Argonne NATIONAL LABORATORY

Materials Design and Discovery: Catalysis and Electrical Energy Storage

Presenter: Nichols A. Romero, ALCF

ESP post-doc: Anouar Benali, ALCF

PI: Larry Curtiss, ANL MSD and CNM

U.S. DEPARTMENT OF ENERGY

Comments from a reviewer on “Material Design and Discovery” from a proposal

- How could this machine with these programs be used to design a new solar cell? Or a new cure for AIDS? Or a new high-T superconductor? This is not intended as a trivial question. The present method of DISCOVERY relies on the trained human mind (insight) and experiment (serendipity). Computational science so far has not delivered any new discoveries because it lacks the possibility of serendipity. The greatest success of computational chemistry has been improved insight into the way material behaves at the atomic level.
- The use of these tools in DESIGN seems more likely. The engineer who is merely trying to optimize an established material, reaction, etc could well use a model where she could tune some parameters to optimize conditions. In fact, information technology tools could be used to mechanically optimize a set of parameters to optimize a given response. This is not done now because predictive computational chemistry tools are too slow to be used in this way for complex systems with millions of atoms. An exascale computer with software that also has reached 10^6 speedup over existing methods (and linear scaling with the number of atoms!) could be used in this engineering design of improved energy devices.

Challenges in Atomistic Simulation, Theory, and Modeling

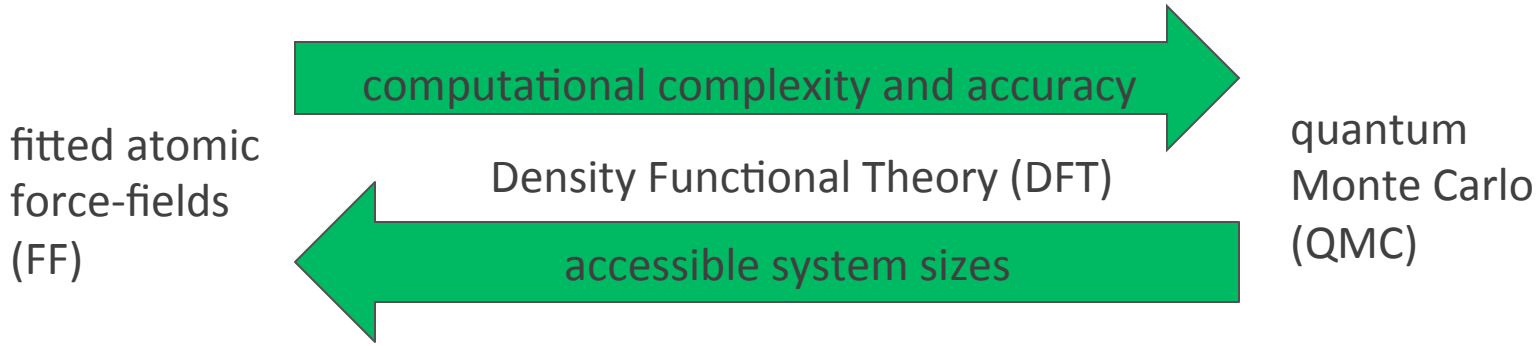
- How to compute macroscopic property Y for material X?
  - "If there is an quantum mechanical <operator>, there is a way."
  - Microscopic properties are moderate (ground state) to difficult (excited state), e.g. binding energies, forces, band gap, phonons, etc.
  - Macroscopic properties not obvious, e.g. friction, flammability, detonation, etc.
- Can I find material X that has target property Z?
  - Inverse design problem
- Can I make material X?
  - Kinetics
  - Thermodynamics
- Should I make the material X?
  - Economics
  - Safety and the environment

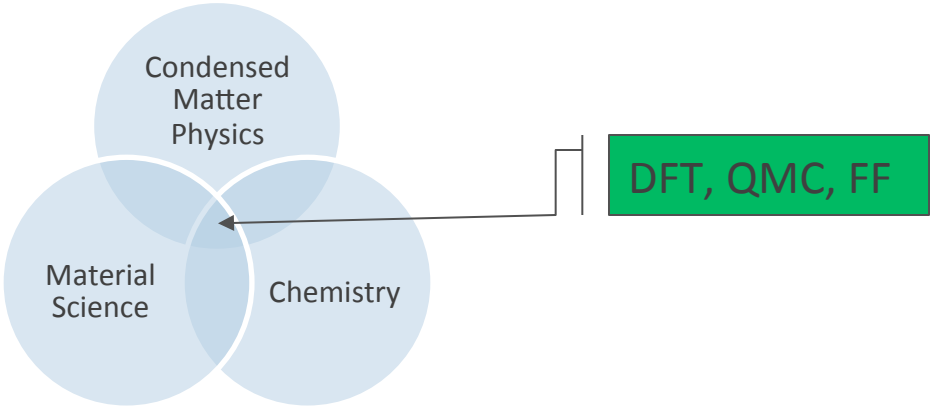
Problems in **material design**, discovery, synthesis, safety (humans and the environment), and cost effective.

Safety is really really important!

Atomic simulations methods in a nutshell

Primarily for ground-state simulations in material science, chemistry, and condensed matter physics





Atomistic simulation methods for advancing materials discovery

- **Multi-method approach with respect to level of theory:**
  - Empirical force-field based molecular dynamics (number of atoms ~ 1,000,000,000 – 1,000,000,000,000)
  - Density Functional Theory (number of valence electrons ~ 10,000)
    - Much more with an O(N) method, but with lots of caveats
  - quantum Monte Carlo (number of valence electrons ~ 1,000)
- **Multi-method approach to modeling :**
  - High-throughput (using “brute force” approach or more sophisticated machine learning)
  - Thermodynamic sampling (i.e. NVT, NPT)
  - Image methods (e.g. nudge elastic bands, phonons)
  - Geometry relaxation and non-equilibrium molecular dynamics

Quantum Monte Carlo needs petascale and beyond

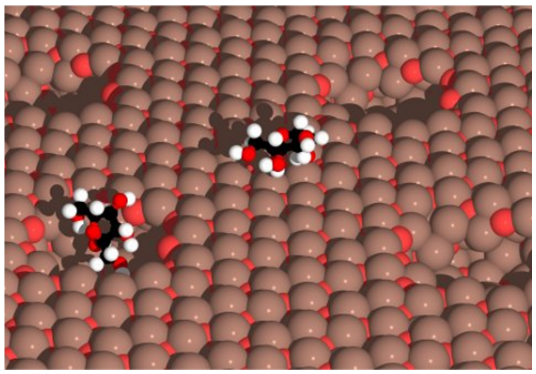
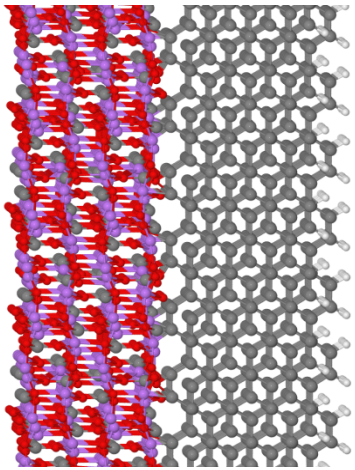
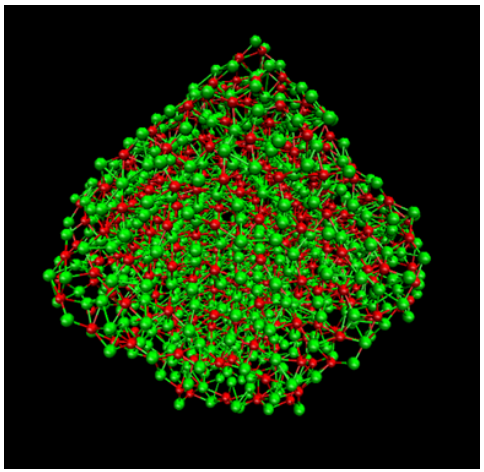
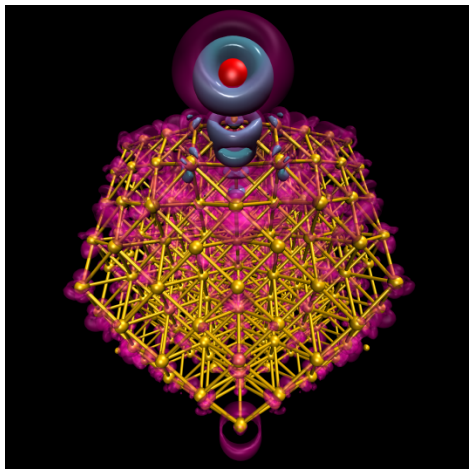
- People do not understand how computational expensive, it is **really** expensive. Consider a rather simple system, 32-water molecules in a box with p.b.c.
  - Classical MD needs an iPhone5 (~1 Gflop)
  - Density Functional Theory needs a few nodes of a commodity cluster (~1 TFlop)
  - Quantum Monte Carlo needs JaguarPF (~1 PFlop)
- Simple reason for needing exascale, more accurate quantum Monte Carlo (QMC)
  - Another 10X in flop rate to calculate forces
  - Another 10X in flop rate to calculate include effects of core electrons (e.g., projector augmented wave method)
  - Another 10-100X in flop rate to study real materials
- QMC will need exascale resources for high accuracy calculations on real materials
  - Is this over kill or are there real applications? Consider defect migration in UO_2.
  - Spin-polarized + QMC + projector augmented wave method + forces + linear scaling + beyond scalar relativistic ? + (probably some type of sampling or image method)
    - Theory not there yet
    - May need beyond exascale

Who is involved?

- Scientific leads:
  - Larry Curtiss, ANL
  - Jeff Greeley, Purdue University
- Catalyst: Nichols A. Romero, ANL
- Post-docs:
  - ANL: Anouar Benali (LCF), William Parker (LCF), K. C. Lau (MSD)
  - Stanford University/SLAC: Lin Li
- Code developers:
  - GPAW, DFT code using projector augmented wave method on real-space grids
    - Jens Jørgen Mortensen, Center for Atomic-scale Molecular Design (CAMd)
    - Jussi Enkovaara, CSC, the Finnish IT Center for Science, Ltd.
  - QMCPACK, QMC code using B-spline, plane waves, and localized orbitals
    - Jeongnim Kim, Oak Ridge National Laboratory
- Performance engineers:
  - Vitali Morozov, ANL
  - Lee Killough , now at Appro

Original Early Science Plans

- Investigate Materials: (depicted below, left to right)
  - Biomass energy conversion
  - Electrical energy interfaces
  - Lithium-air batteries
  - Catalysis with transition metal nanoparticles
- DFT calculations on systems containing > 10,0000 valence electrons:
  - GPAW code used on up to 32-racks on Blue Gene/P for single point energy calculations
  - Geometry optimization and MD needed

Revised Early Science Plans

- GPAW-based calculations would encounter non-trivial algorithmic difficulties:
  - Impacts all O(N^3) DFT codes, not just GPAW
  - Canonical DFT has a number issues, arises from non-local representations of Ψ:
    - Dense diagonalization exhibits poor performance (see http://arxiv.org/pdf/1205.2107v1)
    - Instabilities in SCF algorithms (see Phys. Rev. B 64, 121101(R) (2001))
    - O(N^3) "wall"
  - Reduce-scaling methods are needed in quantum mechanical approaches:
    - Fragment-type methods (GAMESS, LS3DF) or localization methods (CONQUEST, MADNESS, SIESTA, and many others)
    - Lots of progress, but also many remaining challenges (metals vs. insulators, precision, etc.)
- Multi-method approach with multiple codes:
  - Explore use of quantum Monte Carlo for materials problems (QMCPACK)
  - Other DFT codes for use large DFT calculations (still an open issue, use CPMD and GPAW to the extent that the computation is tractable)
  - Force-field based molecular dynamiacs (LAMMPS)

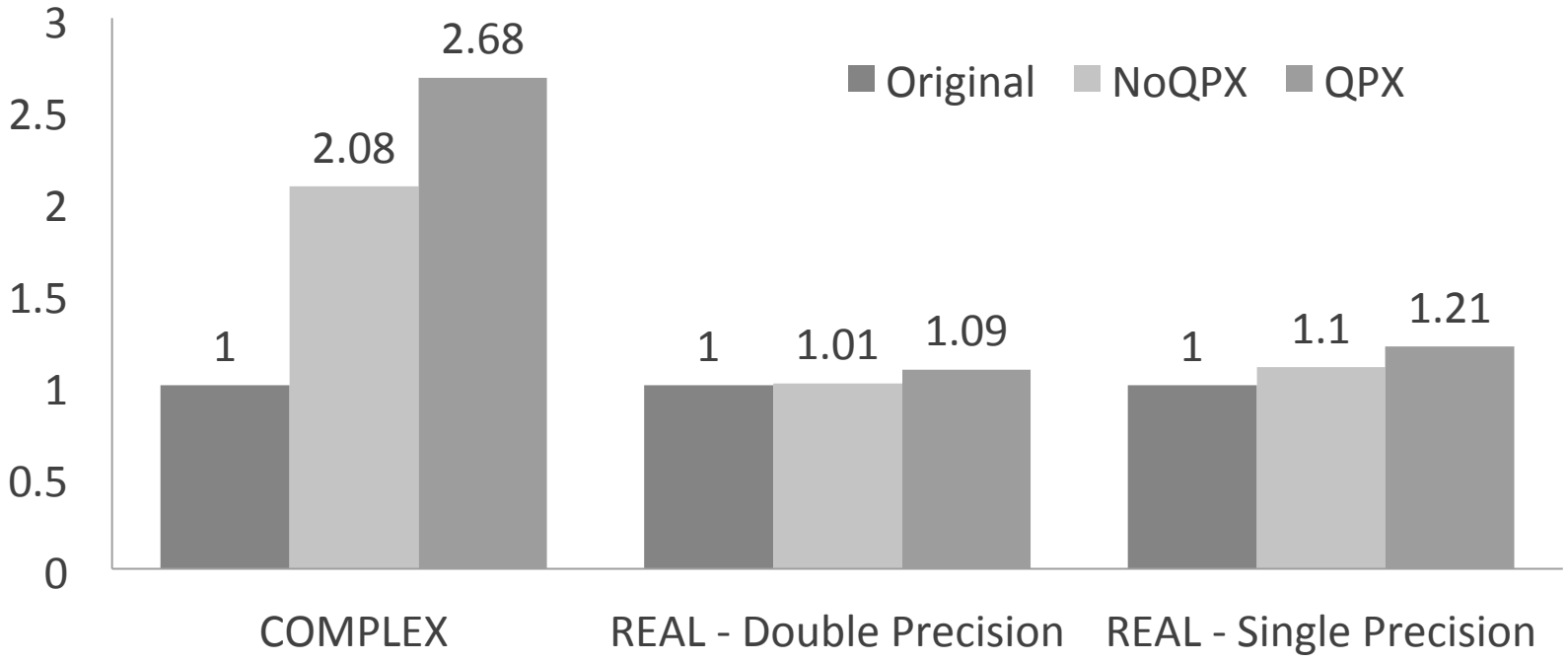
Revised Early Science Plans (cont'd)

- Broad scientific goals mostly unchanged, but focus on fewer problems.
- Investigation of vdW-dominated systems (completed, paper in preparation)
  - Anouar Benali (ALCF, ESP) in collaboration with O. Anatole von Lilienfeld (ALCF) and Luke Shulenburger (SNL)
  - QMC calculation using QMCPACK
  - Nobel gases and anti-cancer agent
- Catalysis with transition metal nanoparticles (early stages)
  - ALCC proposal and BES funding (PI: Greeley)
  - Catalysis study using QMCPACK
  - Work by William Parker (ALCF) in collaboration with Jeongnim Kim (ORNL)
  - Also investigate catalysis on transition metal oxides surface (e.g. ZnO_2) as part collaborative INCITE work
- Lithium-air batteries (early stages)
  - INCITE proposal and EFRC funding (PI: Curtiss)
  - Battery work using DFT and FF MD calculations
  - Lithium peroxide growth at cathode (porous carbon)

QMCPACK in a nutshell

- Variational Monte Carlo (VMC) and Diffusion Monte Carlo (DMC).
  - Supports many basis: PW, B-splines, LCAO
  - Supports many boundary conditions
- Programming languages: C and C++
- Parallelism: MPI and OpenMP parallelism
- When to use QMC instead of DFT:
  - vdW-dominated systems
  - Strong-correlated systems
  - Chemical accuracy is needed
- Performance characterization:
  - Many kernels are memory bandwidth limited (?)
  - Ideal OpenMP scaling
  - Ideal MPI scaling
  - Minimal I/O
  - Replicated data

QMCPACK – performance on Blue Gene/Q



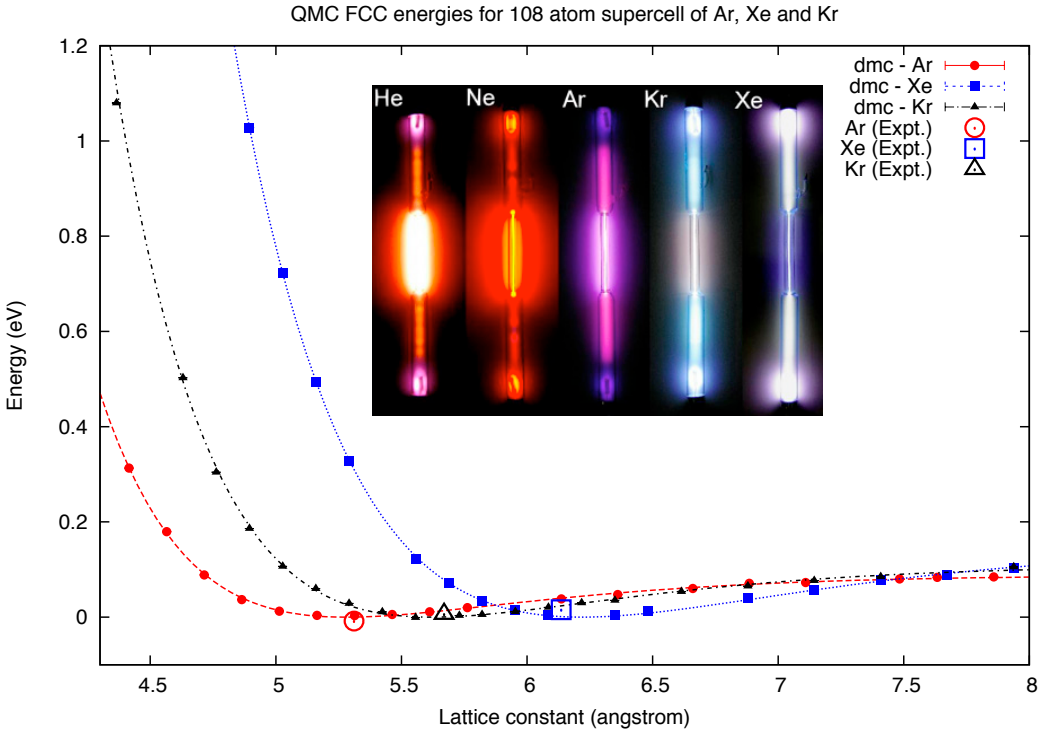
Application speed-up using QPX and prefetching is 2.68X from original Algorithm.

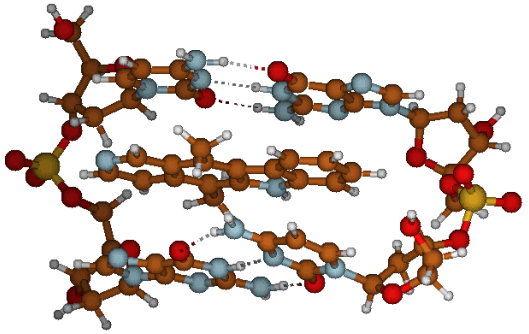
QMCPACK – current status and future developments

- Current status:
  - Production science is already underway
  - QMCPACK has been scaled to all 96-racks of Sequoia as proof-of-principle in the near future (group from Livermore and Sandia)
  - Single precision working allows storage of wave functions for larger simulations
  - Double grid technique allows additional memory savings in cases with vacuum
- Future work (first two are short term, last two are long term)
  - Percentage of peak is low 5%, other systems is about 10%.
  - Presently not competitive with GPU version of the code:
    - Optimize single precision version of Einspline
    - nested OpenMP parallelism
  - More compact representations of the wave function needed
    - LAPW (in progress)
    - Distribute/tessellate wave function without a performance hit
  - Trial wave function currently optimized with serial LAPACK (general non-symmetric EVP)

vdW-dominated systems

- Benali in collaboration with O. A. von Lillenfeld (ALCF) and L. Shulenburger (SNL)
- Pure and hybrid DFT is useless for vdW interactions.



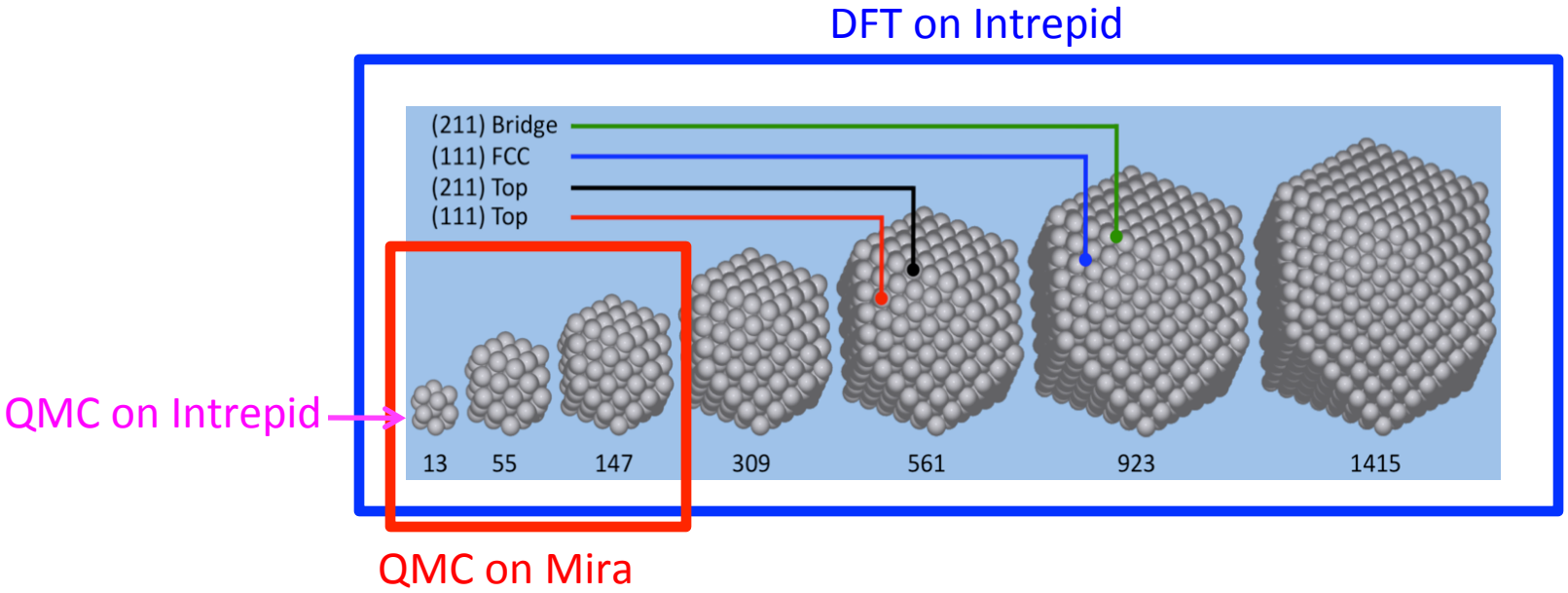


uterine cancer drug

method	Δe_{bind} (kCal/mol)
DFT	+5.2
vdW –TS	-36.6
vdW-TB	-39.1
vdW-MB	-50.7
QMC-DMC	-33.6 +/- 0.98

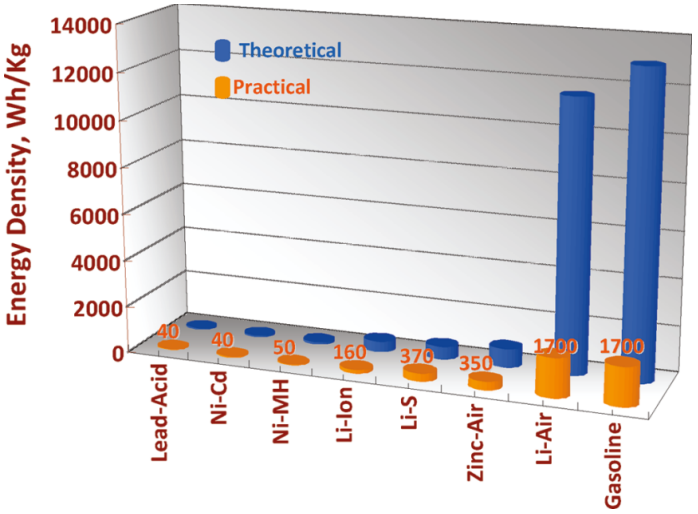
Catalysis on transition nanoparticles

- Synthesis of industrial chemical
- Pollution remediation (carbon monoxide (poisonous) to carbon dioxide (harmless))
- Three papers based on GPAW calculations on Intrepid over 3 years of INCITE
- Gold behaves as expected in the limit of infinite clusters, but not platinum
- Will continue to explore on Mira with QMC



Lithium-air battery

- potential for much larger energy densities than current batteries
- many challenges
- long-term research is essential to provide the breakthroughs needed for this new technology
- High performance computing is needed to provide insight into the complexities of the li-air battery at the molecular level and help design new materials for electrolytes and electrodes



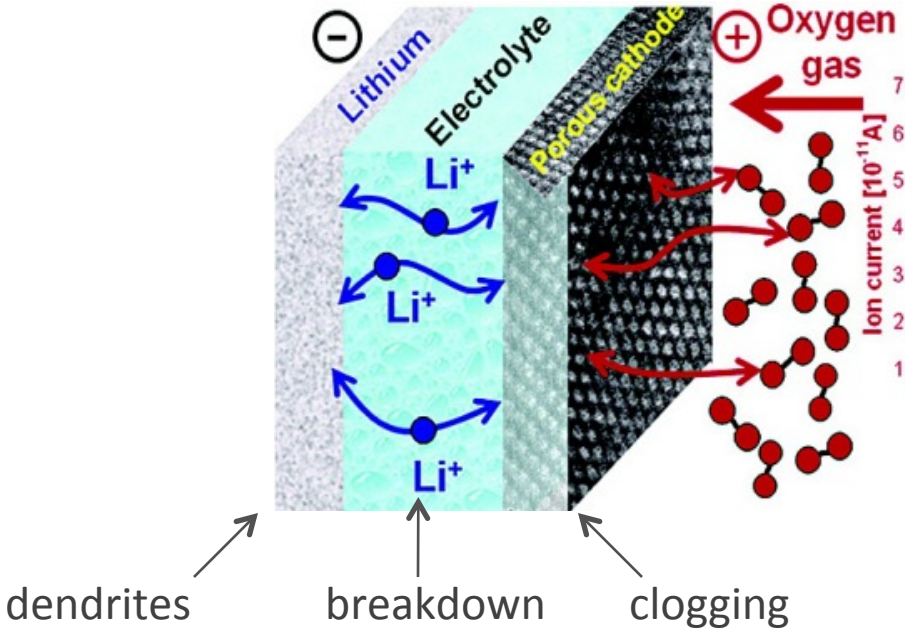
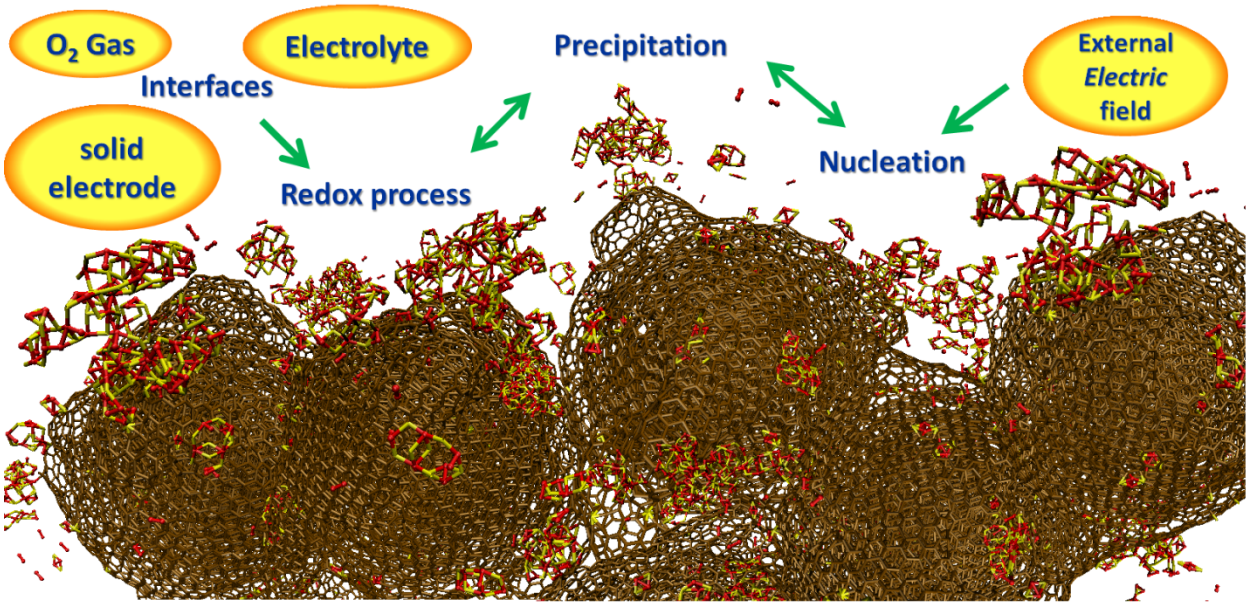
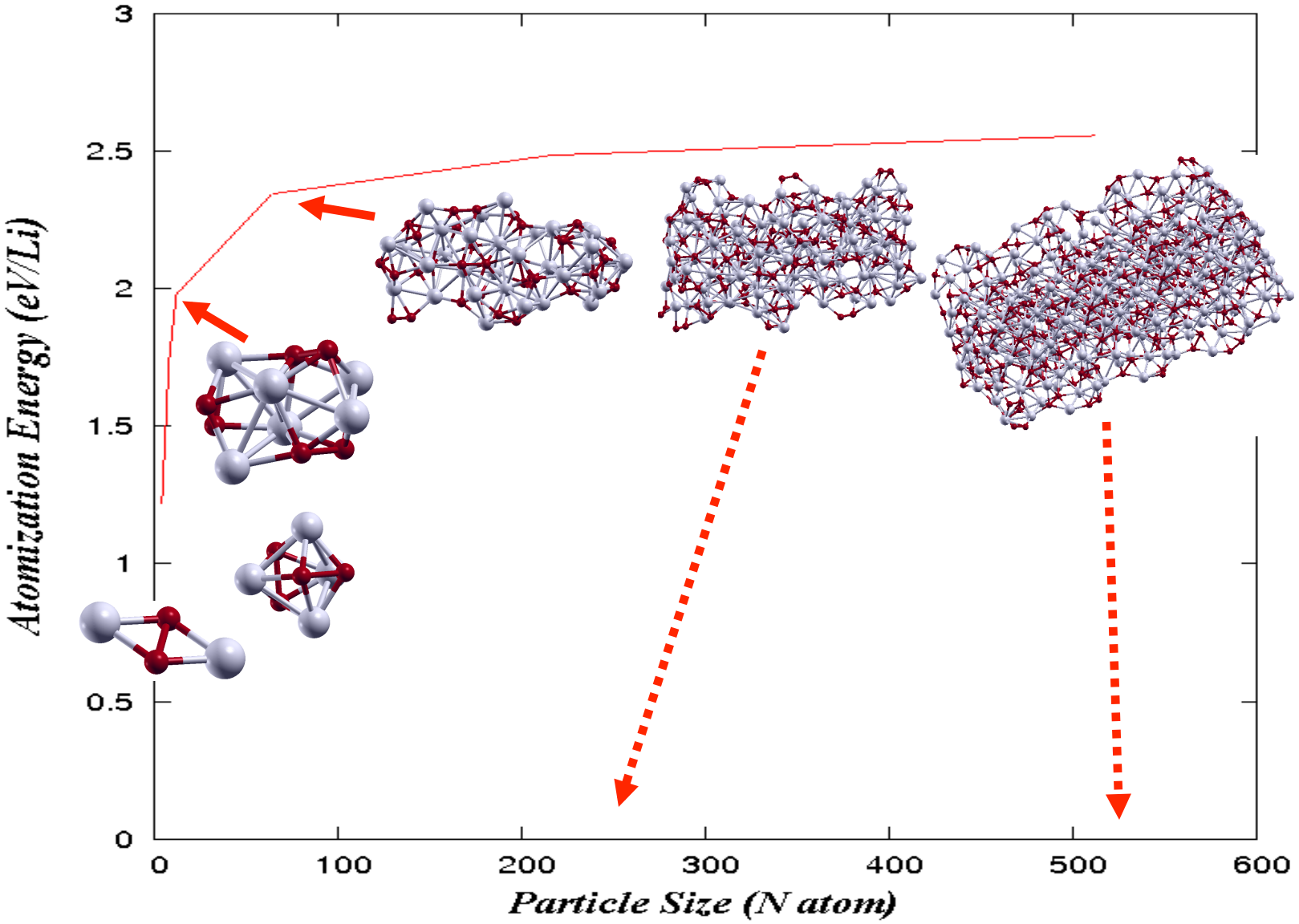


Illustration of processes occurring at Li-air interfaces – to be modeled using Mira



- DFT simulations on Mira will be used to model various processes on in the lithium battery
  - Nucleation and growth of lithium peroxide at interfaces
  - Effect of electrolyte on the growth process
  - Electrocatalysts
  - Larger models for nanoparticles
- MD simulations will be used to simulation nanocrystalline lithium peroxide

Size evolution of $(Li_2O_2)_N$ nanoparticles extend to several thousand atom systems



K. C. Lau

Summary

- IBM early access system was very helpful.
- Allocation: 50 million core hours
- Used: - 5 million core hours
- Miscellaneous: 0.5 million core hours
- GPAW calculations: 4.5 million core hours
  - Relaxed 923 platinum nanoparticle
- QMCPACK calculations: 50 million core hours
  - EOS Ar, Kr, Xe
  - Two and three-body contributions to the many body energy of Ar
  - Pt_{13} nanoparticle
  - Bulk Pt
- Largest production calculation:
  - GPAW – 8-racks
  - QMCPACK – 32-racks
- Papers (in preparation): 1

Acknowledgements

- Technical University of Denmark (DTU), Center for Atomic-scale Molecular Design (CAMd)
  - Jens Jørgen Mortensen (lead developer of GPAW)
- CSC, the Finnish IT Center for Science, Ltd.
  - Jussi Enkovaara
- Argonne National Laboratory
  - Anouar Benali, Vitali Morozov,Lee Killough (now at Appro)
- Oak Ridge National Laboratory:
  - Jeongnim Kim (lead developer of QMCPACK)
- IBM (Rochester, Watson, Sweden, Toronto)
  - Paul Coffman, Bob Walkup, Basil Kenneth, Wang Chen

This research used resources of the Argonne Leadership Computing Facility at Argonne National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under contract DE-AC02-06CH11357.