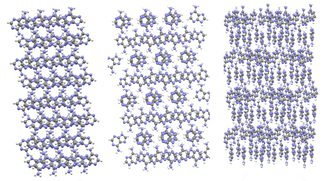
Periodic DFT calculations of vibrational and molecular dynamics on large organic molecular systems using OLCF computers

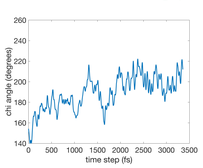
Ada Sedova[1]*, Anup Pandey[2] and Micholas Dean Smith[3]

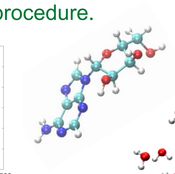
[1]Scientific Computing Group, NCCS, ORNL, [2]Vibrational Spectroscopy Group, Neutron Scattering Division, ORNL, [3]Center for Molecular Biophysics, UT/ORNL

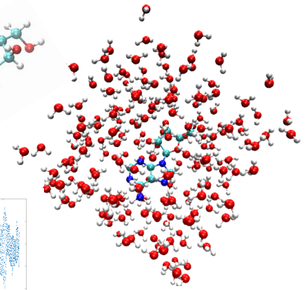
Abstract: Density functional theory (DFT) calculations on increasingly large systems have become possible in recent years with HPC-based parallel programs. Time-to-solution and scaling benchmarks for vibrational normal-mode calculations and molecular dynamics simulations using the periodic VASP and CP2K programs are presented. Accuracy of vibrational calculations is assessed using comparisons to incoherent neutron vibrational spectra collected on the VISION instrument at SNS. The first ab initio molecular dynamics (AIMD) simulations of fully solvated small biomolecules using the linear-scaling SCF routine in CP2K are also presented, and results are compared to classical molecular dynamics simulations. We show the possibility of AIMD as a new tool for biomolecular simulation with this increasing performance on HPC systems.

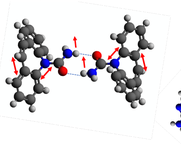
Introduction: DFT calculations provide a computationally less expensive method of solving the quantum many-body problem for a system of electrons than wave-function methods. Drawing on the the Hohenberg-Kohn theorems, and using the Born-Oppenheimer approximation, the time-independent Schrödinger equation is solved for the electrons in a set of atoms using functionals of the electron density, computed using single-electron non-interacting reference orbitals known as the Kohn-Sham orbitals. This greatly simplifies the calculation, but introduces an empirical element to the solution method because the form of functionals is not known and must be guessed. A first approximation uses a method known as the local density approximation (LDA), based on a uniform electron gas, and can achieve adequate results for some systems. The next level of approximation is known as the Generalized Gradient Approximation (GGA), which uses the density as well as the gradient of the density. The Perdew-Burke-Ernzerhof (PBE) functional is a popular GGA functional that has achieved better results for molecular bond energies and cohesive energies of solids than LDA. Periodic DFT codes approximate a bulk condensed phase like a crystal by using periodic boundary conditions to create an infinite system, via Fourier space methods. Many of these codes can be used to perform simulations of vibrational dynamics using a harmonic normal-mode analysis, and to perform molecular dynamics simulations using the Born-Oppenheimer description. These types of calculations do not require parameterization besides what is required by the functional forms, as opposed to classical models which use empirical energy functions parameterized by an extensive fitting procedure.

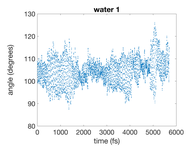


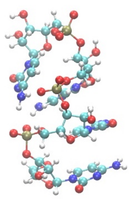


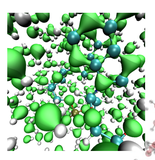


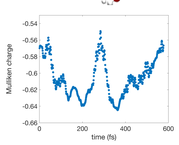


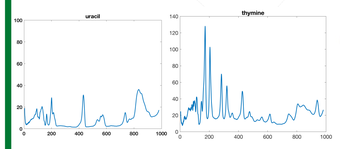


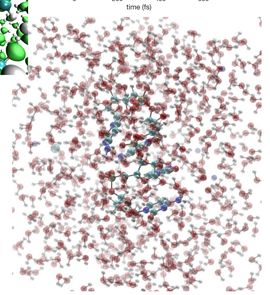


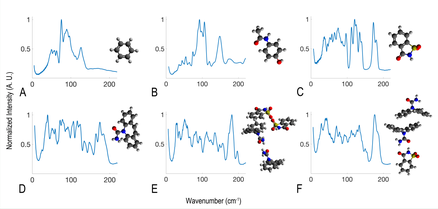


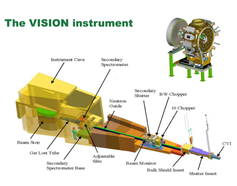




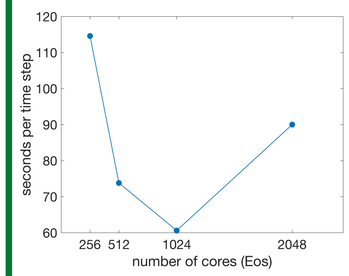


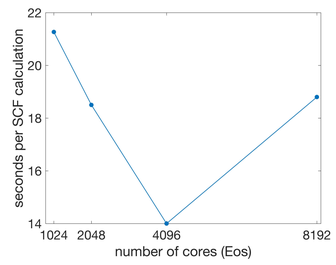


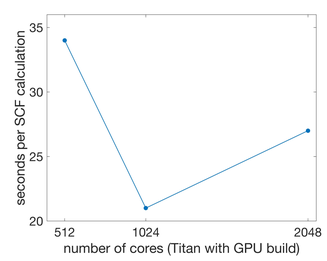




Performance of VASP and CP2K on OLCF computers: B-O molecular dynamics



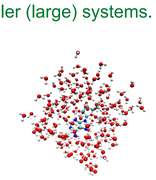


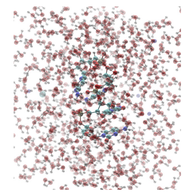


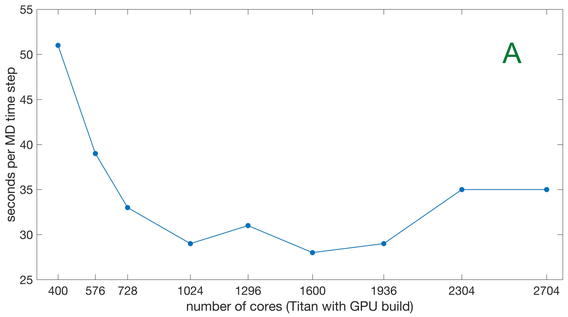
VASP MD: 480 atoms, 1600 electrons, PW cutoff 400 eV, GGA (PBE) functional, gamma-point, SCF cutoff 1E-7.

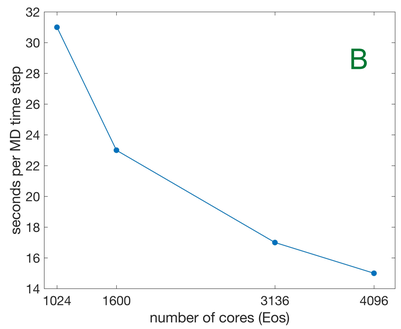
Published benchmark settings, CP2K MD with linear-scaling SCF calculation: 768 atoms, 2048 electrons, PW cutoff 300 Ry, DZVP Gaussian basis set, LDA (Padé) functional, gamma-point, SCF cutoff of 1E-4 and filter cutoff of 1E-5.

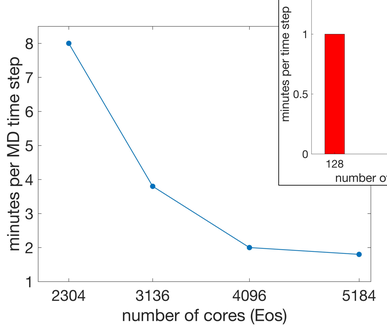
CP2K employs heavy use of scientific libraries like scalapack, fftw, and blacs, and performs well on hyper-threaded Intel Xeons, and when MKL with threaded fftw and blacs is available. Currently the GPU portion of the code seems to improve the performance and scaling less, for smaller (large) systems.

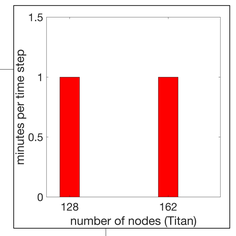






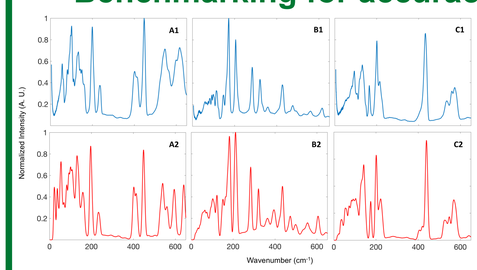




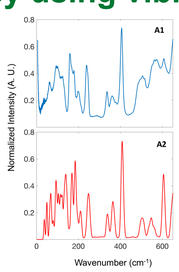


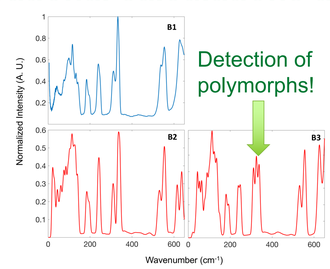
Actual production settings, 623 atom ribonucleotide system with 300 explicit waters, CP2K MD with linear-scaling SCF, using PBE, the D3 dispersion correction, a 400 Ry PW cutoff, DZVP, SCF cutoff 1E-5 and filter cutoff 1E-6, A: Titan, B; Eos, and 2207 atom tetranucleotide system with same settings, C and D. Cores used with D were 1024 and 1296, because the GPU build does not permit use of more than 8 cores per node.

Benchmarking for accuracy using vibrational calculations and simulated VISION spectra



Calculated vs. experimental low frequency INS spectra of pyrimidine bases. A1: cytosine, exp. A2: cytosine, calc. B1: thymine, exp. B2: thymine, calc. C1: uracil, exp. C2: uracil, calc.

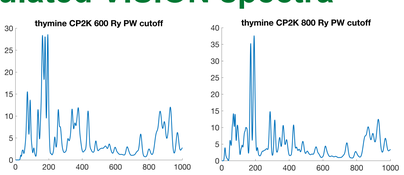




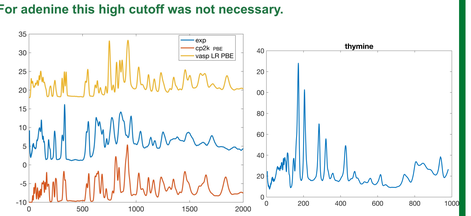
Detection of polymorphs!

Calculated vs. experimental low frequency INS spectra of purine bases. A1: guanine, exp. A2: guanine, calc. B1: adenine, exp. B2: adenine, polymorph I, calc. B3: adenine, polymorph II, calc.

VASP vibrational calculation using 400 eV PW cutoff and PBE functional followed by aCLIMAX post-processing to simulate VISION spectral measurements of nucleic acid bases.



For adenine this high cutoff was not necessary.



Finite-displacement Method Vibrational Calculations, 1600 electron system (adenine 211 supercell)	Hours	Cores	Core-hours
VASP (uses symmetry) with PHONOPY driver for parallelization	6.5	128	832
CP2K (does not use symmetry)	14.9	2048	30,515.2

CP2K vibrational calculations, using 600 and 800 Ry PW cutoff and PBE functional for thymine, vs. 500 PW cutoff for adenine, also followed by aCLIMAX post-processing. This is NOT using the linear scaling SCF with the lower tolerance settings. For the 800 Ry cutoff, ~49,000 core hours were required. Further testing of the CP2K LS-SCF-based molecular dynamics with reduced settings is necessary to determine which dynamical information is described correctly.

References

Marx D, Hutter J. Ab initio molecular dynamics: basic theory and advanced methods. Cambridge University Press; 2009 Apr 30.

Kresse G, Furthmüller J. VASP the Guide; Vienna University of Technology: Vienna, Austria, 2001.

Hutter J, Iannuzzi M, Schiffmann F, VandeVondele J. CP2K: atomistic simulations of condensed matter systems. Wiley Interdisciplinary Reviews: Computational Molecular Science. 2014 Jan 1;4(1):15-25.

VandeVondele J, Borstnik U, Hutter J. Linear scaling self-consistent field calculations with millions of atoms in the condensed phase. Journal of Chemical Theory and Computation. 2012 Mar 14;8(10):3565-73.

Borštnik U, VandeVondele J, Weber V, Hutter J. Sparse matrix multiplication: The distributed block-compressed sparse row library. Parallel Computing. 2014 May 1;40(5-6):47-58.

Schütt O, Messmer P, Hutter J, VandeVondele J. GPU accelerated sparse matrix matrix multiplication for linear scaling density functional theory. Electronic Structure Calculations on Graphics Processing Units: From Quantum Chemistry to Condensed Matter Physics. 2016 Apr 18:173-90.

Mitchell PC. Vibrational spectroscopy with neutrons: with applications in chemistry, biology, materials science and catalysis. World Scientific; 2005.

Ramirez-Cuesta AJ. aCLIMAX 4.0. 1, The new version of the software for analyzing and interpreting INS spectra. Computer Physics Communications. 2004 Mar 1;157(3):226-38.

This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR227525, was supported by the U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering, under Award DE-SC0010419, and made use of the VISION beamline at ORNL's Spallation Neutron Source, which is supported by the Scientific User Facilities Division, Office of Basic Energy Sciences (BES), U.S. Department of Energy (DOE), under Contract DE-AC0500OR22725 with UT Battelle, LLC.

OAK RIDGE National Laboratory

U.S. DEPARTMENT OF ENERGY