NWChem: Past, Present, and Future

E. Apra,[1] E. J. Bylaska,[1] W. A. de Jong,[2] N. Govind,[1] K. Kowalski,[1, a)] T. P. Straatsma,[3] M. Valiev,[1] H. J. J. van Dam,[4] Y. Alexeev,[5] J. Anchell,[6] V. Anisimov,[5] F. W. Aquino,[7] R. Atta-Fynn,[8] J. Autschbach,[9] N. P. Bauman,[1] J. C. Becca,[10] D. E. Bernholdt,[11] K. Bhaskaran-Nair,[12] S. Bogatko,[13] P. Borowski,[14] J. Boschen,[15] J. Brabec,[16] A. Bruner,[17] E. Cauet,[18] Y. Chen,[19] G. N. Chuev,[20] C. J. Cramer,[21] J. Daily,[1] M. J. O. Deegan,[22] T. H. Dunning Jr.,[23] M. Dupuis,[9] K. G. Dyall,[24] G. I. Fann,[11] S. A. Fischer,[25] A. Fonari,[26] H. Fruchtl,[27] L. Gagliardi,[21] J. Garza,[28] N. Gawande,[1] S. Ghosh,[29] K. Glaesemann,[1] A. W. Gotz,[30] J. Hammond,[6] V. Helms,[31] E. D. Hermes,[32] K. Hirao,[33] S. Hirata,[34] M. Jacquelin,[2] L. Jensen,[10] B. G. Johnson,[35] H. Jonsson,[36] R. A. Kendall,[11] M. Klemm,[6] R. Kobayashi,[37] V. Konkov,[38] S. Krishnamoorthy,[1] M. Krishnan,[19] Z. Lin,[39] R. D. Lins,[40] R. J. Little eld,[41] A. J. Logsdail,[42] K. Lopata,[43] W. Ma,[44] A. V. Marenich,[45] J. Martin del Campo,[46] D. Mejia-Rodriguez,[47] J. E. Moore,[6] J. M. Mullin,[48] T. Nakajima,[49] D. R. Nascimento,[1] J. A. Nichols,[11] P. J. Nichols,[50] J. Nieplocha,[1] A. Otero-de-la-Roza,[51] B. Palmer,[1] A. Panyala,[1] T. Pirojsirikul,[52] B. Peng,[1] R. Peverati,[38] J. Pittner,[53] L. Pollack,[54] R. M. Richard,[55] P. Sadayappan,[56] G. C. Schatz,[57] W.A. Shelton,[58] D. W. Silverstein,[59] D. M. A. Smith,[6] T. A. Soares,[60] D. Song,[1] M. Swart,[61] H. L. Taylor,[62] G. S. Thomas,[1] V. Tipparaju,[63] D. G. Truhlar,[21] K. Tsemekhman,[64] T. Van Voorhis,[65] A. Vazquez-Mayagoitia,[5] P. Verma,[66] O. Villa,[67] A. Vishnu,[1] K. D. Vogiatzis,[68] D. Wang,[69] J. H. Weare,[70] M. J. Williamson,[71] T. L. Windus,[72] K. Wolinski,[14] A. T. Wong,[73] Q. Wu,[4] C. Yang,[2] Q. Yu,[74] M. Zacharias,[75] Z. Zhang,[76] Y. Zhao,[77] and R. J. Harrison[78]

[1)]Pacific Northwest National Laboratory, Richland, WA 99352, USA

[2)] Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA

[3)]National Center for Computational Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[4)]Brookhaven National Laboratory, Upton, NY 11973, USA

[5)] Argonne Leadership Computing Facility, Argonne National Laboratory, Argonne, IL 60439,

[6)]Intel Corporation, Santa Clara, CA 95054, USA

[7)]QSimulate, Cambridge, MA 02139, USA

[8)]Department of Physics, The University of Texas at Arlington, Arlington, TX 76019, USA

[9)]Department of Chemistry, University at Buffalo, State University of New York, Buffalo, NY 14260, USA

[10)]Department of Chemistry, The Pennsylvania State University, University Park, PA 16802, USA

[11)]Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA

[12)]Washington University, St. Louis, MO 63130, USA

[13)]4G Clinical, Wellesley, MA 02481, USA

[14)]Faculty of Chemistry, Maria Curie-Skłodowska University in Lublin, 20-031 Lublin, Poland

[15)]Department of Chemistry, Iowa State University, Ames, IA 50011, USA

[16)]J. Heyrovský Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, 18223 Prague 8, Czech Republic

[17)]Department of Chemistry and Physics, University of Tennessee at Martin, Martin, TN 38238, USA

[18)]Service de Chimie Quantique et Photophysique (CP 160/09), Université libre de Bruxelles, B-1050 Brussels, Belgium

[19)]Facebook, Menlo Park, CA 94025, USA

[20)]Institute of Theoretical and Experimental Biophysics, Russian Academy of Science, Pushchino, Moscow Region, 142290, Russia

[21)]Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA

[22)]SKAO, Jodrell Bank Observatory, Macclesfield, SK11 9DL, United Kingdom

[23)] Department of Chemistry, University of Washington, Seattle, WA 98195, USA

[24)] Dirac Solutions, Portland, OR 97229, USA

[25)]Chemistry Division, U. S. Naval Research Laboratory, Washington, DC 20375, USA

26)School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA 30332, USA[b)]

[27)]EaStCHEM and School of Chemistry, University of St. Andrews, St. Andrews KY16 9ST, United Kingdom

[28)]Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, Col. Vicentina, Iztapalapa, C.P. 09340 Ciudad de México, Mexico

29)Department of Chemistry, Chemical Theory Center, and Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA[c)]

[30]San Diego Supercomputer Center, University of California, San Diego, La Jolla, CA 92093, USA

[31]Center for Bioinformatics, Saarland University, D-66041 Saarbrucken,¨ Germany
[32] Combustion Research Facility, Sandia National Laboratories, Livermore, CA 94551, USA
[33]Next-generation Molecular Theory Unit, Advanced Science Institute, RIKEN, Saitama 351-0198, Japan

[34]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[35]Acrobatiq, Pittsburgh, PA 15206, USA
[36]Faculty of Physical Sciences, University of Iceland, Iceland and Department of Applied Physics, Aalto University, FI-00076 Aalto, Espoo, Finland
[37]ANU Supercomputer Facility, Australian National University, Canberra, Australia
[38] Chemistry Program, Florida Institute of Technology, Melbourne, FL 32901, USA
[39]Department of Physics, University of Science and Technology of China, Hefei, China

[40]Aggeu Magalhaes Institute, Oswaldo Cruz Foundation, Recife, Brazil
[41] Zerene Systems LLC, Richland, WA 99354, USA
[42] Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Cardiff, Wales, CF10 3AT,

[43]Department of Chemistry, Louisiana State University, Baton Rouge, LA 70803, USA

[44]Institute of Software, Chinese Academy of Sciences, Beijing, China
45)Department of Chemistry,Chemical Theory Center, and Supercomputing Institute, University of Minnesota, Minneapolis, MN 55455, USA[d)]

[46]Departamento de F´ısica y Qu´ımica Teorica,´ Facultad de Qu´ımica, Universidad Nacional Autonoma´ de Mexico,´ Mexico´ City, Mexico

[47]Quantum Theory Project, Department of Physics, University of Florida, Gainesville, FL 32611, USA
[48]DCI-Solutions, Aberdeen Proving Ground, MD 21005, USA
[49]Computational Molecular Science Research Team, RIKEN Center for Computational Science, Kobe, Hyogo, 650-0047, Japan

[50]Los Alamos National Laboratory, Los Alamos, NM 87545, USA
[51]Departamento de Qu´ımica F´ısica y Anal´ıtica and MALTA-Consolider Team, Facultad de Qu´ımica, Universidad de Oviedo, 33006 Oviedo, Spain
[52]Department of Chemistry, Prince of Songkla University, Hat Yai, Songkhla, 90112,

[53]J. Heyrovsky´ Institute of Physical Chemistry, Academy of Sciences of the Czech Republic, v.v.i., 18223 Prague 8, Czech Republic

[54]StudyPoint, Boston, MA 02114, USA
[55] Ames Laboratory, Ames, IA 50011, USA
[56] School of Computing, University of Utah, Salt Lake City, UT 84112, USA
[57] Department of Chemistry, Northwestern University, Evanston, IL 60208, USA
[58] Cain Department of Chemical Engineering, Louisiana State University, Baton Rouge, LA 70803,

[59]Universal Display Corporation, Ewing, NJ 08618, USA
[60]Dept. of Fundamental Chemistry, Universidade Federal de Pernambuco, Recife, Brazil

[61]ICREA, 08010 Barcelona; Universitat Girona, Institut de Qu´ımica Computacional i Catalisi,` Campus Montilivi, 17003 Girona, Spain

[62]CD-adapco/Siemens, Melville, NY 11747, USA
[63] Cray Inc., Bloomington, MN 55425, USA
[64] Gympass, New York, NY 10013, USA
[65]Department of Chemistry, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[66]1QBit, Vancouver, BC V6E 4B1, Canada
[67] NVIDIA, Santa Clara, CA 95051, USA
[68] Department of Chemistry, University of Tennessee, Knoxville, TN 37996, USA
[69]College of Physics and Electronics, Shandong Normal University, Jinan, Shandong 250014, China

[70]Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093,

[71]Department of Chemistry, Cambridge University, Lensfield Road, Cambridge CB2 1EW, United Kingdom

[72]Department of Chemistry, Iowa State University and Ames Laboratory, Ames, IA 50011, USA

[73]Qwil, San Francisco, CA 94107, USA
[74] AMD, Santa Clara, CA 95054, USA
[75] Department of Physics, Technical University of Munich, 85748 Garching, Germany
[76]Stanford Research Computing Center, Stanford University, Stanford, CA 94305, USA
[77]State Key Laboratory of Silicate Materials for Architectures, International School of Materials Science and Engineering, Wuhan University of Technology, Wuhan 430070, China
[78]Institute for Advanced Computational Science, Stony Brook University, Stony Brook, NY 11794, USA

(Dated: May 12, 2020)

Specialized computational chemistry packages have permanently reshaped the landscape of chemical and materials science by providing tools to support and guide experimental efforts and for the prediction of atomistic and electronic properties. In this regard, electronic structure packages have played a special role by using first-principle-driven methodologies to model complex chemical and materials processes. Over the last few decades, the rapid development of computing technologies and the tremendous increase in computational power have offered a unique chance to study complex transformations using sophisticated and predictive many-body techniques that describe correlated behavior of electrons in molecular and condensed phase systems at different levels of theory. In enabling these simulations, novel parallel algorithms have been able to take advantage of computational resources to address the polynomial scaling of electronic structure methods. In this paper, we briefly review the NWChem computational chemistry suite, including its history, design principles, parallel tools, current capabilities, outreach and outlook.

I. INTRODUCTION

The NorthWest Chemistry (NWChem) modeling software is a popular computational chemistry package that has been designed and developed to work efficiently on massively parallel processing supercomputers[1–3]. It contains an umbrella of modules that can be used to tackle most electronic structure theory calculations being carried out today. Since 2010, the code is distributed as open-source under the terms of the Educational Community License version 2.0 (ECL 2.0).

Electronic structure theory provides a foundation for our understanding of chemical transformations and processes in complex chemical environments. For this reason, accurate electronic structure formulations have already permeated several key areas of chemistry, biology, biochemistry, and materials sciences, where they have become indispensable elements for building synergies between theoretical and experimental efforts and for predictions. Over the last few decades, intense theoretical developments have resulted in a broad array of electronic structure methods and their implementations, designed to describe structures, interactions, chemical reactivity, dynamics, thermodynamics, and spectral properties of molecular and material systems. The success of these computational tools hinges upon several requirements regarding the accuracy of many-body models, reliable algorithms for dealing with processes at various spatial and temporal scales, and effective utilization of ever-growing computational resources. For instance, the predictive power of computational chemistry requires sophisticated quantum mechanical approaches that systematically account for electronic correlation effects. Therefore, the design of versatile electronic structure codes is a major undertaking that requires close collaboration between experts in theoretical and computational chemistry, applied mathematics, and computer science.

NWChem [2–8], like other widely used electronic structure programs, was developed to fully realize the potential of computational modeling to answer key scientific questions. It provides a wide range of capabilities that can be deployed on supercomputing platforms to solve two fundamental equations of quantum mechanics[9–11] - time-independent and time-dependent Schrödinger equations:

$$H|\Psi\rangle = E|\Psi\rangle\,; \qquad (1)$$

$$i\hbar\frac{\partial|\Psi\rangle}{\partial t} = H|\Psi\rangle\,; \qquad (2)$$

and a fundamental equation of Newtonian mechanics

$$m_i a_i = F_i\,; \qquad (3)$$

where forces F_i include information about quantum effects. Given the breadth of electronic structure theory, it does not come as a surprise that equations (1)-(2) can be solved using various representations of quantum mechanics employing wavefunctions ($|\Psi\rangle$), electron densities ($\rho(\vec{r})$), or self-energies ($\Sigma(\omega)$), which comprise the wide spectrum of NWChem's functionalities to compute the electronic wave-functions, densities, and associated properties of molecular and periodic systems. These functionalities include Hartree-Fock[12–15] self-consistent field (SCF) and post-SCF correlated many-body approaches that build on the SCF wavefunction to tackle static and dynamic correlation effects. Among correlated approaches, NWChem offers second-order Møller-Plesset perturbation theory; single-

a) Electronic mail: karol.kowalski@pnnl.gov
b) Current Affiliation: Schrödinger, Inc, New York, NY 10036, USA
c) Current Affiliation: Max Planck Institute für Kohlenforschung, 45470 Mülheim an der Ruhr, Germany
d) Current Affiliation: Gaussian Inc., Wallingford, CT 06492, USA

and multi-reference, ground- and excited-state, and linear-response coupled-cluster (CC) theories; multi-configuration self-consistent field (MCSCF); and selected and full configuration interaction (CI) codes. NWChem provides exten-sive density functional theory[16–18] (DFT) capabilities with Gaussian and plane-wave basis set implementations. Within the Gaussian basis set framework, a broad range of DFT response properties, ground and excited-state molecular dynamics, linear-response (LR) and real-time (RT) time-dependent density functional theory (TDDFT) are available. The plane-wave DFT implementations offer the capability to run scalable ab initio and Car-Parrinello molecular dynamics[19], and band-structure simulations. The plane-wave code supports both norm-conserving[20–22] and projector augmented wave (PAW)[23] pseudopotentials.

For all DFT methods outlined above, both analytical or numerical gradients and Hessians are available for geometry optimization and vibrational analysis. Addition-ally, NWChem is capable of performing classical molec-ular dynamics (MD) simulations using either AMBER or CHARMM force fields. Through its modular design, the ab initio methods can be coupled with the classical MD to per-form mixed quantum mechanics and molecular mechanics simulations (QM/MM). Various solvent models and rela-tivistic approaches are also available, with the spin-orbit contribution being only supported at the Hartree-Fock (HF) and DFT levels of theory and associated response proper-ties. The NWChem functionality described is only a subset of its full capabilities. We refer the reader to the NWChem website[8] to learn about the full suite of functionalities avail-able to the user community.

Currently, NWChem is developed and maintained primarily by researchers at the Department of Energy (DOE) Pacific Northwest National Laboratory (PNNL), with help from researchers at other research institutions. It has a broad user base, and it is being used across the national laboratory system and throughout academia and industry around the world. In this paper, we provide a high-level overview of NWChem's core capabilities, recent developments in electronic methods, and a short discussion of ongoing and future efforts. We also illustrate the strengths of NWChem stemming from the possibility of seamless integration of methodologies at various scales and review scientific results that would not otherwise be obtainable without using its highly-scalable implementations of electronic structure methods.

II. BRIEF HISTORY

The NWChem project[1–7,24,25] started in 1992. It was originally designed and implemented as part of the construction project associated with the EMSL user facility at PNNL. Therefore, the software project started around four years before the EMSL computing center was up and running. This raised challenges for the software developers working on the project, such as predicting the features of future hardware architectures and how to deliver high performing software, while maintaining programmer productivity. Overcoming these challenges led to a design effort that strove for flexibility and extensibility, as well as high-level interfaces to functionality that hid some of the hardware issues from the chemistry software application developer. Over the years, this design and implementation have successfully advanced multiple science agendas, and NWChem's extensive code base of more than 2 million lines provides high-performance, scalable software code with advanced scientific capabilities that are used through-out the molecular sciences community.

NWChem is an example of a co-design effort harnessing the expertise of researchers from multiple scientific disciplines to provide users with computational chemistry tools that are scalable both in their ability to treat large scientific computational chemistry problems efficiently and in their use of computing resources from high-performance parallel supercomputers to conventional workstation clusters. In particular, NWChem has been designed to handle

- biomolecules, nanostructures, interfaces, and solid-state,
- chemical processes in complex environments,
- hybrid quantum/classical simulations,
- ground and excited-states and non-linear optical prop-erties,
- simulations of UV-Vis, photo-electron, X-ray spec-troscopies,
- Gaussian basis functions or plane-waves,
- ab-initio molecular dynamics on the ground and ex-cited states,
- relativistic effects.

The scalability of NWChem has provided a computational platform to deliver new scientific results that would be unobtainable if parallel computational platforms were not used. For example, NWChem's implementation of a non-orthogonally spin adapted CCSD(T) method has been demonstrated to scale to 210,000 processors available at the Oak Ridge National Laboratory's (ORNL) Leadership Com-puting Facilities,[26–28] whereas the plane-wave DFT code has been able to utilize close to 100,000 processor cores on NERSC's Cray-XE6 supercomputer.[29] Although imple-mented only for the perturbative part of coupled-cluster with singles and doubles (CCSD)[30] and triples correction (CCSD(T)),[31] NWChem was one of the first computational chemistry codes to have been ported to utilize graphics pro-cessing units (GPUs).[32] Several parts of the code have also been rewritten to take advantage of the Intel Xeon Phi fam-ily of processors - good scalability and performance have been demonstrated for the ab initio molecular dynamics plane-wave DFT code on the most recent Knights Landing version of the processor.[33,34] The non-iterative triples part of the CCSD(T) method has been demonstrated to scale to 55,200 Intel Phi threads and 62,560 cores through concur-rent utilization of CPU and Intel Xeon Phi Knights Corner accelerators.[35]

III. DESIGN PRINCIPLES

NWChem has a five-tiered modular architecture. The first tier is the Generic Task Interface. This interface (an abstract programming interface, not a user interface) serves as the mechanism that transfers control to the different modules in the second tier, which consists of the Molecular Calculation Modules. The molecular calculation modules are the high-level programming modules that accomplish computational tasks, performing particular operations using the specified theories defined by the user in the input file. These independent modules of NWChem share data only through a disk-resident database, which allows modules to share data or to share access to files containing data. The third tier consists of the Molecular Modeling Tools. These routines provide basic chemical functionality such as symmetry, basis sets, grids, geometry, and integrals. The fourth tier is the Software Development Toolkit, which is the basic foundation of the code. The fifth tier provides the Utility Functions needed by nearly all modules in the code. These include such functionality as input processing, output processing, and timing. [topsep=0pt, partopsep=0pt] The Generic Task Interface controls the execution of NWChem. The flow of control proceeds in the following steps:

1. Identify and open the input file.
2. Complete the initialization of the parallel environ-ment.
3. Process start-up directives.
4. Summarize start-up information and write it to the output file.
5. Open the run-time database.
6. Process the input sequentially (ignoring start-up di-rectives), including the first task directive.
7. Execute the task.
8. Repeat steps 6 and 7 until reaching the end of the input file or encountering a fatal error condition.

The input parser processes the user's input file and trans-lates the information into a form meaningful to the main program and the driver routines for specific tasks.

As mentioned in step 5 of the task flow control, NWChem makes use of a run-time database to store the main computational parameters. This is in the same spirit of check-pointing features available in other quantum chem-istry codes. The information stored in the run-time database can be used at a later time in order to restart a calcula-tion. Restart capabilities are available for most modules. For example, SCF generated files (run-time database and molecular orbitals) can be used either to continue a geome-try optimization or to compute molecular properties. The important second and fourth tiers are discussed as part of the subsequent sections.

IV. PARALLEL TOOLS

The design and early development of Global Arrays[36–39] (GA) toolkit occurred in the same period when the NWChem project started. The GA toolkit, which is the central component of the Software Development Toolkit, was adopted by the NWChem developers as the main ap-proach for the parallelization of the dense matrices present in quantum chemistry methods that make use of local basis functions. In current computer science parlance, Global Ar-rays can be viewed as a Partitioned Global Address Space (PGAS) model that provides a high level of abstraction for the programmer to the dense distributed arrays. In contrast to message passing constructs such as MPI, where the de-veloper has to worry about coordinating send and receive operations, the use of Global Arrays in NWChem requires so-called single-sided functions (e.g. put, get, accumulate) to manipulate data structures in a single operation. The choice of distribution model for sharing a given global ar-ray among the memory available to the processes in use plays a crucial role in efficient parallelization at large scale.

The GA toolkit has been ported to a variety of parallel computer architectures. The porting process has focused in the past in optimizing the ARMCI[40] library. The Aggregate Remote Memory Copy (ARMCI) library optimizes perfor-mance by fully exploiting network characteristics such as latency, bandwidth, and packet injection rate through the use of low-level network protocols (e.g. Infiniband Verbs). More recent porting options make use either of ComEx[41] or of the ARMCI-MPI[42] communication runtimes. Both ComEx and ARMCI-MPI make use of MPI libraries, in-stead of low-level network protocols, albeit with different approaches.

V. MAIN METHODOLOGIES

In this section, we describe the key methods that comprise the Molecular Calculation Modules. We first describe the Gaussian basis HF and DFT implementa-tions for molecular systems. This is followed by the post-SCF wavefunction-based perturbative (MP2), multi-configuration SCF, and high-accuracy (coupled-cluster the-ory) approaches for molecules, including the tensor con-traction engine (TCE). Molecular response properties and relativistic approaches are then described. The plane-wave based DFT implementation for Car-Parrinello molecular dy-namics and periodic condensed phase systems is described next, followed by classical molecular dynamics and hybrid methods.

A. Hartree-Fock

The NWChem SCF module computes closed-shell re-stricted Hartree-Fock (RHF) wavefunctions, restricted high-spin open-shell Hartree-Fock (ROHF) wavefunctions, and spin-unrestricted Hartree-Fock (UHF) wavefunctions. The Hartree-Fock equations are solved using a conjugate gradi-ent method with an orbital Hessian based preconditioner[43].

The most expensive part to compute in the SCF code is the two-electron contribution to the matrix element of the Fock operator (resulting from the sum of Coulomb and Exchange operators). To compute these matrix elements, NWChem developers have implemented parallel algorithms using either a distributed data approach[44] (where the Fock matrix is distributed among the aggregate memory of the processes involved in the calculation) or a replicated data approach (where an entire copy of the Fock matrix is stored in memory of each process).

Several options are available for the initial guess of the

SCF calculations. The default choice uses the eigenvectors of a Fock-like matrix formed from a superposition of the atomic densities. Other options include the use of eigen-vectors of the bare-nucleus Hamiltonian or the one-electron Hamiltonian, the projections of molecular orbital from a smaller basis to a larger one, or molecular orbitals formed by superimposing the orbitals of fragments of the molecule being studied. Symmetry can be used to speed up the Fock matrix construction via the petite-list algorithm. Molec-ular orbitals are symmetry adapted as well in NWChem. The resolution of the identity (RI) four-center, two-electron integral approximation has also been implemented.[45]

In order to avoid full matrix diagonalization, the SCF program uses a preconditioned conjugate gradient (PCG) method that is unconditionally convergent. Basically, a search direction is generated by multiplying the orbital gradient (the derivative of the energy with respect to the orbital rotations) by an approximation to the inverse of the level-shifted orbital Hessian. In the initial iterations, an inexpensive one-electron approximation to the inverse orbital Hessian is used. Closer to convergence, the full orbital Hessian is used, which should provide quadratic convergence. For both the full or one-electron orbital Hes-sians, the inverse-Hessian matrix-vector product is formed iteratively. Subsequently, an approximate line search is performed along the new search direction.

Both all-electron basis sets and effective core potentials (ECPs) can be used. Effective core potentials are a use-ful means of replacing the core electrons in a calculation with an effective potential, thereby eliminating the need for the core basis functions, which usually require a large set of Gaussians to describe them. In addition to replacing the core, they may be used to represent relativistic effects, which will be discussed later.

B. Density Functional Theory

The NWChem DFT module for molecular systems uses a Gaussian basis set to compute closed- and open-shell den-sities and Kohn-Sham orbitals in the local density approxi-mation (LDA), generalized gradient approximation (GGA), t-dependent and Laplacian-dependent meta-generalized gradient approximation (metaGGA), any combination of local and non-local approximations (including exact ex-change and range-separated exchange), and asymptotically corrected exchange-correlation potentials. NWChem con-tains energy-gradient implementations of most exchange-correlation functionals available in the literature, includ-ing a flexible framework to combine different functionals. However, second derivatives are not supported for meta-functionals and third derivatives are supported only for a selected set of functionals. For a detailed description, we refer the reader to the online documentation[46].

The DFT module reuses elements of the Gaussian ba-sis SCF module for the evaluation of the Hartree-Fock exchange and of the Coulomb matrices by using 4-index 2-electron electron repulsion integrals; the formal scaling of the DFT computation can be reduced by choosing to use auxiliary Gaussian basis sets to fit the charge density[47] and use 3-index 2-electron integrals, instead.

The DFT module supports both a distributed data ap-proach and a mirrored arrays[48] approach for the evaluation of the exchange-correlation potential and energy. The mir-rored arrays option, used by default, allows the calculation to hide network communication overhead by replicating the data between processes belonging to the same network node.

In analogy with what is available in the SCF module, the DFT module can perform restricted closed-shell, unre-stricted open-shell, and restricted open-shell calculations. However, in contrast to the SCF module that uses PCG to solve the SCF equation, the DFT module implements di-agonalization with parallel eigensolvers.[49–54] DIIS (direct inversion in the iterative subspace or direct inversion of the iterative subspace)[55], level-shifting[56,57] and density matrix damping can be used to accelerate the convergence of the iterative SCF process. Another technique that can be used to help SCF convergence makes use of electronic smearing of the molecular orbital occupations, by using a gaussian broadening function following the prescription of Warren and Dunlap[58]. Additionally, calculations with fractional numbers of electrons can be performed to analyze the be-havior of exchange-correlation functionals and their impact on molecular excited states and response properties.[59–66]

The Perdew and Zunger[67] method to remove the self-interaction contained in many exchange-correlation func-tionals has been implemented[68] within the Optimized Ef-fective Potential (OEP) method[69,70] and within the Krieger-Li-Iafrate (KLI) approximation.[71,72]

The asymptotic region of the exchange-correlation po-tential can be modified by the van-Leeuwen-Baerends exchange-correlation potential that has the correct $\frac{1}{r}$ asymptotic behavior. The total energy is then computed using the definition of the exchange-correlation functional. This scheme is known to tend to over-correct the deficiency of most uncorrected exchange-correlation potentials[73,74] and can improve TDDFT-based excitation calculations, but it is not variational. A variationally consistent ap-proach to address this issue is via range-separated exchange-correlation functionals and the recently developed nearly correct asymptotic potential or NCAP[75], which are imple-mented in NWChem.

To describe dispersion interactions, both the exchange-hole dipole moment dispersion model (XDM)[76] and Grimme's DFT-D3 dispersion correction (both zero-damped and BJ-damped variants) for DFT functionals[77,78] are available. In many cases, one can obtain reasonably accurate non-covalent interaction energies at van der Waals distances with meta-functionals in NWChem even without adding extra dispersion terms.[79]

Numerical integration is necessary for the evaluation of the exchange-correlation contribution to the density func-tional when Gaussian basis functions are used. The three-dimensional molecular integration problem is reduced to a sum of atomic integrations by using the approach first pro-posed by Becke[80]. NWChem implements a modification of the Stratmann algorithm[81], where the polynomial partition function $w_A(r)$ is replaced by a modified error function erfn (where n can be 1 or 2).

$$w_A(r) = \prod_{B \neq A} \frac{1}{2}\left(1 - \mathrm{erf}(\mu'_{AB})\right)$$

$$\mu'_{AB} = \frac{1}{\alpha} \frac{\mu_{AB}}{(1 - \mu_{AB}^2)^n}$$

$$\mu_{AB} = \frac{r_A - r_B}{|r_A - r_B|}$$

The default quadrature used for the atomic centered nu-merical integration is an Euler-MacLaurin scheme for the radial components (with a modified Mura-Knowles[82] trans-formation) and a Lebedev[83] scheme for the angular compo-nents.

On top of the petite-list symmetry algorithm used in the same fashion as in the SCF module, the evaluation of the exchange-correlation kernel incurs additional time savings when the molecular symmetry is a subset of the O_h point group, exploiting the octahedral symmetry of the Lebedev angular grid.

NWChem also has an implementation of a variational treatment of the one-electron spin-orbit operator within the DFT framework. Calculations can be performed either with an all-electron relativistic approach (for example, ZORA) or with an ECP and a matching spin-orbit (SO) potential.

Other capabilities built on the DFT module include the electron transfer (ET)[84,85], constrained DFT (CDFT)[86–88], and frozen density embedding (FDE)[89–91] modules, respec-tively.

1. Time-Dependent Density Functional Theory

a. Linear-Response Time-Dependent Density Func-tional Theory: NWChem supports a spectrum of single excitation theories for vertical excitation energy calcula-tions, namely, configuration interaction singles (CIS)[92], time-dependent Hartree-Fock (TDHF or also known as random-phase approximation RPA), time-dependent den-sity functional theory (TDDFT)[93–95], and Tamm-Dancoff approximation[96] to TDDFT. These methods are imple-mented in a single framework that invokes Davidson's trial vector algorithm (or its modification for a non-Hermitian eigenvalue problem). An efficient special symmetric Lanc-zos algorithm and kernel polynomial method has also been implemented.[97]

In addition to valence vertical excitation energies, core-level excitations[98] and emission spectra[99,100] can also be computed. Analytical first derivatives of vertical excitation energies with a selected set of exchange-correlation func-tionals can also be computed,[101] which allows excited-state optimizations and dynamics. Origin-independent optical rotation and rotatory strength tensors can also be calculated with the LR-TDDFT module within the gauge including atomic orbital (GIAO) basis formulation.[62,102–104] Exten-sions to compute excited-state couplings are currently underway and will be available in a future release.

b. Real-Time Time-Dependent Density Functional The-ory: Real-time time-dependent density functional theory (RT-TDDFT) is a DFT-based approach to electronic ex-cited states based on integrating the time-dependent Kohn-Sham (TDKS) equations in time. The theoretical underpin-nings, strengths, and limitations are similar to traditional linear-response (LR) TDDFT methods, but instead of a frequency domain solution to the TDKS equations, RT-TDDFT yields a full time-resolved, potentially non-linear solution. Real-time simulations can be used to compute not only spectroscopic properties (e.g., ground and excited-state absorption spectra, polarizabilities, etc.)[98,105–108], but also the time and space-resolved electronic response to arbi-trary external stimuli (e.g., electron charge dynamics after laser excitation)[105,109] and non-linear spectroscopies.[110,111] RT-TDDFT has the potential to be efficient for computing spectra in systems with a high density of states[112] as, in principle, an entire absorption spectrum can be computed from only one dynamics simulation.

This functionality is developed on the Gaussian basis set DFT module for both restricted and unrestricted cal-culations and can be run with essentially any combina-tion of basis set and exchange-correlation functional in NWChem. A number of time propagation algorithms have been implemented[113] within this module, with the default being the Magnus propagator.[114] Unlike LR-TDDFT, which requires second derivatives, RT-TDDFT can be used with all the functionals since only first derivatives are needed for the propagation. The current RT-TDDFT implementation assumes frozen nuclei and no dissipation.

2. Ab Initio Molecular Dynamics

This module leverages the Gaussian basis set methods to allow for seamless molecular dynamics of molecular systems. The nuclei are treated as classical point parti-cles and their motion is integrated via the velocity Verlet algorithm.[115,116] In addition to being able to perform sim-ulations in the microcanonical ensemble, we have imple-mented several thermostats to control the kinetic energy of the nuclei. These include the stochastic velocity rescaling approach of Bussi, Donadio, and Parrinello[117], Langevin dynamics according to the implementation of Bussi and Parrinello[118], the Berendsen thermostat[119], and simple ve-locity rescaling.

The potential energy surface upon which the nuclei move can be provided by any level of theory implemented within NWChem, including DFT, TDDFT, MP2, and the corre-lated wavefunction methods in the TCE module. If analyt-ical gradients are implemented for the specified method, these are automatically used. Numerical gradients will be used in the event that analytical gradients are not available at the requested level of theory. This module has been used to demonstrate how the molecular dynamics based deter-mination of vibrational properties can complement those determined through normal mode analysis, therefore allow-ing to achieve a deeper understanding of complex dynamics and to help interpret complex experimental signatures.[120] Extensions to include non-adiabatic dynamics have been implemented in a development version and will be available in a future release.

C. Wavefunction Formulations

The wavefunction-based methods play a special role in all electronic structure packages. Their strengths origi-nate in the possibility of introducing, using either various orders of perturbation theory or equivalently through the linked cluster theorem (for example, see Refs. 121 and 122) various ranks of excitations, a systematic hierarchy of elec-tron correlation effects. NWChem offers implementations of several correlated wavefunction approaches including many-body perturbation theory approaches and coupled-cluster methods.

1. Perturbative Formulations

a. MP2: Three algorithms are available in NWChem to compute the Møller-Plesset (or many-body) perturbation theory second-order correction[123] to the Hartree-Fock en-ergy (MP2). They vary in capability, the size of the system that can be treated and use of other approximations

Semi-direct MP2 is recommended for most large ap-plications on parallel computers with significant disk I/O capability. Partially transformed integrals are stored on disk, multi-passing as necessary. RHF and UHF references may be treated including computa-tion of analytic derivatives. The initial semi-direct code was later modified to use aggregate memory instead of disk to store intermediate, therefore not requiring any I/O operation.

Fully-direct[124] MP2. This is of utility if only lim-ited I/O resources are available (up to about 2800 functions). Only RHF references and energies are available.

The resolution of the identity (RI) approximation MP2 (RI-MP2)[125] uses the RI approximation and is, therefore, only exact in the limit of a complete fitting basis. However, with some care, high accuracy may be obtained with relatively modest fitting basis sets. An RI-MP2 calculation can cost over 40 times less than the corresponding exact MP2 calculation. RHF and UHF references with only energies are available.

2. Multi-con gurational Self-Consistent Field (MCSCF)

A large-scale parallel multi-configurational self-consistent field (MCSCF) method has been developed in NWChem by integration of the serial LUCIA program of Olsen[126,127]. The generalized active space approach is used to partition large configuration interaction (CI) vectors and generate a sufficient number of nearly equal batches for parallel distribution. This implementation allows the execution of complete active space self-consistent field (CASSCF) calculations with non-conventional active spaces. An unprecedented CI step for an expansion composed of almost one trillion Slater determinants has been reported[127].

3. Coupled-Cluster Theory

The coupled-cluster module of NWChem contains two classes of implementations (a) parallel implementation of the CCSD(T) formalism[31] for closed-shell systems, and (b) a wide array of CC formalisms for arbitrary reference functions. The latter class of implementations automati-cally generated by Tensor Contraction Engine[128,129] is an example of a successful co-design effort.

a. Closed-Shell CCSD(T): The coupled-cluster method was introduced to chemistry by Čížek[130] (see also Ref. 131), and is a post-Hartree-Fock electron correlation method. Development of the canonical coupled-cluster code in NWChem commenced in 1995 under a collab-oration with Cray Inc to develop a massively parallel coupled-cluster program designed to run on a Cray T3E. Full details of the implementation are given in Kobayashi and Rendell[132].

The coupled-cluster wavefunction is written as an expo-nential of excitation operators acting on the Hartree-Fock reference:

$$|\Psi_{CC}\rangle = e^{T}|\Phi\rangle \qquad (4)$$

where $T = T_1 + T_2 + \cdots$ is a cluster operator represented as a sum of its many-body components, i.e., singles T_1, doubles T_2, etc. and $|\Phi\rangle$ is the so-called reference function (usually chosen as a reference determinant). In practical applications the above sum is truncated at some excitation rank. For ex-ample, the CCSD method[30] is defined by including singles and doubles, i.e., $T \simeq T_1 + T_2$. Introducing the exponential ansatz (4) into the Schrödinger equation, premultiplying both sides by e^{-T}, using the Hausdorff formula, and pro-jecting onto the subspace of excitation functions, gives a set of coupled non-linear equations that are solved iteratively to yield the coupled-cluster energy and amplitudes. For example, for the CCSD formulation one obtains

$$\langle\Phi|(H_N e^{T_1+T_2})_C|\Phi\rangle = \Delta E_{CCSD} \qquad (5)$$

$$\langle\Phi_i^a|(H_N e^{T_1+T_2})_C|\Phi\rangle = 0\,; \qquad (6)$$

$$\langle\Phi_{ij}^{ab}|(H_N e^{T_1+T_2})_C|\Phi\rangle = 0\,; \qquad (7)$$

where H_N is the electronic Hamiltonian in normal product form ($H_N = H - \langle\Phi|H|\Phi\rangle$), subscript C represents a con-nected part of a given operator expression, and ΔE_{CCSD} is CCSD correlation energy. The closed-shell CCSD imple-mentation employs the optimized form of the CC equations discussed by Scuseria et al.[133] as was programmed in the TITAN program[134]. The nature of the Cray T3E hardware required significant rewriting of earlier coupled-cluster al-gorithms to take into account the limited memory available per core (8 MW) and the prohibitive penalty of I/O oper-ations. Of the various four indexed quantities, those with four occupied indices were replicated in local memory (i.e. the memory associated with a single core), and those with one or two virtual indices were distributed across the global memory of the machine (i.e. the sum of the memory of all the processors), and accessed in computational batches. The terms involving integrals with three and four virtual or-bital indices still proved too costly for the available memory and to circumvent this problem, these terms were evaluated in a "direct" fashion. This structure distinguishes NWChem from most other coupled-cluster programs. Thus, to make effective use of the available memory, as much as possible should be allocated, by using global arrays, with the bare minimum for the arrays replicated in local memory.

The canonical CCSD implementation in NWChem also contains the perturbative triples correction, denoted (T), of Raghavachari et al.[31]. This correction is an estimate from Møller-Plesset perturbation theory[123] and evaluates the triples contribution to MP4 using the optimized cluster amplitudes at the end of a CCSD calculation. The CCSD(T) method is commonly referred to as the gold standard for ab initio electronic structure theory calculations. Its computational cost scales as n^7, making it considerably more expensive than a CCSD calculation. However, the triples are non-iterative and only require two-electron integrals with at most three virtual orbital indices, hence avoiding the previous memory and I/O issues and so the correction was easily adapted from the "aijkbc algorithm" of an earlier work by Rendell et al[135].

In recent years, a great deal of effort was invested to enhance the performance of the iterative and non-iterative parts of the CCSD(T) workflow. Performance tuning of the iterative part resulted in scaling the code up to 223,200 processors of the ORNL Jaguar computer.[26,136] Significant speedups for the CCSD iterative part were achieved by introducing efficient optimization techniques to alleviate the communication bottlenecks caused by a copious amount of communication requests introduced by a large class of low-dimensionality tensor contractions. This optimization provided a significant 2- to 5-fold performance increase in the CCSD iteration time depending on the problem size and available memory, and improved the CCSD scaling to 20,000 nodes of the NCSA Blue Waters supercomputer[137].

b. Tensor Contraction Engine and High-Accuracy Formulations: NWChem implements a large number of high-rank electron-correlation methods for the ground, excited, and electron-detached/attached states as well as for molecular properties. The underlying ansatzes span configuration interaction (CI), coupled-cluster (CC), many-body perturbation theories (MBPT), and various combinations thereof. A distinguishing feature of these implementations is their uniquely forward-looking development strategy. These parallel-executable codes, as well as their formulations and algorithms, were computer-generated by the symbolic algebra program[138] called Tensor Contraction Engine (TCE).[128] TCE was one of the first attempts to provide a scalable tensor library for parallel implementations of many-body methods, which extends the ideas of automatic CC code generation introduced by Janssen and Schaefer,[139] Li and Paldus,[140] and Nooijen and co-workers.[141,142]

The merits of such a symbolic system are many: (1) It expedites otherwise time-consuming and error-prone derivation and programming processes; (2) It facilitates parallelization and other laborious optimizations of the synthesized programs; (3) It enhances the portability, maintainability, extensibility, and thus the lifespan of the whole program module; (4) It enables new or higher-ranked methods to be implemented and tested rapidly which are practically impossible to write manually. TCE is, therefore, one of the earliest examples[139] of an expert system that lifts the burden of derivation/programming labor so that computational chemists can focus on imagining new ansatz—a development paradigm embraced quickly by other chemistry software developers.[143–145]

The working equations of an ab initio electron-correlation method are written with sums-of-products of matrices, whose elements are integrals of operators in the Slater determinants. For many methods, the matrices have the general form:[146]

$$\langle F_i | \hat{L}_j^\dagger \hat{H} \exp(\hat{T}_k) \hat{R}_l | F_m \rangle_{C/L}; \qquad (8)$$

where F_i is the whole set of the i-electron excited (or electron-detached/attached) Slater determinants, $\hat{H}$ is the Hamiltonian operator, $\hat{T}_k$ is a k-electron excitation operator, $\hat{R}_l$ is an l-electron excitation (or electron detachment/attachment) operator, and $\hat{L}_j^\dagger$ is a j-electron de-excitation (or electron detachment/attachment) operator. Subscript 'C/L' means that the operators can be required to be connected and/or linked diagrammatically. For example, the so-called T_2-amplitude equation of coupled-cluster singles and doubles (CCSD) is written as

$$0 = \langle F_2 | \hat{H} \exp(\hat{T}_1 + \hat{T}_2) | F_0 \rangle_C : \qquad (9)$$

With the ansatz of a method given in terms of Eq. (8), TCE first (1) evaluates these operator-determinant expressions into sums-of-products of matrices (molecular integrals and excitation amplitudes) using normal-ordered second quantization and Wick's theorem, second (2) transforms the latter into a computational sequence (algorithm), which consists in an ordered series of binary matrix multiplications and additions, and third (3) generates parallel-execution programs implementing these matrix multiplications and additions, which can be directly copied into appropriate directories of the NWChem source code and which are called by a short, high-level driver subroutine humanly written (see Fig. 1).

In step (2), TCE finds the (near-)minimum cost path of evaluating sums-of-products of matrices by solving the matrix-chain problem (defining the so-called "intermediates") and by performing common subexpression elimination and intermediate reuse. In step (3), the computer-generated codes perform dynamically-load-balanced parallel matrix multiplications and additions, taking advantage of spin, spatial, and index-permutation symmetries. The parallelism, symmetry usage, and memory/disk space management are all achieved by virtue of TCE's data structure: every matrix (molecular integrals, excitation amplitudes, intermediates, etc.) is split into spin- and spatial-symmetry-adapted tiles, whose sizes are determined at runtime so that the several largest tiles can fit in core memory. Only symmetrically-unique, non-zero tiles are stored gapless (with their storage addresses recorded in hash tables which are also auto-generated by TCE) and used in parallel tile-wise multiplications and additions, which are dynamically distributed to idle processors on a first-come, first-served basis. NWChem's parallel middleware, especially Global Arrays, was essential for making the computer-generated parallel codes viable.

TCE is a part of the NWChem source-code distribution, and a user is encouraged to implement their own ansatzes

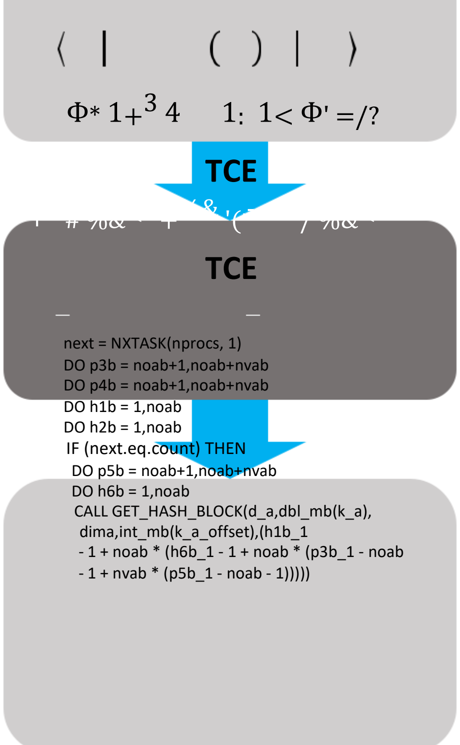

Figure 1. A schematic representation of TCE workflow (see text for details).

into high-quality parallel codes. Therefore TCE has paved the way for quick development of various implementations of coupled-cluster methods that would take disproportion-ately longer time if hand-coded. Additionally, TCE pro-vided a new testing ground for several novel parallel al-gorithms for accurate many-body methods and has been used to generate a number of canonical implementations of single reference CC methods for ground- and excited-state calculations for arbitrary reference function including: RHF, ROHF, UHF, and multi-reference cases. Below we listed basic components of the TCE infrastructure in NWChem:

- various perturbative methods ranging from second (MBPT(2)/MP2) to fourth-order (MBPT(4)/MP4) of Møller-Plesset perturbation theory,
- single reference iterative (CCD,[130] CCSD,[30] CCSDT,[147–149] CCSDTQ[150,151]) and non-iterative (CCSD(T),[31] CR-CCSD(T),[152] LR-CCSD(T),[153] CCSD(2),[154–156] CCSD(2)T ,[156] CCSDT(2)Q)[156] CC approximations for ground-state calculations,
- single reference iterative (EOMCCSD,[157,158] EOMCCSDT,[159,160] EOMCCSDTQ[146,161]) and non-iterative (CR-EOMCCSD(T)[162]) Equation-of-motion CC (EOMCC) approximations[163] for excited-state calculations,
- Ionization potential and electron affinity EOMCC (IP/EA-EOMCC) methods,[164–170]
- linear-response CC (LR-CC) methods for calculating static and frequency-dependent polarizabilities and static hyperpolarizabilities at the CCSD and CCSDT levels of approximation,[171]
- state-specific multi-reference CC (MR-CC) methods for quasi-degenerate systems.[172–178]

The TCE infrastructure has also been used in exploring new parallel algorithms and algorithms for emerging computer architectures. The most important examples include:

- parallel algorithms for excited-state CR-EOMCCSD(T) calculations with demonstrated scalability across 210,000 cores of Jaguar Cray XT5 system at the Oak Ridge Leadership Computing Facility (OLCF)[28],
- new CC algorithms for GPU and Intel MIC architectures (single-reference CC and MR-CC theories),[32,34,35,179,180]
- new algorithms for multi-reference CC methods uti-lizing processor groups and multiple levels of paral-lelism (the so-called reference-level of parallelism of Refs.[181,182]) with demonstrated scalability across 80,000 cores of Jaguar Cray XT5 system,[182]
- new execution models for the iterative CCSD and EOMCCSD models.[28]

With TCE, one can perform CC calculations for closed-and open-shell systems characterized by 1,000-1,300 or-bitals. Some of the most illustrative examples of TCE calculations are (1) static and frequency-dependent polariz-abilities for the C60 molecule,[183] excited state simulations for p-conjugated chromophores,[184] and IP-EOMCCSD cal-culations for ferrocene with explicit inclusion of solvent molecules. One cutting edge application of TCE CC was the early application of EOMCC methodologies in excited-state studies of functionalized forms of porphyrin[28]. Ad-ditionally, TCE has also served as a development plat-form for early implementations of the coupled-cluster Green's function formalism.[185–188] The TCE development has since been followed by several other efforts towards enabling scalable tensor libraries. This includes Super In-struction Assembly Language SIAL,[144,189] Cyclop Ten-sor Framework (CTF),[190] TiledArray framework,[191] and Libtensor,[192] which have been used to develop scalable implementations of CC methods.

D. Relativistic Methods

Methods which include treatment of relativistic effects are based on the Dirac equation[193], which has a four-component wavefunction. The solutions to the Dirac equa-tion describe both positrons (the "negative energy" states) and electrons (the "positive energy" states), as well as both spin orientations, hence the four components. The wavefunction may be broken down into two-component functions traditionally known as the large and small components; these may further be broken down into the spin components.[194–197]

The implementation of approximate all-electron relativistic methods in quantum chemical codes requires the removal of the negative energy states and the factoring out of the spin-free terms. Both of these may be achieved using a transformation of the Dirac Hamiltonian known in general as a Foldy-Wouthuysen (FW) transformation. Unfortunately, this transformation cannot be represented in closed form for a general potential, and must be approximated. One popular approach is the Douglas and Kroll[198] method developed by Hess[199,200]. This approach decouples the positive and negative energy parts to second-order in the external potential (and also fourth-order in the fine structure constant, a). Other approaches include the zeroth order regular approximation (ZORA)[201–204], modification of the Dirac equation by Dyall[205], which involves an exact FW transformation on the atomic basis set level[206,207] and the exact 2-component (X2C) formulation, which is a catch-all for a variety of methods that arrive at an exactly decoupled two-component Hamiltonian using matrix algebra.[197,208–211] NWChem contains released implementations of the DKH, ZORA, and Dyall approaches, while the X2C method is available in a development version.[209,211]

Since these approximations only modify the integrals, they can, in principle, be used at all levels of theory. At present, the Douglas-Kroll, ZORA and X2C implementations can be used at all levels of theory, whereas Dyall's approach is currently available at the Hartree-Fock level.

a. Douglas-Kroll Approximation: NWChem contains three second-order Douglas-Kroll approximations termed as FPP, DKH, and DKHFULL. The FPP is the approximation based on free-particle projection operators[199], whereas the DKH and DKFULL approximations are based on external-field projection operators[200]. The latter two are considerably better approximations than the former. DKH is the Douglas-Kroll-Hess approach and is the approach that is generally implemented in quantum chemistry codes. DKFULL includes certain cross-product integral terms ignored in the DKH approach (see for example, Haberlen¨ and Rosch¨[212]). The third-order Douglas-Kroll approximation (DK3) implements the method by Nakajima and Hirao[213,214].

b. Zeroth Order Regular Approximation (ZORA): The spin-free and spin-orbit versions of the one-electron zeroth order regular approximation (ZORA) have been implemented. Since the ZORA correction depends on the potential, it is not gauge invariant. This is addressed by using the atomic approximation of van Lenthe and coworkers.[215,216] Within this approximation, the ZORA corrections are calculated using the superposition of densities of the atoms in the system. As a result, only intra-atomic contributions are involved, and no gradient or second derivatives of these corrections need to be calculated. In addition, the corrections need only to be calculated once at the start of the calculation and stored. The ZORA approach is implemented in two ways in NWChem, one where the ZORA potential components are directly computed on an all-electron grid[204] and a second approach, where the ZORA potential is computed using the model potential approach due to van Wullen¨ and co-workers.[217,218]

c. Dyall's Modified Dirac Hamiltonian Approximation: The approximate methods described in this section are all based on Dyall's modified Dirac Hamiltonian. This Hamiltonian is entirely equivalent to the original Dirac Hamiltonian, and its solutions have the same properties. The modification is achieved by a transformation on the small component. This gives the modified small component the same symmetry as the large component. The advantage of the modification is that the operators now resemble those of the Breit-Pauli Hamiltonian, and can be classified in a similar fashion into spin-free, spin-orbit, and spin-spin terms. It is the spin-free terms which have been implemented in NWChem, with a number of further approximations. Negative energy states are removed by a normalized elimination of the small component (NESC), which is equivalent to an exact Foldy-Wouthuysen (EFW) transformation. Both one-electron and two-electron versions of NESC (NESC1E and NESC2E, respectively) are available, and both have analytic gradients.[205–207]

E. Molecular Properties

A broad array of simple and response-based molecular properties can be calculated using the HF and DFT wavefunctions in NWChem. These include: natural bond analysis, dipole, quadrupole, octupole moments, Mulliken population analysis and bond order analysis, Lowdin¨ population analysis, electronic couplings for electron transfer,[84,85] Raman spectroscopy,[219,220], electrostatic potential (diamagnetic shielding) at nuclei, electric field and field gradient at nuclei, electric field gradients with relativistic effects[221], electron and spin density at nuclei, GIAO-based NMR properties like shielding, hyperfine coupling (Fermi-Contact and Spin-Dipole expectation values), indirect spin-spin coupling,[222–224] G-shift,[225] EPR, paramagnetic NMR parameters,[226,227] and optical activity.[102,103,228,229] Note that only linear-response is supported and for single frequency, electric field, and mixed electric-magnetic field perturbations. Ground state and dynamic dipole polarizabilities for molecules can be calculated at the CCSD, CCSDT, and CCSDTQ levels using the linear-response formalism.[230] For additional information, we refer the reader to the online manual.[8]

F. Periodic Plane-Wave Density Functional Theory

The NWChem plane-wave density functional theory (NWPW) module contains two programs:

PSPW - A pseudopotential and projector augmented (PAW) plane-wave G-point code for calculating molecules, liquids, crystals, and surfaces,

BAND - A pseudopotential plane-wave band structure code for calculating crystals and surfaces with small band gaps (e.g. semi-conductors and metals),

These programs use a common infrastructure for carrying out operations related to plane-wave basis sets that is parallelized with the MPI and OpenMP libraries[29,33,34,231–235] The NWPW module can be used to carry out many different kinds of simulations. In addition to the standard simulations implemented in other modules, e.g. energy, optimize, and freq, there are additional capabilities specific to PSPW and BAND that can be used to carry out NVE and NVT[236] Car–Parrinello[19] and Born-Oppenheimer molecular dynamics simulations, hybrid ab initio molecular dynamics and molecular mechanics (AIMD-MM) simulations[234,237], Gaussian/Fermi/Marzari-Vanderbilt smearing, Potential-of-Mean-Force (PMF)[238]/Metadynamics[239,240]/Temperature-Accelerated-Molecular-Dynamics (TAMD)[241,242]/Weighted-Histogram-Analysis-Method (WHAM)[243] free energy simulations, AIMD-EXAFS simulations using open-source versions of the FEFF software[244–246] that have been parallelized, electron transfer calculations[247], unit cell optimization, op-timizations with space group symmetry, Monte-Carlo NVT and NPT simulations, phonon calculations, simulations with spin-orbit corrections, Wannier[248] and rank reducing density matrix[249] localization calculations, Mulliken[250] and Blochl·[251] charge analysis, Gaussian cube file generation, periodic dipole and infrared (AIMD-IR) simulations, band structure plots, density of states. Calculations can also be run using a newly developed i-PI[252] interface, and more direct interfaces to ASE[253], nanoHUB[254], and EMSL Arrows[255] simulation tools are currently being implemented.

A variety of exchange-correlation functionals have been implemented in both codes, including the local den-sity approximation (LDA) functionals, generalized gradi-ent approximation (GGA) functionals, full Hartree-Fock and screened exchange, hybrid DFT functionals, self-interaction correction (SIC) functionals[256], localized ex-change method, DFT+U method, and Grimme dispersion corrections[77,78], as well as recently implemented vdW dis-persion functionals[257], and meta-generalized gradient ap-proximation (metaGGA) functionals. The program contains several codes for generating pseudopotentials, including Hamann[20] and Troulier-Martin[21], and PAW[23] potentials. These codes have the option for generating potentials with multiple projectors and semi-core corrections. It also contains codes for reading in HGH[258], GTH[259], and norm-conserving pseudopotentials in the CPI and TETER formats. Codes for reading Optimized Norm-Conserving Vanderbilt (ONCV) pseudopotentials[260,261] and USPP PAW potentials will become available in future releases of NWChem.

The pseudopotential plane-wave DFT methods implemented in NWChem are a fast and efficient way to calculate molecular and solid-state properties using DFT[16,17,19,29,235,262–270]. In these approaches, the fast varying parts of the valence wavefunctions inside the atomic core regions and the atomic core wavefunctions are re-moved and replaced by pseudopotentials[20–22,271–274]. Pseudopotentials are chosen such that the resulting pseudoatoms have the same scattering properties as the original atoms. The rationale for this approach is that the changes in the electronic structure associated with making and breaking bonds only occur in the interstitial region outside the atomic core regions (see Fig. 2). Therefore, removing the core regions should not affect the bonding of the system. For this approach to be useful, it is necessary for the pseudopoten-tials to be smooth in order for plane-wave basis sets to be used. As the atomic potential becomes stronger the core re-gion becomes smaller and the pseudopotential grows steep. As a result, the pseudopotential can become very stiff, re-quiring large plane-wave basis sets (aka cutoff energies), for the first-row transition metals atoms, the lanthanide atoms, and towards the right-hand side of the periodic table (fluorine).

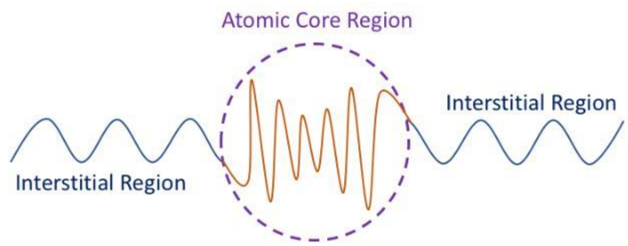

Figure 2. Illustration of the atomic core and interstitial regions in a valence wavefunction. Bonding takes place in the interstitial re-gion and the atomic core regions change very little from molecule to molecule. Figure from Bylaska et al.[234].

The projected augmented plane-wave method (PAW)[23,232,275–277] is another related approach that removes many of the problems of the somewhat ad hoc nature of the pseudopotentials approach. However, in the PAW approach, instead of discarding the rapidly varying parts of the electronic functions, these are projected onto a local basis set (e.g., a basis of atomic functions), and no part of the electron density is removed from the problem. Another key feature of PAW is that by maintaining a local description of the system, the norm-conservation condition (needed for proper scattering from the core) can be relaxed, which facilitates the use of smaller plane-wave basis sets (aka cutoff energies) then for many standard pseudopotentials. Historically, the PAW method was implemented as a separate program in the NWPW module, rather than being fully integrated into the PSPW and BAND codes. This separation significantly hindered its development and use. As of NWChem version 6.8 (released in 2017), the PAW approach has been integrated into the PSPW code, and it is currently being integrated into the BAND code. It will become available in future releases of NWChem.

In recent years, with advances in High-Performance Computing (HPC) algorithms and computers, it is now possible to run AIMD simulations up to 1 ns for non-trivial system sizes. As a result, it is now possible to effec-tively use free-energy methods with AIMD and AIMD/MM approaches. Free energy approaches are useful for simu-lating reactions where traditional quantum chemistry ap-proaches can be difficult to use and often require the exper-tise of a very experienced quantum chemist, e.g. reactions that are complex with concerted or multi-step components and/or interact strongly with the solvent. Recent examples include solvent coordination and hydrolysis of actinides metals[197,278–281] (see Fig. 3), hydrolysis of explosives[234], and ion association in AlCl3[237]. To help users learn how to use these new techniques, we developed a tutorial on

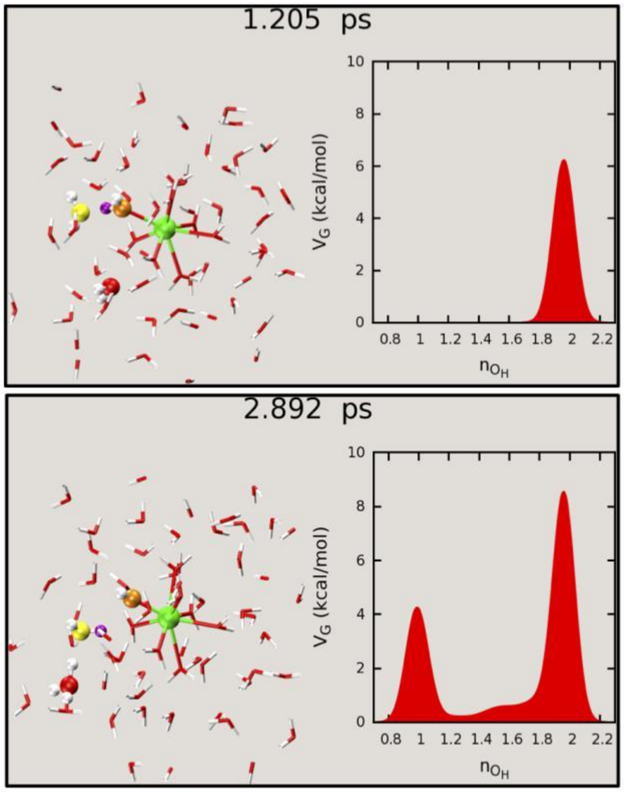

Figure 3. Snapshots from a metadynamics simulation of the hydrolysis of the U^{4+} aqua ion[278]. During the simulation a proton jumps from a first shell water molecule to a second shell water molecule and then subsequently to other water molecules via a Grotthuss mechanism.

carrying out finite temperature free energy calculations in NWChem[282].

The NWPW module continues to be actively developed. There are on-going developments for RPA and GW-RPA methods, an electron transfer MCSCF method, Raman and Mossbauer¨ spectroscopy, and a hybrid method that integrates classical DFT[283] into ab initio molecular dynamics (AIMD-CDFT). In addition to these developments, we are actively developing the next generation of plane-wave codes as part of the NWChemEx project. These new codes, which are being completely written from scratch, will contain all the features currently existing in the NWPW module. Be-sides implementing fast algorithms to use an even larger number of cores and new algorithms to run efficiently on GPUs, it includes a more robust infrastructure to facilitate the implementation of an O(N) DFT code based on the work of Fattebert et al.[284]

G. Optimization, Transition State, and Rate theory Approaches

A variety of drivers and interfaces are available in NWChem to perform geometry minimization and transition state optimizations. The default algorithms in NWChem for performing these optimizations are quasi-Newton meth-ods with line searches. These methods are fairly robust, and they can be used to optimize molecules, clusters, and periodic unit cells and surfaces. They can also be used in conjunction with both point group and space group sym-metries, excited state TDDFT surfaces, as well as with a variety of external fields, such as external point charges, COSMO[285] or SMD[286] Model. The default methods also work seamlessly with electronic structure methods that do not have nuclear gradients implemented by automatically using finite difference gradients. NWChem also contains default methods for calculating harmonic vibrational fre-quencies and phonon spectra for periodic systems. These methods are able to make use of analytic Hessians if they are available, otherwise a finite difference approach is used. A vibrational self-consistent field[287] (VSCF) method is also available in NWChem and it can be used to calculate anharmonic contributions to specified vibrational modes. There is also an interface called DIRDYVTST[288] that uses NWChem to compute energies, gradients, and Hessians for direct dynamics calculations with POLYRATE[289].

A variety of external packages, such as ASE[253,290] and Sella[291,292], can also be used for finding energy minima, saddle points on energy surfaces, and frequencies using either python scripting or a newly developed i-PI[252] in-terface. Python programs may be directly embedded into the NWChem input and used to control the execution of NWChem. The python scripting language provides useful features, such as variables, conditional branches, and loops, and is also readily extended. Other example applications for which it could be used include scanning potential en-ergy surfaces, computing properties in a variety of basis sets, optimizing the energy with respect to parameters in the basis set, computing polarizabilities with a finite field, simple molecular dynamics, and parallel in time molecular dynamics[293].

NWChem also contains an implementation of the nudged elastic band (NEB) method of Jonsson´ and co-workers[294–297] and the zero-temperature string method of vanden Eijden et al.[298] Both these methods can be used to find minimum energy paths. Currently, a quasi-Newton algorithm is used for the NEB optimization. A better approach for this kind of optimization is to use a non-linear multi-grid algorithm, such as the Full Approximation Scheme (FAS)[299]. A new implementation of NEB based on FAS is available on Bitbucket[300], and an integrated version will soon be available in NWChem.

H. Classical Molecular Dynamics

The integration of a molecular dynamics (MD) module in NWChem enables the generation of time evolution trajectories based on Newton's equation of motion of molecular systems in which the required forces can origi-nate from a classical force field, any implemented quantum mechanical method for which spatial derivatives have been implemented, or hybrid quantum mechanical/molecular mechanical (QM/MM) approaches. The method is based on the ARGOS molecular dynamics software, originally designed for vector processors,[301] but later redesigned for massively parallel architectures.[25,302–304]

a. System Preparation: The preparation of a molecular system is done by a separate prepare module that reads the molecular structure and assembles a topology from the databases with parameters for the selected force field. The topology file contains all static information for the system. In addition, this module generates a so-called restart file with all dynamic information. The prepare module has a wide range of capabilities that include the usual functions of placing counter-ions and solvation with any solvent defined in the database. The prepare module is also used to define Hamiltonian changes for free energy difference calculations, and the definition of those parts of the molecular systems that will be treated quantum mechanically in QM/MM simulations. Some of the more unique features include setting up a system for proton hopping (QHOP) simulations,[305,306] and the setup of biological membranes from a single lipid-like molecule. This last capability has been successfully used for the first extensive simulation studies of complex asymmetric lipopolysaccharide membranes of Gram-negative microbes[307–311] and their role in the capture of recalcitrant environmental heavy metal ions,[312] microbial adhesion to geochemical surfaces,[313–316] and the structure and dynamics of trans-membrane proteins including ion transporters[317–319](Fig. 4).

b. Force Fields: The force field implemented in NWChem consists of harmonic terms for bonded, angle and out of plane bending interactions, and trigonometric terms for torsions. Non-bonded van der Waals and elec-trostatic interactions are represented by Lennard-Jones and Coulombic terms, respectively. Non-bonded terms are eval-uated using charge groups and subject to a user-specified cutoff radius. Electrostatic interaction corrections beyond the cutoff radius are estimated using the smooth particle mesh Ewald method.[320] Parameter databases are provided for the AMBER[321] and CHARMM[322] force fields.

Even for purely classical MD simulations, the integration with the electronic structure methods provides a convenient way of determining electrostatic parameters for missing fragments in standard force field databases, through the use of restrained electrostatic potential fitting[323,324] to which a variety of additional constraints and restraints can be applied.

c. Simulation Capabilities: Ensemble types available in NWChem are NVE, NVT, and NPT, using the Berendsen thermostat and barostat.[119] Newton's equations of motion are integrated using the standard leap-frog Verlet or velocity Verlet algorithms. A variety of fundamental properties are evaluated by default during any molecular dynamics simulation. Parallel execution time analysis is available to determine the parallel efficiency.

The MD module has extensive free energy simulation capabilities,[325–330] which are implemented in a so-called multi-configuration approach. For each incremental change of the Hamiltonian to move from the initial to the final state, sometimes referred to as a window, a full molecular simulation is carried out. This allows for a straightforward evaluation of statistical and systematic errors where needed, including a correlation analysis.[331] Based on the ARGOS code[301] it has some unique features, such as the separation-shifted scaling technique to allow atoms to appear from or disappear to dummy atoms.[332] One of the advantages of the integration of MD into the electronic structure methods framework in NWChem is the ability to carry out hybrid QM/MM simulations (discussed in the next section). The preparation of molecular systems for the MD module allows for flexibly specifying parts of the molecular system to be treated by any of the implemented electronic structure methods capable of evaluating positional gradients.

A unique feature in the NWChem MD module is the optional specification of protonatable sites on both solute and solvent molecules. Pairs of such sites can dynamically change between protonated or unprotonated state, effectively exchanging a proton. Transitions are governed by a Monte Carlo type stochastic method to determine when transitions occur. This so-called QHOP approach was developed by the research group of Helms.[306]

d. Analysis Capabilities: The NWChem MD capability includes two analysis modules. The original analysis module, analyze, analyzes trajectories in a way that reads individual structures one time step at a time and distributes the data in a domain decomposition fashion as in the molecular simulation that generated the data. The second data-intensive analysis module, diana, reads entire trajectories and distributes the data in the time domain. This is especially effective for analyses that require multiple passes through a trajectory, but requires the availability of potentially large amounts of memory.[333,334] An example of such analyses is the Essential Dynamics Analysis, a principal component analysis (PCA) based calculation to determine the dominant motions in molecular trajectories.

e. Parallel Implementation Strategy: The most effective way of distributing a system with large numbers of particles is through the use of domain decomposition of the physical space. The implementation in NWChem, facilitated through the use of the Global Arrays (GA) toolkit, partitions the simulation space into rectangular cells that are assigned to different processes ranks or threads. Each of these ranks carries out the calculation of intra-cell atomic energies and forces of the cells assigned. Inter-cell ener-gies and forces are evaluated by one of the ranks that was assigned one or the other of the cell pairs.

Two load balancing methods have been implemented in NWChem, both based on measured computation time. In the first one, the assignment of inter-node cell pair cal-culations is redefined such that assignments move from the busiest node to the less busy node. This scheme re-quires minimal additional communication, and since only two nodes are involved in the redistribution of work, the communication is local, i.e. node to node. In the second scheme, the physical size of the most time-consuming cell is reduced, while all other cells are made slightly larger. This scheme requires communication and redistribution of atoms on all nodes. In practice, the first scheme is used until performance no longer improves, after which the second scheme is used once followed by returning to use the first scheme. This approach has been found to improve load bal-ancing even in systems with a very asymmetric distribution

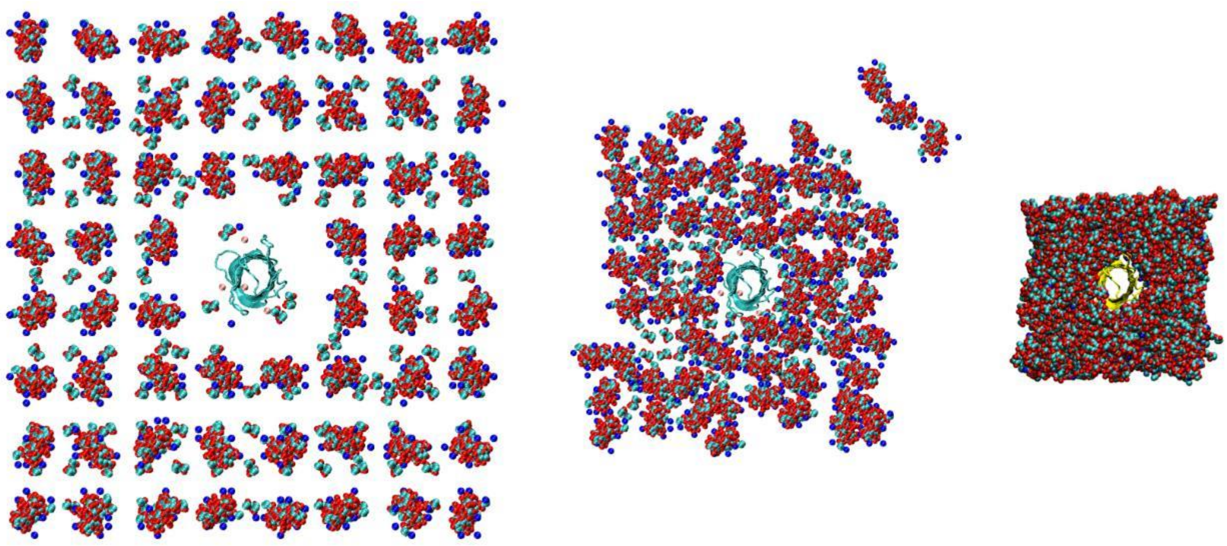

Figure 4. The NWChem MD Prepare utility facilitates the setup of trans-membrane proteins in complex asymmetric membrane environments in a semi-automated procedure. Shown here are the top views of step 1 in which membrane lipopolysaccharide molecules with the necessary counter ions are placed on a rectangular grid around a trans-membrane protein, in which each membrane lipid molecule is randomly rotated around the principal molecular axis (left panel), step 2 in which each cluster of a lipid molecule is translated towards the center of the transmembrane protein such that no steric clashed occur (center panel), and step 3 in which the system is equilibrated using strict restraint potentials to keep the lipid molecules aligned along the normal of the membrane and the lipid head groups in the plane of the membrane (right panel). After this procedure, the system would be solvated and equilibrated while slowly removing the positional restraint potentials.

of computational intensity.[335]

VI. HYBRID METHODS

We define hybrid methods as those coupling different levels of description to provide an efficient calculation of a chemical system, which otherwise may be outside the scope of conventional single-theory approaches. The physical motivation for such methods rests on the observation that, in the majority of complex chemical systems, the chemical transformation occurs in localized regions surrounded by an environment, which can be considered chemically inert to a reasonable approximation. Since hybrid methods require the combination of multiple theoretical methods in a single simulation, the diversity of simulation methodologies available in NWChem makes it a platform particularly apt for this purpose.

One common example involves chemical transformations in a bulk solution environment, forming the foundations of wide variety of spectroscopic measurements (UV-vis, NMR, EPR, etc.). The reactive region, referred to as the "solute", involves electronic structure degrees of freedom and thus requires the quantum mechanical (QM) based description, such as DFT or more complex wavefunction methods. In the conventional approach, such QM description would be necessarily extended to the entire system making the problem a heroic computational task. In a hybrid approach, the treatment of a surrounding environment ("solvent") would be delegated to a much simpler description, such as the continuum model (CM), for example. The latter is supported in NWChem via two models - COSMO[285] (COnductor-like Screening MOdel) and SMD[286] (Solvation Model based on Density) Model. The resulting QM/CM approaches are particularly well suited for accurate and efficient calculation of solvation free energies, geometries in solution, and spectroscopy in solution. The SMD model employs the Poisson equation with non-homogeneous dielectric constant for bulk electrostatic effects, and solvent-accessible-surface tensions for cavitation, dispersion, and solvent-structure effects, including hydrogen bonding. For spectroscopy in solution, the Vertical Excitation (or Emission) Model (VEM) has also been implemented for calculating the vertical excita-tion (absorption) or vertical emission (fluorescence) energy in solution according to a two-time-scale model of solvent polarization[336].

For systems where an explicit solvation environment treatment is needed (for example, heterogeneous systems like a protein matrix), NWChem provides a solution in terms of combined quantum mechanics/molecular mechan-ics (QM/MM) approach.[337,338] Here, the environment is described at the classical molecular mechanics level. This offers more fidelity compared with a continuum solvent de-scription, while still keeping the computational costs down. The total energy of the system in QM/MM approach can be represented as a sum of the energies corresponding to QM and MM regions:

$$E(r;R;y) = E_{qm}(r;R;y) + E_{mm}(r;R) \tag{10}$$

where y denotes electronic degrees of freedom, and r;R refer to nuclear coordinates of QM and MM regions cor-respondingly. The QM energy term can be further decom-posed into internal and external parts

$$E_{qm}[r;R;y] = E_{qm}^{int}[r;y] + E_{qm}^{ext}[r;R;r] \tag{11}$$

where r is the electron density.

As a generic module, the QM/MM implementation can utilize any of the Gaussian basis set based QM modules available in NWChem and supports nearly all the task func-tionalities. The calculation of QM energy remains the main computational expense in the QM/MM approach. This issue is more pronounced compared with the continuum coupling case, because of the additional atomistic degrees of freedom associated with MM description. The latter comes into play because any change in the MM degrees of freedom will, in general, trigger the recalculation of the QM energy ($E_{qm}(r;R;y)$). To alleviate these issues during the optimization, the QM/MM module offers the option of alternating relaxation of QM and MM regions. During the latter phase, the user may utilize an approximation where the QM degrees of freedom are kept frozen until the next cycle of QM region relaxation, offering significant com-putational savings. A similar technique can be utilized in the dynamical equilibration of the MM region and calcula-tions of reaction pathways and free energies. In addition to the native MD module, the NWChem QM/MM mod-ule can also utilize the external AMBER MD code[339] for running the classical part of the calculations. In this case, QM/MM simulations involve two separate NWChem and AMBER calculations with data exchange mediated through files written to disk.

Additionally, the QM/MM capability in NWChem has resulted in the development and refinement of force-field parameters, that can, in turn, be used in classical molecular dynamics simulations. Over the last two decades, classi-cal parameters obtained using NWChem have been em-ployed to address the underlying mechanisms of a variety of novel complex biological systems and their interactions (e.g., lipopolysaccharide membranes, carbohydrate moi-eties, mineral surfaces, radionuclides, organophosphorous compounds)[307,308,311–313,340–344] which has led to a signifi-cant expansion of the database of AMBER- and Glycam-compatible force fields and the GROMOS force field for lipids, carbohydrates and nucleic acids.[345–351]

For cases where a classical description of the environ-ment is deemed insufficient, NWChem offers an option to perform an ONIOM type calculation.[352] The latter differs from QM/MM in that the lower level of theory is not re-stricted to its region but also encompasses regions from all the higher levels of description. For example, in the case of the two-level description, the energy is written as

$$E(R) = E^L(R)+(E^H(R^H) - E^L(R^H)) \qquad (12)$$

where subscripts H;L refer to high and low levels of theory correspondingly. The high-level treatment is restricted to a smaller portion of the system (R^H), while the low level of theory goes over the entire space (R). The second term in the above equation takes care of overcounting. The NWChem ONIOM module implements two- and three-layer ONIOM models for use in energy, gradient, geometry optimization, and vibrational frequency calculations with any of the pure QM methods within NWChem.

A new development in hybrid method capabilities of NWChem involves classical density functional theory (cDFT).[353–355] The latter represents a classical variant of electronic structure DFT, where the main variable is the classical density of the atoms.[356,357] Conceptually, this type of description lies between continuum and classical force field models, providing orders of magnitude improvements over classical MD simulations. The approach is based on in-corporating important structural features of the environment in the form of classical correlation functions. This allows for efficient and reliable calculations of thermodynamical quantities, providing an essential link between electronic structure description at the atomistic level and phenomena observed at the macroscopic scale.

VII. PARALLEL PERFORMANCE

The design and development of NWChem from the out-set was driven by parallel scalability and performance to enable large scale calculations and achieve fast time-to-solution by using many CPUs where possible. The parallel tools outlined in section IV provided the programming framework for this.

The advent of new architectures such as the GPU[358] platforms have required the parallel coding strategy within NWChem to be revisited. At present, the coupled-cluster code within TCE can utilize both the CPU and GPU hard-ware at a massive scale[32,359]. The emergence of many-core processors in the last ten years provided the opportunity for starting a collaborative effort with Intel corporation to opti-mize NWChem on this new class of computer architecture. As part of this collaboration, the TCE implementation of the CCSD(T) code was ported to the Intel Xeon Phi line of many-core processors[35] using a parallelization strategy based on a hybrid GA-OpenMP approach. The ab initio plane-wave molecular dynamics code (section V F) has also been optimized to take full advantage of these Intel many-core processors[33,231].

In the rest of this section, we will discuss the paral-lel scalability and performance of the main capabilities in NWChem.

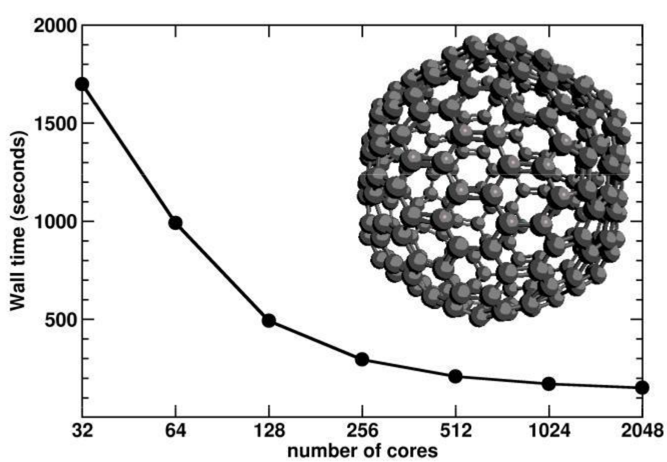

Figure 5. C_{240} DFT benchmark.

a. Gaussian Basis Density Functional Theory: In Fig. 5, we report the parallel performance of the Gaussian basis set DFT module in NWChem. This calculation involved performing a PBE0 energy calculation (four SCF iterations in direct mode) on the C_{240} molecule with the 6-31G* basis set (3600 basis functions) without symmetry. These calculations were performed on the Cascade supercomputer located at PNNL.

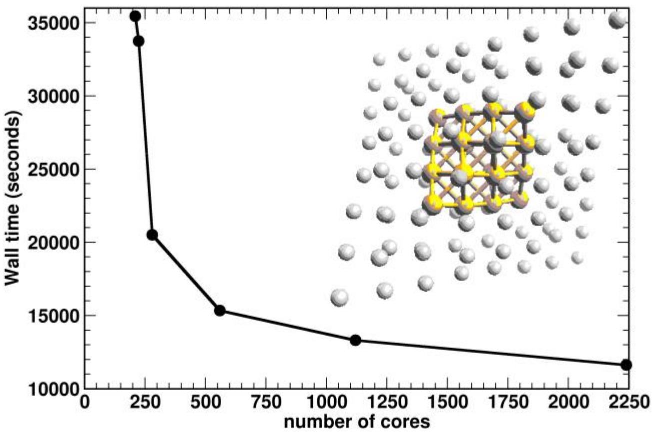

Figure 6. LR-TDDFT benchmark for the Au_{20} molecule in a neon matrix.

b. Time-Dependent Density Functional Theory: In Fig. 6, we report the parallel performance of the Gaussian basis set LR-TDDFT module in NWChem. This calculation involved computing 100 excitation energies, requiring 11 Davidson iterations, for the Au_{20} molecule surrounded by a matrix of 80 Ne atoms[360] (1840 basis functions) with D_2 symmetry, using the B3LYP functional. These calculations were performed on the Cascade supercomputer located at PNNL.

c. Closed-shell CCSD(T): The parallel implemen-tation of the CCSD(T) approach by Kobayashi and Rendell[132], employing the spin adaptation scheme based on the unitary group approach (UGA)[133] within NWChem, was one of the first scalable implementations of the CC formalism capable of taking advantage of several hundred processors. This implementation was used in simulations involving tera- and peta-scale architectures where chemical accuracy is required to describe ground-state potential energy surfaces. One of the best illustrations of the performance of the CCSD(T) implementation is provided by calculations for water clusters[26]. In the largest calcula-tion, $(H_2O)_{24}$, sustained performance of 1.39 PetaFLOP/s (double precision) on 223,200 processors of ORNL's Jaguar system was documented. This impressive performance was mostly attributed to the (T)-part characterized by $n_o^3 n_u^4$ numerical overhead (where n_o and n_u refer to the total numbers of correlated occupied and virtual orbitals) and its relatively low communication footprint.

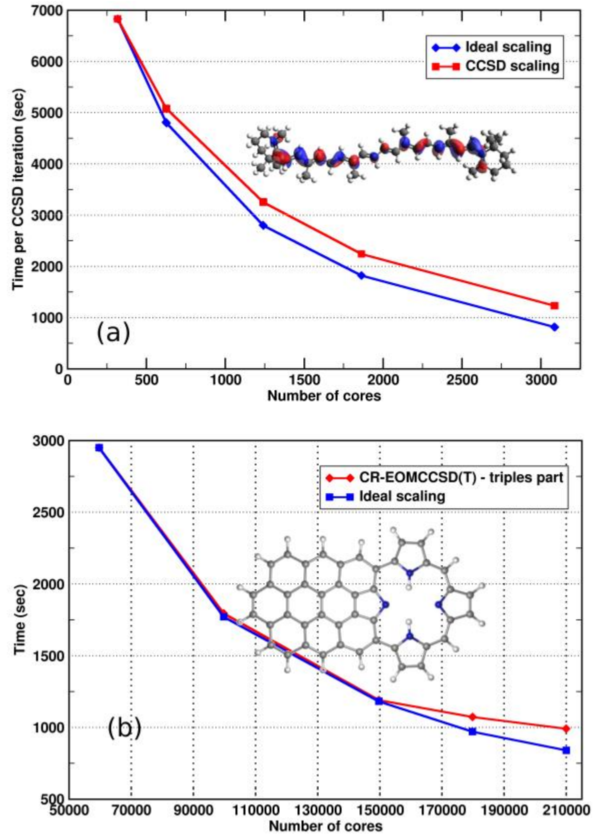

Figure 7. Benchmark EOMCC scalability tests for : (a) beta-carotene and (b) free-base porphyrin (FBP) fused coronene. Timings for CR-EOMCCSD(T) approach for the coronene fused free-base porphyrin in the AVTZ basis set were determined from calculations on the ORNL's Jaguar Cray XT5 computer system.

d. Tensor Contraction Engine: The TCE has enabled parallel CC/EOMCC/LR-CC calculations for closed- and open-shell systems characterized by 1,000-1,300 orbitals. Some of the most illustrative examples include calcu-lations for static and frequency-dependent polarizabili-ties for polyacenes and C_{60} molecule,[183,361] excited state simulations for p-conjugated chromophores,[184] and IP-EOMCCSD calculations for carbon nanotubes.[362] A good illustration of the scalability of the TCE module is provided by the application of GA-based TCE implementations of the iterative (CCSD/EOMCCSD) and non-iterative (CR-EOMCCSD(T)) methods in studies of excited states of b -carotene[363] and functionalized forms of porphyrin[28] (see Fig.7(a) and (b), respectively). While non-iterative meth-ods are much easier to scale across a large number of cores (Fig.7 (b)), scalability of the iterative CC methods is less easy to achieve. However, using early task-flow algorithms for TCE CCSD/EOMCCSD methods[28] it was possible to achieve satisfactory scalability in the range of 1,000-8,000 cores.

e. Recent Implementation of Plane-Wave DFT AIMD for Many-Core Architectures: The very high degree of parallelism available on machines with many-core pro-cessors is forcing developers to carefully revisit the im-plementation of their programs in order to make use of this hardware efficiently. In this section, after a brief overview of the computational costs and parallel strategies for AIMD, we present our recent work[33] on adding thread-level parallelism to the AIMD method implemented in NWChem[3,29,364].

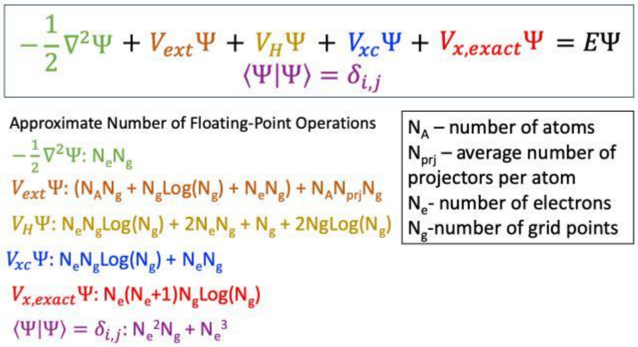

Figure 8. Operation count of Hyi in a plane-wave DFT simulation. Figure from Ref. 231

The main computational costs of an energy minimization or AIMD simulation are the evaluation of the electronic gradient d E_{total} =d y_i = Hy_i and algorithms used to maintain orthogonality. These costs are illustrated in Figure 8. Due to their computational complexity, the electron gradient Hy_i and orthogonalization need to be calculated as efficiently as possible. The main parameters that determine the cost of a calculation are N_g, N_e, N_a, and $N_{pro\ j}$, where N_g is the size of the three-dimensional FFT grid, N_e is the number of occupied orbitals, N_a is the number of atoms, $N_{pro\ j}$ is the number of projectors per atom, and N_{pack} is the size of the reciprocal space.

The evaluation of the electron gradient (and orthogonal-ity) contains three major computational pieces that need to be efficiently parallelized:

- applying V_H and V_{xc}, involving the calculation of 2N_e 3D FFTs;
- calculating the non-local pseudopotential, V_{NL}, dom-inated by the cost of the matrix multiplications W = P^T Y , and Y_2 = PW , where P is an N_{pack} ($N_{pro\ j}$ N_a) matrix, Y and Y_2 are N_{pack} N_e matrices, and W is an ($N_{pro\ j}N_a$) N_e matrix;
- enforcing orthogonality, where the most expensive matrix multiplications are S = Y^T Y and Y_2 = Y S, where Y and Y_2 are N_{pack} N_e matrices, and S is an N_e N_e matrix. In this work, Lagrange multi-plier kernels are used for maintaining orthogonality of Kohn-Sham orbitals[29,365–368].

In Fig. 9 the timing results for a full AIMD simulation of 256 water molecules on 16, 32, 64, 128, 256, and 1024 KNL nodes are shown. The “Cori” system at NERSC was used to run this benchmark. This benchmark was taken from Car Parrinello simulations of 256 H_2O with an FFT grid of N_g = 180^3 (N_e=2056) using the plane-wave DFT module (PSPW) in NWChem. In these timings, the number of threads per node was 66. The size of this benchmark sim-ulation is about 4 times larger than many mid-size AIMD simulations carried out in recent years, e.g. in recent work

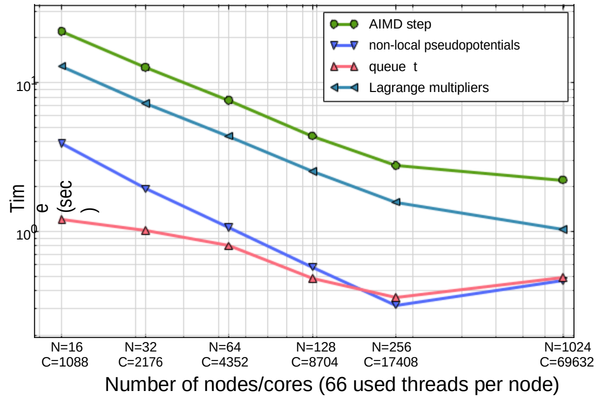

Figure 9. Scalability of major components of an AIMD step on the Xeon Phi partition for a simulation of 256 H_2O molecules. Figure from Bylaska et al.[33].

by Bylaska and co-workers.[279,280,369–372] The overall timings show strong scaling up to 1024 KNL nodes (69632 cores) and the timings of the major kernels, the pipelined 3D FFTs, non-local pseudopotential, and Lagrange multi-plier kernels all displayed significant speedups.

f. Classical Molecular Dynamics: The molecular dynamics module in the current NWChem release is based on the distribution of cells over available ranks in the cal-culation. Simulations exhibit good scalability when cells only require communication with immediately neighboring cells. When the combination of cell size and cutoff radius is such that interactions with atoms in cells beyond the imme-diate neighbors are required, performance is significantly affected. This limits the number of ranks that can effectively be used. For example, a system with 500,000 atoms will only scale well up to 1000 ranks. In future implementations, the cell-cell pair-list will be distributed over the available ranks. While this leads to additional communication for ranks that do not “own” a cell, the implementation of a new communication scheme that avoids global communication has been demonstrated to improve scalability by at least an order of magnitude.[304]

VIII. OUTREACH

Given the various electronic structure methods available in NWChem, it does not come as a surprise that many of these functionalities have been integral to various projects focused on extensions of quantum chemical capabilities to exa-scale architectures and emerging quantum computing (see Fig. 10 for a pictorial representation of recent developments). Below we describe several examples of such a synergy.

a. Interfacing with Other Software: Over the years, many open-source and commercial developers have been using NWChem as a resource for their capability development, and building add-on tools to increase the code’s usability. Various open-source and commercial platforms provide user interface capabilities to set up and analyze the results of calculations that can be performed with NWChem.[253,255,373–380] NWChem initially developed its own graphical user interface called the Extensible Computational Chemistry Environment[381], which is cur-rently supported by a group of open-source developers. In addition, multiple codes use quantities from the NWChem simulation, such as wavefunctions as input for the calculation of additional properties not directly available in the code.[382–391] NWChem is able to export electrostatic potential and charge densities with the Gaussian cube format[392] and can use the Molden format[393] to write or read molecular orbitals. This allows codes[394–398] to utilize NWChem's data to, for example, display charge densities and electrostatic potentials. NWChem can also generate AIM wavefunction files that have been used by a variety of codes to calculate various properties.[76,399–401] Recently, NWChem has also been interfaced with the SEMIEMP code[402], which can be used to perform real-time electronic dynamics using the INDO/S Hamiltonian.[403,404]

b. Common Component Architecture: It is an attrac-tive idea to encapsulate complex scientific applications as components with standardized interfaces. The components interact only through these well-defined interfaces and can be combined into full applications. The main motivation is to be able to reuse and swap components as needed and seamlessly create complex applications. There have been a few attempts to introduce this approach to the scientific software development community. The most notable DOE-led effort was the Common Component Architecture (CCA) Forum[405], which was launched in 1998 as a scientific com-munity effort to create components designed specifically for the needs of high-performance scientific computing. A more recent development is the rise of Simulation Develop-ment Environment (SDE) framework[406], which has features that are related to the components of CCA.

NWChem developers have participated in CCA and SDE effort resulting in the creation of the NWChem component. As an example, the NWChem CCA component was used in the building applications for molecular geometry opti-mization from multiple quantum chemistry and numerical optimization packages[407], combination of multiple theoret-ical methods to improve multi-level parallelism[408], demon-stration of multi-level parallelism[409], and standardization of integral interfaces in quantum chemistry[410]. In the end, the CCA framework was too cumbersome to use for de-velopers, requiring significant efforts to develop interfaces and making components to work together. It resulted in the retirement of CCA Forum in 2010, but the work done on standardization of interfaces is continuing to benefit the quantum chemistry community to this day.

c. NWChemEx: The NWChemEx project is a natural extension of NWChem to overcome the scalability challenges associates with migrating the current code base to exa-scale platforms. NWChemEx is being developed to address two outstanding problems in advanced biofuels research: (i) development of a molecular understanding of proton controlled membrane transport processes and (ii) development of catalysts for the efficient conversion of

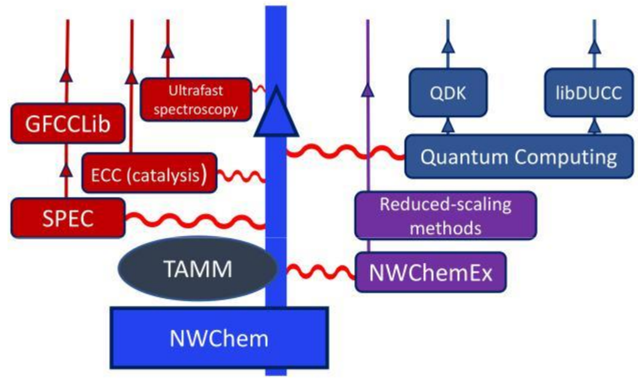

Figure 10. A "connected diagram" describing ongoing efforts to extend computational chemistry models to exa-scale and quantum computing. In each case, NWChem provides a testing and devel-opment platform. A significant role in these projects is played by Tensor Algebra for Many-body Methods (TAMM) library. The ECC acronym stands for the Exa-scale Catalytic Chemistry project supported by BES.[411] The QDK-NWChem interface with the lib-DUCC library is used for downfolding electronic Hamiltonians.[412]

biomass-derived intermediates into biofuels, hydrogen, and other bioproducts; therefore the main focus is on enabling scalable implementations of the ground-state canonical CC formalisms utilizing Cholesky decomposed form of the two-electron integrals[413–418] as well as linear scaling CC formulations based on the domain-based local pair natural orbital CC formulations (DLPNO-CC).[419–421] and embedding methods.

d. Scalable Predictive methods for Excitations and Correlated phenomena (SPEC): The main focus of the SPEC software project is to provide the users with a new generation of methodologies to simulate excited states and excited-state processes using existing peta- and emerging exa-scale architectures. These new capabilities will play an important role in supporting the experimental efforts at light source facilities, which require accurate and reliable modeling tools. The existing NWChem capabilities are being used to verify and validate SPEC implementations including excitation energy, ionization potential, and electron affinity variants of the EOMCC theory as well as hierarchical Green's function formulations ranging from the lower order GW+Bethe-Salpeter equation (GW+BSE)[422] to hierarchical coupled-cluster Green's function (GFCC) methods[185–188,423] and multi-reference CC methods.

e. Quantum Information Sciences: Quantum comput-ing not only offers the promise of overcoming exponential computational barriers of conventional computing but also in achieving the ultimate level of accuracy in studies of chal-lenging processes involving multi-configurational states in catalysis, biochemistry, photochemistry, and materials sci-ence to name only a few areas where quantum information technologies can lead to the transformative changes in the way how quantum simulations are performed. NWChem, with its computational infrastructure to characterize second-quantized forms of electronic Hamiltonians in various basis sets (Gaussian and plane-waves) and with wavefunction methodologies to provide an initial characterization of the ground- and excited-state wavefunctions, can be used as a support platform for various types of quantum simulators. The recently developed QDK-NWChem interface[424] (QDK designates Quantum Development Kit developed by Microsoft Research team) for quantum simulations and libraries for CC downfolded electronic Hamiltonians for quantum computing[412] are good illustrations of the utilization of NWChem in supporting the quantum computing effort.

IX. DATA AVAILABILITY STATEMENT

The data that support the findings of this study are avail-able from the corresponding author upon reasonable re-quest.

X. CONCLUSIONS

The NWChem project is an example of a successful co-design effort that harnesses the expertise and experience of researchers in several complementary areas, including quan-tum chemistry, applied mathematics, and high-performance computing. Over the last three decades, NWChem has evolved into a code that offers a unique combination of computational tools to tackle complex chemical processes at various spatial and time scales.

In addition to the development of new methodologies, NWChem is being continuously upgraded, with new algorithms, to take advantage of emerging computer architectures and quantum information technologies. We believe the community model of NWChem will continue to spur exciting new developments well into the future.

SUPPLEMENTARY MATERIAL

See supplementary material for tutorial slides showing examples of NWChem input files.

ACKNOWLEDGMENTS

The authors wish to acknowledge the important contri-butions to the NWChem project of the following deceased researchers: Ricky A. Kendall, Jarek Nieplocha and Daniel W. Silverstein.

We would like to thank people that have helped the progress of the NWChem project by providing valuable feedback, scientific direction, suggestions, discussions and encouragement. This (incomplete) list includes: A. Ander-sen, J. Andzelm, R. Baer, A. Bick, V. Blum, D. Bowman, E. A. Carter, A. Chakraborty, G. Cisneros, D. Clerc, A. J. Cohen, L. R Corrales, A. R. Felmy, A. Fortunelli, J. Fulton, D. A. Dixon, P. Z. El-Khoury, D. Elwood, G. Fanourgakis, D. Feller, B. C. Garrett, B. Ginovska, E. Glendening, V. A. Glezakou, M. F. Guest, M. Gutowski, M. Hackler, M. Hanwell, W. L. Hase, A. C. Hess, W. M. Holden, C. Huang, W. Huang, E. S. Ilton, E. P. Jahrman, J. Jakowski, J. E. Jaffe, J. Ju, S. M. Kathmann, R. Kawai, M. Khalil, X. Krokidis, L. Kronik, P. S. Krstic,´ R. Kutteh, X. Long, A. Migliore, E. Miliordos, M. Nooijen, J. Li, L. Lin, S. Liu, N. Maitra, M. Malagoli, B. Moore II, P. Mori, S. Mukamel, C. J. Mundy, V. Meunier, V. Murugesan, D. Neuhauser, S. Niu, R. M. Olson, S. Pamidighantam, M. Pavanello, M. P. Prange, C. D. Pemmaraju, D. Prendergast, S. Raugei, J. J. Rehr, A. P. Rendell, R. Renslow, T. Risthaus, K. M. Rosso, R. Rousseau, J. R. Rustad, G. Sandrone, G. K. Schenter, G. T. Seidler, H. Sekino, P. Sherwood, C. Skylaris, H. Song, L. Subramanian, B. G. Sumpter, P. Sushko, S. Tretiak, R. M. Van Ginhoven, B. E. Van Kuiken, J. H. van Lenthe, B. Veeraraghavan, E. Vorpagel, X. B. Wang, J. Warneke, Y. A. Wang, J. C. Wells, S. S. Xantheas, W. Yang, J. Zador, Y. Zhang.

The core development team acknowledges support from the following projects at the Pacific Northwest National Laboratory. Pacific Northwest National Laboratory is operated by Battelle Memorial Institute for the U.S. Department of Energy under Contract DE-AC05-76RL01830: (i) Environmental and Molecular Sciences Laboratory (EMSL), the Construction Project, and Operations, the Office of Biolog-ical and Environmental Research, (ii) the Office of Basic Energy Sciences, Mathematical, Information, and Com-putational Sciences, Division of Chemical Sciences, Geo-sciences, and Biosciences (CPIMS, AMOS, Geosciences, Heavy Element Chemistry, BES Initiatives: CCS-SPEC, CCS-ECC, BES-QIS, BES-Ultrafast), (iii) the Office of Advanced Scientific Computing Research through the Sci-entific Discovery through Advanced Computing (SciDAC), Exascale Computing Project (ECP): NWChemEx. Addi-tional funding was provided by the Office of Naval Re-search, the U.S. DOE High Performance Computing and Communications Initiative and industrial collaborations (Cray, Intel, Samsung).

The work related to the development of QDK-NWChem interface was supported by the “Embedding Quantum Computing into Many-body Frameworks for Strongly Correlated Molecular and Materials Systems” project, which is funded by the U.S. Department of Energy(DOE), Office of Science, Office of Basic Energy Sciences, the Division of Chemi-cal Sciences, Geosciences, and Biosciences, and Quantum Algorithms, Software, and Architectures (QUASAR) Initia-tive at Pacific Northwest National Laboratory (PNNL).

S. A. Fischer acknowledges support from the U.S. Of-fice of Naval Research through the U.S. Naval Research Laboratory.

A. J. Logsdail acknowledges support from the UK EP-SRC under the “Scalable Quantum Chemistry with Flexible Embedding” Grants EP/I030662/1 and EP/K038419/1.

A. Otero-de-la-Roza acknowledges support from the Spanish government for a Ramon´ y Cajal fellowship (RyC-2016-20301) and for financial support (projects PGC2018-097520-A-100 and RED2018-102612-T).

J. Autschbach acknowledges support from the U.S. Department of Energy, Office of Basic Energy Sciences, Heavy Element Chemistry program (Grant No. DE-SC0001136, relativistic methods & magnetic resonance parameters) and the National Science Foundation (Grant No. CHE-1855470, dynamic response methods).

D. Mejia Rodriguez acknowledges support from the U.S. Department of Energy Grant No. DE-SC0002139.

D. G. Truhlar acknowledges support from the NSF under grant no. CHE–1746186.

E. D. Hermes was supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Chemi-cal Sciences, Geosciences and Biosciences Division, as part of the Computational Chemistry Sciences Program (Award Number: 0000232253). Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Se-curity Administration under contract DE-NA0003525.

Z. Lin acknowledges support from the National Natu-ral Science Foundation of China (Grants 11574284 and 11774324).

J. Garza thanks CONACYT for the project FC-2016/2412.

K. Lopata gratefully acknowledges support by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Early Career Program, under Award No. DE-SC0017868 and U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Award No. DE-SC0012462

L. Gagliardi and C. J. Cramer acknowledge support from by the Inorganometallic Catalyst Design Center, an Energy Frontier Research Center funded by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences (BES) (DE-SC0012702). They acknowledge the Minnesota Supercomputing Institute (MSI) at the University of Min-nesota for providing computational resources.

This work used resources provided by EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory (PNNL) and PNNL Institutional Computing (PIC). PNNL is operated by Bat-telle Memorial Institute for the United States Department of Energy under DOE Contract No. DE-AC05-76RL1830.

The work also used resources provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

This work also used resources provided by Oak Ridge Leadership Computing Facility (OLCF) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

REFERENCES

[1] D. E. Bernholdt, E. Apra, H. A. Fruchtl,¨ M. F. Guest, R. J. Harrison, R. A. Kendall, R. A. Kutteh, X. Long, J. B. Nicholas, J. A. Nichols, H. L. Taylor, A. T. Wong, G. I. Fann, R. J. Littlefield, and J. Nieplocha, International Journal of Quantum Chemistry 56, 475 (1995).

[2] R. A. Kendall, E. Apra,` D. E. Bernholdt, E. J. Bylaska, M. Dupuis, G. I. Fann, R. J. Harrison, J. Ju, J. A. Nichols, J. Nieplocha, T. P. Straatsma, T. L. Windus, and A. T. Wong, Computer Physics Communications 128, 260 (2000).

[3] M. Valiev, E. J. Bylaska, N. Govind, K. Kowalski, T. P. Straatsma, H. J. J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T. L. Windus, and W. A. de Jong, Computer Physics Communications 181, 1477 (2010).

[4] R. J. Harrison, J. A. Nichols, T. P. Straatsma, M. Dupuis, E. J. By-laska, G. I. Fann, T. L. Windus, E. Apra, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, W. A. de Jong, M. Deegan, K. Dyall, D. Elwood, H. Fruchtl, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nieplocha, A. Rendall, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, K. Wolin-ski, A. Wong, and Z. Zhang, NWChem, A Computational Chemistry Package for Parallel Computers, version 4.1, Pacific Northwest Na-tional Laboratory, Richland, Washington (2000).

[5] T. P. Straatsma, E. Apra, T. Windus, E. Bylaska, W. de Jong, S. Hirata, M. Valiev, M. Hackler, L. Pollack, R. Harrison, M. Dupuis, D. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, A. Auer, E. Brown, G. Cis-neros, G. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z. Zhang, NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.6 (2004), Pacific Northwest National Laboratory, Richland, Washington (2004).

[6] E. Apra, T. Windus, T. P. Straatsma, E. Bylaska, W. de Jong, S. Hi-rata, M. Valiev, M. Hackler, L. Pollack, K. Kowalski, R. Harrison, M. Dupuis, D. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, A. Auer, E. Brown, G. Cisneros, G. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Dee-gan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Little-field, X. Long, B. Meng, T. Nakajima, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z. Zhang, NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7, Pacific Northwest National Laboratory, Richland, Wash-ington (2005).

[7] E. J. Bylaska, W. A. de Jong, N. Govind, K. Kowalski, T. P. Straatsma, M. Valiev, D. Wang, E. Apra, T. L. Windus, J. Hammond, M. Hackler, Y. Zhao, R. Harrison, M. Dupuis, D. M. A. Smith, A. A. Auer, E. Brown, G. Cisneros, G. I. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. A. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, L. Pollack, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z. Zhang, NWChem, A Computational Chemistry Package for Parallel Computers, Version 5.1, Pacific Northwest National Laboratory, Richland, Washington (2007).

[8]"NWChem - High-Performance Computational Chemistry Software," https://github.com/nwchemgit/nwchem/wiki (2020).

[9]A. Szabo and N. S. Ostlund, Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory (Courier Corporation, 2012).

[10]C. J. Cramer, Essentials of Computational Chemistry: Theories and Models (John Wiley & Sons, 2013).

[11]F. Jensen, Introduction to Computational Chemistry (John Wiley & Sons, 2017).

12 V. Fock, Zeitschrift fur¨ Physik 61, 126 (1930).

13 J. C. Slater, Physical Review 35, 210 (1930).

14 V. Fock, Zeitschrift fur¨ Physik 62, 795 (1930).

15 D. R. Hartree and W. Hartree, Proceedings of the Royal Society of London. Series A - Mathematical and Physical Sciences 150, 9 (1935).

16 P. Hohenberg and W. Kohn, Physical Review 136, B864 (1964).

17 W. Kohn and L. J. Sham, Physical Review 140, A1133 (1965).

[18]R. G. Parr and W. Yang, Density-Functional Theory of Atoms and Molecules (International Series of Monographs on Chemistry) (Oxford University Press, USA, 1994).

19 R. Car and M. Parrinello, Physical Review Letters 55, 2471 (1985).

20 D. R. Hamann, Physical Review B 40, 2980 (1989).

21 N. Troullier and J. L. Martins, Physical Review B 43, 1993 (1991).

22 G. B. Bachelet, D. R. Hamann, and M. Schluter,¨ Physical Review B 26, 4199 (1982).

23 P. E. Blochl,¨ Physical Review B 50, 17953 (1994).

[24]J. Anchell, E. Apra, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, M. Dupuis, K. Dyall, G. Fann, H. Fruchtl, M. Gutowski, R. Harrison, A. Hess, J. Jaffe, R. Kendall, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, J. Nichols, J. Nieplocha, A. Rendall, M. Stave, T. P. Straatsma, H. Taylor, G. Thomas, K. Wolinski, and A. Wong, NWChem, Version 3.2, High Performance Computational Chemistry Group, Pacific Northwest Na-tional Laboratory, Richland, WA (1998).

25 T. P. Straatsma, E. J. Bylaska, H. J. J. van Dam, N. Govind, W. A. de Jong, K. Kowalski, and M. Valiev, Annual Reports of Computational Chemistry 7, 151 (2011).

[26]E. Apra,` A. P. Rendell, R. J. Harrison, V. Tipparaju, W. A. de Jong, and S. S. Xantheas, in Proceedings of the conference on high performance computing networking, storage and analysis (ACM, 2009) p. 66.

27V. Tipparaju, E. Apra,` W. Yu, and J. S. Vetter, in Proceedings of the 7th ACM International Conference on Computing Frontiers (ACM, 2010) pp. 207–216.

[28]K. Kowalski, S. Krishnamoorthy, R. M. Olson, V. Tipparaju, and E. Apra, in Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis (ACM, 2011) p. 72.

29 E. J. Bylaska, K. Tsemekhman, N. Govind, and M. Valiev, in Compu-tational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology, edited by J. R. Reimers (Wiley Online Library, 2011) pp. 77–116.

30 G. D. Purvis and R. J. Bartlett, The Journal of Chemical Physics 76, 1910 (1982).

31 K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chemical Physics Letters 157, 479 (1989).

32 W. Ma, S. Krishnamoorthy, O. Villa, and K. Kowalski, Journal of Chemical Theory and Computation 7, 1316 (2011).

[33]E. J. Bylaska, M. Jacquelin, W. A. de Jong, J. Hammond, and M. Klemm, in High Performance Computing, edited by J. M. Kunkel, R. Yokota, M. Taufer, and J. Shalf (Springer International Publishing, Cham, 2017) pp. 404–418.

[34]E. J. Bylaska, E. Apra,` K. Kowalski, M. Jacquelin, W. A. de Jong, A. Vishnu, B. Palmer, J. Daily, T. P. Straatsma, and J. R. Hammond, "Transitioning NWChem to the Next Generation of Manycore Ma-chines," in Exascale Scientific Applications: Scalability and Perfor-mance Portability (CRC Press, 2017) p. 165.

35 E. Apra, M. Klemm, and K. Kowalski, in Proceedings of the Inter-national Conference for High Performance Computing, Networking, Storage and Analysis (IEEE Press, 2014) pp. 674–684.

[36]J. Nieplocha, R. Harrison, and R. Littlefield, SIAM News 28, 12 (1995).

37 J. Nieplocha, R. J. Harrison, and R. J. Littlefield, The Journal of Supercomputing 10, 169 (1996).

[38]J. Nieplocha, B. Palmer, V. Tipparaju, M. Krishnan, H. Trease, and E. Apra,` The International Journal of High Performance Computing Applications 20, 203 (2006).

[39]M. Krishnan, B. Palmer, A. Vishnu, S. Krishnamoorthy, J. Daily, and D. Chavarria, The Global Arrays User Manual, Pacific Northwest National Laboratory, Richland, Washington (2012).

40 J. Nieplocha, V. Tipparaju, M. Krishnan, and D. K. Panda, The Inter-national Journal of High Performance Computing Applications 20, 233 (2006).

41 J. Daily, A. Vishnu, B. Palmer, H. van Dam, and D. Kerbyson, in 2014 21st International Conference on High Performance Computing (HiPC) (IEEE, 2014) pp. 1–10.

42 J. Dinan, P. Balaji, J. R. Hammond, S. Krishnamoorthy, and V. Tippa-raju, in 2012 IEEE 26th International Parallel and Distributed Process-ing Symposium (IEEE, 2012) pp. 739–750.

43 A. T. Wong and R. J. Harrison, Journal of Computational Chemistry 16, 1291 (1995).

44 I. T. Foster, J. L. Tilson, A. F. Wagner, R. L. Shepard, R. J. Harri-son, R. A. Kendall, and R. J. Littlefield, Journal of Computational Chemistry 17, 109 (1996).

45 R. A. Kendall and H. A. Fruchtl,¨ Theoretical Chemistry Accounts 97, 158 (1997).

46 See https://tinyurl.com/nwchem-xc for Summary table of XC functionals available in NWChem.

47 B. I. Dunlap, J. W. D. Connolly, and J. R. Sabin, The Journal of Chemical Physics 71, 3396 (1979).

48B. Palmer, J. Nieplocha, and E. Apra,` in 2003 Proceedings IEEE International Conference on Cluster Computing (2003) pp. 420–428.

[49]G. I. Fann and R. J. Littlefield, in Proc. 6th SIAM Conference on Parallel Processing for Scientific Computing (SIAM, Philadelphia, PA, USA, 1993).

[50]G. Fann, R. Littlefield, and D. Elwood, in Proc. High Performance Computing '95, Simulation MultiConference (Society for Computer Simulation, San Diego, CA, USA, 1995).

[51]I. S. Dhillon, G. Fann, and B. N. Parlett, in SIAM Conference on Parallel Processing for Scientific Computing (SIAM, 1997).

[52]L. S. Blackford, J. Choi, A. Cleary, E. D'Azeuedo, J. Demmel, I. Dhillon, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley, ScaLAPACK User's Guide, edited by J. J. Dongarra (Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997).

[53]T. Auckenthaler, V. Blum, H.-J. Bungartz, T. Huckle, R. Johanni, L. Kramer,¨ B. Lang, H. Lederer, and P. R. Willems, Parallel Com-puting 37, 783 (2011).

[54]A. Marek, V. Blum, R. Johanni, V. Havu, B. Lang, T. Auckenthaler, A. Heinecke, H.-J. Bungartz, and H. Lederer, The Journal of Physics: Condensed Matter 26, 213201 (2014).

55 P. Pulay, Chemical Physics Letters 73, 393 (1980).

56 V. R. Saunders and I. H. Hillier, International Journal of Quantum Chemistry 7, 699 (1973).

57 M. F. Guest and V. R. Saunders, Molecular Physics 28, 819 (1974).

58 R. W. Warren and B. I. Dunlap, Chemical Physics Letters 262, 384 (1996).

59 S. Refaely-Abramson, S. Sharifzadeh, N. Govind, J. Autschbach, J. B. Neaton, R. Baer, and L. Kronik, Physical Review Letters 109, 226405 (2012).

60T. Stein, J. Autschbach, N. Govind, L. Kronik, and R. Baer, The Journal of Physical Chemistry Letters 3, 3740 (2012).

61 M. Srebro and J. Autschbach, Journal of Chemical Theory and Compu-tation 8, 245 (2012).

62 B. Moore, M. Srebro, and J. Autschbach, Journal of Chemical Theory and Computation 8, 4336 (2012).

63 J. Autschbach and M. Srebro, Accounts of Chemical Research 47, 2592 (2014).

64 H. Sun and J. Autschbach, ChemPhysChem 14, 2450 (2013).

65 H. Sun and J. Autschbach, Journal of Chemical Theory and Computa-tion 10, 1035 (2014).

66 B. Moore, H. Sun, N. Govind, K. Kowalski, and J. Autschbach, Journal of Chemical Theory and Computation 11, 3305 (2015).

67 J. P. Perdew and A. Zunger, Physical Review B 23, 5048 (1981).

68 J. Garza, R. Vargas, J. A. Nichols, and D. A. Dixon, The Journal of Chemical Physics 114, 639 (2001).

69 R. T. Sharp and G. K. Horton, Physical Review 90, 317 (1953).

70 J. D. Talman and W. F. Shadwick, Physical Review A 14, 36 (1976).

71 Y. Krieger, J. B .and Li and G. J. Iafrate, Physical Review A 45, 101

(1992).
72 Y. Li, J. B. Krieger, and G. J. Iafrate, Physical Review A 47, 165 (1993).
73 M. E. Casida and D. R. Salahub, The Journal of Chemical Physics 113, 8918 (2000).
74 S. Hirata, C.-G. Zhan, E. Apra,` T. L. Windus, and D. A. Dixon, The Journal of Physical Chemistry A 107, 10154 (2003).
75 J. Carmona-Esp´ındola, J. L. Gazquez,´ A. Vela, and S. B. Trickey, Journal of Chemical Theory and Computation 15, 303 (2018).
76 A. Otero-De-La-Roza and E. R. Johnson, The Journal of Chemical Physics 138, 204109 (2013).
77 S. Grimme, J. Antony, S. Ehrlich, and H. Krieg, The Journal of Chem-ical Physics 132, 154104 (2010).
78 S. Grimme, S. Ehrlich, and L. Goerigk, Journal of Computational Chemistry 32, 1456 (2011).
79 Y. Zhao and D. G. Truhlar, Chemical Physics Letters 502, 1 (2011).
80 A. D. Becke, The Journal of Chemical Physics 88, 2547 (1988).
81 R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, Chemical Physics Letters 257, 213 (1996).
82 M. E. Mura and P. J. Knowles, The Journal of Chemical Physics 104, 9848 (1996).
83 V. I. Lebedev and D. Laikov, Doklady Mathematics 59, 477 (1999), translated from Doklady Akademii Nauk, Vol. 36, No. 6, 1999, pp. 741–745.
84 A. Farazdel, M. Dupuis, E. Clementi, and A. Aviram, Journal of the American Chemical Society 112, 4206 (1990).
85 K. M. Rosso, D. M. A. Smith, and M. Dupuis, The Journal of Chemical Physics 118, 6455 (2003).
86 Q. Wu and T. Van Voorhis, Physical Review A 72, 024502 (2005).
87 Q. Wu and T. Van Voorhis, Journal of Chemical Theory and Computa-tion 2, 765 (2006).
88 Q. Wu and T. Van Voorhis, The Journal of Physical Chemistry A 110, 9212 (2006).
89 T. A. Wesołowski and A. Warshel, The Journal of Physical Chemistry 97, 8050 (1993).
90 T. A. Wesołowski, Physical Review A 77, 012504 (2008).
91 Y. Lu, M. R. Farrow, P. Fayon, A. J. Logsdail, A. A. Sokol, C. R. A. Catlow, P. Sherwood, and T. W. Keal, Journal of Chemical Theory and Computation 15, 1317 (2019).
92 J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, The Journal of Physical Chemistry 96, 135 (1992).
93 C. Jamorski, M. E. Casida, and D. R. Salahub, The Journal of Chemical Physics 104, 5134 (1996).
94 R. Bauernschmitt and R. Ahlrichs, Chemical Physics Letters 256, 454 (1996).
95 R. Bauernschmitt, M. Haser,¨ O. Treutler, and R. Ahlrichs, Chemical Physics Letters 264, 573 (1997).
96 S. Hirata and M. Head-Gordon, Chemical Physics Letters 314, 291 (1999).
97 E. Vecharynski, J. Brabec, M. Shao, N. Govind, and C. Yang, Computer Physics Communications 221, 42 (2017).
98 K. Lopata, B. E. Van Kuiken, M. Khalil, and N. Govind, Journal of Chemical Theory and Computation 8, 3284 (2012).
99 Y. Zhang, S. Mukamel, M. Khalil, and N. Govind, Journal of Chemical Theory and Computation 11, 5804 (2015).
100 Y. Zhang, U. Bergmann, R. Schoenlein, M. Khalil, and N. Govind, The Journal of Chemical Physics 151, 144114 (2019).
101 D. W. Silverstein, N. Govind, H. J. Van Dam, and L. Jensen, Journal of Chemical Theory and Computation 9, 5490 (2013).
102 M. Srebro, N. Govind, W. A. de Jong, and J. Autschbach, The Journal of Physical Chemistry A 115, 10930 (2011).
103 J. Autschbach, ChemPhysChem 12, 3224 (2011).
104 Y. Zhang, J. R. Rouxel, J. Autschbach, N. Govind, and S. Mukamel, Chemical Science 8, 5969 (2017).
105 K. Lopata and N. Govind, Journal of Chemical Theory and Computa-tion 7, 1344 (2011).
106 S. Tussupbayev, N. Govind, K. Lopata, and C. J. Cramer, Journal of Chemical Theory and Computation 11, 1102 (2015).
107 S. A. Fischer, C. J. Cramer, and N. Govind, Journal of Chemical Theory and Computation 11, 4294 (2015).
108 D. N. Bowman, J. C. Asher, S. A. Fischer, C. J. Cramer, and N. Govind, Physical Chemistry Chemical Physics 19, 27452 (2017).
109 A. Bruner, S. Hernandez, F. Mauger, P. M. Abanador, D. J. LaMaster, M. B. Gaarde, K. J. Schafer, and K. Lopata, The journal of physical chemistry letters 8, 3991 (2017).
110 D. Cho, J. R. Rouxel, M. Kowalewski, P. Saurabh, J. Y. Lee, and S. Mukamel, The Journal of Physical Chemistry Letters 9, 1072 (2018).
111 A. Bruner, S. M. Cavaletto, N. Govind, and S. Mukamel, Journal of Chemical Theory and Computation 15, 6832 (2019).
112 Y. Wang, K. Lopata, S. A. Chambers, N. Govind, and P. V. Sushko, The Journal of Physical Chemistry C 117, 25504 (2013).
113 A. Castro, M. A. L. Marques, and A. Rubio, The Journal of Chemical Physics 121, 3425 (2004).
114 W. Magnus, Communications on Pure and Applied Mathematics 7, 649 (1954).
115 L. Verlet, Physical Review 159, 98 (1967).
116 W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson, The Journal of Chemical Physics 76, 637 (1982).
117 G. Bussi, D. Donadio, and M. Parrinello, The Journal of Chemical Physics 126, 014101 (2007).
118 G. Bussi and M. Parrinello, Physical Review E 75, 056707 (2007).
119 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, The Journal of Chemical Physics 81, 3684 (1984).
120 S. A. Fischer, T. W. Ueltschi, P. Z. El-Khoury, A. L. Mifflin, W. P. Hess, H. F. Wang, C. J. Cramer, and N. Govind, Journal of Physical Chemistry B 120, 1429 (2016).
121 I. Lindgren and J. Morrison, Atomic many-body theory, Vol. 3 (Springer Science & Business Media, 2012).
122 I. Shavitt and R. J. Bartlett, Many-body methods in chemistry and physics: MBPT and coupled-cluster theory (Cambridge university press, 2009).
123 C. Møller and M. S. Plesset, Physical Review 46, 618 (1934).
124 A. T. Wong, R. J. Harrison, and A. P. Rendell, Theoretica chimica acta 93, 317 (1996).
125 D. E. Bernholdt and R. J. Harrison, Chemical Physics Letters 250, 477 (1996).
126 J. Olsen, P. Jørgensen, and J. Simons, Chemical Physics Letters 169, 463 (1990).
127 K. D. Vogiatzis, D. Ma, J. Olsen, L. Gagliardi, and W. A. de Jong, The Journal of Chemical Physics 147, 184111 (2017).
128 S. Hirata, The Journal of Physical Chemistry A 107, 9887 (2003).
129 G. Baumgartner, A. Auer, D. E. Bernholdt, A. Bibireata, V. Choppella, D. Cociorva, Xiaoyang Gao, R. J. Harrison, S. Hirata, S. Krishnamoor-thy, S. Krishnan, Chi-chung Lam, Qingda Lu, M. Nooijen, R. M. Pitzer, J. Ramanujam, P. Sadayappan, and A. Sibiryakov, Proceedings of the IEEE 93, 276 (2005).
130 J. C´ızek,ˇ The Journal of Chemical Physics 45, 4256 (1966).
131 J.C´ızekˇ and J. Paldus, International Journal of Quantum Chemistry 5, 359 (1971).
132 R. Kobayashi and A. P. Rendell, Chemical Physics Letters 265, 1 (1997).
133 G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, The Journal of Chemical Physics 89, 7382 (1988).
134 T. J. Lee, A. P. Rendell, and J. E. Rice, TITAN – a set of electronic structure programs.
135 A. P. Rendell, T. J. Lee, A. Komornicki, and S. Wilson, Theoretica Chimica Acta 84, 271 (1993).
136 S. Yoo, E. Apra,` X. C. Zeng, and S. S. Xantheas, The Journal of Physical Chemistry Letters 1, 3122 (2010).
137 V. M. Anisimov, G. H. Bauer, K. Chadalavada, R. M. Olson, J. W. Glenski, W. T. C. Kramer, E. Apra,` and K. Kowalski, Journal of Chemical Theory and Computation 10, 4307 (2014).
138 S. Hirata, Theoretical Chemistry Accounts 116, 2 (2006).
139 C. L. Janssen and H. F. Schaefer, Theoretica Chimica Acta 79, 1 (1991).
140 X. Li and J. Paldus, The Journal of Chemical Physics 101, 8812 (1994).
141 M. Nooijen and V. Lotrich, Journal of Molecular Structure: THEOCHEM 547, 253 (2001).
142 M. Nooijen, International Journal of Molecular Sciences 3, 656 (2002).
143 J. A. Parkhill, K. Lawier, and M. Head-Gordon, The Journal of Chemi-cal Physics 130, 084101 (2009).
144 E. Deumens, V. F. Lotrich, A. Perera, M. J. Ponton, B. A. Sanders, and R. J. Bartlett, Wiley Interdisciplinary Reviews: Computational Molecular Science 1, 895 (2011).
145 M. K. MacLeod and T. Shiozaki, The Journal of Chemical Physics 142, 051103 (2015).
146 S. Hirata, The Journal of Chemical Physics 121, 51 (2004).

147 J. Noga and R. J. Bartlett, The Journal of Chemical Physics 86, 7041 (1987).

148 J. Noga and R. J. Bartlett, The Journal of Chemical Physics 89, 3401 (1988).

149 G. E. Scuseria and H. F. Schaefer, Chemical Physics Letters 152, 382 (1988).

150 N. Oliphant and L. Adamowicz, The Journal of Chemical Physics 95, 6645 (1991).

151 S. A. Kucharski and R. J. Bartlett, Theoretical Chemistry Accounts 80, 387 (1991).

152 K. Kowalski and P. Piecuch, The Journal of Chemical Physics 113, 18 (2000).

153 K. Kowalski and P. Piecuch, The Journal of Chemical Physics 122, 074107 (2005).

154 S. R. Gwaltney and M. Head-Gordon, Chemical Physics Letters 323, 21 (2000).

155 S. R. Gwaltney and M. Head-Gordon, The Journal of Chemical Physics 115, 2014 (2001).

156 S. Hirata, P.-D. Fan, A. A. Auer, M. Nooijen, and P. Piecuch, The Journal of Chemical Physics 121, 12197 (2004).

157 J. Geertsen, M. Rittby, and R. J. Bartlett, Chemical Physics Letters 164, 57 (1989).

158 D. C. Comeau and R. J. Bartlett, Chemical Physics Letters 207, 414 (1993).

159 K. Kowalski and P. Piecuch, The Journal of Chemical Physics 115, 643 (2001).

160 S. A. Kucharski, M. Włoch, M. Musiał, and R. J. Bartlett, The Journal of Chemical Physics 115, 8263 (2001).

161 M. Kallay´ and J. Gauss, The Journal of Chemical Physics 121, 9257 (2004).

162 K. Kowalski and P. Piecuch, The Journal of Chemical Physics 120, 1715 (2004).

163 J. F. Stanton and R. J. Bartlett, The Journal of Chemical Physics 98, 7029 (1993).

164 M. Nooijen and R. J. Bartlett, The Journal of Chemical Physics 102, 3629 (1995).

165 J. F. Stanton and J. Gauss, The Journal of Chemical Physics 103, 1064 (1995).

166 M. Musiał and R. J. Bartlett, The Journal of Chemical Physics 119, 1901 (2003).

167 M. Musiał, S. A. Kucharski, and R. J. Bartlett, The Journal of Chemical Physics 118, 1128 (2003).

168 J. R. Gour and P. Piecuch, The Journal of Chemical Physics 125, 234107 (2006).

169 M. Kamiya and S. Hirata, The Journal of Chemical Physics 125, 074111 (2006).

170 M. Kamiya and S. Hirata, The Journal of Chemical Physics 126, 134112 (2007).

171 H. Koch and P. Jørgensen, The Journal of Chemical Physics 93, 3333 (1990).

172 J. Pittner, P. Nachtigall, P. Carsky, J. Masik, and I. Hubac, The Journal of Chemical Physics 110, 10275 (1999).

173 J. Pittner, The Journal of Chemical Physics 118, 10876 (2003).

174 U. S. Mahapatra, B. Datta, and D. Mukherjee, Mol. Phys. 94, 157 (1998).

175 U. S. Mahapatra, B. Datta, B. Bandyopadhyay, and D. Mukher-jee, "State-specific multi-reference coupled cluster formulations: Two paradigms," (Academic Press, 1998) pp. 163–193.

176 U. S. Mahapatra, B. Datta, and D. Mukherjee, The Journal of Chemical Physics 110, 6171 (1999).

177 F. A. Evangelista, W. D. Allen, and H. F. Schaefer III, The Journal of Chemical Physics 127, 024102 (2007).

178 F. A. Evangelista, A. C. Simmonett, W. D. Allen, H. F. Schaefer III, and J. Gauss, The Journal of Chemical Physics 128, 124104 (2008).

179 K. Bhaskaran-Nair, W. Ma, S. Krishnamoorthy, O. Villa, H. J. J. van Dam, E. Apra,` and K. Kowalski, Journal of Chemical Theory and Computation 9, 1949 (2013).

180 E. Apra` and K. Kowalski, Journal of Chemical Theory and Computation 12, 1129 (2016).

181 J. Brabec, J. Pittner, H. J. J. van Dam, E. Apra,` and K. Kowalski, Journal of Chemical Theory and Computation 8, 487 (2012).

182 K. Bhaskaran-Nair, J. Brabec, E. Apra, H. J. J. van Dam, J. Pittner, and K. Kowalski, The Journal of Chemical Physics 137, 094112 (2012).

183 K. Kowalski, J. R. Hammond, W. A. de Jong, and A. J. Sadlej, The Journal of Chemical Physics 129, 226101 (2008).

184 K. Kowalski, R. M. Olson, S. Krishnamoorthy, V. Tipparaju, and E. Apra,` Journal of Chemical Theory and Computation 7, 2200 (2011).

185 M. Nooijen and J. G. Snijders, International Journal of Quantum Chem-istry 44, 55 (1992).

186 M. Nooijen and J. G. Snijders, International Journal of Quantum Chem-istry 48, 15 (1993).

187 M. Nooijen and J. G. Snijders, The Journal of Chemical Physics 102, 1681 (1995).

188 L. Meissner and R. J. Bartlett, International Journal of Quantum Chem-istry 48, 67 (1993).

189 E. Deumens, V. F. Lotrich, A. S. Perera, R. J. Bartlett, N. Jindal, and B. A. Sanders, in Annual Reports in Computational Chemistry, Vol. 7 (Elsevier, 2011) pp. 179–191.

190 E. Solomonik, D. Matthews, J. R. Hammond, J. F. Stanton, and J. Dem-mel, Journal of Parallel and Distributed Computing 74, 3176 (2014).

191 C. Peng, J. A. Calvin, F. Pavosevi˘c,` J. Zhang, and E. F. Valeev, The Journal of Physical Chemistry A 120, 10231 (2016).

192 E. Epifanovsky, M. Wormit, T. Kus,´ A. Landau, D. Zuev, K. Khistyaev, P. Manohar, I. Kaliman, A. Dreuw, and A. I. Krylov, Journal of Com-putational Chemistry 34, 2293 (2013).

193 P. A. M. Dirac, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character 117, 610 (1928).

194 K. G. Dyall and K. Fægri Jr, Introduction to relativistic quantum chem-istry (Oxford University Press, 2007).

195 T. Saue, ChemPhysChem 12, 3077 (2011).

196 M. Reiher and A. Wolf, Relativistic Quantum Chemistry, 2nd ed. (Wiley-VCH, 2015).

197 J. Autschbach, N. Govind, R. Atta-Fynn, E. J. Bylaska, J. W. Weare, and W. A. de Jong, "Computational tools for predictive modeling of properties in complex actinide systems," in Computational Methods in Lanthanide and Actinide Chemistry, edited by M. Dolg (John Wiley & Sons, Ltd, 2015) Chap. 12, pp. 299–342.

198 M. Douglas and N. M. Kroll, Annals of Physics 82, 89 (1974).

199 B. A. Hess, Physical Review A 32, 756 (1985).

200 B. A. Hess, Physical Review A 33, 3742 (1986).

201 C. Chang, M. Pelissier, and P. Durand, Physica Scripta 34, 394 (1986).

202 E. van Lenthe, The ZORA equation, Ph.D. thesis, University of Amster-dam (1996).

203 S. Faas, J. G. Snijders, J. H. van Lenthe, E. van Lenthe, and E. J. Baerends, Chemical Physics Letters 246, 632 (1995).

204 P. Nichols, N. Govind, E. J. Bylaska, and W. A. de Jong, Journal of Chemical Theory and Computation 5, 491 (2009).

205 K. G. Dyall, The Journal of Chemical Physics 100, 2118 (1994).

206 K. G. Dyall, The Journal of Chemical Physics 106, 9618 (1997).

207 K. G. Dyall and T. Enevoldsen, The Journal of Chemical Physics 111, 10000 (1999).

208 D. Peng and M. Reiher, Theoretical Chemistry Accounts 131, 1081 (2012).

209 J. Autschbach, D. Peng, and M. Reiher, Journal of Chemical Theory and Computation 8, 4239 (2012).

210 J. Autschbach, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 372, 20120489 (2014).

211 J. Autschbach, Journal of Chemical Theory and Computation 13, 710 (2017).

212 O. D. Haberlen¨ and N. Rosch,¨ Chemical Physics Letters 199, 491 (1992).

213 T. Nakajima and K. Hirao, Chemical Physics Letters 329, 511 (2000).

214 T. Nakajima and K. Hirao, The Journal of Chemical Physics 113, 7786 (2000).

215 J. H. Van Lenthe, S. Faas, and J. G. Snijders, Chemical Physics Letters 328, 107 (2000).

216 J. H. Van Lenthe and J. N. J. Van Lingen, International Journal of Quantum Chemistry 106, 2525 (2006).

217 C. van Wullen,¨ The Journal of Chemical Physics 109, 392 (1998).

218 C. van Wullen¨ and C. Michauk, The Journal of Chemical Physics 123, 204113 (2005).

219 J. M. Mullin, J. Autschbach, and G. C. Schatz, Computational and Theoretical Chemistry 987, 32 (2012).

220 F. Aquino and G. C. Schatz, Journal of Physical Chemistry A 118, 517

(2014).
221 F. Aquino, N. Govind, and J. Autschbach, Journal of Chemical Theory and Computation 6, 2669 (2010).
222 R. Ditchfield, Molecular Physics 27, 789 (1974).
223 M. Dupuis, Computer Physics Communications 134, 150 (2001).
224 P. Verma and J. Autschbach, Journal of Chemical Theory and Computation 9, 1932 (2013).
225 F. Aquino, N. Govind, and J. Autschbach, Journal of Chemical Theory and Computation 7, 3278–3292 (2011).
226 J. Autschbach, S. Patchkovskii, and B. Pritchard, Journal of Chemical Theory and Computation 7, 2175 (2011).
227 F. Aquino, B. Pritchard, and J. Autschbach, Journal of Chemical Theory and Computation 8, 598 (2012).
228 C. Shen, M. Srebro-Hooper, T. Weymuth, F. Krausbeck, J. T. L. Navarrete, F. J. Ram´ırez, B. Nieto-Ortega, J. Casado, M. Reiher, J. Autschbach, et al., Chemistry–A European Journal 24, 15067 (2018).
229 F. Krausbeck, J. Autschbach, and M. Reiher, The Journal of Physical Chemistry A 120, 9740 (2016).
230 J. R. Hammond, N. Govind, K. Kowalski, J. Autschbach, and S. S. Xantheas, The Journal of Chemical Physics 131, 214103 (2009).
231 M. Jacquelin, W. A. de Jong, and E. J. Bylaska, in 2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS) (IEEE Computer Society, 2017) pp. 234–243.
232 E. J. Bylaska, M. Valiev, R. Kawai, and J. H. Weare, Computer Physics Communications 143, 11 (2002).
233 E. J. Bylaska, K. Tsemekhman, S. B. Baden, J. H. Weare, and H. Jonsson, Journal of Computational Chemistry 32, 54 (2011).
234 E. J. Bylaska, in Annual Reports in Computational Chemistry, Vol. 13 (Elsevier, 2017) pp. 185–228.
235 M. Valiev, E. J. Bylaska, A. Gramada, and J. H. Weare, in Reviews in Modern Quantum Chemistry: A Celebration of the Contributions of Robert G. Parr, edited by K. D. Sen (World Scientific, Singapore, 2002).
236 P. E. Blochl¨ and M. Parrinello, Physical Review B 45, 9413 (1992).
237 E. Cauet,¨ S. A. Bogatko, E. J. Bylaska, and J. H. Weare, Inorganic Chemistry 51, 10856 (2012).
238 B. Roux, Computer Physics Communications 91, 275 (1995).
239 G. Bussi, A. Laio, and M. Parrinello, Physical Review Letters 96, 090601 (2006).
240 A. Barducci, M. Bonomi, and M. Parrinello, Wiley Interdisciplinary Reviews: Computational Molecular Science 1, 826 (2011).
241 L. Maragliano and E. Vanden-Eijnden, The Journal of Chemical Physics 128, 184110 (2008).
242 L. Maragliano and E. Vanden-Eijnden, Chemical Physics Letters 426, 168 (2006).
243 S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, Journal of Computational Chemistry 13, 1011 (1992).
244 J. J. Rehr and A. L. Ankudinov, Journal of Synchrotron Radiation 8, 61 (2001).
245 J. J. Rehr and R. C. Albers, Physical Review B 41, 8139 (1990).
246 A. L. Ankudinov, B. Ravel, J. J. Rehr, and S. D. Conradson, Physical Review B 58, 7565 (1998).
247 E. J. Bylaska and K. Rosso, Journal of Chemical Theory and Computation 14, 4416 (2018).
248 P. L. Silvestrelli, Physical Review B 59, 9703 (1999).
249 A. Damle, L. Lin, and L. Ying, Journal of Chemical Theory and Computation 11, 1463 (2015).
250 R. Kawai and J. H. Weare, The Journal of Chemical Physics 95, 1151 (1991).
251 P. E. Blochl,¨ The Journal of Chemical Physics 103, 7422 (1995).
252 V. Kapil, M. Rossi, O. Marsalek, R. Petraglia, Y. Litman, T. Spura, B. Cheng, A. Cuzzocrea, R. H. Meißner, D. M. Wilkins, et al., Com-puter Physics Communications 236, 214 (2019).
253 A. H. Larsen, J. J. Mortensen, J. Blomqvist, I. E. Castelli, R. Chris-tensen, M. Dułak, J. Friis, M. N. Groves, B. Hammer, C. Hargus, E. D. Hermes, P. C. Jennings, P. B. Jensen, J. Kermode, J. R. Kitchin, E. L. Kolsbjerg, J. Kubal, K. Kaasbjerg, S. Lysgaard, J. B. Maronsson, T. Maxson, T. Olsen, L. Pastewka, A. Peterson, C. Rostgaard, J. Schiøtz, O. Schutt,¨ M. Strange, K. S. Thygesen, T. Vegge, L. Vil-helmsen, M. Walter, Z. Zeng, and K. W. Jacobsen, Journal of Physics: Condensed Matter 29, 273002 (2017).
254 G. Klimeck, M. McLennan, S. P. Brophy, G. B. Adams III, and M. S. Lundstrom, Computing in Science & Engineering 10, 17 (2008).
255 E. J. Bylaska, "EMSL Arrows," https://arrows.emsl.pnnl.gov/ api (2020).
256 E. J. Bylaska, K. Tsemekhman, and F. Gao, Physica Scripta T124, 86 (2006).
257 D. C. Langreth, M. Dion, H. Rydberg, E. Schroder,¨ P. Hyldgaard, and B. I. Lundqvist, International Journal of Quantum Chemistry 101, 599 (2005).
258 C. Hartwigsen, S. Gœdecker, and J. Hutter, Physical Review B 58, 3641 (1998).
259 S. Goedecker, M. Teter, and J. Hutter, Physical Review B 54, 1703 (1996).
260 D. Hamann, Physical Review B 88, 085117 (2013).
261 M. Schlipf and F. Gygi, Computer Physics Communications 196, 36 (2015).
262 R. G. Parr and W. Yang, Horizons of Quantum Chemistry (Oxford university press, 1994).
263 W. E. Pickett, Reviews of Modern Physics 61, 433 (1989).
264 J. Ihm, A. Zunger, and M. L. Cohen, Journal of Physics C: Solid State Physics 12, 4409 (1979).
265 M. C. Payne, M. P. Teter, D. C. Allan, T. Arias, and J. D. Joannopoulos, Reviews of Modern Physics 64, 1045 (1992).
266 D. K. Remler and P. A. Madden, Molecular Physics 70, 921 (1990).
267 G. Kresse and J. Furthmuller,¨ Physical Review B 54, 11169 (1996).
268 D. Marx and J. Hutter, "Ab initio molecular dynamics: Theory and implementation," in Modern methods and algorithms of quantum chem-istry, Vol. 1, edited by J. Grotendorst (NIC, 2000) pp. 301–449.
269 R. M. Martin, Electronic Structure: Basic Theory and Practical Meth-ods (Cambridge University Press, 2004).
270 Y. Chen, E. J. Bylaska, and J. H. Weare, in Molecular Modeling of Geochemical Reactions: An Introduction, edited by J. R. Kubicki (John Wiley & Sons, Ltd, 2016) pp. 107–149.
271 J. C. Phillips, Physical Review 112, 685 (1958).
272 J. C. Phillips and L. Kleinman, Physical Review 116, 287 (1959).
273 B. J. Austin, V. Heine, and L. J. Sham, Physical Review 127, 276 (1962).
274 M. T. Yin and M. L. Cohen, Physical Review B 25, 7403 (1982).
275 N. A. W. Holzwarth, G. E. Matthews, R. B. Dunning, A. R. Tackett, and Y. Zeng, Physical Review B 55, 2005 (1997).
276 G. Kresse and D. Joubert, Physical Review B 59, 1758 (1999).
277 M. Valiev and J. H. Weare, The Journal of Physical Chemistry A 103, 10588 (1999).
278 R. Atta-Fynn, E. J. Bylaska, and W. A. De Jong, MRS Online Proceed-ings Library Archive 1383, mrsf11–1383 (2012).
279 R. Atta-Fynn, E. J. Bylaska, and W. A. de Jong, The Journal of Physical Chemistry Letters 4, 2166 (2013).
280 S. O. Odoh, E. J. Bylaska, and W. A. de Jong, Journal of Physical Chemistry A 117, 12256 (2013).
281 R. Atta-Fynn, E. J. Bylaska, and W. A. de Jong, The Journal of Physical Chemistry A 120, 10216 (2016).
282 R. Atta-Fynn, E. J. Bylaska, and W. A. de Jong, " Finite tem-perature free energy calculations in NWChem:Metadynamics and Umbrella Sampling-WHAM," https://github.com/nwchemgit/ nwchem/wiki/nwchem-new-pmf.pdf.
283 D. Meng, B. Zheng, G. Lin, and M. L. Sushko, Communications in Computational Physics 16, 1298 (2014).
284 D. Osei-Kuffuor and J.-L. Fattebert, Physical Review Letters 112, 046401 (2014).
285 A. Klamt and G. Schurmann, Journal of the Chemical Society, Perkin Transactions 2 , 799 (1993).
286 A. V. Marenich, C. J. Cramer, and D. G. Truhlar, Journal of Physical Chemistry B 113, 6378 (2009).
287 G. M. Chaban, J. O. Jung, and R. B. Gerber, The Journal of Chemical Physics 111, 1823 (1999).
288 Y.-Y. Chuang, D. G. Truhlar, R. A. Kendall, B. C. Garrett, and T.L. Windus, "DIRDYVTST," https://github.com/nwchemgit/ nwchem/wiki/Interface (2004).
289 J. Zheng, J. L. Bao, R. Meana-Paneda,˜ S. Zhang, B. J. Lynch, J. C. Cor-chado, Y.-Y. Chuang, P. L. Fast, W.-P. Hu, Y.-P. Liu, G. C. Lynch, K. A. Nguyen, C. F. Jackels, A. Fernandez Ramos, B. A. Elling-son, V. S. Melissas, J. Villa,` I. Rossi, E. L. Coitino,˜ J. Pu, T. V. Albu, A. Ratkiewicz, R. Steckler, B. C. Garrett, and D. G. Isaac-son, A. D.and Truhlar, "Polyrate 17-C: Computer Program for the Calculation of Chemical Reaction Rates for Polyatomics," https:

//comp.chem.umn.edu/polyrate (2017), [Online; accessed 15-February-2020].
290 "Atomic Simulation Environment," https://wiki.fysik.dtu.dk/ ase/.
291 E. D. Hermes, K. Sargsyan, H. N. Najm, and J. Zador,´ Journal of Chemical Theory and Computation 15, 6536 (2019).
292 "Sella: a utility for finding first order saddle points," https://github.com/zadorlab/sella.
293 E. J. Bylaska, J. Q. Weare, and J. H. Weare, The Journal of Chemical Physics 139, 074114 (2013).
294 H. Jonsson,´ G. Mills, and K. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions," in Classical and Quantum Dynamics in Condensed Phase Simulations, edited by B. J. Berne, G. Ciccotti, and D. F. Coker (World Scientific, 1998) pp. 385– 404.
295 G. Henkelman and H. Jonsson,´ The Journal of Chemical Physics 113, 9978 (2000).
296 G. Henkelman, B. P. Uberuaga, and H. Jonsson,´ The Journal of Chemical Physics 113, 9901 (2000).
297 S. Smidstrup, A. Pedersen, K. Stokbro, and H. Jonsson,´ The Journal of Chemical Physics 140, 214106 (2014).
298 W. E, W. Ren, and E. Vanden-Eijnden, The Journal of Chemical Physics 126, 164103 (2007).
299 V. E. Henson, in Computational Imaging, Vol. 5016, edited by C. A. Bouman and R. L. Stevenson, International Society for Optics and Photonics (SPIE, 2003) pp. 36–48.
300 E. J. Bylaska, "Python-NEB," https://bitbucket.org/ ebylaska/python-neb.
301 T. P. Straatsma and J. A. McCammon, Journal of Computational Chemistry 11, 943 (1990).
302 T. P. Straatsma, M. Philippopoulos, and J. A. McCammon, Computer Physics Communications 128, 377 (2000).
303 T. P. Straatsma, Journal of Physics: Conference Series 16, 287 (2005).
304 T. P. Straatsma and D. G. Chavarr´ıa-Miranda, Computer Physics Communications 184, 2634 (2013).
305 T. P. Straatsma and V. Helms, "Molecular Dynamics Module of NWChem - Design and Application in Protein Simulations," (World Scientific Pub. Co., 2000) pp. 70–82.
306 W. Gu, T. Frigato, T. P. Straatsma, and V. Helms, Angewandte Chemie International Edition 46, 2939 (2007).
307 R. D. Lins and T. P. Straatsma, Biophysical Journal 81, 1037 (2001).
308 R. M. Shroll and T. P. Straatsma, Biopolymers 65, 395 (2002).
309 T. P. Straatsma, Abstracts of Papers of the American Chemical Society 230, U1295 (2005).
310 T. A. Soares and T. P. Straatsma, Molecular Simulation 34(3), 295 (2008).
311 T. A. Soares, T. P. Straatsma, and R. D. Lins, Journal of the Brazilian Chemical Society 19(2), 312 (2008).
312 R. D. Lins, E. R. Vorpagel, M. Guglielmi, and T. P. Straatsma, Biomacromolecules 9, 29 (2008).
313 R. M. Shroll and T. P. Straatsma, Molecular Simulation 29, 1 (2003).
314 R. M. Shroll and T. P. Straatsma, Biophysical Journal 84, 1765 (2003).
315 T. P. Straatsma, Abstracts of Paper of the American Chemical Society 225, U805 (2003).
316 A. R. Felmy, C. Liu, and T. P. Straatsma, Geochimica et Cosmochimica Acta 69(10), A171 (2005).
317 T. P. Straatsma, "Lipopolysaccharide Membranes and Membrane Pro-teins of Pseudomonas aeruginosa studied by computer simulation," in Recent Progress in Computational Sciences and Engineering, Lecture Series on Computer and Computational Sciences, Vol. 7B, edited by T. Simos and G. Maroulis (VSP Brill, 2006) pp. 519–525.
318 T. A. Soares and T. P. Straatsma, AIP Conference Proceedings 963, 1375 (2007).
319 T. P. Straatsma and T. A. Soares, Proteins: Structure, Function, and Bioinformatics 74, 475 (2009).
320 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, The Journal of Chemical Physics 103, 8577 (1995).
321 J. W. Ponder and D. A. Case, in Protein Simulations, Advances in Protein Chemistry, Vol. 66 (Academic Press, 2003) pp. 27 – 85.
322 K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. E. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and J. A. MacKerell, Journal of Computational Chemistry 31, 671 (2010).
323 C. I. Bayly, P. Cieplak, W. Cornell, and P. A. Kollman, The Journal of Physical Chemistry 97, 10269 (1993).
324 W. D. Cornell, P. Cieplak, C. I. Bayly, and P. A. Kollman, Journal of the American Chemical Society 115, 9620 (1993).
325 T. P. Straatsma and J. A. McCammon, The Journal of Chemical Physics 95, 1175 (1991).
326 T. P. Straatsma and J. A. McCammon, Methods in Enzymology 22, 497 (1991).
327 T. P. Straatsma and J. A. McCammon, Annual Review of Physical Chemistry 43, 407 (1992).
328 T. P. Straatsma, M. Zacharias, and J. A. McCammon, "Free energy difference calculations in biomolecular systems," in Computer Simu-lation of Biomolecular Systems, edited by W. F. van Gunsteren et al. (ESCOM, Leiden, 1993) pp. 349–367.
329 T. P. Straatsma, "Free energy by molecular simulation," in Reviews in Computational Chemistry, Vol. IX, edited by K. B. Lipkowitz and D. B. Boyd (John Wiley & Sons, Ltd, 2007) pp. 81–127.
330 T. P. Straatsma, "Free energy simulations," in Encyclopedia of Compu-tational Chemistry, edited by P. von Rague Schleyer (American Cancer Society, 2002) pp. 1083–1089.
331 T. P. Straatsma, H. J. C. Berendsen, and A. J. Stam, Mol. Phys. 57, 89 (1986).
332 M. Zacharias, T. P. Straatsma, and J. A. McCammon, The Journal of Chemical Physics 100, 9025 (1994).
333 T. P. Straatsma, AIP Conference Proceedings 963, 1379 (2007).
334 E. S. Peterson, E. G. Stephan, A. L. Corrigan, R. D. Lins, T. A. Soares, R. E. Scarberry, L. K. Williams, S. J. Rose, C. Lai, T. J. Critchlow, and T. P. Straatsma, "Northwest trajectory analysis capability: A platform for enhancing computational biophysics analysis," in Proceedings of the International Conference on Bioinformatics & Computational Biology, Las Vegas, NV, July 14-17, 2008 (2008).
335 T. P. Straatsma and J. A. McCammon, IBM Systems Journal 40, 328 (2001).
336 A. V. Marenich, C. J. Cramer, D. G. Truhlar, C. A. Guido, B. Mennucci, G. Scalmani, and M. J. Frisch, Chemical Science 2, 2143 (2011).
337 M. Valiev, B. C. Garrett, M.-K. Tsai, K. Kowalski, S. M. Kathmann, G.K. Schenter, and M. Dupuis, The Journal of Chemical Physics 127, 051102 (2007).
338 M. Valiev, E. J. Bylaska, M. Dupuis, and P. G. Tratnyek, The Journal of Physical Chemistry A 112, 2713 (2008).
339 D. Case, I. Ben-Shalom, S. Brozell, D. Cerutti, T. Cheatham, III, V. Cruzeiro, R. D. T.A. Darden, D. Ghoreishi, G. Giambasu, T. Giese, M. Gilson, H. Gohlke, A. Goetz, N. H. D. Greene, R Harris, Y. Huang, S. Izadi, A. Kovalenko, R. Krasny, T. Kurtzman, T. Lee, S. LeGrand, P. Li, J. L. C. Lin, T. Luchko, R. Luo, V. Man, D. Mermelstein, K. Merz, Y. Miao, G. Monard, C. Nguyen, H.Nguyen, A. Onufriev, F. Pan, R. Qi, D. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L.Simmerling, J. Smith, J. Swails, R. Walker, J. Wang, H. Wei, L. Wilson, R. Wolf, X. Wu, L. Xiao, Y.Xiong, D. York, and P. Kollman, AMBER 2019, University of California, San Francisco (2019).
340 T. A. Soares, M. A. Osman, and T. P. Straatsma, Journal of Chemical Theory and Computation 3, 1569 (2007).
341 D. E. B. Gomes, R. D. Lins, P. G. Pascutti, C. Lei, and T. A. Soares, The Journal of Physical Chemistry B 115, 15389 (2011).
342 D. E. B. Gomes, R. D. Lins, P. G. Pascutti, C. Lei, and T. A. Soares, The Journal of Physical Chemistry B 114, 531 (2010).
343 B. H. Lower, R. D. Lins, Z. Oestreicher, T. P. Straatsma, M. F. Hochella, L. Shi, and S. K. Lower, Environmental Science & Technology 42, 3821 (2008).
344 J.-F. Boily and R. D. Lins, The Journal of Physical Chemistry C 113, 16568 (2009).
345 I. Chandrasekhar, M. Kastenholz, R. D. Lins, C. Oostenbrink, L. D. Schuler, D. P. Tieleman, and W. F. van Gunsteren, European Biophysics Journal 32, 67 (2003).
346 R. D. Lins and P. H. Hunenberger,¨ Journal of Computational Chemistry 26, 1400 (2005).
347 L. Pol-Fachin, V. H. Rusu, H. Verli, and R. D. Lins, Journal of Chemi-cal Theory and Computation 8, 4681 (2012).
348 L. Pol-Fachin, H. Verli, and R. D. Lins, Journal of Computational Chemistry 35, 2087 (2014).
349 T. A. Soares, P. H. Hunenberger,¨ M. A. Kastenholz, V. Krautler,¨ T. Lenz, R. D. Lins, C. Oostenbrink, and W. F. van Gunsteren, Journal of Computational Chemistry 26, 725 (2005).
350 E. F. Franca, L. C. G. Freitas, and R. D. Lins, Biopolymers 95, 448

(2011).
351 E. F. Franca, R. D. Lins, L. C. G. Freitas, and T. P. Straatsma, Journal of Chemical Theory and Computation 4, 2141 (2008).
352 M. Svensson, S. Humbel, R. D. J. Froese, T. Matsubara, S. Sieber, and K. Morokuma, The Journal of Physical Chemistry 100, 19357 (1996).
353 G. N. Chuev, M. V. Fedotova, and M. Valiev, The Journal of Chemical Physics 152, 041101 (2020).
354 M. Valiev and G. N. Chuev, J. Stat. Mech.: Theory Exp. 2018, 093201 (2018).
355 G. N. Chuev, M. Valiev, and M. V. Fedotova, Journal of Chemical Theory and Computation 8, 1246 (2012).
356 D. Chandler, J. D. McCoy, and S. J. Singer, The Journal of Chemical Physics 85, 5971 (1986).
357 D. Chandler, J. D. McCoy, and S. J. Singer, The Journal of Chemical Physics 85, 5977 (1986).
358 J. D. Owens, D. Luebke, N. Govindaraju, M. Harris, J. Kruger,¨ A. E. Lefohn, and T. J. Purcell, Computer Graphics Forum 26, 80 (2007).
359 N. Gawande, K. Kowalski, B. Palmer, S. Krishnamoorthy, E. Apra, J. Manzano, V. Amatya, and J. Crawford, in 2019 IEEE/ACM Workshop on Exascale MPI (ExaMPI) (2019) pp. 29–38.
360 C. Yu, W. Harbich, L. Sementa, L. Ghiringhelli, E. Apra,´ M. Stener, A. Fortunelli, and H. Brune, The Journal of Chemical Physics 147, 074301 (2017).
361 J. R. Hammond, K. Kowalski, and W. A. de Jong, The Journal of Chemical Physics 127, 144105 (2007).
362 B. Peng, N. Govind, E. Apra,` M. Klemm, J. R. Hammond, and K. Kowalski, The Journal of Physical Chemistry A 121, 1328 (2017).
363 H.-S. Hu, K. Bhaskaran-Nair, E. Apra,` N. Govind, and K. Kowalski, The Journal of Physical Chemistry A 118, 9087 (2014).
364 W. A. de Jong, E. Bylaska, N. Govind, C. L. Janssen, K. Kowalski, T. Muller,¨ I. M. B. Nielsen, H. J. J. van Dam, V. Veryazov, and R. Lindh, Physical Chemistry Chemical Physics 12, 6896 (2010).
365 J. S. Nelson, S. J. Plimpton, and M. P. Sears, Physical Review B 47, 1765 (1993).
366 J. Wiggs and H. Jonsson, Computer Physics Communications 87, 319 (1995).
367 A. Canning and D. Raczkowski, Computer Physics Communications 169, 449 (2005).
368 F. Gygi, IBM Journal of Research and Development 52, 137 (2008).
369 T. W. Swaddle, J. Rosenqvist, P. Yu, E. J. Bylaska, B. L. Phillips, and W. H. Casey, Science 308, 1450 (2005).
370 J. R. Rustad and E. J. Bylaska, Journal of the American Chemical Society 129, 2222 (2007).
371 R. Atta-Fynn, D. F. Johnson, E. J. Bylaska, E. S. Ilton, G. K. Schenter, and W. A. de Jong, Inorganic Chemistry 51, 3016 (2012).
372 J. L. Fulton, E. J. Bylaska, S. Bogatko, M. Balasubramanian, E. Cauet,¨ G. K. Schenter, and J. H. Weare, The Journal of Physical Chemistry Letters 3, 2588 (2012).
373 M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek, and G. R. Hutchison, Journal of Cheminformatics 4, 17 (2012).
374 See http://www.jmol.org/ for Jmol: an open-source Java viewer for chemical structures in 3D.
375 See https://www.chemcraftprog.com for Chemcraft – Graphical software for visualization of quantum chemistry computations.
376 See http://mocalc2012.sourceforge.net/ for MoCalc2012.
377 N. M. O'boyle, A. L. Tenderholt, and K. M. Langner, Journal of Computational Chemistry 29, 839 (2008).
378 See http://www.chemissian.com/ for Chemissian.
379 See http://www.scienomics.com for Scienomics, MAPS Plat-form.
380 See http://www.biomolecular-modeling.com/Ascalaph/ for Ascalaph.
381 G. Black, K. Schuchardt, D. Gracio, and B. Palmer, in Computa-tional Science — ICCS 2003, edited by P. M. A. Sloot, D. Abramson, A. V. Bogdanov, Y. E. Gorbachev, J. J. Dongarra, and A. Y. Zomaya (Springer Berlin Heidelberg, Berlin, Heidelberg, 2003) pp. 122–131.
382 See http://www-stone.ch.cam.ac.uk/programs.html for CamCASP (Cambridge package for Calculation of Anisotropic Site Properties).
383 see http://www.chemshell.org forChemShell, a Computational Chemistry Shell.
384 T. Y. Nikolaienko, L. A. Bulavin, and D. M. Hovorun, Computational and Theoretical Chemistry 1050, 15 (2014).
385 S. Canneaux, F. Bohr, and E. Henon, Journal of Computational Chem-istry 35, 82 (2014).
386 U. Lourderaj, R. Sun, S. C. Kohale, G. L. Barnes, W. A. de Jong, T. L. Windus, and W. L. Hase, Computer Physics Communications 185, 1074 (2014).
387 See http://perso.neel.cnrs.fr/xavier.blase/fiesta/ index.html for Fiesta.
388 See https://pupil.sourceforge.net for PUPIL, Program for User Package Interface and Linking.
389 J. Wehner, L. Brombacher, J. Brown, C. Junghans, O. C¸aylak, Y. Kha-lak, P. Madhikar, G. Tirimbo,` and B. Baumeier, Journal of Chemical Theory and Computation, Journal of Chemical Theory and Computa-tion 14, 6253 (2018).
390 See https://github.com/KristapsE/PyDP4 for PyDP4.
391 A. Supady, V. Blum, and C. Baldauf, Journal of Chemical Information and Modeling 55, 2338 (2015).
392 See https://gaussian.com/cubegen/ for Gaussian cube.
393 G. Schaftenaar, E. Vlieg, and G. Vriend, Journal of Computer-Aided Molecular Design 31, 789 (2017).
394 see https://ugovaretto.github.io/molekel/ for Molekel 5.4.
395 D. L. Bergman, L. Laaksonen, and A. Laaksonen, Journal of Molecular Graphics and Modelling 15, 301 (1997).
396 W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graph-ics 14, 33 (1996).
397 See https://sourceforge.net/projects/jbonzer/ for Jam-beroo.
398 K. Momma and F. Izumi, Journal of Applied Crystallography 41, 653 (2008).
399 J. Contreras-Garc´ıa, E. R. Johnson, S. Keinan, R. Chaudret, J.-P. Pique-mal, D. N. Beratan, and W. Yang, Journal of Chemical Theory and Computation 7, 625 (2011).
400 See http://www.quimica.urv.es/XAIM for XAIM.
401 T. Lu and F. Chen, Journal of Computational Chemistry 33, 580 (2012).
402 S. Ghosh, "SEMIEMP: Open source code for semiempirical quna-tum chemistry calculation," http://github.com/SoumenChem/ semiemp.
403 S. Ghosh, A. Andersen, L. Gagliardi, C. J. Cramer, and N. Govind, Journal of Chemical Theory and Computation 13, 4410 (2017).
404 S. Ghosh, J. C. Asher, L. Gagliardi, C. J. Cramer, and N. Govind, The Journal of Chemical Physics 150, 104103 (2019).
405 D. E. Bernholdt, B. A. Allan, R. Armstrong, F. Bertrand, K. Chiu, T. L. Dahlgren, K. Damevski, W. R. Elwasif, T. G. W. Epperly, M. Govin-daraju, D. S. Katz, J. A. Kohl, M. Krishnan, G. Kumfert, J. W. Larson, S. Lefantzi, M. J. Lewis, A. D. Malony, L. C. McInnes, J. Nieplocha, B. Norris, S. G. Parker, J. Ray, S. Shende, T. L. Windus, and S. Zhou, The International Journal of High Performance Computing Applica-tions 20, 163 (2006).
406 R. M. Richard, C. Bertoni, J. S. Boschen, K. Keipert, B. Pritchard, E. F. Valeev, R. J. Harrison, W. A. de Jong, and T. L. Windus, Computing in Science Engineering 21, 48 (2019).
407 J. P. Kenny, S. J. Benson, Y. Alexeev, J. Sarich, C. L. Janssen, L. C. Mcinnes, M. Krishnan, J. Nieplocha, E. Jurrus, C. Fahlstrom, and T. L. Windus, Journal of Computational Chemistry 25, 1717 (2004).
408 T. P. Gulabani, M. Sosonkina, M. S. Gordon, C. L. Janssen, J. P. Kenny, H. Netzloff, and T. L. Windus, in Proceedings of the 2009 Spring Simulation Multiconference, SpringSim '09 (Society for Computer Simulation International, San Diego, CA, USA, 2009).
409 M. Krishnan, Y. Alexeev, T. L. Windus, and J. Nieplocha, in SC '05: Proceedings of the 2005 ACM/IEEE Conference on Supercomputing (IEEE, 2005) pp. 23–23.
410 J. P. Kenny, C. L. Janssen, E. F. Valeev, and T. L. Windus, Journal of Computational Chemistry 29, 562 (2008).
411 E. Mutlu, K. Kowalski, and S. Krishnamoorthy, in Proceedings of the 6th ACM SIGPLAN International Workshop on Libraries, Languages and Compilers for Array Programming (ACM, 2019) pp. 46–56.
412 N. P. Bauman, E. J. Bylaska, S. Krishnamoorthy, G. H. Low, N. Wiebe, C. E. Granade, M. Roetteler, M. Troyer, and K. Kowalski, The Journal of Chemical Physics 151, 014107 (2019).
413 J. Bostrom,¨ M. Pitonak, F. Aquilante, P. Neogrady, T. B. Pedersen, and R. Lindh, Journal of Chemical Theory and Computation 8, 1921 (2012).
414 T. B. Pedersen, A. M. J. Sanchez´ de Meras,´ and H. Koch, The Journal of Chemical Physics 120, 8887 (2004).

415 E. Epifanovsky, D. Zuev, X. Feng, K. Khistyaev, Y. Shao, and A. I. Krylov, The Journal of Chemical Physics 139, 134105 (2013).
416 X. Feng, E. Epifanovsky, J. Gauss, and A. I. Krylov, The Journal of Chemical Physics 151, 014110 (2019).
417 C. Peng, J. A. Calvin, and E. F. Valeev, International Journal of Quan-tum Chemistry 119, e25894 (2019).
418 S. D. Folkestad, E. F. Kjønstad, and H. Koch, The Journal of Chemical physics 150, 194112 (2019).
419 C. Riplinger and F. Neese, The Journal of Chemical Physics 138, 034106 (2013).
420 C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, The Journal of Chemical Physics 144, 024109 (2016).
421 F. Pavoseviˇc,` C. Peng, P. Pinski, C. Riplinger, F. Neese, and E. F. Valeev, The Journal of Chemical Physics 146, 174108 (2017).
422 G. Onida, L. Reining, and A. Rubio, Reviews of Modern Physics 74, 601 (2002).
423 B. Peng and K. Kowalski, Journal of Chemical Theory and Computation 14, 4335 (2018).
424 G. H. Low, N. P. Bauman, C. E. Granade, B. Peng, N. Wiebe, E. J. Bylaska, D. Wecker, S. Krishnamoorthy, M. Roetteler, K. Kowalski, M. Troyer, and N. A. Baker, arXiv preprint arXiv:1904.01131 (2019).