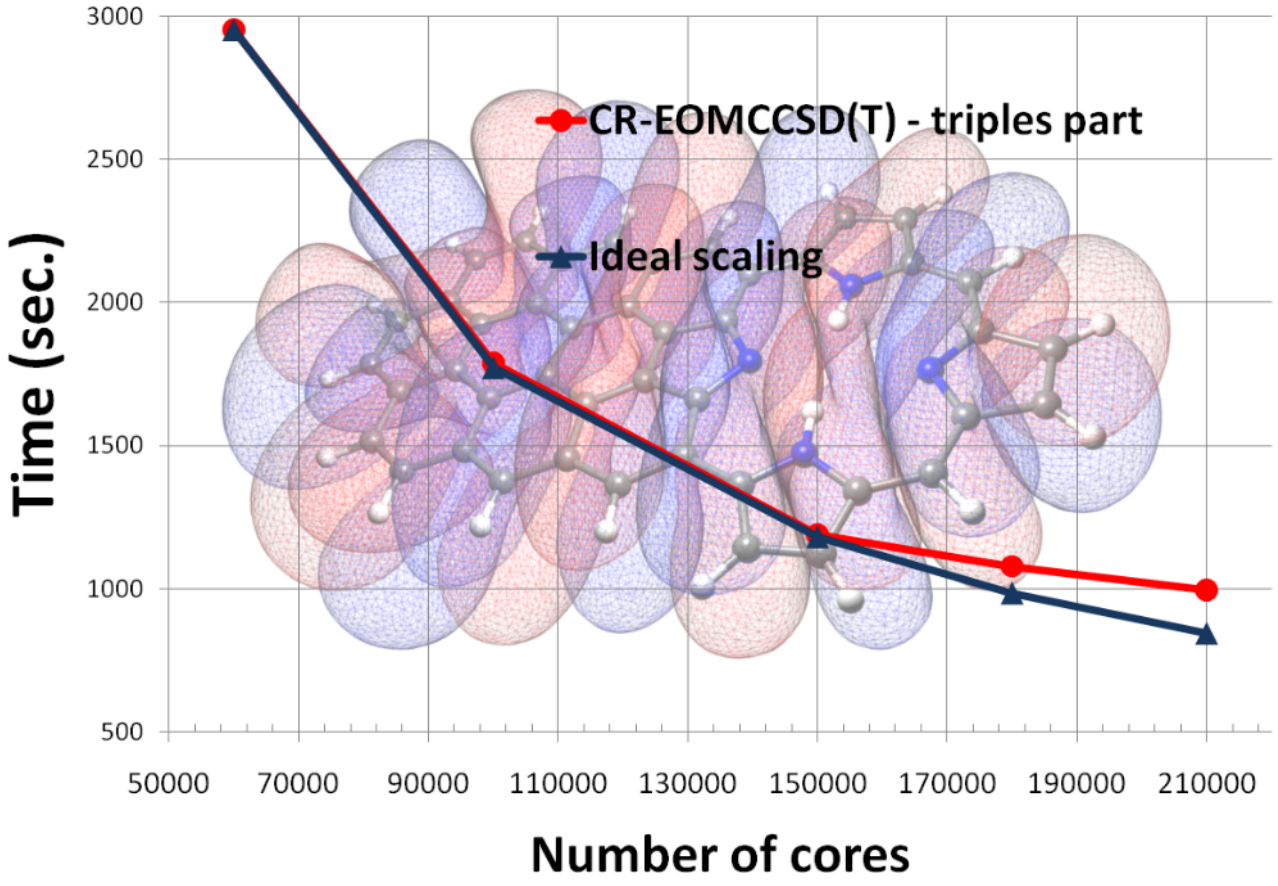


CCSD(T) benchmark on Cray XT5 : 24 water molecules



Scalability of the non-iterative EOMCC code

94 %parallel efficiency using 210,000 cores



- Scalability of the triples part of the CR-EOMCCSD(T) approach for the FBP-f-coronene system in the AVTZ basis set. Timings were determined from calculations on the Jaguar Cray XT5 computer system at NCCS/ORNL in 2011

GPU Benchmarks: CCSD(T)

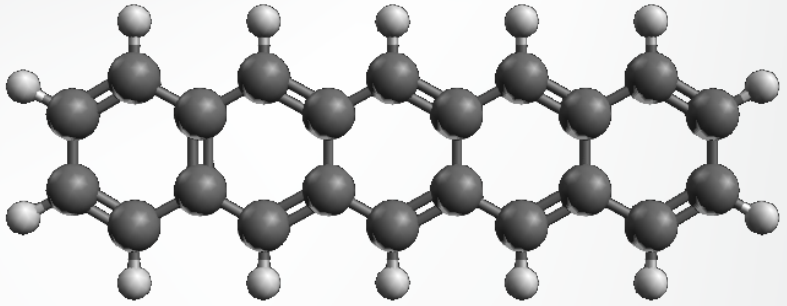
E. Aprà, K. Bhaskaran-Nair
K. Kowalski
W. Ma, O. Villa

Titan Cray XK7 system at ORNL

CCSD(T)

378 basis set functions, C1 symmetry;

98 nodes: 8 cores per node + 1GPU)



Pentacene $C_{22}H_{14}$

Using 0 GPU device(s) per node wall time / sec 9240.3

Using **1 GPU** device per node wall time / sec 1630.7

Speed-up GPU vs CPU = 9240.3 /1630.7 ~ **5.6x**.

What worked in NWChem

- Initial years: good balance of staff members & post-docs
- Collaboration of chemists, computer scientist and applied mathematicians
- Close integration of HW upgrades and SW development stages

What almost worked in NWChem

- Code structure, SW portability
- Parallelization approach
- Collaborators (developers & users) network
- Stable methodological development process

What did not work in NWChem

- Code structure able to accommodate new parallelization approaches?
- Need to abandon the monolithic structure to
  - accommodate wider range of methodologies
  - make life easier for new developers/contributors
  - Interface NWChem with other codes (e.g. LAMMPS, Amber)

TCE in NWChem: Strengths and weaknesses

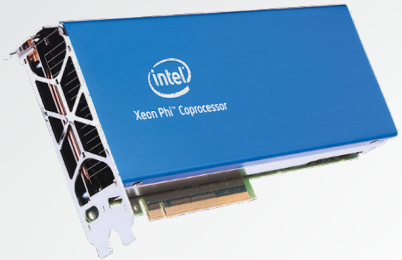
EMSL

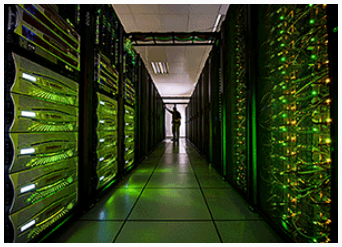
- DSLs separate mathematical and physical models from the implementation
- Rapid/less error prone protocol for implementing complex theoretical models
- Current version of TCE defined by a monolithic structure
  - Most efficient serial & parallel code is obtained by hand coding
  - Hard to exploit data-locality
  - Profiling …

Path forward

- Size of code O(10M) lines
- Should we use an evolutionary or revolutionary approach (if it somehow works, why should I change anything …?)
- Do we need to re-write the code from scratch?
- Risks associated with SW over-design (keep it simple …)
- Open the code to make it more accessible to interfaces to other modeling codes, workflows, etc ...

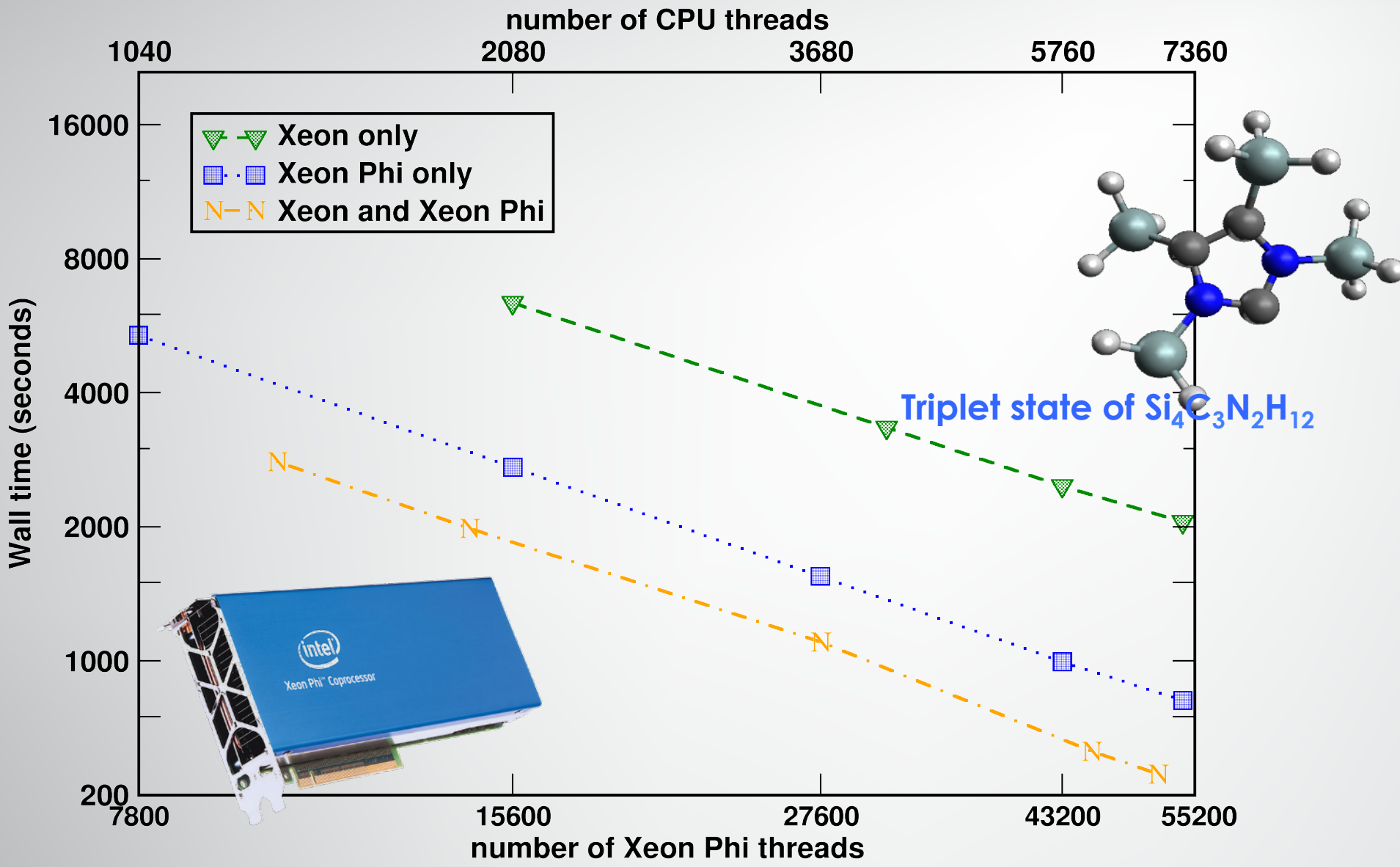
EMSL cascade Hardware specs



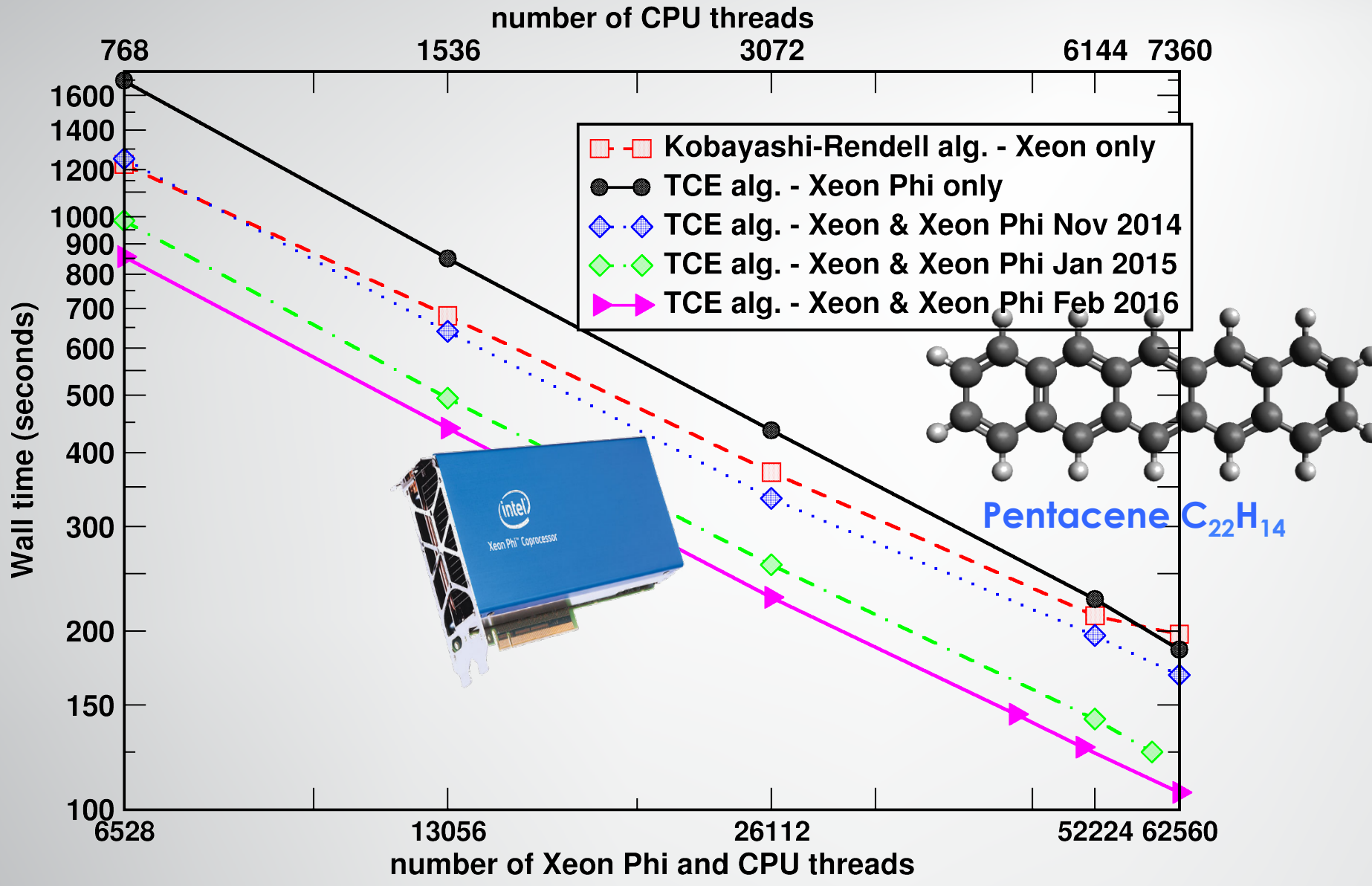


- Theoretical peak: 3.4 Pflops, Linpack run: 2.5 Pflops
- 1440 compute nodes, compute nodes, on each node
  - Two 2.6GHz Intel Xeon E5-2670 8-core processors
  - Two Intel Xeon Phi 5110P
  - 128 GB DDR3 memory
- FDR Infiniband network
- 2.7 petabyte shared parallel filesystem (60 Gb/sec)

CCSD(T) on Intel Xeon Phi



CCSD(T) on the Intel Xeon Phi Coprocessor



Thanks