

Breakthrough Simulations of Condensed Matter Systems

David Ceperley

NCSA & Department of Physics

University of Illinois at Urbana-Champaign

- Why Quantum Monte Carlo?
- Various QMC methods
- Challenges at the petascale
- Materials under extreme conditions: hydrogen and helium
- Accurate description of water
- Materials design as a peta-apps

Jeongnim Kim will discuss QMCPACK and performance

"The general theory of quantum mechanics is now almost complete. The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."

Dirac, 1929

$$\hat{H} = -\sum_{i=1}^N \frac{\hbar^2}{2m_i} \nabla_i^2 + \sum_{i<j} \frac{e_i e_j}{r_{ij}} \quad \hat{H}\Psi(r_1, r_2, \dots) = E\Psi(r_1, r_2, \dots)$$

Maxwell, Boltzmann and Schrödinger gave us the model (at least for condensed matter physics.) Hopefully, all we must do is numerically solve the mathematical problem and determine the properties. (*first principles or ab initio methods*) Without numerical calculations, the predictive power of quantum mechanics is limited.

The curse of dimensionality: Direct quantum methods are slow!

Suppose we represent the complete **N**-body wavefunction and treat it as strictly a problem in linear algebra—find the exact solution.

If each dimension takes **100** complex numbers

Then **N** particles in 3 dimensions will take **10^{6N}** numbers..

Even with computer time and memory increasing exponentially, the size of system we can treat will only grow linearly in time.

2 particle scattering was done on earliest computers ~1950

3 particle scattering (Fadeev eq., coupled channels) after 1980's.

4 particle scattering (generic 12 dimensional problem) is still very hard.

Expect progress on 5 particle scattering in 2020.

There is no way out of this argument--**except to change the problem.**

We don't always need the wavefunction! Experimentalists can't measure it!

Stochastic methods (simulations) don't scale this way.

Nature sets a very high standard for accuracy!

energy of hydrogen atom = 157,887K

1eV=11,600K Room temperature = 300K=0.025eV

Modeling of processes relevant to materials, chemistry, biology, ...
needs to be accurate at the level of 0.01eV \sim 0.4mH

Examples for small molecules: error in binding energy

Figure from Shiwei Zhang

$O_3, H_2O_2, C_2, F_2, Be_2, \dots$

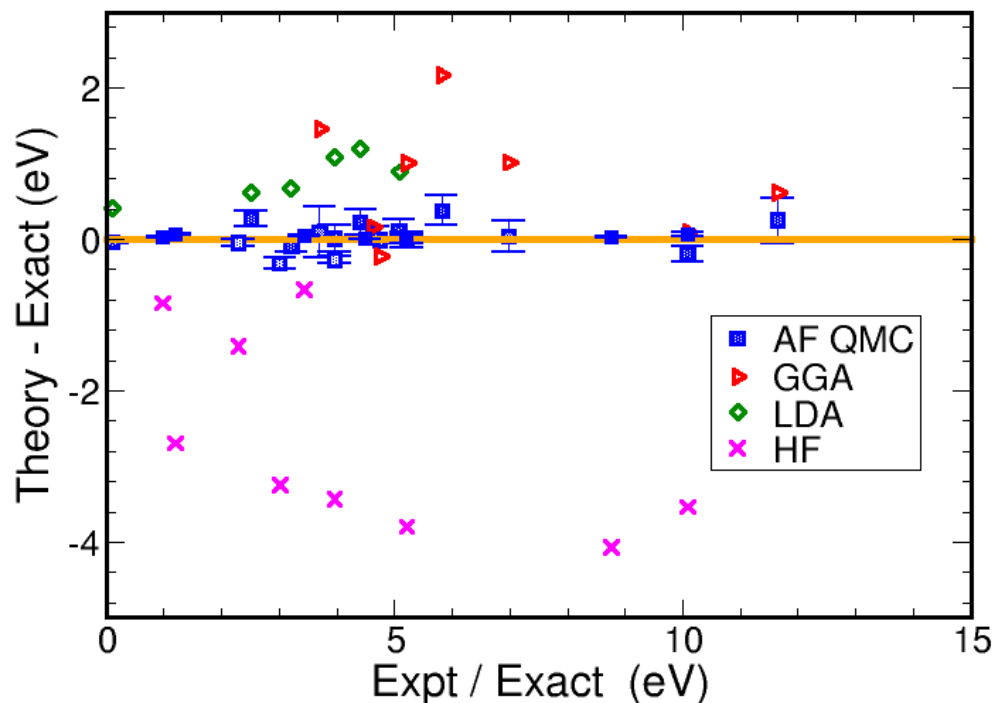
Si_2, P_2, S_2, Cl_2

As_2, Br_2, Sb_2

TiO, MnO

Current methods (e.g. DFT) are inadequate

QMC typically accurate to better than 0.1eV



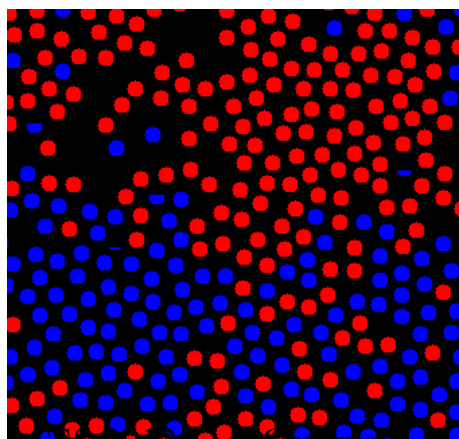
Quantum Monte Carlo

- Premise: need to use simulation techniques to “solve” many-body quantum problems just as you need them classically.
- Both the wavefunction and expectation values are determined by the simulations. Correlation built in from the start.
- QMC gives most accurate method for general quantum many-body systems.
- QMC electronic energy is a standard for approximate DFT calculations. (3rd largest citation in Physical Review Letts.)
- provides a new understanding of quantum phenomena and a practical tool
- A continuum of QMC methods:
 - Variational Monte Carlo (VMC)
 - Projector Monte Carlo methods for $T=0$:
 - Diffusion Monte Carlo (DMC)
 - Reptation MC (RQMC)
 - Auxiliary field QMC (AFQMC)
 - Path Integral Monte Carlo for $T>0$ (PIMC)
 - Coupled electron-ion Monte Carlo $T>0$ (CEIMC)

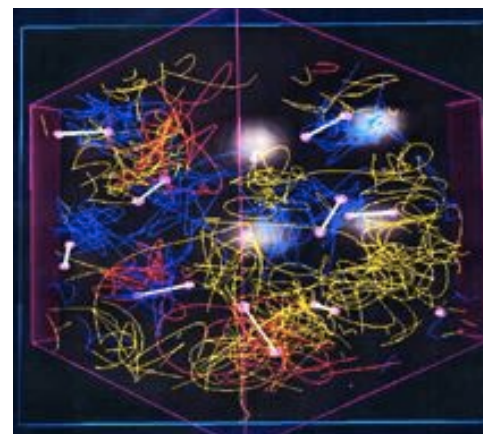
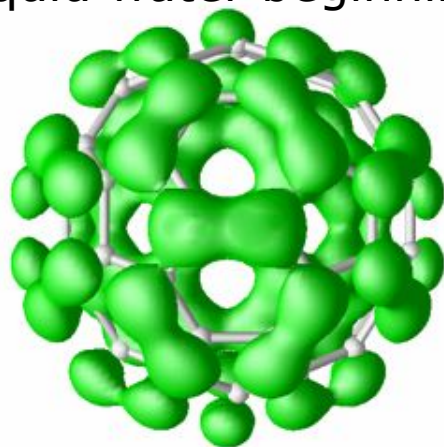
Goal is NOT large N , but higher accuracy and new capabilities

Breakthrough Quantum Monte Carlo simulations:

- Hard-core bosons on a CDC 6600 (1974)
- Electron gas on CRAY-1 (1980)
- Superfluid helium (1984)
- Ground state of solid hydrogen at high pressures, CRAY XMP and CYBER 205 (1987)
- Electronic and structure properties of carbon/silicon clusters on HP 9000/715 cluster and Cray Y-MP (1995)
- Coupled Electron-Ion Monte Carlo simulations of dense hydrogen on Linux Clusters (2000s)
- Direct simulation of liquid water beginning in 2009 (Jaguar-pf)



ACS 2012 Ceperley



Variational Monte Carlo (VMC)

(McMillan 1965)

- Put correlation directly into the wavefunction.
- Integrals are hard to do: need MC.
- Take sequence of increasingly better wavefunctions. Stochastic optimization is important!
- **Can we make arbitrarily accurate functions?** Method of residuals says how to do this.
- Recent progress with “backflow” and other forms and with optimization of trial function parameters
- VMC has no sign problem, and with classical complexity.

- Posit a wavefunction $f(\mathbf{R}, \mathbf{a})$
- sample $|f(\mathbf{R}, \mathbf{a})|^2$ with random walk.
- minimize energy or variance of $f(\mathbf{R}, \mathbf{a})$ with respect to \mathbf{a}

$R \equiv (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \text{"walker"}$

$$\Psi_2(R) = \text{Det}\{\phi_i(\mathbf{r}_j)\} e^{-\sum_{i<j} u_{ij}(r_{ij})}$$

$$\Psi_{n+1}(R) \approx \Psi_n(R) e^{-\underbrace{\langle \phi_n^{-1} H \phi_n \rangle}_{\text{smoothing}}}$$

smoothing

Projector Monte Carlo

e.g. Diffusion Monte Carlo (DMC)

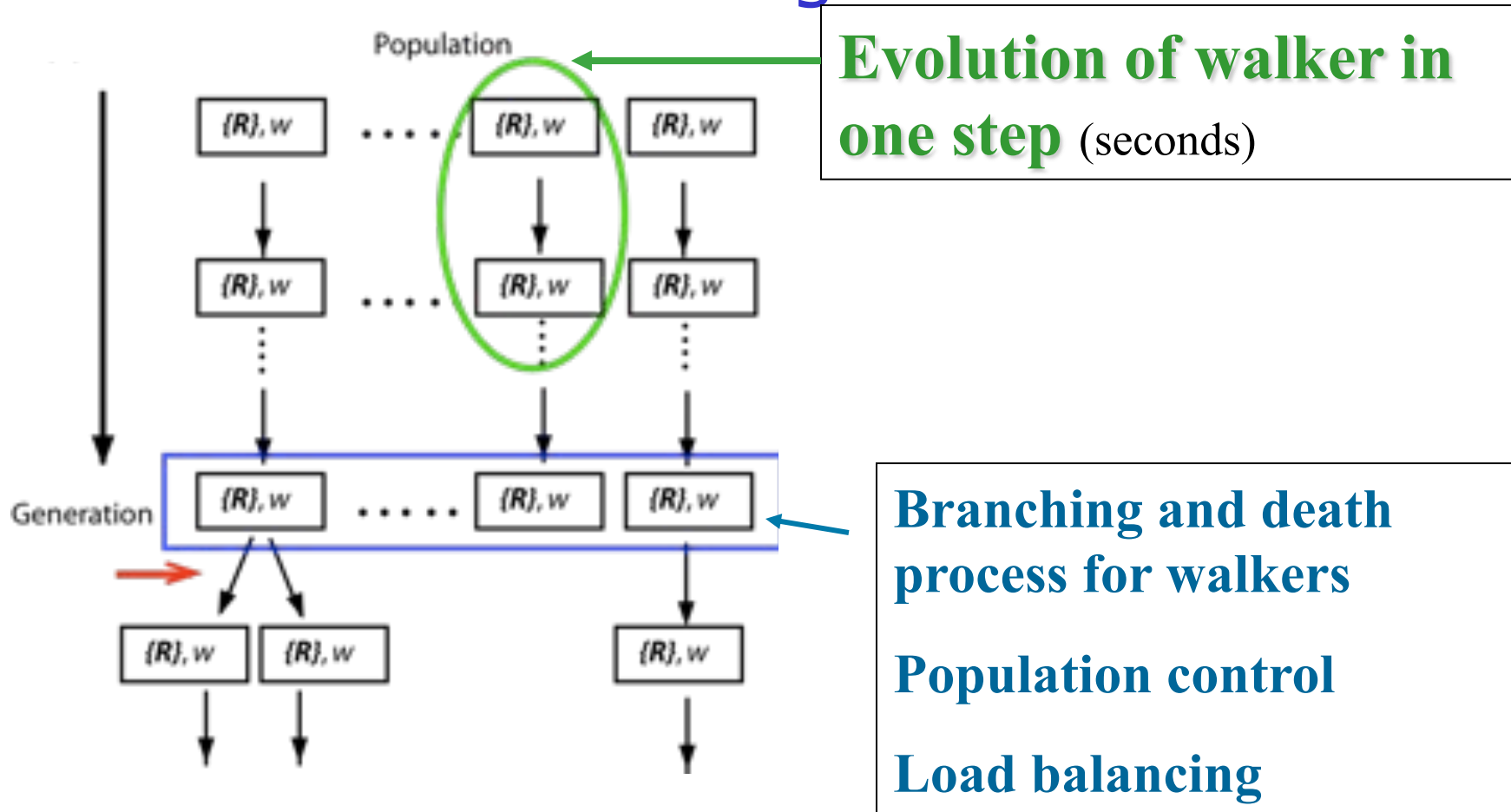
- Automatic way to get better wavefunctions.
- Project single state using the Hamiltonian

$$\phi(t) = e^{-(H-E)t} \phi(0)$$

- This is a diffusion + branching operator.
- Very scalable: each walker gets a processor.
- **But is this a probability?**
- **Yes!** for bosons since ground state can be made real and non-negative. **But** all excited states must have sign changes.
- In **exact** methods one carries along the sign as a weight and samples the modulus. This leads to the famous sign problem

$$\phi(t) = e^{-(H-E)t} \text{sign}(\phi(R,0)) |\phi(R,0)|$$

DMC algorithm



QMC is a “naturally” parallel application

It has many ways to use asynchronous computation:

- statistics (many copies of the same job) 100-fold
- parameter studies (density, T , N ) 100-fold
- imaginary time slices (path integrals) 10-fold
- boundary conditions on the phase 100-fold
- walkers in branching random walks 1000-fold
- averages over random environment 100-fold

Not all are relevant for all problems.

QMCPACK

- Open-source library and application package to perform Quantum Monte Carlo (QMC) Simulations
- Implements various QMC algorithms: VMC, DMC, RQMC
 - Generic representations of the physical entities and models
 - Object-oriented implementation of QMC algorithms (C++)
 - Generic programming of computational Kernels
- Designed for **large-scale QMC simulations** of molecules, solids and nanostructures: OpenMP/MPI Hybrid parallelization, effective for multi-core systems
- Standard open-source libraries and utilities for development, compilation and executions
- Adopts XML/HDF5 for I/O
- Developed at Illinois, now a consortium

Principal author: Jeongnim Kim @UIUC/NCSA/ORNL

Evolution of MD and MC Simulations

Y
C
C
C
C
C
C
C
C
C
C

- Hard sphere MD/MC ~1953 (Metropolis, Alder)
- Empirical potentials (e.g. Lennard-Jones) ~1960 (Verlet, Rahman)
- Potentials from density functional theory ~1985 (Car-Parrinello)
- Potentials from Quantum Monte Carlo (CEIMC) ~2000

- Initial simulations used semi-empirical potentials.
- Much progress with “ab initio” molecular dynamics simulations where the effects of electrons are solved for each step.
- However, the potential surface as determined by density functional theory is not always accurate enough

How can we use computation to make a better model?

The Phases of Hydrogen and Helium at High Pressure as revealed by simulations

- Computational methods:
 - DFT - MD
 - Quantum Monte Carlo
 - Structure prediction

WHO DID THE WORK?

Miguel Morales, Livermore

Carlo Pierleoni: L'Aquila, Italy

Jeff McMahon U of Illinois

+help from others

DOE DE-FG52-06NA26170

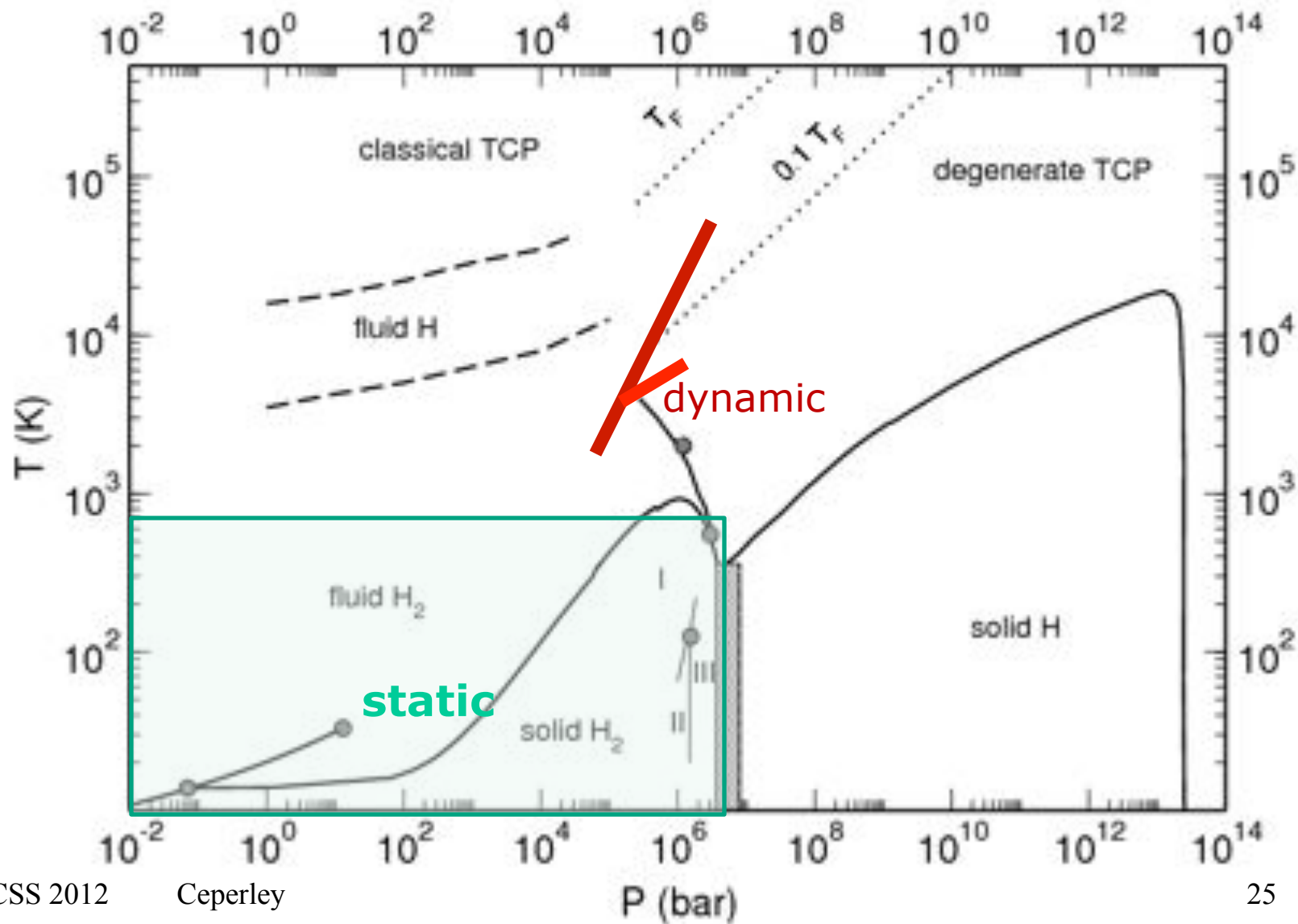
INCITE award for computer time



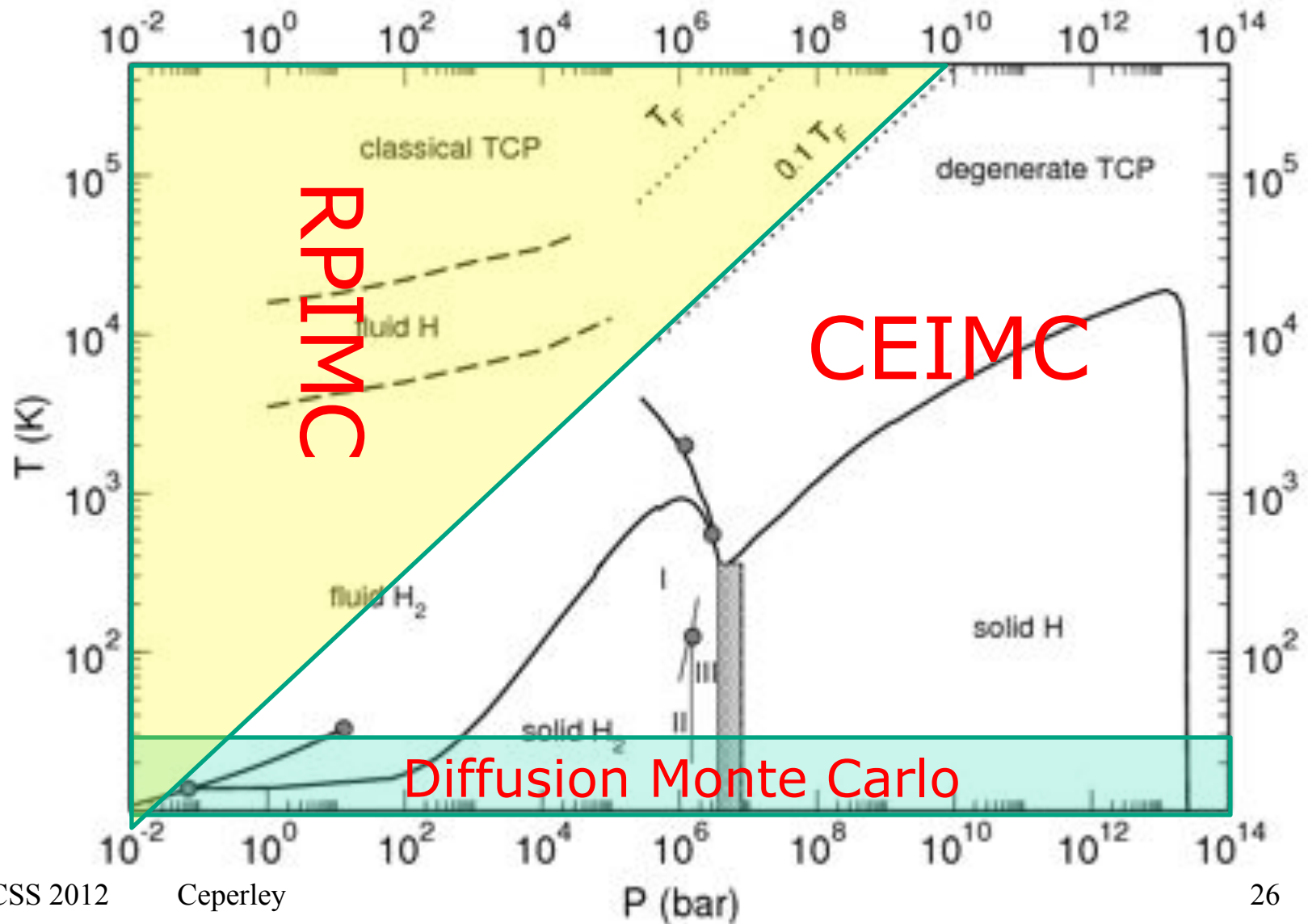
Why study dense Hydrogen & Helium?

- Hydrogen is the most abundant element in the universe, the major component of gas giants, yet there is very little understanding of its high-pressure phases
- Helium is the second most abundant element. What are the properties of H-He mixtures?
- Fundamental physics: unusual quantum states:
 - A metallic superfluid at 0K, the only known liquid metal at 0K
 - A superconducting superfluid, pairing of both protons and electrons
 - Solid H may be a very high- T_c superconductor
- Ideal test bed for algorithms. If we can't do dense hydrogen!!!

Hydrogen Phase Diagram: experiment



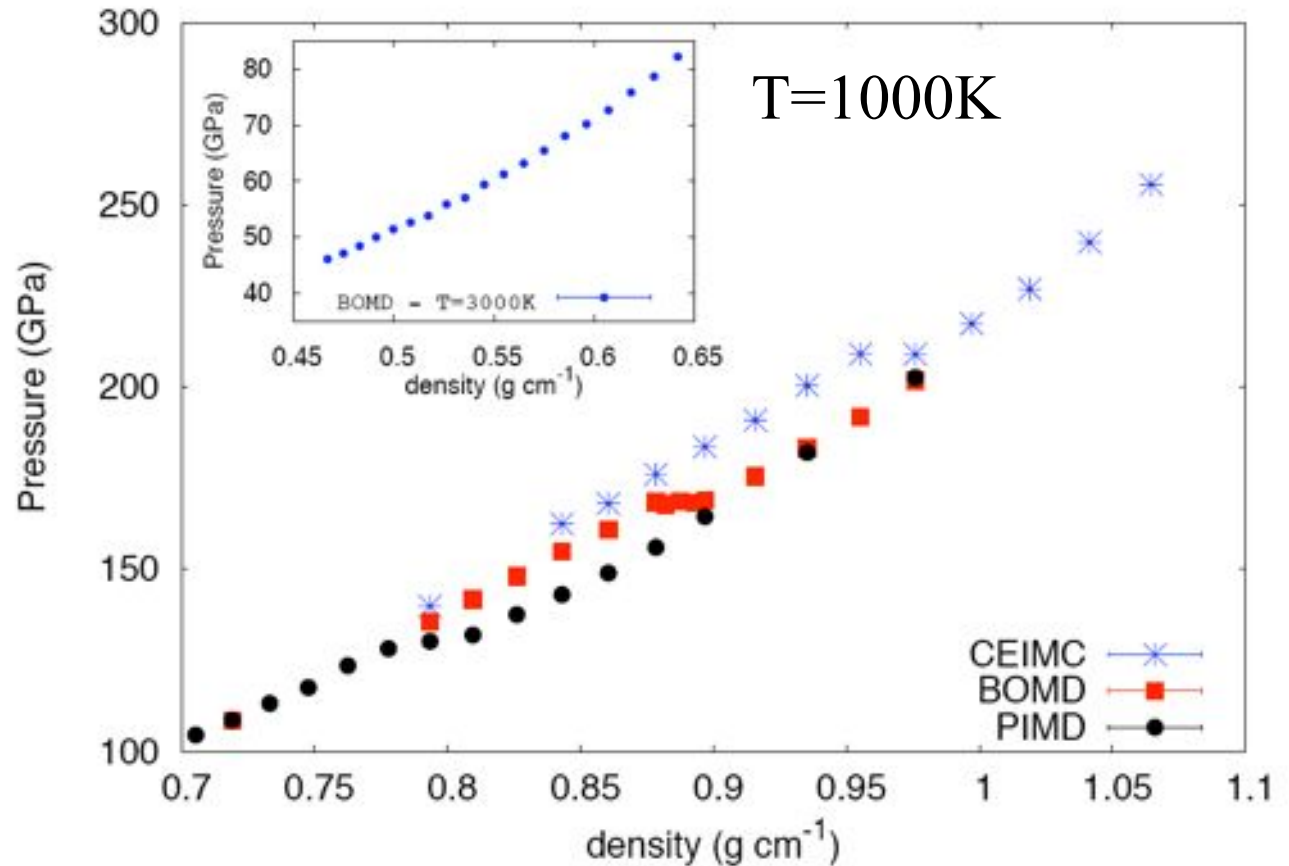
Regimes for Quantum Monte Carlo

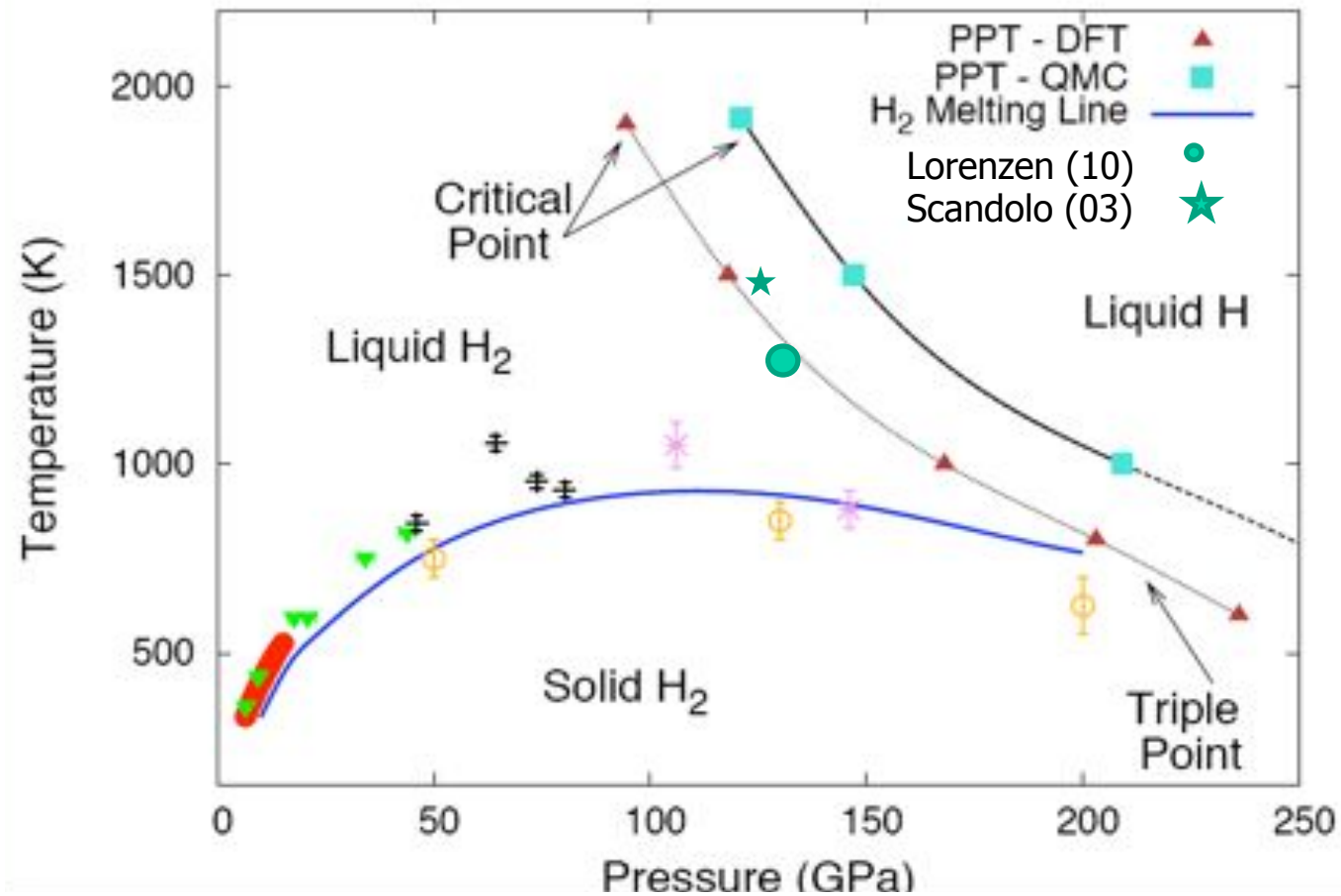


Liquid-Liquid Transition

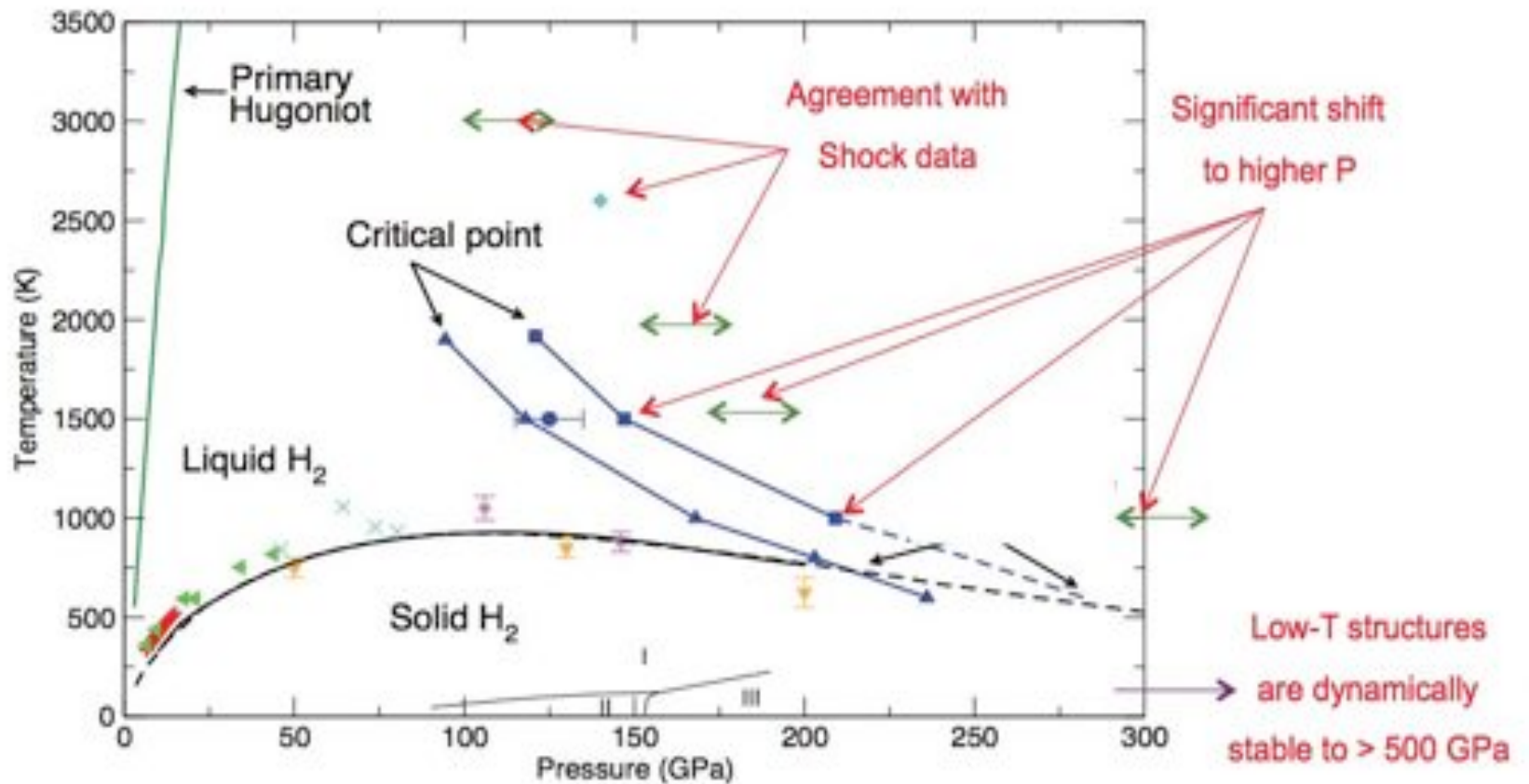
Morales, Pierleoni, Schwegler, DMC, PNAS 2010.

- Pressure plateau at low temperatures ($T < 2000\text{K}$) - signature of a 1st order phase transition
- Seen in CEIMC and BOMD at different densities
- Many previous results!
 - Finite size effects are very important
 - Narrow transition ($\sim 2\%$ width in V)
 - Low critical temperature



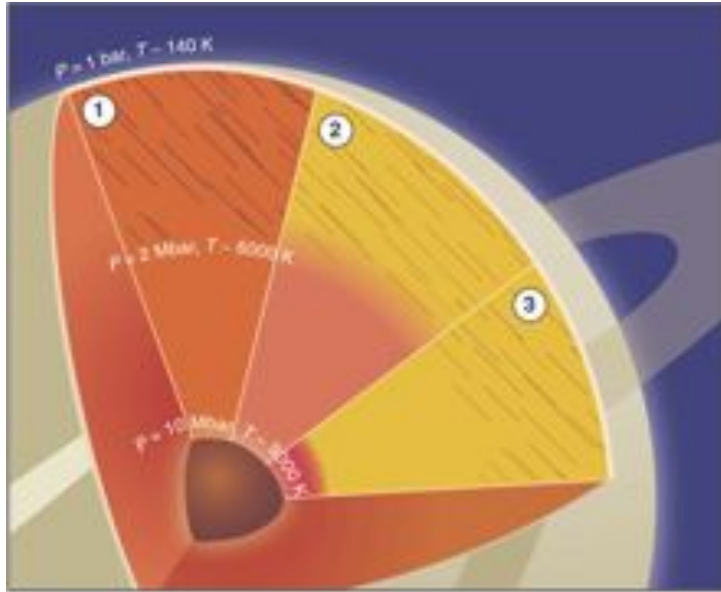


- DFT underestimates band gaps → transition too early.
- Liquid-liquid line for QMC with quantum protons lies near the DFT line.
- quantum protons ~ band gap problem of DFT
- *Can experiment see the triple point on the melting line?*



- DFT functional with van der Waals interaction pushes transition to higher pressures—in agreement with shock experiments
- Current work is to understand solid hydrogen phases at $T < 500\text{K}$.

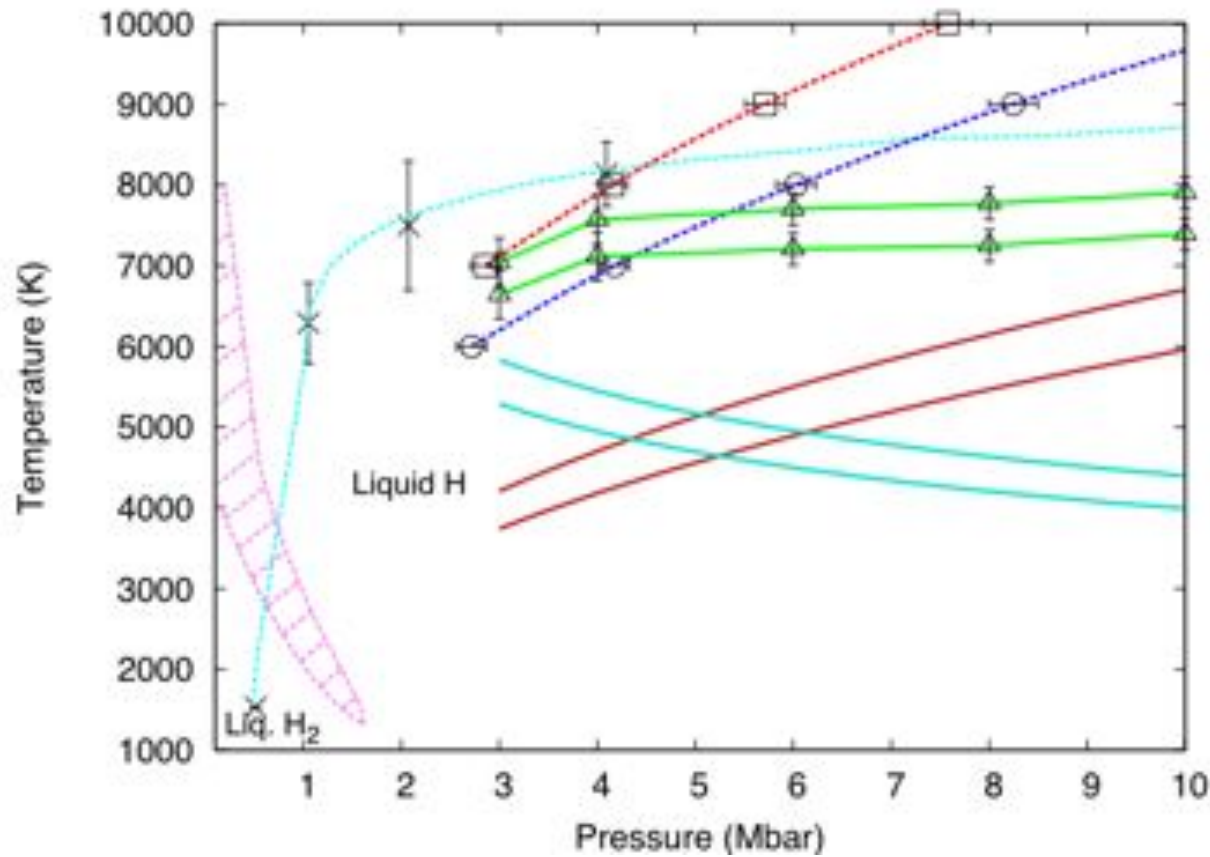
Hydrogen-Helium Miscibility



Taken from: Fortney J. J., *Science* **305**, 1414 (2004).

- Giant Planets
 - Primary components are H and He
 - $P(\rho, T, x_i)$ closes set of hydrostatic equations
 - Interior models depend very sensitively on EOS and phase diagram
- Saturn's Luminosity
 - Homogeneous evolutionary models do not work for Saturn
 - Additional energy source in planet's interior is needed
 - Does it come from Helium segregation (rain) as suggested by Smoluchowski (1965)

Mixing Phase Diagram



--- Jupiter Isentrope
--- Saturn Isentrope

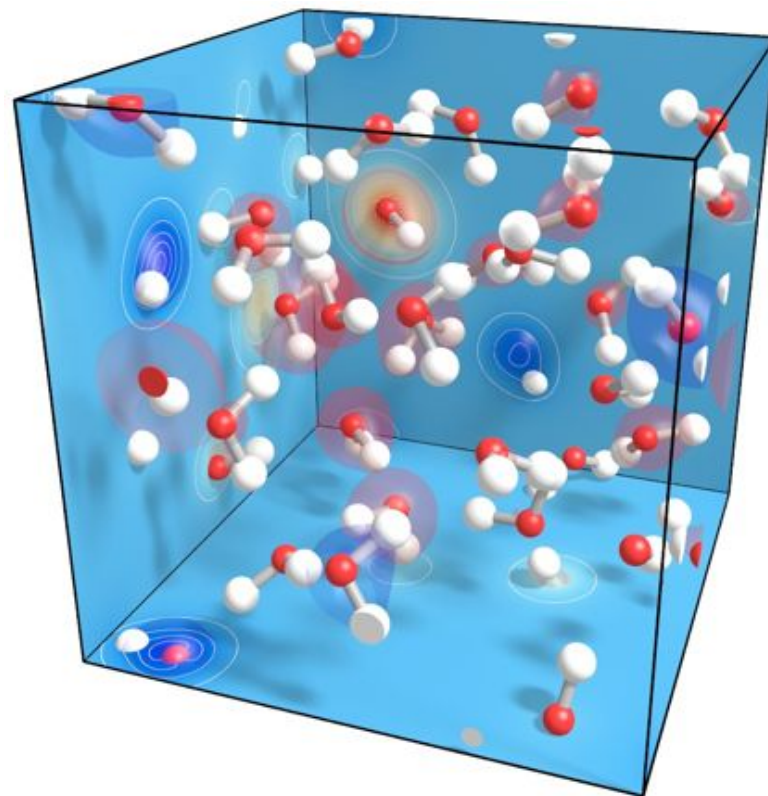
[$x=0.07&0.067$]

Demixing Temperatures:

— This work:
— Redmer, et al.
— Hubbard - DeWitt
— Pfaffenzeller, et al.

Simulation of water is a grand challenge

- H_2O will form 4 hydrogen bonds with tetrahedral ordering
- In the liquid, tetrahedral ordering is broken by thermal fluctuations.
- Random hydrogen bond network
- Frustrated system with many local minima
- Hindered rotations
- Quantum ZPM of protons important
- Current “force fields” and DFT functionals are inadequate.



Challenges for First Principles Simulation of Liquid Water

Long autocorrelation time

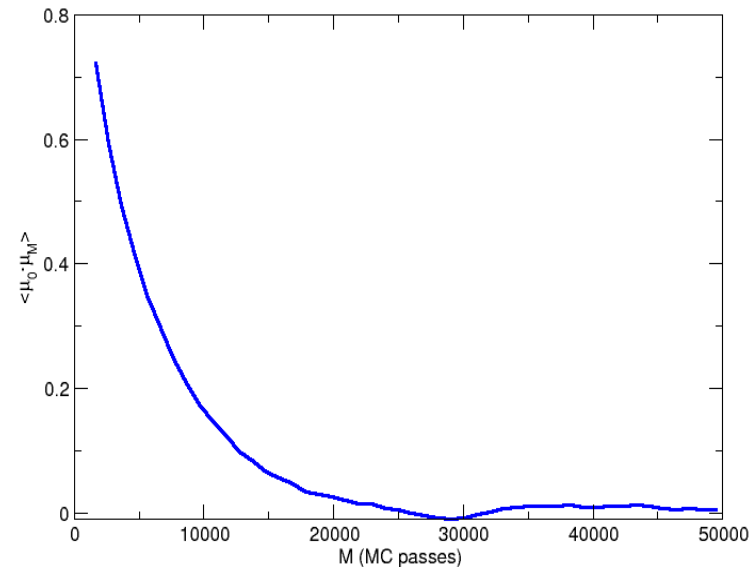
- order 1ps hydrogen bond lifetime
- May lead to unconverged simulations

Quantum Effects

- Zero point motion of protons is very important
- Standard path integral formulation increases computational cost by ~50

Electronic Structure

- How accurate is DFT-GGA?

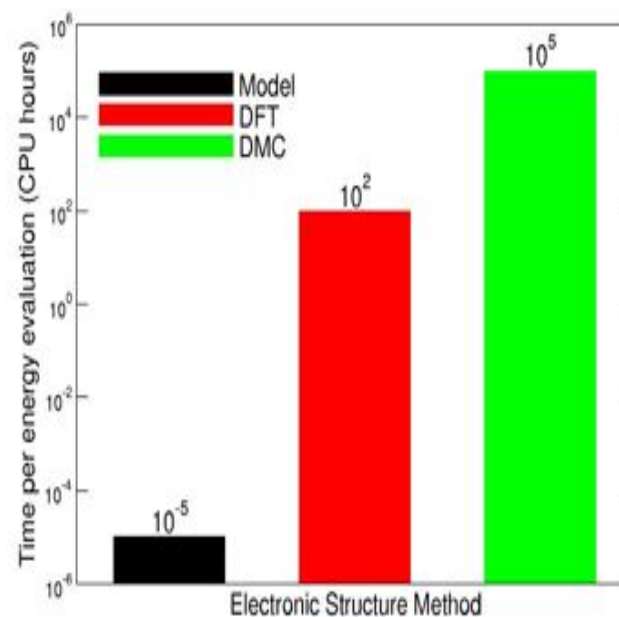
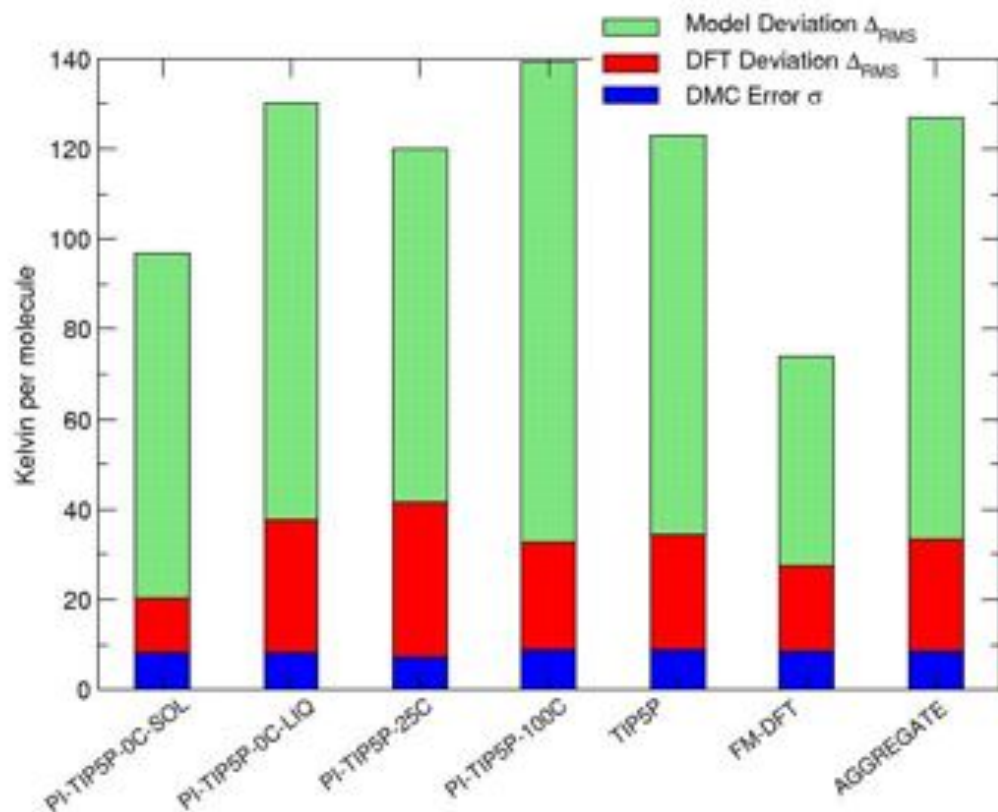


Obtaining uncorrelated samples requires of order 10000 MC passes (equivalent to 10 ps MD trajectory)

QMC Studies of Liquid Water

- Generated water configurations (mostly 32 molecules) using DFT and model potentials at several temperatures ($T = 0, 25, 100$ C, 1 atm)
- For each configuration, orbitals from self-consistent DFT-GGA (PBE) plus VMC optimization of correlated trial function.
- Ran Diffusion Monte Carlo for each configuration.
- Errors on energy for 32 molecules are less than 1 mHartree (or 9K/molecule). (part per million accuracy!)
- 50K hours of run time per configuration with several hundred configurations.
- Early access allocation on Jaguar-pf at Oak Ridge National Laboratory.

Comparison of Potential Energy Surfaces



- DMC RMS error $\sigma = 0.8$ mHartree, or 8 K/mol.
- Avg. DFT deviation: 33 K/mol. (4σ)
- Avg. model pot. deviation: 100 K/mol/ (12σ)

- Simulations needed for larger ice and water systems (96 molecules)
- Can we get the melting temperature right?
- Make database for water “energy surface” to construct new model for water
- Full *ab initio* Coupled Electron - Ion MC on water is very expensive but feasible: exascale level application.
- CEIMC will be an important application for petascale systems. It can provide necessary accuracy for disordered systems with ~ 1000 atoms.

Materials Design

- Problem: find a material having properties A, B and C and a low cost to manufacture.
- Combinatorics leads to a very large number of possible compounds to search. [$>92^k$] where k is complexity of unit cell.
- Replace more complex, unreliable algorithms with brute force or stochastic search (e.g. genetic algorithms).

Such studies are starting to take place

- Computational for each compound was a PhD thesis 15 years ago
- Now search among 10^5 combinations for the optimal material.
- But we need methods that properly treat correlation!

Plenty of work at the peta-scale and exa-scale

SUMMARY

- No existing methods are perfect but QMC today is competitive with other methods and usually much more accurate.
- Progress in *ab initio* simulations in last 40 years, coming from both
 - Computer power
 - Algorithmic power
- We are now in position to do much more accurate simulations
- Our petascale goal: water and transition metal oxides (e.g. MnO), ...
- Ab initio computation of electronic system is a great problem to solve. Intellectually **and** technologically very important. More work needed in algorithms to get higher accuracy, treat larger systems, heavier elements allowing:
 - benchmarking to validate cheaper approaches
 - replace more approximate approaches.