Quantum Molecular Dynamics in the Post–Petaflops Era

Nichols A. Romero, Argonne National Laboratory

Aiichiro Nakano, University of Southern California

Katherine M. Riley, Argonne National Laboratory

Fuyuki Shimojo, Kumamoto University

Rajiv K. Kalia and Priya Vashishta, University of Southern California

Paul C. Messina, Argonne National Laboratory

As the scale of quantum molecular dynamics simulations has grown in time and system size, QMD codes must increase intranode and instruction–level parallelism to take advantage of emerging supercomputer architectures. The authors present one promising parallelization approach and illustrate its success on one of the world's most powerful systems.

Quantum molecular dynamics (QMD) simulations are a critical tool in various scientific domains including condensed-matter physics, materials science, chemistry, and biology. As the scale of QMD simulations has grown beyond the canonical metrics of time and system size, there has been a paradigm shift to more "naturally" parallel methods such as divide-and-conquer, equilibrium and nonequilibrium statistical sampling, and high-throughput screening. The scientific community has also needed to adapt to the short five-year life cycle of new computer architectures.

We explore how QMD simulations have evolved over the years and then focus on a novel approach that extends divide-and-conquer. We also present two examples of successes achieved with this method on Mira (www.alcf.anl.gov/mira), the 10-petaflops IBM Blue Gene/Q system at Argonne National Laboratory (ANL) currently ranked fifth on the TOP500 supercomputer list.

CALCULATING ELECTRONIC STRUCTURE

Electronic structure theory is concerned with approximately solving the Schrödinger equation (SE) for atoms, molecules, and condensed-phase materials. The SE is a nonrelativistic quantum-mechanical equation of motion that governs the interaction of matter via electromagnetic forces, most notably the Coulombic interaction. The properties of everyday materials are

given by solutions to the SE, Ψ, which are called wave functions.

The time-dependent SE (TDSE) for a material's electrons and nuclei is given by

$$(\widehat{T} + \widehat{V})\Psi = i\frac{\partial \Psi}{\partial t}, \text{ where} \qquad (1)$$

$$\Psi = \Psi\left(\boldsymbol{r}_1, \boldsymbol{r}_2, \ldots \boldsymbol{r}_{N_e}, \boldsymbol{R}_1, \boldsymbol{R}_2, \ldots \boldsymbol{R}_{N_I}\right). \qquad (2)$$

$\widehat{T}$ and $\widehat{V}$ are the quantum-mechanical operators for the kinetic energy and potential, respectively, of all the electrons and nuclei; $\boldsymbol{r}$, $\boldsymbol{R}$, N_e, and N_I are the coordinates of the electrons, coordinates of the nuclei, total number of electrons, and total number of nuclei, respectively. The TDSE is a high-dimensional linear partial differential equation that relates the total energy of the wave function to its partial time derivative.

If the potential operator is time independent, which is the case for many materials of interest, then the TDSE can be simplified to the time-independent SE (TISE):

$$(\widehat{T} + \widehat{V})\Psi = E\Psi, \qquad (3)$$

where E is the total energy of all the electrons and nuclei.

Solving the SE is further complicated by symmetry constraints imposed by quantum mechanics. The wave functions must be either symmetric or antisymmetric under an odd number of quantum particle permutations. This combinatorial operation means that exact SE solutions are $O(N!)$ in computational complexity. As a result, exact solutions are computationally intractable for all but the simplest atoms even with modern massively parallel supercomputers. However, methods with varying degrees of accuracy can be used to calculate approximations.

Electronic structure can be calculated using a vast array of methods with esoteric acronyms such as QMC (quantum Monte Carlo), DFT (density functional theory), and CCSD(T) (coupled cluster with single, double, and perturbative triplets), just to name a few. These methods can be broadly categorized into *single-particle* and *many-body* methods. The former deal with a

set of N one-electron wave functions, $\Psi = \Psi(\boldsymbol{r}_i; \boldsymbol{r}_{j \neq i}, \boldsymbol{R})$, that are calculated self-consistently with a "mean field" produced by all the electrons. The latter methods use the many-electron wave function in Equation 2 to achieve greater accuracy but at higher computational cost.

High-performance computing (HPC) has been critical in enabling electronic structure calculations due to their high order of computational complexity. For example, the computational complexity of DFT is $O(N^3)$ and that of CCSD(T) is $O(N^7)$.

DFT is one of the most widely used methods because of its accessibility to nonexperts, its computational tractability, and its qualitative and quantitative predictive power. With DFT, all of a material's electronic structural properties can be obtained from a universal energy functional of the ground-state density. This significantly reduces the calculation's dimensionality from a $3N$-dimensional many-electron wave function to a 3D density function.[1]

There are many DFT implementations, based on equations developed in the mid-1960s by Walter Kohn and Lu Jeu Sham (KS),[2] that provide an ansatz for this energy functional. Using an approximation for the core electrons known as the pseudopotential, these algorithms iteratively solve the equations by calculating what are known as KS orbitals.

QUANTUM MOLECULAR DYNAMICS

The computational workhorse of materials science is molecular dynamics (MD), which simulates the dynamic properties of materials by computing the motion of all the atoms due to the interactions of the forces between them. Aneesur Rahman performed the

> DENSITY FUNCTIONAL THEORY SIGNIFICANTLY REDUCES THE DIMENSIONALITY OF ELECTRONIC STRUCTURE CALCULATIONS.

first MD simulation at ANL in 1964,[3] the same year Kohn and Pierre Hohenberg introduced DFT. It was not until the 1990s, however, that computational power was sufficient to couple MD with DFT.[4] These simulations treat the nuclei as particles propagated by classical equations of motion, such as Newtonian mechanics, and the electrons quantum mechanically. However, other equations of motion can be used, most notably the Car–Parrinello method[5] that substitutes a fictitious mass for the electrons to avoid computing the DFT ground state at each time step. DFT-based MD methods go by many names including ab initio MD, first-principles MD, DFT-MD, or simply QMD.

The sidebar "Density Functional Theory Calculations on High-Performance Computing Architectures" describes the increasing scale of DFT calculations enabled by advances in supercomputing technology.

This article focuses on QMD techniques based on pseudopotentials and the plane-wave method. In this formulation, each KS orbital is enumerated by band index n, spin index σ, Brillouin zone index $\boldsymbol{k}$ (the k-point), and plane-wave index G:

$$\psi_{nk\sigma}\left(\boldsymbol{r}\right)=\sum_{G}c_{nk\sigma}\left(\boldsymbol{G}\right)e^{-iGr} \;. \qquad (4)$$

The total number of KS orbitals solved for are $N_b \times N_k \times N_\sigma$, which is the product of the total number of bands, total number of k-points, and total number of spins, respectively.

LINEAR-SCALING DFT

In ordinary conditions, a wave function's high dimensionality only manifests itself at short distances (across a few atoms). Researchers have proposed various $O(N)$ DFT techniques to exploit this data-locality principle, known
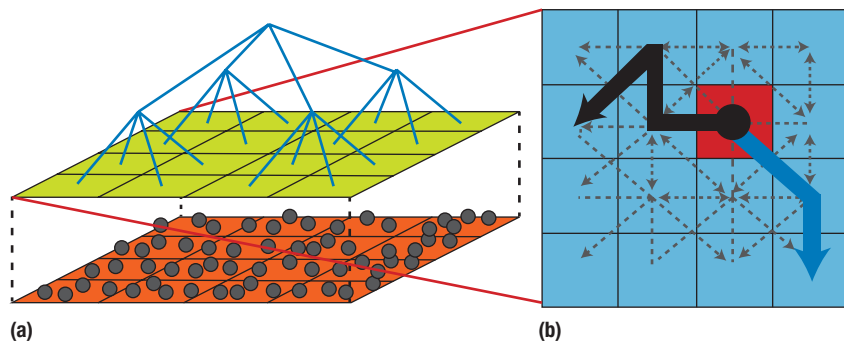

FIGURE 1. Divide–conquer–recombine (DCR) algorithm. (a) In the divide-and-conquer (DC) phase, the algorithm abstracts atoms (spheres in the bottom plane) within each DC domain (parallelograms in the bottom plane) through collective variables such as a finite-difference representation of the charge distribution. It combines the DC domains recursively to form a tree data structure of progressively coarser cells and, upon reaching the entire simulation volume at the tree's root, computes $O(N)$ by traversing the tree upward and downward. (b) Blue and black lines show typical triplet and quadruplet computations, respectively, involving a DC domain in the recombine phase.

as *quantum nearsightedness*.[6] Among these techniques is the divide-and-conquer (DC) DFT algorithm pioneered by Weitao Yang.[7] DC in general is highly scalable on massively parallel computers and has been applied successfully to broad computational problems ranging from the formally $O(N^2)$ classical N-body problem to the $O(N^3)$ eigenvalue problem, to the exponentially complex quantum N-body problem. DC-DFT subdivides the physical system into spatially localized domains embedded in a global electrostatic potential, which is computed in $O(N)$ operations using a tree-based multigrid. Specifically, the algorithm numerically obtains local KS orbitals in the domains (calculated using the preconditioned conjugate gradient method) and the global electrostatic potential using a global–local self-consistent-field iteration.

Our own lean DC-DFT (LDC-DFT) algorithm minimizes the $O(N)$ prefactor by optimizing DC computational parameters as well as density-adaptive boundary conditions at the domain peripheries.[8,9] We have implemented LDC-DFT on parallel computers by employing multiple levels of parallelism with hierarchical band–spatial decomposition. At the coarser level, we use task decomposition among

domains; at the finer level, we parallelize the plane-wave-based calculations within each domain by combining spatial and band decompositions.

Even with $O(N)$ DFT techniques, predicting a material system's long-time dynamics is computationally difficult, as the sequential bottleneck of time precludes straightforward parallelization. Here, another locality principle—temporal locality—comes to the rescue. Namely, the system retains local minimum-energy states most of the time, except during rare transitions between states. In such cases, transition state theory allows reformulation of the sequential long-time dynamics as a computationally more efficient parallel search for low activation-barrier transition events.[10]

METASCALABLE DCR

While computationally efficient, the tree topology of DC computations cannot capture rich physics, including high-order intermolecular correlations. We therefore propose an extension of DC called *divide-conquer-recombine* (DCR).[8,9]

DCR in space

As Figure 1 shows, DCR operates in two phases. The algorithm first constructs globally informed local solutions in

DENSITY FUNCTIONAL THEORY CALCULATIONS ON HIGH-PERFORMANCE COMPUTING ARCHITECTURES

Density functional theory (DFT) calculations were carried out in the 1970s[1] on computers that used punch cards as inputs, just a few years after the pioneering work on DFT by Pierre Hohenberg, Walter Kohn, and Lu Jeu Sham. The supercomputers of the 1990s enabled more complex DFT formulations. Two early parallel DFT calculations were carried out by teams led by Ivan Štich[2] and Karl Brommer[3] on the silicon (111) surface. Stich's group used the 64-node Meiko i860 Computing Surface at Edinburgh University and the Intel iPSC/960 hypercube at Daresbury Laboratory, while Brommer's team used the Connection Machines at the Pittsburgh Supercomputer Center and at Los Alamos National Laboratory as well as a Cray YMP at the San Diego Supercomputer Center. Both calculations appeared to be data and task parallel with respect to plane-wave coefficients.

In discussing their code's parallelization, Štich and his colleagues[4] noted three possible data decomposition strategies: k-point, band index, and plane-wave coefficients and real space—the latter necessitated by use of the dual-space formalism.[5] After considering these strategies' benefits and drawbacks both in isolation and in combination, they argued for data parallelism with respect to plane-wave coefficients and real space, henceforth referred to as *spatial decomposition*. This choice was motivated by their systems' architectures: the i860 and iPSC/960 were message-passing machines with only 16 Mbytes of memory per node, which was insufficient to store the entire 3D mesh for a single electronic-wave function. The researchers chose a 1D decomposition into 2D slabs to minimize communication during the 3D fast Fourier transform (FFT).

Parallel DFT codes became more commonplace after the electronic structure community adopted the MPI (message passing interface) standard, though some codes were instead based on SHMEM (symmetric hierarchical memory access). All codes were data and task parallel with respect to spatial decomposition, which was necessary because of memory constraints on nodes. The availability of symmetric multiprocessing clusters encouraged developers to parallelize on the spin and k-point indices—it was likewise simple to express this parallelism with MPI collectives. The community was also fortunate that many DFT compute kernels could be formulated as FFTs and BLAS3 operations. This permitted a substantial degree of vectorization (instruction-level parallelism) because developers could simply write their DFT codes based on well-established APIs and link their executables against vendor-optimized libraries. Using these established APIs and libraries is an application development best practice, but developers still must plan for compiler-based performance when writing code. Because compiler technology is not able to detect all vectorization opportunities, overrelying on compilers to identify all instruction-level parallelism can impede application performance.

After focusing primarily on MPI parallelization with respect to plane-wave coefficients, k-points, and spin indices for most of the 1990s, researchers doing DFT calculations encountered the first of what turned out to be several dense linear algebra bottlenecks in the late 1990s and early 2000s. As the number of atoms and electrons increased with more ambitious scientific investigations, dense linear algebra equations with computational complexity $O(N^3)$ became dominant in

the *DC phase* and then in the *recombine phase* synthesizes these into a global solution—for example, delocalized molecular orbital and charge-carrier dynamics[8]—encompassing large spatiotemporal scales. The recombine phase typically involves range-limited

n-tuple computations to account for higher interdomain correlations.

To implement DCR efficiently on massively parallel supercomputers, we combined a global real-space multigrid solver that computes interdomain electrostatic potential with a fast Fourier

transform (FFT)-based spectral method that obtains the intradomain plane-wave functions. Our parallel QMD code, implemented with hybrid message passing (MPI) and multithread programming (OpenMP), achieved both global scalability and local efficiency:

iterative DFT solvers. While parallel dense linear algebra packages such as ScaLAPACK had existed since 1995, adoption was slow for both technical and nontechnical reasons.

Perhaps the most widely recognized DFT code to combine all three distinct levels of data parallelization is QBox. With this code, Francois Gygi and his colleagues[6] performed a plane-wave DFT calculation on 1,000 molybdenum atoms (12,000 electrons with 8 k-points) using the entire Blue Gene/L system at Lawrence Livermore National Laboratory. Their application achieved 367 teraflops on 65,526 compute nodes and earned them the 2006 ACM Gordon Bell Prize (GBP). Later DFT codes including NWChem (www.nwchem-sw.org), VASP (Vienna Ab initio Simulation Package; www.vasp.at), CPMD (www.cpmd.org), Quantum ESPRESSO (www.quantum-espresso.org), and GPAW (https://wiki.fysik.dtu.dk/gpaw) implement a similar parallelization strategy.

Linear-scaling DFT methods outperform traditional $O(N^3)$ approaches on large-scale material systems by using the nearsightedness of quantum mechanics to exploit locality. While the mathematical machinery underlying these methods varies, they all transform the DFT calculation of a large system to the sum of DFT calculations of the system's smaller constituent parts. This can be accomplished by wave-function localization, density matrix minimization, or explicit divide-and-conquer methods.[7] The linear-scaling 3D fragment (LS3DF) method developed by Lin-Wang Wang and his colleagues[8] is an example of an explicit divide-and-conquer method. Wang's team from Lawrence Berkeley National Laboratory was awarded the 2008 GBP for achieving

107.5 teraflops on Argonne National Laboratory's Blue Gene/P system (163,840 cores). LS3DF was showcased on a 13,824-atom oxygen-doped zinc telluride alloy and ran more than 400 times faster than a traditional DFT method.

References

1. J.F. Janak, V.L. Moruzzi, and A.R. Williams, "Ground-State Thermomechanical Properties of Some Cubic Elements in the Local-Density Formalism," *Physical Rev. B*, vol. 12, no. 4, 1975, pp. 1257–1261.

2. I. Štich et al., "*Ab Initio* Total-Energy Calculations for Extremely Large Systems: Application to the Takayanagi Reconstruction of Si(111)," *Physical Rev. Letters*, vol. 68, no. 9, 1992, pp. 1351–1354.

3. K.D. Brommer et al., "*Ab Initio* Theory of the Si(111)-(7x7) Surface Reconstruction: A Challenge for Massively Parallel Computation," *Physical Rev. Letters*, vol. 68, no. 9, 1992, pp. 1355–1358.

4. L.J. Clarke, I. Štich, and M.C. Payne, "Large-Scale *Ab Initio* Total Energy Calculations on Parallel Computers," *Computer Physics Comm.*, vol. 72, no. 1, 1992, pp. 14–28.

5. J.L. Martins and M.L. Cohen, "Diagonalization of Large Matrices in Pseudopotential Band-Structure Calculations: Dual-Space Formalism," *Physical Rev. B*, vol. 37, no. 11, 1988, pp. 6134–6138.

6. F. Gygi et al., "Large-Scale Electronic Structure Calculations of High-Z Metals on the Blue Gene/L Platform," *Proc. ACM/IEEE Conf. Supercomputing* (SC 06), 2006; doi:10.1145/1188455.1188502.

7. D.R. Bowler and T. Miyazaki, "$O(N)$ Methods in Electronic Structure Calculations," *Reports on Progress in Physics*, vol. 75, no. 3, 2012, doi:10.1088/0034-4885/75/3/036503.

8. L.W. Wang et al., "Linearly Scaling 3D Fragment Method for Large-Scale Electronic Structure Calculations," *Proc. ACM/IEEE Conf. Supercomputing* (SC 08), 2008; doi:10.1109/SC.2008.5218327.

a 50.3-million-atom QMD benchmark exhibited 98.4 percent of the perfect parallel speedup as well as 50.5 percent of the peak flops on all 786,432 cores of Mira.[8,9] We expect that this global–local separation within the DCR algorithm is metascalable—that is, it will

continue to scale on future parallel supercomputers.

DCR in time

One way to sample rare events is through *parallel replica dynamics* (PRD).[10] PRD reduces sequential

long-time dynamics, as shown in Figure 2a, to statistically independent parallel trajectory runs (or replicas), as shown in Figure 2b. To handle widely disparate activation barriers in complex reaction pathways, an extension of PRD called super-state PRD (SPRD)
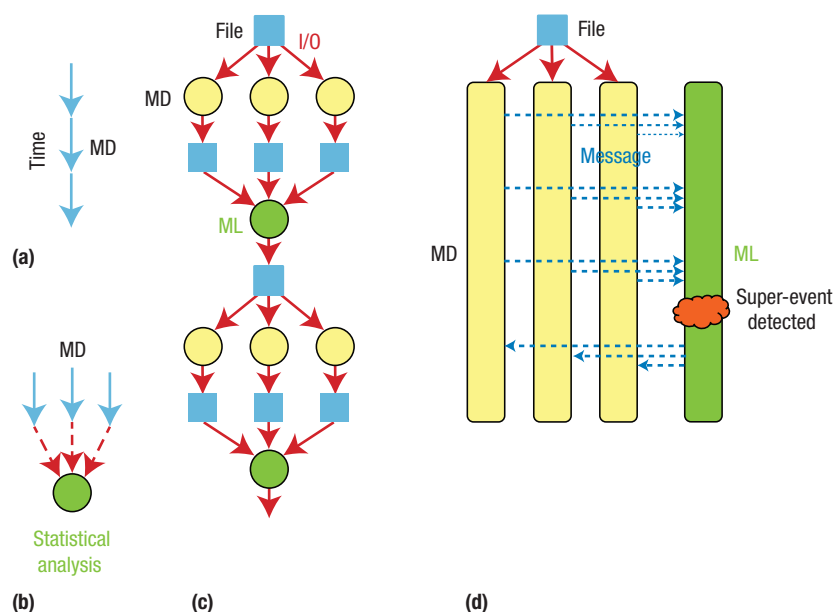
FIGURE 2. DCR in time. (a) Molecular dynamics (MD) simulations have sequential time dependence. (b) Parallel replica dynamics (PRD) predicts long–time behavior through statistical analysis of multiple parallel MD trajectories. (c) Conventional file–based and (d) new in situ PRD simulations. ML represents machine–learning tasks.

groups microscopic states into a super-state and handles only rare transitions between those super-states (or super-events) with PRD. Super-states are usually predefined based on some a priori knowledge. In the case of complex dynamics, for which such a priori definitions are unavailable, we instead employ machine-learning approaches to automatically detect super-events on the fly.[11] Conventional SPRD simulations run for a block of MD time steps (for example, 1,000), followed by a separate postprocessing phase that applies cross-time-series statistical analysis to stored simulation trajectory data to detect super-events, as Figure 2c shows. We are currently developing an in situ SPRD method that carries out correlation analysis as data is produced to reduce the need to store data in files for postprocessing, as Figure 2d shows.

EXAMPLE QMD SIMULATIONS USING DCR
Researchers use QMD to conduct time-resolved ab initio simulations of how nanometer-scale features such as defects and interfaces affect material dynamics and thus physical properties, providing atomistic insight into structure–property relationships that could lead to improved designs of advanced materials and devices. Here, we present two example QMD simulations using DCR: on-demand hydrogen production from water using metallic-alloy particles for renewable energy, and photoexcited charge-carrier dynamics in low-cost, high-efficiency solar cells.

Hydrogen on demand
In a two-step thermochemical cycle, an exothermic reaction between metal and water produces hydrogen gas, followed by solar energy–assisted reduction of the metal-oxide product to regenerate metal fuel. A potential application of this technology is on-board hydrogen production for hydrogen-powered vehicles, but conventional metal–water reaction kinetics is too slow to make such on-demand hydrogen production commercially viable.

A DCR-based QMD simulation involving 16,611 atoms on all 786,432 cores of Mira revealed that alloying aluminum particles with lithium achieves orders-of-magnitude-faster reactions with higher yields.[12] As

Figure 3 shows, the simulations revealed a key nanostructural design expected to scale up to industrially relevant particle sizes.

Photoexcitation dynamics in solar cells
In solar cells, absorption of a photon creates a pair of a negatively charged electron and a positively charged hole. The photoexcited electron–hole pair, called an *exciton*, can only exist in a special spin configuration known as a *singlet*. Singlet fission (SF), the process whereby a spin-singlet exciton splits into two spin-triplet excitons, could double the photo-induced electric current to boost power conversion efficiency. Unfortunately, SF has only been observed in high-quality crystals that are very expensive to grow. Realizing a high SF yield in mass-produced disordered solids could revolutionize low-cost fabrication of efficient solar cells.

Recently, researchers made an experimental breakthrough when they observed SF in amorphous diphenyl tetracene (DPT). However, atomistic mechanisms that enable efficient SF in amorphous molecular solids remain elusive, largely due to the required large quantum-mechanical calculations that capture nanostructural features.

As Figure 4 shows, our DCR-based approach involving 6,400-atom non-adiabatic QMD (NAQMD) and kinetic Monte Carlo (KMC) simulations identified the key molecular geometry and exciton-flow network topology for SF "hot spots" in amorphous DPT, where fission occurs preferentially.[13] We first performed the DC-NAQMD simulations to calculate the hopping of photoexcited excitons between neighboring spatial domains, then recombined

the first-principles hopping rates in the KMC simulations to predict global exciton dynamics. The simulations revealed the molecular origin of experimentally observed two-time scales in exciton population dynamics, and could pave a way to nanostructural design of efficient solar cells from first principles.

QMD PROGRAMMING ISSUES

For the past decade, supercomputing hardware has been delivering more computational capability by growing node complexity; clock-cycle speeds have not increased. Massively threaded or hybrid nodes with accelerators are, and will continue to be, the norm on the largest systems. We can expect escalating node parallelism by an order or magnitude or more for each five-year generation. Per-node storage capacity might increase, but not through traditional DRAM—more likely, it will occur through a storage hierarchy with each tier using different technology and providing different performance. To use these architectures, QMD applications must evolve.

Fortran and C/C++ with MPI are expected to continue as the primary QMD programming models, with an additional abstraction (MPI+X) to aid with node-level parallelism. OpenMP and OpenACC are strong candidates for X on massively threaded nodes and might help on accelerators such as GPGPUs. Codes written in CUDA and OpenCL are unlikely to port to massively threaded nodes and thus could be limited to hybrid-node systems. MPI+MPI is another potential solution; new standards aim to address node-level parallelism through a shared-memory-like model. Currently, most community attention
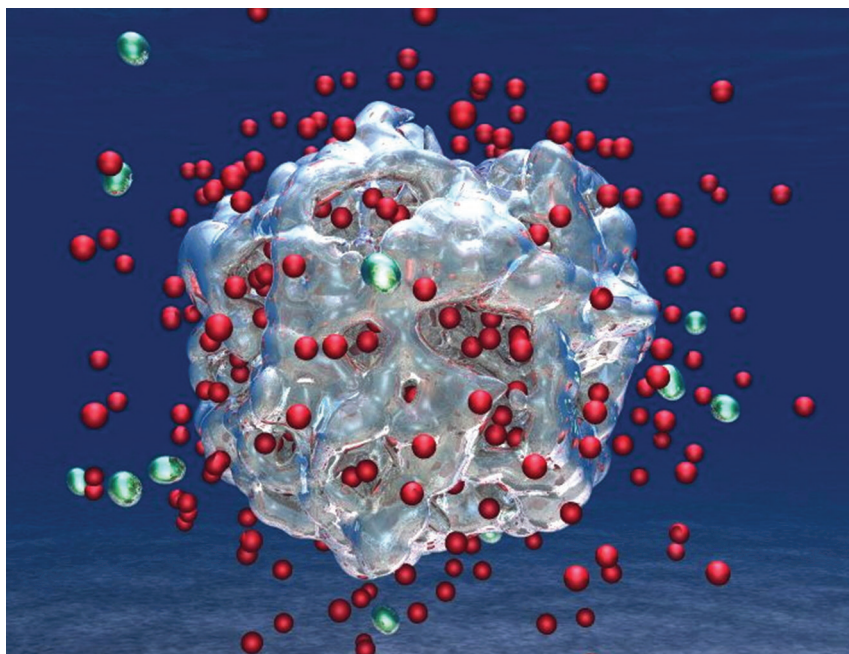

FIGURE 3. DCR-based quantum molecular dynamics (QMD) simulation of hydrogen production from water using an aluminum–lithium alloy particle. The valance electron density (silver isosurface) is centered around the aluminum atoms, whereas some of the lithium atoms (red spheres) are dissolved in water. Green ellipsoids represent produced hydrogen molecules. Water molecules are not shown for clarity. (Reprinted with permission from K. Shimamura et al., "Hydrogen-on-Demand Using Metallic Alloy Nanoparticles in Water," *Nano Letters*, vol. 14, no. 7, 2014, pp. 4090–4096. Copyright 2015, ACS Publishing, LLC.)
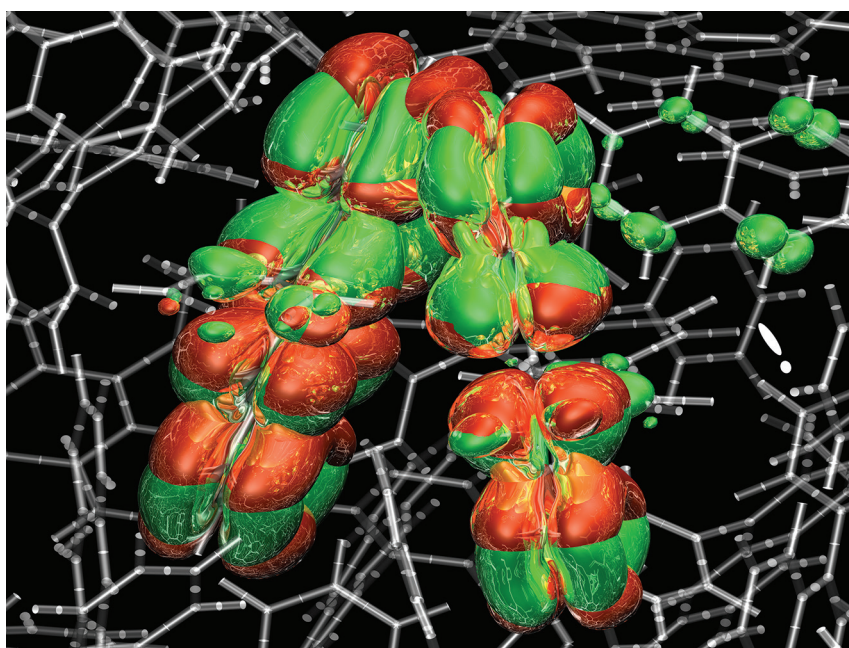

FIGURE 4. DCR-based nonadiabatic QMD simulation of photo-induced excitons in amorphous diphenyl tetracene (DPT). Orange and green isosurfaces represent quasi-electron and quasi-hole wave functions, and gray rods represent DPT molecules. (Reprinted with permission from W. Mou et al., "Nanoscopic Mechanisms of Singlet Fission in Amorphous Molecular Solid," *Applied Physics Letters*, vol. 102, no. 17, 2013; http://dx.doi.org/10.1063/1.4795138. Copyright 2015, AIP Publishing, LLC.)

ABOUT THE AUTHORS

NICHOLS A. ROMERO is a principal project specialist in computational science and team lead at the Argonne Leadership Computing Facility (ALCF) at Argonne National Laboratory (ANL). His primary research interest is massively parallel electronic-structure calculations in condensed-matter physics. Romero received a PhD in physics from the University of Illinois at Urbana–Champaign. He is a member of the American Physical Society (APS). Contact him at naromero@alcf.anl.gov.

AIICHIRO NAKANO is a professor of computer science, physics and astronomy, chemical engineering and materials science, and biological sciences at the University of Southern California (USC). His research focuses on scalable parallel algorithms for materials science. Nakano received a PhD in physics from the University of Tokyo. He is a Fellow of the APS. Contact him at anakano@usc.edu.

KATHERINE M. RILEY is a principal applications engineer and Director of Science at ALCF. Her primary research interests are in scientific application architectures for high-performance computing (HPC) and programming environments for supercomputers. She was trained in computational astrophysics and large-scale multiphysics codes at the University of Chicago, where she is currently pursuing a PhD in scientific computing. Contact her at riley@alcf.anl.gov.

FUYUKI SHIMOJO is a professor of physics at Kumamoto University. His main research interests are condensed-matter physics, computational materials science, and parallel algorithms for large-scale quantum-molecular simulations. Shimojo received a PhD in physics from Niigata University. He is a member of the APS and the Physical Society of Japan. Contact him at shimojo@kumamoto-u.ac.jp.

RAJIV K. KALIA is a professor of physics and astronomy, chemical engineering and materials science, and computer science at USC. His research focuses on multiscale simulations involving atomistic, mesoscale, and continuum approaches on parallel supercomputers. Kalia received a PhD in physics from Northwestern University. He is a Fellow of the APS. Contact him at rkalia@usc.edu.

PRIYA VASHISHTA is a professor of chemical engineering and materials science, physics and astronomy, and computer science at USC. His research interests include HPC and visualization for large multiscale simulations of materials and processes. Vashishta received a PhD in physics from the Indian Institute of Technology Kanpur. He is a Fellow of the APS. Contact him at priyav@usc.edu.

PAUL C. MESSINA is a Senior Computational Scientist and Distinguished Fellow at ANL. His research focuses on scientific computing environments. Messina received a PhD in mathematics from the University of Cincinnati. Contact him at messina@anl.gov.

vector parallelization opportunities. At a minimum, application developers must provide directives for restructuring their implementations to reflect vector parallelism. With hundred-way node-level parallelism here and thousand-way parallelism on the way, developers must identify all potential parallelism to make best use of these platforms.

Domain-specific languages (DSLs) might help expose such parallelism. Although DSLs are not widely used, that could change as developers gain more experience in their design and implementation. For example, the Tensor Contraction Engine has been used to generate two million lines of code for NWChem.[14] In addition, new scientific application frameworks could provide easier abstractions for levels of parallelism. For example, MADNESS (Multiresolution Adaptive Numerical Evaluation for Scientific Simulation) aims to facilitate application development by providing a higher-level abstraction for node parallelism.[15] To see wide adoption, these approaches must be scalable on systems of interest, have committed availability on supercomputers, and be well supported.

DCR is a scalable approach for analyzing large material systems but is not without disadvantages. One obvious shortcoming, inherent to all known linear-scaling DFT methods, is the inability to treat both insulating and metallic systems within a single, unified reduced-scaling ($O(N^x)$, $x < 3$) approach. This is a nontrivial methodological problem that arises from the distinct decay properties of the density matrix for insulators versus metals.

More problematic are DFT's inherent deficiencies, which can only be

is focused on ensuring capability and performance in OpenMP.

Regardless of the specific implementation for node-level parallelism, application codes will have to increase intranode and instruction-level parallelism. Vectorized cores are already common on supercomputers, but most application codes rely on compilers, which currently cannot detect many

overcome by either improved functionals or many-body methods such as QMC and CCSD(T). The systematic accuracy of many-body methods makes them particularly promising. With significantly higher computational costs than DFT, these methods demand exascale computers. One can imagine, in the not-too-distant future, such systems performing QMD using QMC instead of DFT on 10,000-electron systems or high-throughput screening (>100,000) of candidate materials using CCSD(T).

Although DCR can simulate material processes efficiently in a given structure, synthesizing a structure remains a challenge. Such "synthesis by design" will require new computational software to model various gas-phase and solution processes and discover how to control material processes at the level of electrons, design atom-precise nanostructures with tailored properties, and extract function–property–structure relationships. This will require simulation methods that can describe the dynamics of electrons and atoms far from equilibrium. In addition, we might need to climb further down the algorithmic-complexity ladder. Thanks to large systems' self-averaging properties, stochastic DFT algorithms can achieve sublinear scaling ($O(N^\epsilon)$ ($\epsilon < 1$)).[16] **C**

REFERENCES

1. P. Hohenberg and W. Kohn, "Inhomogeneous Electron Gas," *Physical Rev.*, vol. 136, no. 3B, 1964, pp. B864–B871.

2. W. Kohn and L.J. Sham, "Self-Consistent Equations Including Exchange and Correlation Effects," *Physical Rev.*, vol. 140, no. 4A, 1965, pp. A1133–A1138.

3. A. Rahman, "Correlations in the Motion of Atoms in Liquid Argon," *Physical Rev.*, vol. 136, no. 2A, 1964, pp. A405–A411.

4. M.C. Payne et al., "Iterative Minimization Techniques for *Ab Initio* Total-Energy Calculations: Molecular Dynamics and Conjugate Gradients," *Reviews of Modern Physics*, vol. 64, no. 4, 1992, pp. 1045–1097.

5. R. Car and M. Parrinello, "Unified Approach for Molecular Dynamics and Density-Functional Theory," *Physical Rev. Letters*, vol. 55, no. 22, 1985, pp. 2471–2474.

6. E. Prodan and W. Kohn, "Nearsightedness of Electronic Matter," *Proc. Nat'l Academy of Sciences*, vol. 102, no. 33, 2005, pp. 11635–11638.

7. W. T. Yang, "Direct Calculation of Electron Density in Density-Functional Theory," *Physical Rev. Letters*, vol. 66, no. 11, 1991, pp. 1438–1441.

8. F. Shimojo et al., "A Divide-Conquer-Recombine Algorithmic Paradigm for Large Spatiotemporal Quantum Molecular Dynamics Simulations," *J. Chemical Physics*, vol. 140, no. 18, 2014; http://dx.doi.org/10.1063/1.4869342.

9. K. Nomura et al., "Metascalable Quantum Molecular Dynamics Simulations of Hydrogen-on-Demand," *Proc. Int'l Conf. High Performance Computing, Networking, Storage and Analysis* (SC14), 2014, pp. 661–673.

10. D. Perez et al., "Accelerated Molecular Dynamics Methods: Introduction and Recent Developments," *Ann. Reports in Computational Chemistry*, vol. 5, R. Wheeler and D. Spellmeyer, eds., Elsevier, 2009, pp. 79–98.

11. A. Nakano et al., "Divide-Conquer-Recombine: An Algorithmic Pathway toward Metascalability," *Proc. 20 Years of Beowulf Workshop in Honor of Thomas Sterling's 65th Birthday* (Beowulf 14), 2014, pp. 17–27.

12. K. Shimamura et al., "Hydrogen-on-Demand Using Metallic Alloy Nanoparticles in Water," *Nano Letters*, vol. 14, no. 7, 2014, pp. 4090–4096.

13. W. Mou et al., "Nanoscopic Mechanisms of Singlet Fission in Amorphous Molecular Solid," *Applied Physics Letters*, vol. 102, no. 17, 2013; http://dx.doi.org/10.1063/1.4795138.

14. G. Baumbartner et al., "Synthesis of High-Performance Parallel Programs for a Class of *Ab Initio* Quantum Chemistry Models," *Proc. IEEE*, vol. 93, no. 2, 2005, pp. 276–292.

15. G.I. Fann et al., "MADNESS Applied to Density Functional Theory in Chemistry and Nuclear Physics," *J. Physics: Conf. Series*, vol. 78, 2007; doi:10.1088/1742-6596/78/1/012018.

16. R. Baer, D. Neuhauser, and E. Rabani, "Self-Averaging Stochastic Kohn-Sham Density-Functional Theory," *Physical Rev. Letters*, vol. 111, no. 10, 2013; http://dx.doi.org/10.1103/PhysRevLett.111.106402.