


Enabling particle applications for exascale computing platforms

The International Journal of High Performance Computing Applications 2021, Vol. 35(6) 572–597
© The Author(s) 2021
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10943420211022829
journals.sagepub.com/home/hpc


Susan M Mniszewski¹ , James Belak², Jean-Luc Fattebert³, Christian FA Negre¹, Stuart R Slattery³, Adetokunbo A Adedoyin¹, Robert F Bird¹, Choongseok Chang⁴, Guangye Chen¹, Stéphane Ethier⁴, Shane Fogerty¹, Salman Habib⁵, Christoph Junghans¹, Damien Lebrun-Grandié³, Jamaludin Mohd-Yusof¹, Stan G Moore⁶, Daniel Osei-Kuffuor², Steven J Plimpton⁶, Adrian Pope⁵, Samuel Temple Reeve³, Lee Ricketson², Aaron Scheinberg⁷, Amil Y Sharma⁴ and Michael E Wall¹

Abstract

The Exascale Computing Project (ECP) is invested in co-design to assure that key applications are ready for exascale computing. Within ECP, the Co-design Center for Particle Applications (CoPA) is addressing challenges faced by particle-based applications across four “sub-motifs”: short-range particle–particle interactions (e.g., those which often dominate molecular dynamics (MD) and smoothed particle hydrodynamics (SPH) methods), long-range particle–particle interactions (e.g., electrostatic MD and gravitational N-body), particle-in-cell (PIC) methods, and linear-scaling electronic structure and quantum molecular dynamics (QMD) algorithms. Our crosscutting co-designed technologies fall into two categories: proxy applications (or “apps”) and libraries. Proxy apps are vehicles used to evaluate the viability of incorporating various types of algorithms, data structures, and architecture-specific optimizations and the associated trade-offs; examples include ExaMiniMD, CabanaMD, CabanaPIC, and ExaSP2. Libraries are modular instantiations that multiple applications can utilize or be built upon; CoPA has developed the Cabana particle library, PROGRESS/BML libraries for QMD, and the SWFFT and fftMPI parallel FFT libraries. Success is measured by identifiable “lessons learned” that are translated either directly into parent production application codes or into libraries, with demonstrated performance and/or productivity improvement. The libraries and their use in CoPA’s ECP application partner codes are also addressed.

Keywords

Cabana particle toolkit, Co-design for exascale, particle applications, performance portability across architectures, PROGRESS/BML for electronic structure

1. Introduction

The US DOE Exascale Computing Project (ECP) Co-design Center for Particle Applications (CoPA) provides contributions to enable application readiness as we move toward exascale architectures for the “motif” of particle-based applications (Alexander et al., 2020). Co-design Center for Particle Applications focuses on co-design for the following “sub-motifs”: short-range particle–particle interactions (e.g., those which often dominate molecular dynamics (MD) and smoothed particle hydrodynamics (SPH) methods), long-range particle–particle interactions (e.g., electrostatic MD and gravitational N-body), particle-in-cell

¹Los Alamos National Laboratory, Los Alamos, NM, USA

²Lawrence Livermore National Laboratory, Livermore, CA, USA

³Oak Ridge National Laboratory, Oak Ridge, TN, USA

⁴Princeton Plasma Physics Laboratory, Princeton, NJ, USA

⁵Argonne National Laboratory, Lemont, IL, USA

⁶Sandia National Laboratories, Albuquerque, NM, USA

⁷Jubilee Development, Washington, DC, USA

Corresponding author:

Susan M Mniszewski, Computer, Computational and Statistical Sciences Division, Los Alamos National Laboratory, P. O. Box 1663, MS B214, Los Alamos, NM 87545-1663, USA.

Email: smm@lanl.gov

(PIC) methods, and $O(N)$ complexity electronic structure and quantum molecular dynamics (QMD) algorithms.

Particle-based simulations start with a description of the problem in terms of particles and commonly use a single-program multiple-data domain decomposition paradigm for internode parallelism. Because particles in each domain interact with particles outside its domain, a list of these outside particles must be maintained and updated through internode communication. This list of outside particles is commonly kept in a set of ghost cells or a ghost region on each node. Intra-node parallelism is commonly performed through work decomposition. From a description of the neighborhood (neighbor list), each particle's forces are calculated to propagate the particles to new positions. The particles are then resorted to begin the next time step. The main components of a time step across the sub-motifs are shown in Figure 1. PIC is unique in that particles are used to solve continuum field problems on a grid. Quantum molecular dynamics solves the computationally intensive electronic structure for problems where details of interatomic bonding are particularly important. Shared and specific functionality are highlighted in Figure 1. The compute, memory, and/or communication challenges requiring optimization on modern computer architectures are identified, extracted, and assembled into

libraries and proxy applications during the CoPA co-design process.

Co-design Center for Particle Application's co-design process of using proxy applications (or apps) and libraries has grown out of a predecessor project, the Exascale Co-Design Center for Materials in Extreme Environments (ExMatEx) (Germann et al., 2013). Two main library directions have emerged, the Cabana Particle Simulation Toolkit and the PROGRESS/BML QMD Libraries, each described in later sections. Each strive for performance portability, flexibility, and scalability across architectures with and without GPU acceleration by providing optimized data structure, data layout, and data movement in the context of the sub-motifs they address. Cabana is focused on short-range and long-range particle interactions for MD, PIC, and N-body applications, while PROGRESS/BML is focused on $O(N)$ complexity algorithms for electronic structure and QMD applications. This split is primarily motivated by the difference in sub-motifs: QMD is computationally dominated by matrix operations, while the other sub-motifs share particle and particle-grid operations. The locations for the open-source CoPA libraries and proxy apps are noted in the sections in which they are described.

The particle motif is used by many application codes to describe physical systems, including MD simulations using

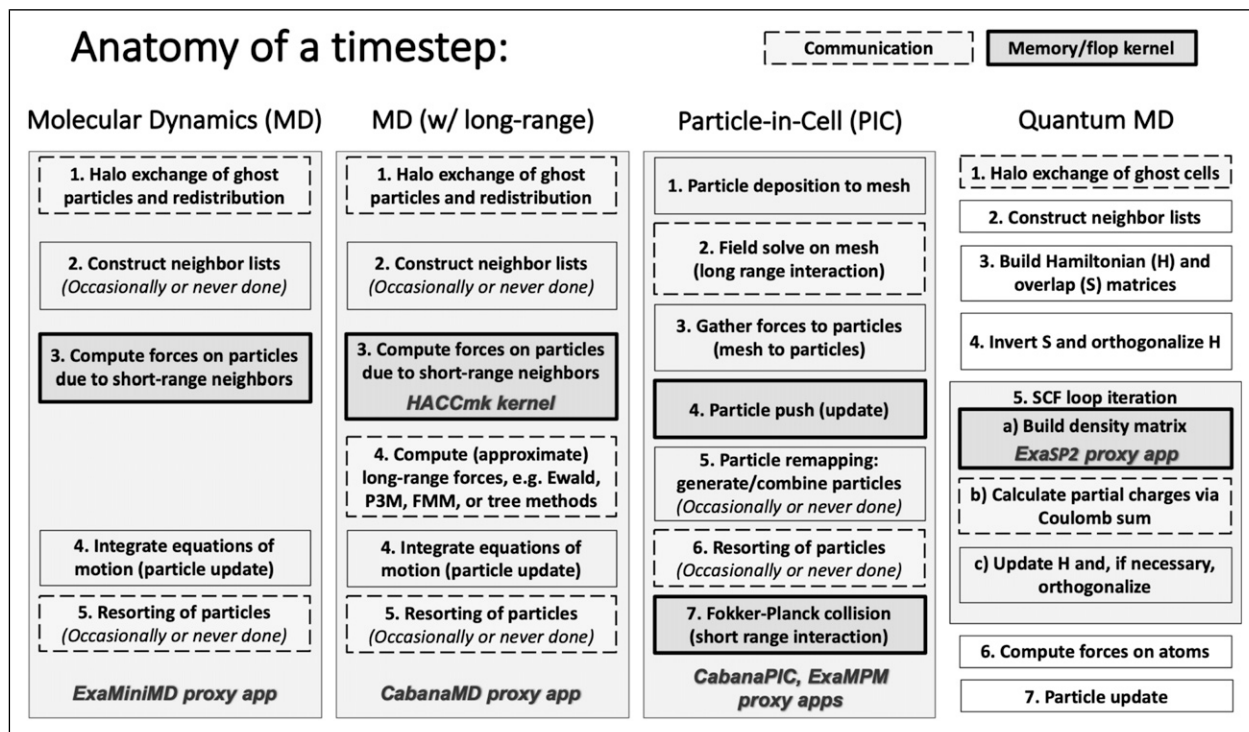


Figure 1. Anatomy of a time step is shown for each of the particle application sub-motifs addressed in Co-design Center for Particle Applications. Communication intensive steps and compute/memory intensive steps are shown in yellow and blue, respectively.

empirical models or the underlying quantum mechanics for particle interactions, cosmological simulations in which the particle may represent an object (e.g., a star) or a cluster of objects and the particle interaction is through gravity, and plasma simulations on grids within a PIC framework to solve the interaction of particles with the electrodynamic field. The computational motifs associated with these application codes depend on the nature of the particle interactions. Short-ranged interactions rely heavily on the creation of a list of neighbors for direct interactions, while long-range interactions use particle-grid methods and fast Fourier transforms (FFTs) to solve the long-range field problem. Details are described in the section on the Cabana toolkit. The quantum mechanics in QMD problems is often expressed as a matrix problem. Quantum molecular dynamics based on localized orbitals in density functional theory (DFT) and tight-binding models are reliant on sparse matrix solvers. Details are described in the section on the PROGRESS/BML Libraries.

Relevant particle applications are represented within CoPA and help drive the co-design process. Exascale Computing Project application projects such as EXAALT (LAMMPS-SNAP), WDMApp (XGC), ExaSky (HACC/SWFFT), and ExaAM (MPM) serve as application partners as shown in Figure 2, as well as non-ECP applications. Details of these engagements are described in the section on Application Partners.

We present descriptions of the Cabana Particle Simulation Toolkit and PROGRESS/BML QMD libraries, followed by PIC algorithm development, and co-design examples with our application partners: XGC, HACC, and LAMMPS-SNAP. We conclude with a summary of our lessons learned and impact on the broader community.

2. Cabana particle simulation toolkit

The Cabana toolkit is a collection of libraries and proxy applications which allows scientific software developers targeting exascale machines to develop scalable and portable particle-based algorithms and applications. The toolkit is an open-source implementation of numerous particle-based algorithms and data structures applicable to a range of application types including (but not limited to) PIC and its derivatives, MD, SPH, and N-body codes (Hockney and Eastwood (1989); Liu and Liu, 2003) and is usable by application codes written in C++, C, and FORTRAN. Notably, this covers the first three sub-motifs (see Figure 1). Cabana is designed as a library particularly because so many computational algorithms are shared across particle applications in these sub-motifs: neighbor list construction, particle sorting, multi-node particle redistribution and halo exchange, etc. This effectively separates shared capabilities from the specific application physics within individual steps of Figure 1, for example, the per-atom force computation in MD and the particle-grid interpolation for PIC. Cabana is available at <https://github.com/ECP-CoPA/Cabana>.

The toolkit provides both particle algorithm implementations and user-configurable particle data structures. Users of Cabana can leverage the algorithms and computational kernels provided by the toolkit independent of whether or not they are also utilizing the native data structures of the toolkit through memory-wrapping interfaces. The algorithms themselves span the space of particle operations necessary to support each relevant application type, spanning across all sub-motifs. This includes intra-node (local and threaded) operations on particles and inter-node (communication between nodes) operations to form

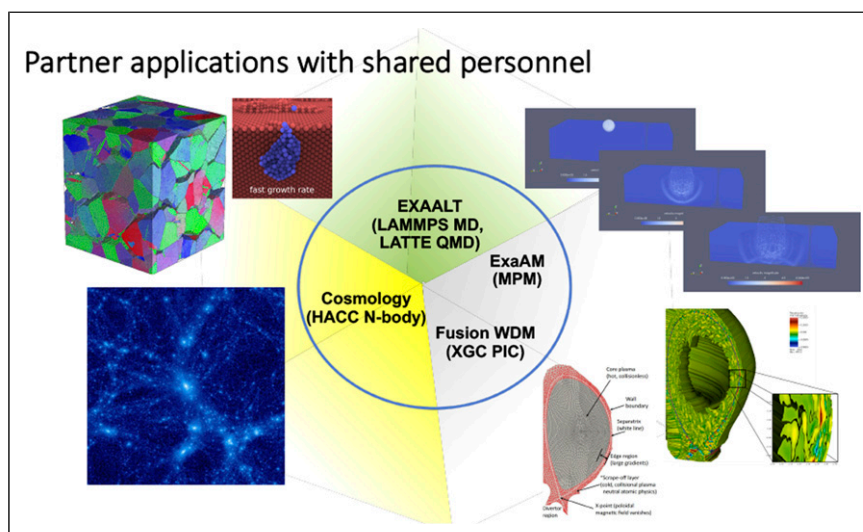


Figure 2. Partner “particle motif” applications with shared personnel are shown. These applications represent all the Co-design Center for Particle Applications sub-motifs.

a hybrid parallel capability. Cabana uses the Kokkos programming model for on-node parallelism (Edwards et al., 2014), providing performance and portability on pre-exascale and anticipated exascale systems using current and future DOE-deployed architectures, including multi-core CPUs and GPUs. Within Cabana, Kokkos is used for abstractions to memory allocation, array-like data structures, and parallel loop concepts which allow a single code to be written for multiple architectures.

While the only required dependency for Cabana is Kokkos, the toolkit is also intended to be interoperable with other ECP scientific computing libraries which the user may leverage for functionality not within the scope of the toolkit. Use of the library in concert with other ECP libraries can greatly facilitate the composition of scalable particle-based application codes on new architectures. Current library

dependencies are shown in Figure 3 with libraries developed by the ECP Software Technology (ST) projects, including hypre for preconditioners and linear solvers (Falgout and Yang, 2002), heFFTe for 3D distributed FFTs (Ayala et al., 2019), and ArborX (Lebrun-Grandié et al., 2019) for threaded and distributed search algorithms.

Cabana includes both particle and particle-grid operations which are critical across the particle sub-motifs. We next review each of these capabilities.

2.1. Particle abstractions

We first summarize the major particle-centric abstractions and functionality of the toolkit including the underlying data structure and the common operations which can be applied to the particles.

2.1.1. Data structures. Particles in Cabana are represented as tuples of multidimensional data. They are general and may be composed of fields of arbitrary types and dimensions (e.g. a mass scalar, a velocity vector, a stress tensor, and an integer id number) as defined by the user’s application. Considering the tuple to be the fundamental particle data structure (struct), several choices exist for composing groups of structs as shown in Figure 4. A simple list of structs, called an Array-of-Structs (AoS) is a traditional choice for particle applications, especially those not targeting optimization for vector hardware. All of the data for a single particle are encapsulated in a contiguous chunk of memory, thereby improving memory locality for multiple fields within a particle. An AoS also offers simplicity for basic particle operations, such as sorting or communication, as the memory associated with a given particle may be manipulated in a single operation. An AoS, however, also has a downside accessing the same data component in multiple particles concurrently requires strided (non-coalesced)

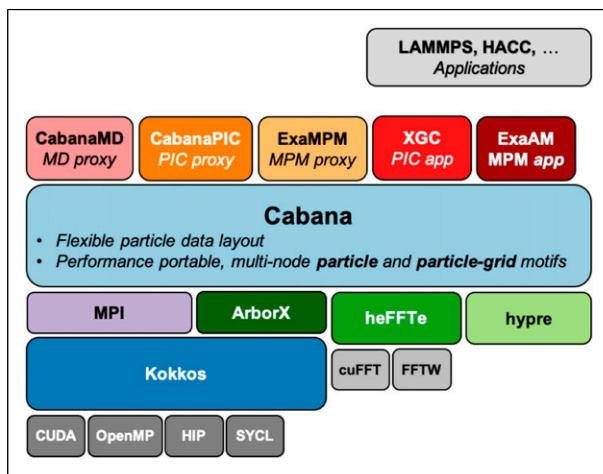


Figure 3. Cabana software stack including dependencies, proxy apps, and production apps.

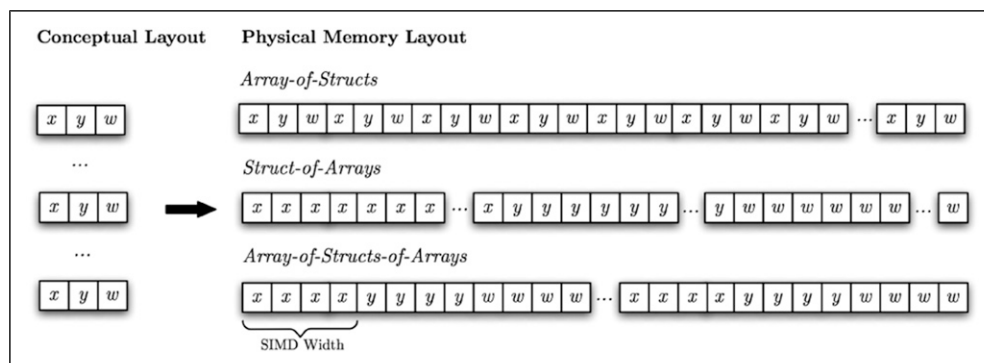


Figure 4. Particles in Cabana are stored in an Array-of-Structs-of-Arrays (AoSoA). Compared to an Array-of-Structs, an AoSoA will provide similar memory locality benefits during data access while also coalescing data access over Single Instruction Multiple Data-length elements when possible. Compared to a Struct-of-Arrays, an AoSoA will provide similar coalescing data access benefits while also introducing additional memory locality.

memory accesses which incur a significant performance penalty on modern vector-based machines.

This penalty for strided access can be alleviated through the use of the Struct-of-Arrays (SoA) memory layout. In an SoA, each particle field is stored in a separate array, with the values of an individual field stored contiguously for all particles. This structure allows for a high-performance memory access pattern that maps well to modern vector-based architectures. The drawback of this approach, however, is two-fold: first, the hardware has to track a memory stream for each particle property used within a kernel and second, the programmer and hardware may have a harder time efficiently operating on all of the data together for a given particle. In light of these features, it is typically favorable to use SoA when effective use of vector-like hardware is vital (as is the case with GPUs), or when a subset of particle fields are used in a given kernel.

Cabana offers a zero-cost abstraction to these memory layouts and further implements a hybrid scheme known as Array-of-Structs-of-Arrays (AoSoA). An AoSoA attempts to combine the benefits of both AoS and SoA by offering user-configurable sized blocks of contiguous particle fields. This approach means a single memory load can fetch a coalesced group of particle field data from memory, while retaining high memory locality for all fields of a given particle. A key performance tuning parameter exposed by Cabana, the added AoSoA dimension enables the user to configure the memory layout of a given particle array at compile time or to have it automatically selected based on the target hardware.

2.1.2. Particle sorting. Particle sorting is a functionality requirement of all currently identified user applications. In plasma PIC, fluid/solid PIC, MD, N-body, and SPH calculations, particle sorting serves as a means of improving memory access patterns based on some criteria which can provide an improvement in on-node performance. This could be, for example, placing particles that access the same grid cell data near each other in memory, or grouping particles together by material type such that particles adjacent in memory will be operated on by the same computational kernel in the application. The frequency of sorting often depends on the application, as well as the given problem. Slowly evolving problems may need to sort particles less frequently as local particle properties that affect memory access (e.g., grid location or nearest neighbors) will be relatively stable. Efficient memory access objectives should be based on the target computing platform (e.g., GPUs or multi-CPU devices). Particle sorting is applicable to locally owned particles or to both locally owned and ghost particles. Cabana provides the ability to sort particles by spatial location through geometric binning or by an arbitrary key value, which can include either a particle property or another user-provided value. Cabana uses the

bin sort functionality in Kokkos, with plans for additional options through Kokkos in the future.

2.1.3. Neighbor list creation. Particle-in-cell, MD, N-body, and SPH can all benefit from the efficient construction of particle neighbor lists. These neighbor lists are typically generated based on distance criteria where a physical neighborhood is defined or instead based on some fixed number of nearest neighbors. In both MD and SPH simulations, the neighbor list is a critical data structure and is computed more frequently than most other particle operations up to the frequency of every time step. In fluid/solid PIC applications, the neighbor list is an auxiliary data structure for computational convenience that, when used, is similarly computed up to the frequency of every time step. In the fluid/solid PIC case the background grid can be used to accelerate the neighbor search, whereas in MD, N-body, and SPH, a grid may need to be created specifically for this search acceleration. Cabana provides multiple variations of neighbor lists, including traditional binning methods used in many MD applications (Verlet lists), as well as tree-based algorithms from the ECP ST ArborX library (Lebrun-Grandié et al., 2019), using compressed and dense storage formats, and thread-parallel hierarchical list creation; all of these options can improve performance for different architectures and particle distributions.

2.1.4. Halo exchange and redistribution. Many user applications require parallel communication of particles between compute nodes when spatial domain decomposition is used and particle data must be shared between adjacent domains. Some applications require halo exchange operations on particles and some, in particular grid-free methods (e.g., MD), additionally require ghost particle representations to complete local computations near domain boundaries. In many cases, the halo exchange is executed at every single time step of the simulation with a communication pattern that may also be computed at every time step or at some larger interval and reused between constructions. In addition, particles need to be redistributed to new compute nodes in many algorithms either as a result of a load balancing operation or because advection has moved particles to a new region of space owned by another compute node. The toolkit provides implementations for ghost particle generation and halo exchange, including both gather and scatter operations, as well as a migration operation to redistribute particles to new owning domains.

2.1.5. Parallel loops. Cabana adds two main extensions to the Kokkos parallel constructs which handle portable threaded parallelism and mapping hierarchical and nested parallelism to up to three levels on the hardware. First, Single Instruction Multiple Data (SIMD) parallel loops are directly connected to the AoSoA data structures and provide the user

simple iteration over the added inner vector dimension, for threaded parallelism over both particle structs and vector, with potential performance improvements by exposing coalesced memory operations. Second, neighbor parallel loops provide both convenience and flexibility for any particle codes which use a neighbor list (see above) to iterate over both particles and neighboring particles (including first- and/or second-level neighbors). Cabana handles all neighbor indexing and the user kernel deals only with application physics. This also enables applications to easily change the parallel execution policy and use the appropriate threading over the neighbor list structures (or serial execution) for the kernel, problem, and hardware.

2.2. Particle-grid abstractions

In addition to pure particle abstractions, the toolkit also contains optional infrastructure for particle-grid concepts which we present next.

2.2.1. Long-range solvers. Among the user applications, long-range solvers encapsulate a wide variety of kernels and capabilities, but many include critical kernels that can apply to many applications. For example, embedded within the long-range solve of a PIC operation are kernels for interpolating data from the particles to the grid and from the grid to the particles, as well as possibly grid-based linear solvers. Other simulations, such as MD and SPH calculations, compose the long-range solvers with particle-grid operations but instead use other algorithms such as FFTs to complete the long-range component of the solve. A variety of libraries provide FFT capabilities, including high-performance scalable FFT libraries being developed for large systems. Cabana currently implements direct use of the ECP ST heFFTe library for performance portable FFTs (Ayala et al., 2019) and Cabana’s flexibility will enable use of other ECP FFT libraries in the future, including FFTX (Franchetti et al., 2020), SWFFT (Pope et al., 2017), and fftMPI (Plimpton et al., 2018). In addition, Cabana interfaces to hypre (Falgout and Yang, 2002) to provide interfaces to linear solvers.

2.2.2. Particle-grid interpolation. Particle-in-cell methods, as well as methods which require long-range solvers, usually need some type of interpolation between particle and grid representations of a field in order to populate the grid data needed by FFTs, linear solvers, or other field operations. The toolkit provides services for interpolation to logically structured grids based on multidimensional spline functions, which are available in multiple orders. By differentiating the spline functions, differential operators may be composed during the interpolation process, allowing users to interpolate gradients, divergences, and other operators of scalar, vector, and tensor fields. Other types of interpolants

can also be added in the future for additional capabilities in PIC applications. As needed, we also envision generalizing the Cabana interpolation infrastructure to support user-defined interpolants on structured grids.

2.3. Proxy apps

Cabana-based proxy apps have been developed in order to demonstrate and improve Cabana functionality as well as to explore new algorithms and ideas when they are deployed in the context of a sub-motif. In addition to those presented here, more proxies are planned to cover additional variations of the algorithmic abstractions represented by application partners.

2.3.1. CabanaMD. CabanaMD is a LAMMPS (Plimpton, 1995) proxy app built on Cabana, developed directly from the ExaMiniMD proxy (“KokkosMD”) (Edwards et al., 2014). CabanaMD represents both the short- and long-range MD sub-motifs; in fact, the MD time step can easily be re-expressed as calls to the Cabana library, as shown in Figure 5. Similar figures could also be created for all sub-motifs in Figure 1 with all the CoPA proxy apps.

CabanaMD is available at <https://github.com/ECP-CoPA/CabanaMD>. MD uses Newton’s equations for the motion of atoms, with various models for interatomic forces and ignoring electrons. In contrast to ExaMiniMD and many applications, only the main physics, the force kernel evaluating the interatomic model and position integration, is entirely retained in the application with everything else handled by Cabana. Main results for demonstrating Cabana capabilities include performance implications of combining or separating particle properties in memory, changing the particle property AoSoA memory layouts, and using the available options for each algorithm (e.g., data layout,

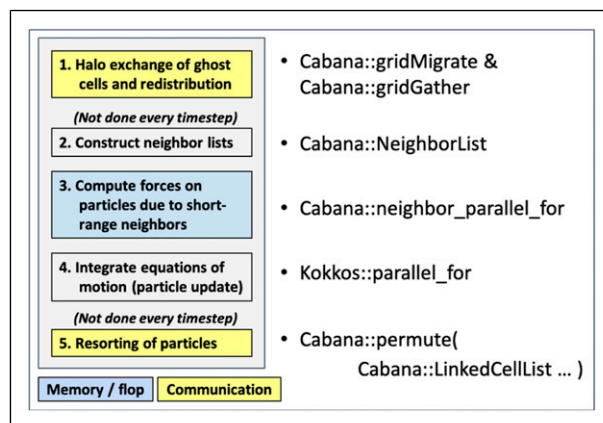


Figure 5. Single molecular dynamics time step re-expressed with the Cabana/Kokkos API. We have purposefully mapped our API to the main algorithmic components of a time step.

hierarchical creation, and hierarchical traversal for neighbor lists). This has been done primarily with the Lennard-Jones short-range force benchmark kernel. In addition, CabanaMD enabled speedup on the order of $3\times$ (one Nvidia V100 GPU compared to a full IBM Power 9 CPU node) in new neural network interatomic models (Behler and Parrinello, 2007) by re-implementing with Kokkos and Cabana, rewriting the short-range force kernels, and exposing threaded parallelism (Desai et al., 2020).

CabanaMD also includes long-range forces using Cabana data structures and particle-grid algorithms, covering the second sub-motif. The smooth particle mesh Ewald (Essmann et al., 1995) method is implemented with Cabana grid structures and spline kernels to spread particle charge onto a uniform grid, 3D FFTs using the Cabana interface to heFFTe (Ayala et al., 2019) for reciprocal space energies and forces, and Cabana gradient kernel to gather force contributions back from the grid to atoms. Real-space energy and force calculations use the Cabana neighbor parallel iteration options, as with short-range interactions. Continuing work for CabanaMD will include benchmarking long-range performance and using it as a vehicle to implement and improve performance portable machine learned interatomic models.

2.3.2. CabanaPIC. CabanaPIC is a relativistic PIC proxy app using Cabana, capable of modeling kinetic plasma and representing the PIC sub-motif. CabanaPIC is available at <https://github.com/ECP-CoPA/CabanaPIC>. It has strong ties with the production code VPIC (Bowers et al., 2009) but is able to act as a representative proxy for all traditional electromagnetic PIC codes which use a structured grid. It implements a typical Boris pusher, as well as a finite-difference time-domain field solver.

CabanaPIC focuses on short-range particle-grid interactions, and its performance is heavily dependent on techniques to martial conflicting writes to memory (such as atomics). It employs Cabana's particle sorting techniques, as well as offers examples of how to use Cabana for simple MPI-based particle passing.

2.3.3. ExaMPM. ExaMPM is a Cabana-based proxy application for the material point method (MPM) which is being used as part of high-fidelity simulations of additive manufacturing with metals in the ExaAM project in ECP. ExaMPM also represents the PIC sub-motif and is available at <https://github.com/ECP-CoPA/ExaMPM>. Material point method, a derivative of PIC, is used to solve the Cauchy stress form of the Navier–Stokes equations including terms for mass, momentum, and energy transport where particles track the full description of the material being modeled in a Lagrangian and continuum sense. The MPM simulations in ExaMPM model the interaction of a laser with metal powders, the melting of the powder due to heating from the

laser, and the solidification of the melted substrate after the laser is turned off. When modeled at a very high fidelity, using particles to track the free surface interface of the molten metal and the liquid–solid interface during phase change will allow for both empirical model generation in tandem with experiments, as well as reduced-order model generation to use with engineering-scale codes in the ExaAM project.

ExaMPM largely implements the base algorithmic components of the MPM model including an explicit form of time integration, a free surface formulation with complex moving interfaces, and higher-order particle-grid transfer operators which reduce dissipation. As an example, Figure 6 shows a water column collapse modeled with ExaMPM. This problem has a dynamic moving surface structure resembling the dynamics of the molten metal in the high-fidelity ExaAM simulations as well as particle populations which change rapidly with respect to the local domain. Scaling up this problem through larger particle counts and larger computational meshes will allow us to study better techniques for interface tracking, more scalable communication and load balancing algorithms to handle the moving particles, and sorting routines to improve locality in particle-grid operations.

2.4. Cabana applications

A significant metric for the impact of a given software library is its adoption. First, Cabana is being used as the basis for a new production application closely related to the ExaMPM proxy described in the previous section. Another notable usage of Cabana is in the transition of an existing application; the section on XGC-Cabana below details the process of converting XGC, initially using Cabana through FORTRAN interfaces, eventually to full C++ with Cabana.

In addition, the Cabana library has and will continue to influence production applications and related libraries. This includes LAMMPS and HACC (also described in later sections), where the performance of an algorithm, data layout, etc. can be demonstrated with Cabana and/or its proxy apps and migrated to the separate application; similarly, interactions between Cabana and the libraries it depends on can improve each for a given application.

3. PROGRESS/BML quantum molecular dynamics libraries

This section focuses on the solvers addressing the quantum part of QMD, the sub-motif listed in the fourth column of Figure 1. Quantum molecular dynamics uses electronic structure (ES)–based atomic forces to advance the position of classical particles (atoms) in the Born–Oppenheimer approximation (Marx and Hutter, 2009). There are many benefits of this technique as compared to classical methods.

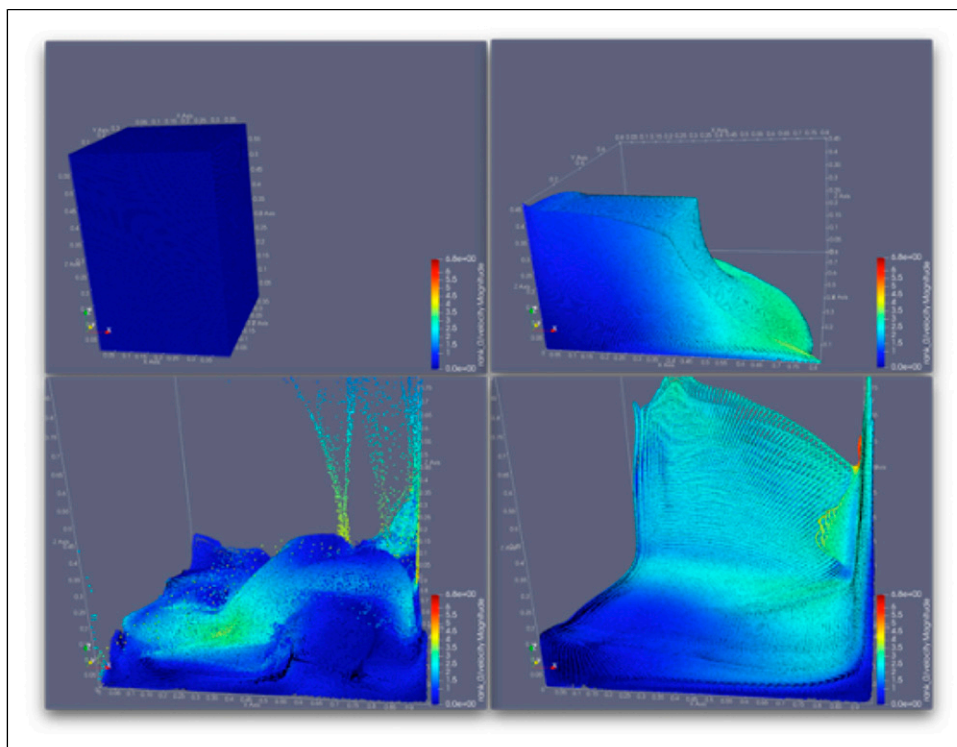


Figure 6. Snapshots of a 3D water column collapse modeled with ExaMPM (ordered clockwise starting from top left). Complex moving interfaces and dynamic local particle populations are being used to study improved algorithms for load balancing, communication, and particle sorting.

These benefits include independence of the results associated with the choice of a particular force field; enabling the formation and breaking of bonds (chemical reactions) as the simulation proceeds; and the possibility of extracting information from the electronic structure throughout the simulation. The counterpart to these benefits is the large computational cost associated with having to determine the ES of the system before advancing the particle coordinates. In order to perform practical MD simulations, the strong scaling limit becomes important, as time-to-solution needs to be as small as possible to enable long simulations with tens of thousands of time steps, and achieve a good sampling of the phase space. Determining the ES is the main bottleneck of QMD, where the so-called single-particle density matrix (DM) needs to be computed from the Hamiltonian matrix. The latter requires a significant amount of arithmetic operations, and it typically scales with the cube of the number of particles. Quantum molecular dynamics is hence characterized by this unique critical step that sets it apart from all the particle simulation methods within the scope of the Cabana toolkit (described in the previous section); therefore, additional libraries are needed for increasing its performance and portability.

Another significant challenge surrounding the development of QMD codes is that solvers used to compute the

ES are strongly dependent on the chemical systems (atom types and bonds), which implies that it is necessary to develop and maintain several different algorithms that are suitable for each particular system. For instance, the single-particle DM for insulators (with a wide energy gap between the highest-occupied and lowest-unoccupied state) is essentially a projector onto the subspace spanned by the eigenfunctions associated with the lowest eigenvalues of the Hamiltonian matrix (occupied states). For metals, on the other hand, the DM is not strictly a projector since a temperature-dependent weight between 0 and 1 is associated with each eigenstate (Fermi–Dirac distribution). Nevertheless, in both cases, the DM can be computed from the knowledge of the eigenpairs of the Hamiltonian which are computationally expensive to determine.

In general, we can say that the construction of the DM ρ can be expressed as a matrix function of the Hamiltonian H . Such a function $\rho = f(H)$ can be computed exactly by diagonalizing matrix H . The function then becomes $\rho = Cf(\xi)C^T$, where ξ is a diagonal matrix containing the eigenvalues of H , and C is a unitary transformation where its columns contain the eigenvectors of H . f is the Fermi distribution functions for finite temperatures.

In order to increase productivity in the implementation and optimization of these algorithms, we have adopted a

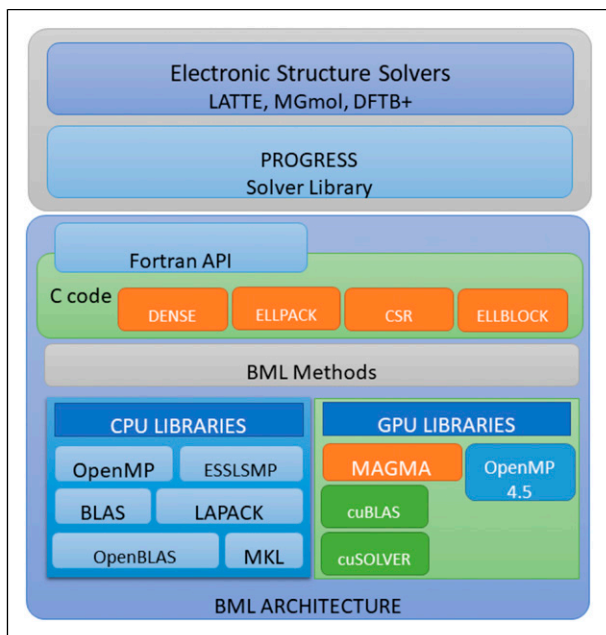


Figure 7. PROGRESS/BML software stack including dependencies and production apps.

framework in which we clearly separate the matrix operations from the solver implementations. The framework relies on two main libraries: “Parallel, Rapid $O(N)$, and Graph-based Recursive Electronic Structure Solve” (PROGRESS), and the “basic matrix library” (BML). The software stack can be seen in Figure 7. At the highest level, electronic structure codes call the solvers in the PROGRESS library which, in turn, rely on BML. The BML provides basic matrix data structure and operations. These consist of linear algebra matrix operations which are optimized based on the format of the matrix and the architecture where the program will run. Applications can also directly implement specific algorithms based on BML when those are not available routines in PROGRESS. Both libraries use travisci and codecov.io for continuous integration and code coverage analysis, respectively. Every commit is tested over a set of compiler and compiler options. Our overarching goal is to construct a flexible library ecosystem that helps to quickly adapt and optimize electronic structure applications on exascale architectures. Alternative libraries that overlap with the matrix formats and algorithms implemented in PROGRESS and BML include DBCSR (The CP2K Developers Group, 2020) and NTPoly (Dawson and Nakajima, 2018), both focusing also on electronic structure applications.

3.1. Basic matrix library

The increase in availability of heterogeneous computer platforms is the motivation behind the development of the

BML software library. Multiple data storage formats (both for sparse and dense) and programming models (distributed, threaded, and task-based) complicate the testing and optimization of electronic structure codes.

The BML package contains the essential linear algebra operations that are relevant to electronic structure problems. The library is written in C, which allows for straightforward implementation on exascale machines. The library also exposes a Fortran interface, with Fortran wrappers written around C functions. This facilitates its usage by a wide variety of codes since many applications codes in this community are written in Fortran. One of the main advantages of BML is that the APIs are the same for all matrix types, data types, and architectures, enabling users to build unified solvers that work for multiple matrix formats. Low-level implementations within the BML library are tailored to particular matrix formats and computer architectures. The formats that can be handled so far are as follows: dense, ELLPACK-R (Mniszewski et al., 2015), compressed sparse row (CSR) (Saad, 2003), and ELLBLOCK (see Figure 8). Here, dense is used to refer to a typical two-dimensional array. It is the most suitable format for treating systems that have a high proportion of nonzero values in the DM. ELLPACK-R is a sparse matrix format constructed using three arrays: a one-dimensional array used to keep track of the number of nonzeros per row for each row; a two-dimensional array used to keep track of the column indices of the nonzero values within a row; and finally, another two-dimensional array used to store the nonzero values. The row data are zero-padded to constant length, so row data are separated by constant stride. ELLBLOCK is a block version of ELLPACK-R. In a nutshell, a matrix is decomposed into blocks that are either considered full of zeroes and not stored or dense blocks that are treated as all nonzeros. Loops over nonzero matrix elements are replaced by loops over nonzero blocks, and dense linear algebra operations are done on blocks. The CSR format in its typical implementation utilizes a floating point array to store the nonzero entries of the matrix in row-major order and an integer array to store the corresponding column indices. In addition, an array which indexes the beginning of each row in the data arrays is needed for accessing the data. Unlike the ELLPACK-R format, which stores entries in a two-dimensional array with a fixed width for all the nonzero entries in the rows of the matrix, the CSR format keeps the variability in the number of nonzeros per row, thus avoiding the need for zero padding. The implementation of the CSR format in BML follows an Array-of-Structs-of-Arrays (AoSoA) approach. A matrix is represented as an array of CSRs, where each compressed row stores only the nonzero entries and the associated column indices of the corresponding row of the matrix. In this approach, the additional array of indexes to the beginning of each row is no longer needed. The matrix stored in this way is extensible, allowing the matrix to grow

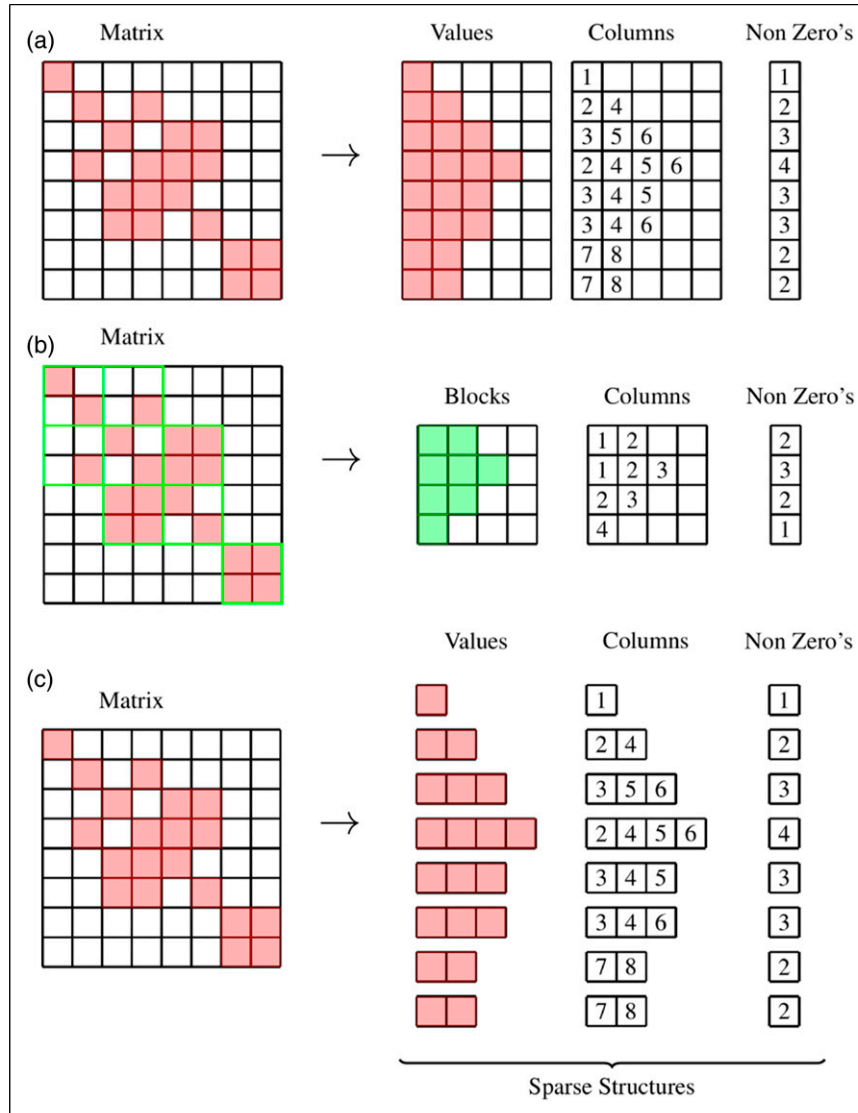


Figure 8. Three sparse matrix formats currently available in BML. (a) ELLPACK-R: A 2D array containing the compressed rows; a second 2D array containing the column indices; and a third 1D array containing the maximum nonzeros per row. (b) ELLBLOCK: Block version of ELLPACK-R where a matrix is decomposed into blocks that are either considered full of zeroes and not stored or dense blocks that are treated as all nonzeros. (c) CSR: A matrix is represented as an array of compressed sparse rows, where each compressed row stores only the nonzero entries and the associated column indices of the corresponding row of the matrix. BML: basic matrix library; CSR: compressed sparse row.

by simply adding new entries without the need to destroy the matrix.

Basic matrix library dense matrix functions are typically wrappers on top of a vendor optimized library, such as BLAS/LAPACK implementations. For NVIDIA GPUs, we use the MAGMA library for dense matrices (Dongarra et al., 2014). We use its memory allocation functions and many of its functionalities. One exception is the dense eigensolver for which we use the NVIDIA cuSOLVER which performs substantially better than the MAGMA solver at the moment. In the case of the sparse formats, each BML function is

specifically written for that particular format. Performance portability is achieved by keeping one codebase with a high flexibility for configuring and building. Basic matrix library was compiled and tested with multiple compilers (GNU, IBM, Intel, etc.) on several pre-exascale architectures.

BML ELLPACK-R matrix functions are implemented with OpenMP, both on CPU and GPU, the latter using target offload. The algorithm implemented utilizes a work array of size (N) per row which is larger than GPU cache for matrix sizes of interest, leading to poor performance on GPU. Previous work by Mohd-Yusof et al. (2015) demonstrated

the performance of a novel merge-based implementation of sparse matrix multiply on GPU, implemented in CUDA. Future implementations will utilize a mix of OpenMP offload and native CUDA kernels to enable performance while retaining a consistent interface with the existing OpenMP implementations. Benchmarking indicates this should allow a speedup of $\sim 8\times$ on an Nvidia V100 compared to IBM Power 9 (using all 21 cores of one socket).

Basic matrix library offers support for four data types: single precision, double precision, complex single precision, and complex double precision. The source code for all these data types is the same for most functions, with C macros that are preprocessed at compile time to generate functions for the four different formats. All BML function names are prefixed with `bml_`. The code listing in Figure 9 shows the use of the BML API on one of our $O(N)$ complexity algorithms, the “second-order spectral projection” (SP2) (Niklasson et al., 2003). We show how BML matrices are allocated passing the matrix type (variable “ellpack” in this case), the element kind (a real kind indicated with the predefined variable `bml_element_real`), and the precision, (in this case a double passed with the variable `dp`). More information about how to use the BML API can be found at <https://lanl.github.io/bml/API/index.html>. Our implementation of various matrix formats is quite mature for CPUs, including their threaded versions. GPU implementation efforts are ongoing. Future developments include a distributed version of BML for various matrix types.

A recent effort Adedoyin et al. (2019), focused on optimizing at the multi-threaded level as opposed to modifying the data structures for performance at the SIMD scale. Several active and passive directives/pragmas were incorporated that aid to inform the compiler on the nature of the data structures and algorithms. Although these optimizations targeted multicore architectures, most are also applicable to many-core architectures present on modern heterogeneous platforms. Herein, we refer to multicore systems as readily available HPC nodes customarily configured with approximately 10–24 cores per socket at high clock speed (2.4–3.9 GHz) and many-core systems as accelerators or GPUs. Several optimization strategies were introduced including (1) strength reduction, (2) memory alignment for large arrays, (3) non-uniform memory access (NUMA) aware allocations to enforce data locality, and (4) the appropriate thread affinity and bindings to enhance the overall multi-threaded performance.

A more in-depth description of the BML library and its functionalities can be found in Bock et al. (2018), and the code is available at <https://github.com/lanl/bml>.

3.2. PROGRESS library

The computational cost of solving this eigenvalue problem to compute the DM, scales as $O(N^3)$, where N is the number

```

!declare BML matrices
type (bml_matrix_t) :: ham, rho, x2
!n:matrix size, m:max. nnz/row
integer :: n, m, dp
real (8) :: thld, tol, nel

dp = kind(1.0d0)

!allocate double precision BML matrices
call bml_zero_matrix("ellpack", &
&bml_element_real, dp, n, m, ham)
call bml_zero_matrix("ellpack", &
&bml_element_real, dp, n, m, rho)
call bml_zero_matrix("ellpack", &
&bml_element_real, dp, n, m, x2)
...
trx = bml_trace(rho)
do i = 0, niter
  call bml_multiply_x2(rho, x2, thld)
  trx2 = bml_trace(x2)
  if (trx2 .le. nel) then
    ! rho <- 2 * X - X2
    call bml_add(2., rho, -1., x2, thld)
    trx = 2.0 * trx - trx2
  else
    ! rho <- X2
    call bml_copy(x2, rho)
    trx = trx2
  end if
  if (abs(nel-trx) < tol) exit
end do
...
call bml_deallocate(x2)
...

```

Figure 9. Fortran example of the SP2 implementation using the BML ELLPACK-R format with a dropping threshold `thld`. The algorithm returns the DM `rho`, while its parameters are the number of electrons `nel` and the Hamiltonian `ham`. n and m are the total number of orbitals and the maximum number of nonzeros per row. Only the main operations are shown for brevity. BML: basic matrix library; DM: density matrix.

of atomic orbitals in the system. Recursive methods, however, such as SP2 (Niklasson et al., 2003), can compute the DM in $O(N)$ operations for a sparse Hamiltonian matrix.

PROGRESS is a Fortran library that can be used for general purpose quantum chemistry calculations. It implements several $O(N)$ solvers (Niklasson et al., 2016; Negre et al., 2016; Mniszewski et al., 2019) and is publicly available at <https://github.com/lanl/qmd-progress>. As described above and shown in Figure 7, PROGRESS relies entirely on BML for algebraic operations; hence, while electronic structure algorithms and solvers are outlined in PROGRESS, the mathematical manipulations are all performed in BML. This library is currently used by LATTE,

a tight-binding (TB) code specifically developed to perform QMD simulations (Bock et al., 2008). It can also be used with DFTB+, a widely used density functional tight-binding code (Hourahine et al., 2020). In TB methods, matrix elements are typically obtained empirically from fits to more accurate calculations or to experiments, rather than being explicitly computed from electronic wave functions. However, the BML library also can be used for first-principles codes, in particular for $O(N)$ codes where matrix elements correspond to pairs of localized orbitals (Fattebert et al., 2016).

As was mentioned previously, the appropriate solver to compute the electronic structure depends strongly on the type of chemical system. Metals, for example, are difficult to treat since their electronic structure is hard to converge given the delocalized nature of the electrons. The Hamiltonian and DM have different sparsity patterns that will determine the matrix format to use. Hence, there is room for exploring different matrix formats and solvers depending on the type of system. The SP2 method, as implemented in the PROGRESS library with the possibility of using BML sparse matrix multiplications, computes the DM without diagonalizing the Hamiltonian matrix. An example SP2 algorithm as implemented in PROGRESS is shown in Figure 9. Its computational complexity becomes $O(N)$ for sparse Hamiltonians, provided a proper thresholding is applied at every iteration. Performance of the PROGRESS library is tested using model Hamiltonian matrices that mimic the actual Hamiltonians for different materials such as semiconductors, soft matter, and metals.

Figure 10 shows the performance of two typical PROGRESS routines for constructing the DM on GPU applied to a soft-matter type of Hamiltonian. The standard algorithm for constructing the DM is based on matrix diagonalization (shown in black on the plot of Figure 10). The SP2 algorithm (see Figure 9), instead, is based on matrix multiplications (shown in red on the plot of Figure 10). The computational complexity is $O(N^3)$ for both algorithms due to the nature of the matrix operations involved. In both cases, these operations scale as $O(N^3)$ for dense (unthresholded) matrices. Furthermore, in these cases, the DM is solved exactly since no threshold is used. For systems where the DM becomes dense and where a sparse format cannot be used, the GPU versions of these algorithms are significantly more performant than the corresponding CPU threaded version. We also notice that the DIAG algorithm is slower than SP2 for smaller systems (less than 6000 orbitals). This is because the dense diagonalization algorithm and its implementation on GPUs are not as efficient, in particular for small matrices, while the SP2 algorithm is dominated by matrix–matrix multiplications which can be implemented very efficiently on GPUs. For large systems, however, the DIAG algorithm performs slightly better.

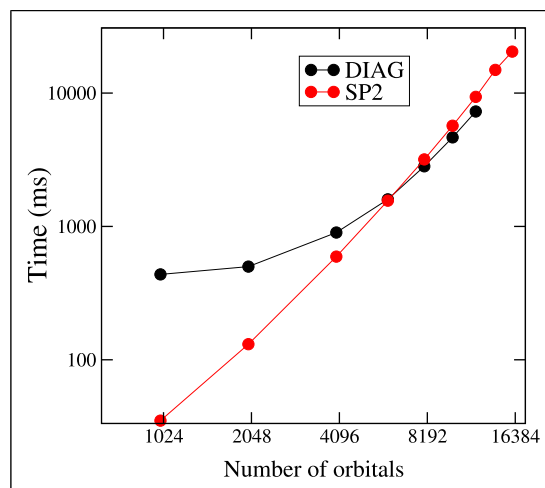


Figure 10. GPU performance comparison of two PROGRESS library routines for constructing the DM: traditional algorithm based of matrix diagonalization (DIAG, upper curve) and the SP2 algorithm (SP2, lower curve) based on matrix multiplications. Diagonalization is using the NVIDIA cuSOLVER library, while SP2 relies on the MAGMA matrix-matrix multiplication function. The plot shows the wall clock time to construct the DM as a function of the number of orbitals. Scaling experiments were run on an OLCF Summit node using one V100 NVIDIA GPU. The dense format is used for all BML matrices. SP2: second order spectral projection; DM: density matrix; OLCF: Oak Ridge Leadership Computing Facility.

For systems leading to a sparse DM, $O(N)$ complexity is achieved by using the SP2 algorithm in combination with a sparse format as is shown in Figure 11. This plot shows the performance of the SP2 algorithm on CPU using different formats (ELLPACK-R, CSR, and ELLBLOCK) applied to a soft-matter type Hamiltonian. In this figure, we notice a large gain in performance obtained by using sparse formats, and the $O(N)$ complexity for the range of system sizes analyzed. In these cases, the DM is not exact and the error depends on the threshold parameter, regardless of which of the three formats is used, and can be chosen to be sufficient for many practical applications.

Depending on the size of system to be simulated and the resulting structure of the matrix, optimal time to solution may be obtained from a particular combination of matrix format and hardware choice. The use of PROGRESS and BML allows these choices to be made at runtime, without changing the underlying code structure.

3.3. PROGRESS/BML applications

Our work on PROGRESS/BML is now being used to develop a capability for all-atom QMD simulations of proteins, extending the impact of our ECP work to biomedical research including studies of SARS-CoV-2 proteins. Current classical biomolecular MD simulations typically

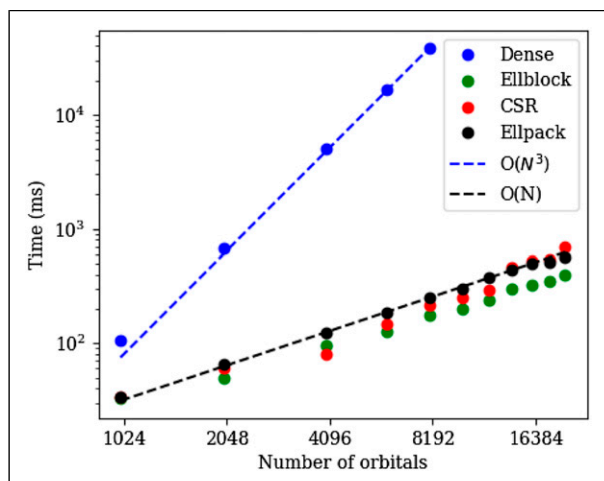


Figure 11. Performance of the SP2 algorithm for the construction of the density matrix as implemented in the PROGRESS library. The plot shows the wall clock time for computing the DM as a function of the number of orbitals. Scaling experiments were run on an OLCF Summit Power 9 node (using one socket, i.e., 21 cores) comparing performance using different matrix formats. The threshold was set to 10^{-5} for all sparse formats. SP2: second-order spectral projection; DM: density matrix; OLCF: Oak Ridge Leadership Computing Facility.

involve $O(10^4-10^5)$ atoms and exceed 100 ns in simulation time. Obtaining nanosecond-duration simulations for such systems is currently beyond reach of QMD codes. The highly scalable NAMD MD code—a popular choice for biomolecular simulations—recently incorporated breakthrough capabilities in hybrid quantum mechanical/molecular mechanical (QM/MM) simulations (Melo et al., 2018), enabling useful simulations with $O(10^1-10^2)$ of the atoms treated using QM. To extend the size of the QM region to a whole protein with $O(10^3)$ atoms, we are now integrating PROGRESS/BML with NAMD, using the LATTE electronic structure code as the QM solver (Bock et al., 2008). LATTE uses DFT in the tight-binding approximation which is an established approach for *ab initio* studies of biomolecular systems (Cui and Elstner, 2014). It combines $O(N)$ computational complexity with extended Lagrangian Born–Oppenheimer MD and has achieved a rate of 2.1 ps of simulation time per day of wall clock time for a solvated Trp cage miniprotein system consisting of 8349 atoms (Mniszewski et al., 2015). The combination of LATTE with NAMD therefore is an excellent choice for pursuing nanosecond-duration whole-protein QM/MM simulations.

Our efforts in integrating LATTE with the PROGRESS and BML libraries have significantly benefited the EXAALT ECP project. Some materials which are the subject of study in EXAALT such as UO_2 required a very involved modification of the LATTE code to increase performance.

This was made possible by the extensive use of the PROGRESS and BML routines.

4. PIC algorithm development

The longevity of any software framework is dictated, at least partially, by its ability to adapt to emerging algorithms that may not have existed, nor been foreseen, at the time of the framework’s development. In this regard, CoPA has been supporting the development of novel PIC algorithms that can improve and accelerate simulations, with an eye toward their implementation in Cabana. In addition to exercising Cabana, this also represents a unique opportunity to rethink and develop new algorithms at scale, potentially shortening the often lengthy gap between algorithmic innovation and subsequent scientific discovery. These algorithms may connect with the PIC codes in XGC and HACC, described in the next sections.

The algorithms being developed build mainly on two recent advances in PIC methodology. The first is the fully implicit PIC algorithm first introduced in Chen et al. (2011) and subsequently expanded upon in Chen et al. (2020), Chen and Chacon (2015), Stanier et al. (2019), Chen et al. (2012), among others. Compared to standard PIC algorithms, these implicit methods enforce exact discrete conservation laws, which are especially important for long-term accuracy of simulations. Their ability to stably step over unimportant but often stiff physical timescales promises tremendous computational speedups in certain contexts. The second recent advance being leveraged is known as “sparse PIC” (Ricketson and Cerfon, 2016, 2018), in which the sparse grid combination technique (SGCT) (Griebel et al., 1992) is used to reduce particle sampling noise in grid quantities. This is achieved by projecting particle data onto various different component grids, each of which is well resolved in at most one coordinate direction. A clever linear combination of these grid quantities recovers near-optimal resolution, but with reduced noise due to the increased size of the cells—and consequently more particles per cell—in the component grids. Quantitatively, use of the SGCT changes the scaling of grid sampling errors from $O((N_p \Delta x^d)^{-1/2})$ to $O((N_p \Delta x)^{-1/2} |\log \Delta x|^{d-1})$ (Ricketson and Cerfon, 2016). Here, N_p is the total number of particles in the simulation, Δx the spatial cell size, and d the spatial dimension of the problem. For comparable sampling errors, sparse PIC thus reduces the required particle number by a factor of $O(1/(\Delta x |\log \Delta x|^{2d-1}))$.

Algorithm development efforts within CoPA have been three-fold. First, a new asymptotic-preserving time integrator for the particle push—PIC item 4 in Figure 1—component of implicit PIC schemes has been developed. This new scheme allows implicit PIC methods to step over the gyroperiod, which often represents a stiff timescale in

strongly magnetized plasmas (e.g., in magnetic confinement fusion devices). Second, implicit PIC schemes have been generalized to handle a broader class of electromagnetic field solvers—PIC item 2 in Figure 1. In particular, we show that exact energy conservation can be implemented using a spectral field solve, while previous studies have been mostly focused on finite-difference schemes. Spectral solvers have much higher accuracy given the same degree of freedom than finite-difference schemes, which can have particular advantages in simulating electromagnetic waves (e.g., in laser–plasma interaction applications). Third, we show that the implicit and sparse PIC methods can be used in tandem, thereby achieving the stability and conservation properties of the former along with the noise-reduction properties of the latter. Use of sparse PIC primarily impacts particle deposition and force gather operations—PIC items 1 and 3 in Figure 1. Each of these advances is described in turn below.

In magnetized plasmas, charged particles gyrate around magnetic field lines with frequency $\Omega_c = qB/m$, where q is the magnitude of the particle’s charge, B the magnetic field strength, and m the particle mass. In many scientific applications, the timescale of this gyration (i.e. Ω_c^{-1}) can be orders of magnitude smaller than the physical timescales of interest. Consequently, it is often too expensive for standard PIC to simulate those applications. Numerous works have circumvented this difficulty by using gyrokinetic models, in which the gyration scale is analytically removed from the governing equations by asymptotic expansion. However, this approach can become difficult, or breaks down if the approximation is only valid in a portion of the problem domain—scientifically relevant examples include the edge of tokamak reactors, magnetic reconnection (Lau and Finn, 1990), and field reversed configurations (Tuszewski, 1988). A more effective approach is to derive a time integrator that recovers the gyrokinetic limit when $\Omega_c \Delta t \gg 1$ while recovering the exact dynamics in the limit $\Delta t \rightarrow 0$.

Our work derives just such a scheme. We present a summary here; the interested reader should refer to Ricketson and Chacón (2020) for more detail. Note that this work represents an inter-project collaboration with the HBPS SciDAC and may be viewed as part of XGC’s efforts toward full-orbit capability.

Our new algorithm builds on prior efforts (Brackbill and Forslund, 1985; Genoni et al., 2010; Parker and Birdsall, 1991; Vu and Brackbill, 1995) that noticed that standard integrators such as Boris and Crank–Nicolson fail to capture the proper limit when $\Omega_c \Delta t \gg 1$. In particular, the Boris algorithm (Parker and Birdsall, 1991) captures magnetic gradient drift motion but artificially enlarges the radius of gyration, while Crank–Nicolson captures the gyroradius but misses the magnetic drift. Some prior efforts cited above derive schemes that capture both but at the cost of large errors in particle energy. These energy errors are particularly

problematic in the longtime simulations enabled by exascale resources.

Our new scheme is the first to capture magnetic drift motion, the correct gyroradius, and conserve energy exactly for arbitrary values of $\Omega_c \Delta t$. The scheme is built on Crank–Nicolson, but adds an additional fictitious force that produces the magnetic drift for larger time steps, and tends to zero for small time steps (thus preserving the scheme’s convergence to the exact dynamics). Energy conservation is preserved by ensuring that this fictitious force is necessarily orthogonal to particle velocity, thereby guaranteeing that it can do no work (i.e., mimicking the effects of the Lorentz force). The scheme shows promising results in various test problems, and implementation in Cabana is expected to help guide the development of effective preconditioning strategies for the necessary implicit solves.

Our second thrust concerns the field solvers (these are examples of *long-range solvers*—see the section on Cabana above and SWFFT below for additional discussion) in implicit PIC schemes. In the references above and all other extant implicit PIC work, these solvers are assumed to be based on second-order finite difference approximations of the underlying partial differential equations. However, there are significant advantages to the use of spectral solvers when treating electromagnetic waves (Vay et al., 2013, 2018).

With these considerations in mind, we have generalized the implicit PIC method to function with spectral solvers without sacrificing the important conservation properties the scheme enjoys. This is done by adapting the mathematical proofs of energy conservation to accommodate spectral solvers. The key necessary features are two integration-by-parts identities that must be satisfied by the solver. A spectral solver can be made to satisfy these identities *if* a binomial filter is applied in a preprocessing step. Such filters are commonly used in PIC schemes to mitigate particle noise, so this requirement is not considered onerous.

The third prong consists of combining sparse grid PIC schemes with implicit PIC. As above, the key here is retaining energy conservation. Because potential energy is computed on the grid and the SGCT introduces a multitude of distinct grids with different resolutions, care is needed even in the definition of a single potential energy quantity. Having taken this care, we have shown that it is indeed possible to conserve energy exactly in the sparse context. The resulting method also leverages the advances above by being compatible with spectral field solvers.

Initial tests have confirmed the theoretically predicted conservation properties, as depicted in Figure 12, which shows energy conservation up to numerical roundoff for the new implicit scheme applied to the diocotron instability (Davidson, 2001). In addition, we illustrate in Figure 13 the ability of the sparse grid scheme to dramatically reduce particle sampling noise in the solution. Future implementation

of this method in Cabana poses unique challenges as its structure is rather different from a typical PIC method. This is due to the various grids required and the need to perform not only particle-grid and grid-particle interpolations but also grid-grid interpolations for post-processing. As a result, it offers a particularly valuable test case for the flexibility of the software infrastructure.

5. Application partners

To enable a deep window into how particle applications use the computational motifs, the CoPA co-design center

established partnerships with several ECP application development projects. Direct application engagement through deep dives and hackathons has resulted in XGC adoption of Cabana/Kokkos, LAMMPS-SNAP GPU algorithm optimization, and the open-source HACC/SWFFT code. Details of these engagements and their impact on exascale readiness are presented below. Kernels can easily be extracted and explored through proxy apps leading to performance improvements. Co-design Center for Particle Application's inter-dependencies with ECP ST library projects, which provide common software capabilities, has also led to improvements and additions. Details of these engagements are described below.

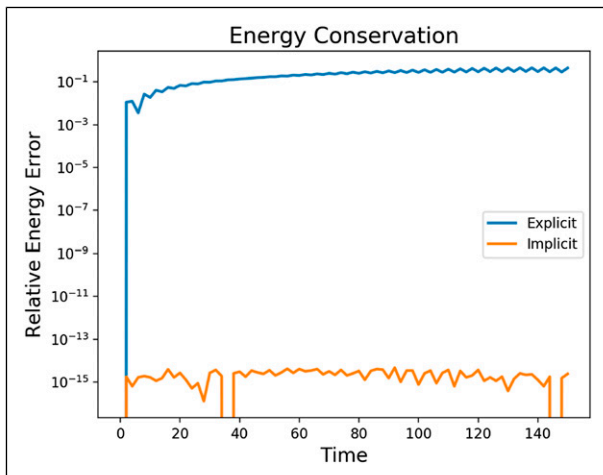


Figure 12. Improved energy conservation properties of implicit sparse PIC (orange, lower curve) compared to traditional standard PIC when applied to the diocotron instability. Note that the implicit scheme conserves energy to machine precision. PIC: particle-in-cell.

5.1. XGC and WDMApp

In this section, we show how the Cabana library has been utilized to enable the fusion WDMApp PIC code XGC to be portable while preserving scalability and performance. In the anatomy of a time step [Figure 1](#), XGC fits into the PIC sub-motif. The particle movement in the particle resorting step has been minimized to avoid MPI communications. Instead, the particle remapping step is heavily utilized in each particle cell independently, which is an embarrassingly parallel operation.

The ECP Whole Device Model Application (WDMApp) project's aim is to develop a high-fidelity model of magnetically confined fusion plasma that can enable better understanding and prediction of ITER and other future fusion devices, validated on present tokamak (and stellarator) experiments. In particular, it aims for a demonstration and assessment of core-edge coupled gyrokinetic physics on sufficiently resolved timescales to study the formation of the pedestal, a physical phenomenon essential to ITER's

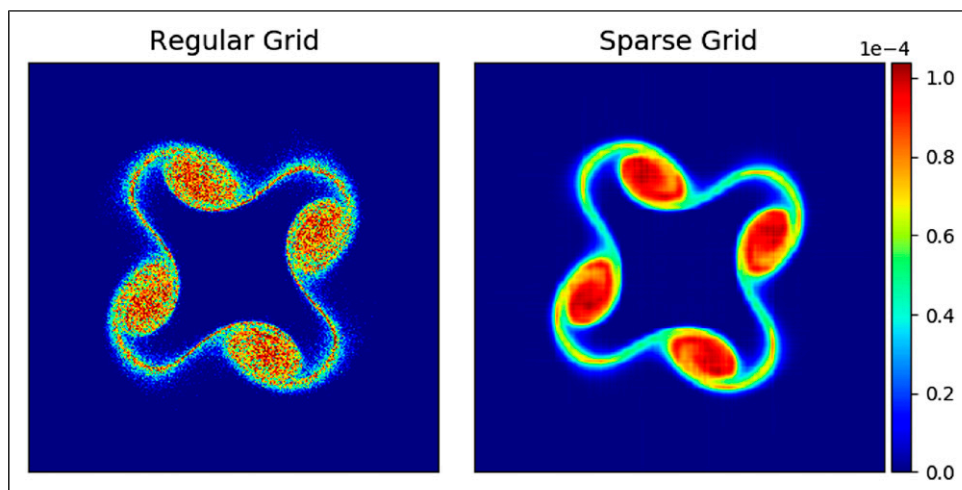


Figure 13. Comparison of electron number density for the diocotron instability computed by two particle-in-cell schemes using the same number of particles (2.6×10^6) and grid resolution (2048^2). On the left, regular grids with explicit time steps are used. On the right, the implicit sparse scheme outlined here is used, with immediately visible reduction in particle sampling noise.

success but whose mechanisms are still not well understood. The WDMapp project involves coupling a less expensive code (GENE continuum code or GEM PIC code, solves for perturbed parts only), which models the tokamak's core, with a more expensive code, XGC (obtains total 5D solution), which is capable of modeling the edge of the device plasma where the computational demands are highest. Performance of the coupled WDMapp code is expected to be dominantly determined by XGC. Performance optimization of XGC is essential to meet the exascale demands.

XGC is a Fortran PIC code used to simulate plasma turbulence in magnetically confined fusion devices (Ku et al., 2018). It is gyrokinetic, a common plasma modeling approach in which velocity is reduced to two dimensions (parallel and perpendicular to the magnetic field), thus reducing total model complexity from 6D to 5D. Markers containing information about the ion and electron particle distribution functions are distributed in this phase space. In a given time step, particle position is used to map charge density onto an unstructured grid. The charge density is solved to determine the global electric field, which in turn is used to update ("push") particle position for the next time step. Particle velocity is also mapped onto an unstructured grid to evaluate the velocity space Coulomb scattering in accordance with the Fokker–Planck operator.

Electron position must be updated with a much smaller time step than ions due to their high relative velocity. They are typically pushed 60 times for every ion step (and field solve), and as a result, the electron push is by far the most expensive kernel in XGC.

In the past, several versions of the electron push kernel were developed that maximized performance on specific hardware. XGC maintained a CUDA Fortran version of the electron push kernel optimized for the previous Oak Ridge supercomputer Titan; an OpenMP version that vectorizes and performs well on CPUs and an unvectorized OpenMP version for use as a cleaner reference.

In addition to the basic time step cycle described above, XGC also has source terms including a Fokker–Planck collision solver on each grid node as briefly mentioned above, which is the second most computationally expensive kernel after the electron push. This kernel offloads work to GPU with OpenACC if available, or uses OpenMP if on CPU. Utilizing multiple offloading programming models in the same simulation poses additional challenges when adapting the code to new platforms and compilers. For example, on Summit only one available compiler PGI supported both OpenACC and CUDA Fortran.

To prepare for exascale architectures, XGC is in the process of significant restructuring. Instead of multiple codebases and offloading programming models, it is being rewritten to use Kokkos and Cabana and to strive toward a

single maintainable, flexible codebase that performs well on all relevant architectures (Scheinberg et al., 2019).

5.1.1. Kokkos/Cabana implementation. Since XGC is written in Fortran, utilizing Kokkos and Cabana posed the additional challenge of Fortran–C++ interfacing. Our initial goal was to use these libraries without significant changes to the main code or to the individual kernels to be off-loaded. We developed an initial such implementation, in which the XGC main code would call a C++ subroutine that wrapped a Kokkos `parallel_for` that launched a kernel that looped over particles and called the necessary Fortran kernel. Kokkos was therefore restricted to a thin interface that managed kernel launching. The Fortran kernel itself had to be modified with preprocessor macros which directed the compiler to compile the code for CPU or GPU as specified; under the hood, CUDA Fortran was still used for GPU offloading.

There were several downsides to this approach. First, it restricted the Kokkos and Cabana features available for use, instead often necessitating custom features for memory management and host-device communication. Second, reliance on Fortran modules often made proper encapsulation difficult. Third, it was unclear if the approach could be easily extended to platforms with AMD or Intel GPUs where no foolproof equivalence to CUDA Fortran would be available. For these reasons, we instead opted to convert XGC into C++, beginning with the major kernels that require offloading, and gradually converting the remaining code. With this new approach, we are better able to utilize the strengths of Kokkos and Cabana by relying on them for memory management, host-device communication, etc.

Due to the piecemeal approach to the code conversion, many data structures on the CPU are still allocated on the Fortran side. At each time step, the particles are rearranged into an array of structures and sent to the GPU as a Cabana AoSoA object. Other data residing in Fortran arrays are wrapped in unmanaged Kokkos Views and can then be copied to Views on the GPU. This method was found to be the least disruptive means of interfacing as the code is gradually converted to C++.

Within kernels that loop over particles, an inner loop is also present, with a range of 1 on GPU and a pre-compiled length (32 by default) on the CPU. These inner loops are mostly vectorized if compiled on CPU, and loop over particles within a single structure of arrays from the AoSoA while the outer loop (the `parallel_for`) loops over structures with OpenMP. The result is a single codebase that vectorizes if compiled for CPU and coalesces if on GPU.

5.1.2. Results. The most expensive operation, the electron push, is now in C++ and offloaded using Kokkos, as well as electron charge deposition and sorting. Since the ion push is independent from the electron push and is still CPU only, it

is performed asynchronously while the electron push is performed on GPU.

A scaling study and comparisons between the different codebases were conducted on both Summit and Cori KNL supercomputers. These tests used simulation parameters and size comparable to those used in scientific production. The new code was found to weak scale well on both machines (Figure 14(a) and (b)). Performance on Cori was found to be similar to the performance of the vectorized Fortran code,

while performance on Summit is also similar to the CUDA Fortran version of the code (Figure 14(c) and (d)). In fact, the Kokkos/Cabana version outperformed previous versions; however, the improvements cannot be entirely attributed to this, since minor algorithmic and structural changes occurred during the porting process.

We conclude that adopting Kokkos and Cabana enabled us to consolidate to a single codebase that is portable to diverse architectures without sacrificing performance.

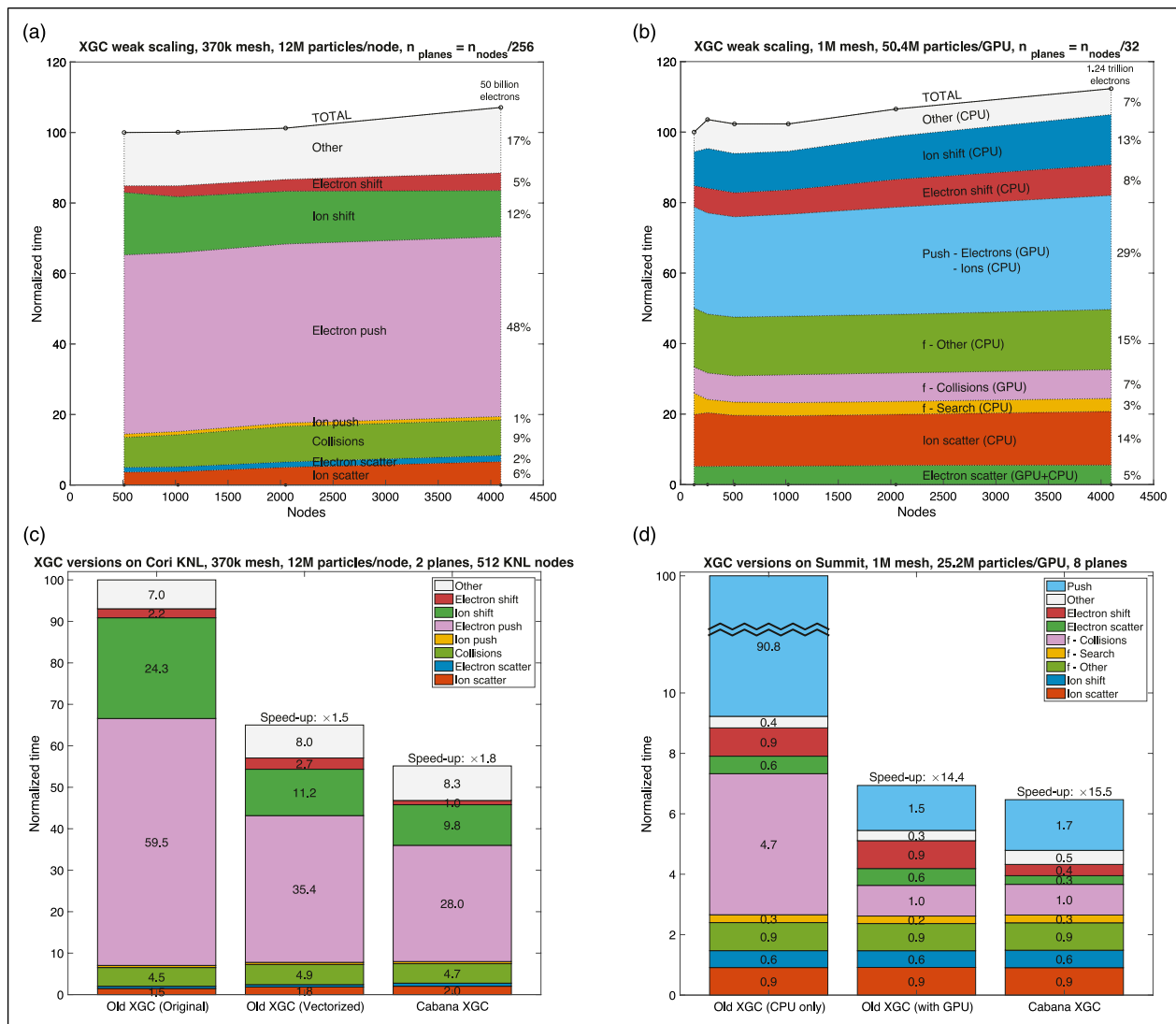


Figure 14. Performance of the whole production XGC code evaluated on the KNL partition of Cori (a) and (c), and on Summit (b) and (d). (a) and (b): Weak scaling studies on both machines demonstrate that supercomputer-scale simulations can be done with the Cabana version without loss of performance. (c) and (d): The Cabana version is compared against previous versions of XGC: an unvectorized OpenMP version and a vectorized OpenMP version on Cori (c), and the OpenMP version and a CUDA Fortran version on Summit (d). We caution here that the colors and legends are different between the KNL partition of Cori (a) and (c) and Summit (b) and (d). In both cases, the Cabana implementation of the expensive electron push kernel performs about as well as the previous, architecture-specific implementations. Ion-push color in (c) is not visible because the wall time of the kernel is negligibly short in the KNL partition of Cori.

5.1.3. Exascale outlook. Conversion of the remaining XGC kernels into C++ is underway. In addition to offloading more XGC kernels, experimentation with more Cabana features (sorting, inter-GPU particle exchange, etc.) will be performed. This may prove useful particularly as more data will be resident on GPU on exascale architectures.

The collision kernel has been converted and offloaded with Kokkos, although performance results are not yet available. With the new collision kernel, OpenACC will no longer be needed and XGC will rely solely on Kokkos for GPU offloading.

5.2. ExaSky

The ExaSky ECP project focuses on extreme-scale cosmological simulations targeted at next-generation sky surveys that observe across multiple wavebands. The simulations follow the development and evolution of cosmic structure in an expanding universe, including not only the effects of gravity but also gas dynamics and a number of astrophysically relevant processes such as radiative cooling, star formation, and various feedback mechanisms, several of which are treated via phenomenological subgrid models.

Cosmological simulations have a vast dynamic range in space, approximately six orders in magnitude, and the corresponding demands on time and density resolution are very severe. ExaSky uses two codes, Hardware/Hybrid Accelerated Cosmology Code (HACC) (Habib et al., 2016) and Nyx (Almgren et al., 2013); HACC uses tracer particles for both dark and ordinary matter (“baryons”), whereas Nyx uses an Eulerian adaptive mesh refinement (AMR)-based method for the gas dynamics. Nyx is strongly coupled to methods being developed by the AMReX ECP co-design center, whereas HACC, because it is essentially a Lagrangian, particle-based code framework, has strong ties to CoPA. In Figure 1, HACC represents a combination of sub-motifs, where PIC methods are used for a Poisson solver to calculate gravitational forces over large distances, and MD-like methods on nearby particles are used to evaluate local contributions to the gravitational force. More details about HACC’s gravitational force splitting are given in the next section.

5.2.1. Hardware/hybrid accelerated cosmology code. Hardware/Hybrid Accelerated Cosmology Code solves the 6D Vlasov–Poisson equation in an expanding universe (Peebles, 1980) and includes gas dynamics via a new SPH scheme, Conservative Reproducing Kernel SPH, an effectively higher-order method that overcomes many of SPH’s known problems while maintaining its advantages (Frontiere et al., 2017). Hardware/Hybrid Accelerated Cosmology Code’s gravity solver splits the gravitational

force computation into two parts, a long-range Poisson solver based on a high-order hybrid spectral method and matched short-range solvers that are designed to be separately optimized for different architectures (direct particle–particle, tree, fast multipole). Hardware/Hybrid Accelerated Cosmology Code’s long-range solver is essentially a PIC method that actively leverages the use of a large, distributed FFT to minimize indirection, reduce particle noise, isotropize the force kernel, and compactly implement higher-order methods for particle deposition and force computation. Time-stepping is performed via an adaptive split-operator, symplectic method that uses subcycling for increased temporal resolution for the dynamics associated with the short-range force. Hardware/Hybrid Accelerated Cosmology Code’s Poisson solver is unusual in that it uses error compensation in the Fourier domain to effectively increase the order of the solver even though the particle-grid interaction is only kept to first nontrivial order (i.e., Cloud-in-Cell deposition and interpolation). Details are given in Habib et al. (2016).

Hardware/Hybrid Accelerated Cosmology Code has its own dedicated, distributed 3D FFT, SWFFT (see below), which has been made publicly available under CoPA. The short-range gravity and hydro solvers comprise the most computationally intensive kernels within HACC and are heavily performance-optimized on a number of architectures. These kernels are highly compact and are excellent candidates to test and exploit the performance portability possibilities using the Cabana framework. As a deliberate result of HACC’s design, 95% of the code does not change as one runs on different platforms (e.g., CPU or CPU + GPU systems), a feature which greatly aids in implementing different performance-portable solutions. Because of the isolation of the computational work into a finite number of compact kernels, a very high level of targeted performance optimization is possible, which would not be the case with the use of generic external libraries. Additionally, the algorithms used are also tied to the architecture as an instance of “software co-design” so the dependencies are not static. Finally, as HACC is often used as a benchmark code on emerging architectures, performant libraries often do not exist on these platforms.

Future work envisaged for HACC is a proxy app based on Cabana and a general long-range solver implemented in Cabana that uses high-order spectral gradients. In addition, as a test of performance portability, we envisage building a short-range gravity kernel in Cabana that can interface with the rest of the HACC code. In this case we can run the full code with a compact, localized modification.

5.2.2. CosmoTools. CosmoTools is the analysis framework associated with HACC. In situ, co-scheduled, and off-line

analyses associated with HACC are complex and computationally demanding in their own right and are as important as running the underlying simulations. Because the analysis methods are diverse, performance portability and especially the ability to use accelerators are both key issues for CosmoTools.

In the ECP context, in situ analysis is of particular importance. Some algorithms in CosmoTools can be built on primitives used by the solver, whereas others, such as neighbor-finding and other clustering-based measurements are unique to CosmoTools; implementation of the latter class of methods often requires the use of efficient graph algorithms. Work is ongoing with the ArborX team (Lebrun-Grandié et al., 2019) to implement new algorithms for clustering analyses (e.g., density-based spatial clustering of applications with noise and N-point correlation functions) on GPUs with promising initial results having been obtained.

5.2.3. SWFFT. Hardware/Hybrid Accelerated Cosmology Code’s performance and scaling requirements involve running very large 3D FFTs (n_g^3 grids, where $n_g \gtrsim 10^4$) distributed across a potentially very large number of MPI ranks ($n_R \gtrsim 10^6$) in order reach the desired dynamic range of the long-range gravitational force via a Poisson solver. A common first approach to a distributed-memory 3D FFT is to divide the grid among ranks along one dimension at a time, creating a 1D “slab” decomposition, but this approach can only work if the number of ranks does not exceed the number of grid vertices along one dimension ($n_R \leq n_g$). In

order to scale to more ranks, HACC’s 3D FFT employs a 2D “pencil” decomposition, where ranks are distributed across one face of the grid at a time, relaxing the constraint on the number of ranks relative to grid size ($n_R \leq n_g^2$). Hardware/Hybrid Accelerated Cosmology Code’s particle operations are sensitive to the ratio between surface area and volume on each rank, so the deposition of particle information onto a grid occurs in a 3D “brick” decomposition. HACC’s 3D FFT requires grid data to be redistributed between the 3D brick decomposition and each of three 2D pencil decompositions ($2D_x, 2D_y, 2D_z$) where the actual 1D FFTs are performed one dimension at a time. HACC’s 3D FFT has implemented this process by going back to the 3D brick decomposition in between all of the pencil decompositions (see Figure 15), so the only communication routines are those that go back and forth between 3D and 2D decompositions. The HACC development team maintains an open-source version of this 3D FFT as the Southwest fast Fourier transform (SWFFT, <https://xgitlab.cels.anl.gov/hacc/SWFFT>). SWFFT is implemented as an out-of-place transform on double-precision complex grid data. The low-level communication is implemented in C, the native high-level FFT interface is implemented as header-only C++, and a Fortran FFT interface is also supported. Currently, there are a few minor differences in the API between SWFFT and HACC’s internal 3D FFT, but the codes are functionally the same.

SWFFT’s implementation and performance characteristics are driven by HACC’s requirements, and the primary goal is excellent weak scaling in memory-limited regimes. An advantage of SWFFT’s communication pattern is that the

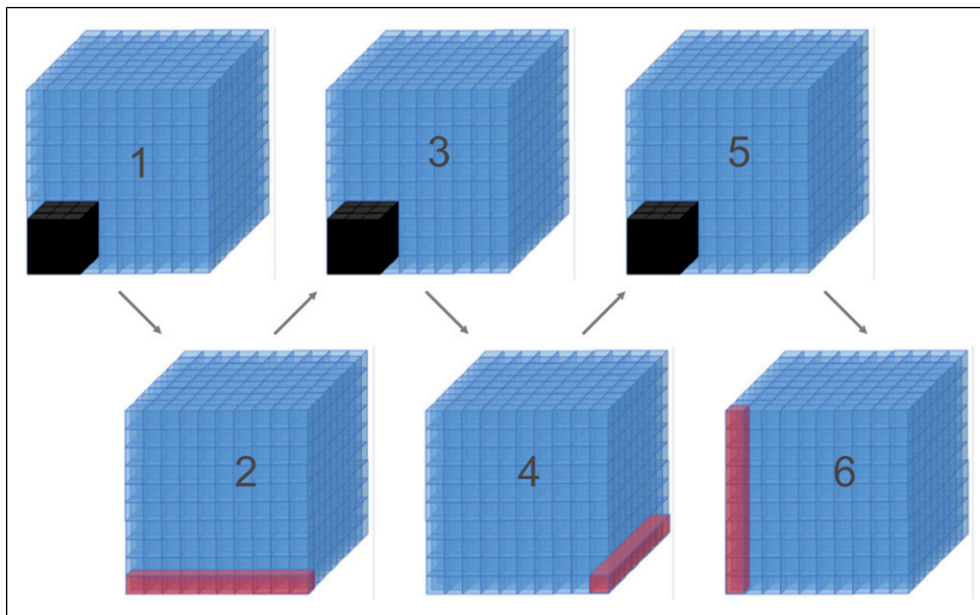


Figure 15. SWFFT decompositions and communication pattern.

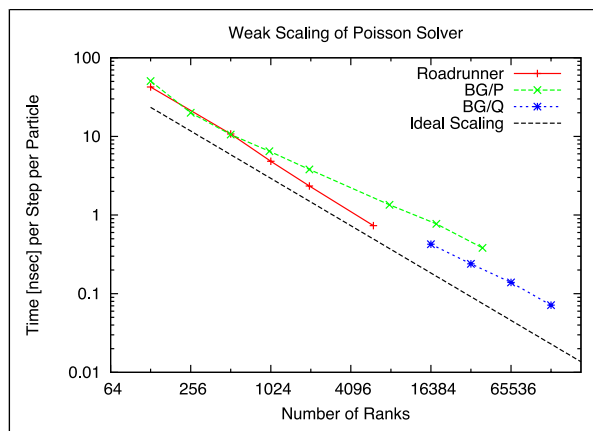


Figure 16. SWFFT weak scaling, reproduced from the SC12 publication [Habib et al. \(2012\)](#).

number of rank pairs that must exchange data scales as the cube root of the total number of ranks ($n_R^{1/3}$). For a communication pattern where data are exchanged directly between pencil decompositions, the number of rank pairs that must exchange data scales as the square root of the total number of ranks ($n_R^{1/2}$). Hardware/Hybrid Accelerated Cosmology Code can maintain a relatively small number of large messages as the number of MPI ranks becomes large, although there are several more communication stages than a direct pencil–pencil communication pattern, so this can emphasize robust weak scaling over absolute minimum latency. [Figure 16](#) shows the scaling of HACC’s Poisson solver, where each Poisson solve involves four 3D FFTs—one forward and three backward for force components using spectral gradients. The largest HACC simulation so far used a $152,30^3$ grid on 1,572,846 MPI ranks on LLNL’s Sequoia IBM Blue Gene/Q system, and each FFT took ~ 10 s to complete. In addition to the source and destination grid memory, SWFFT uses send and receive buffers to reorganize data into messages, but the fractional overhead of those buffer scales as $n_R^{(-1/3)}$ and becomes smaller at larger scales.

The stand-alone SWFFT code was developed to serve as a MiniApp that represented the dominant communication workload in HACC and also as a potential tool for use in solvers in other applications. Through CoPA and ExaSky, an experimental version of Nyx implemented a gravitational solver based on SWFFT. For Nyx, a branch of SWFFT was created with additional flexibility in mapping 3D subvolumes to MPI ranks, and that branch will be re-integrated into the main branch and used to support a new memory-balancing mode in HACC. We are also exploring integrating SWFFT as a backend FFT for solvers written in CoPA’s Cabana framework. SWFFT has already demonstrated scaling up to $\sim 15,000^3$ grids on ~ 1.5 M MPI ranks, and on exascale systems, HACC plans to use $20,000^3 - 30,000^3$

grids. By maintaining the stand-alone open-source SWFFT and participating in ECP, we hope to help other applications and science domains that could benefit from using extremely large FFTs on exascale systems.

5.3. Improving GPU performance of a machine learned potential for MD

The EXAALT project within ECP seeks to extend the accuracy, length, and timescales of material science simulations to model plasma-facing metals used in future fusion reactors like ITER. One method to extend timescales is to run up to millions of small MD simulations (1K to 1M atoms each) and use the parallel replica dynamics (PRD) algorithm as encoded in the ParSplice program ([Perez et al., 2016](#)) to stitch them together into statistically accurate long timescale trajectories. To accurately model defects in metals surfaces bombarded with plasma ions, each replica uses the SNAP machine-learned (ML) interatomic potential ([Thompson et al., 2015](#)), available in the LAMMPS MD code ([Plimpton, 1995](#)) (<https://lammps.sandia.gov>). The ability to run the full-scale model on an exascale machine for long timescales thus depends on the performance of SNAP on one or a few GPUs when simulating a small system (one replica in the PRD ensemble).

A Kokkos version of the SNAP potential was originally implemented in the ExaMiniMD proxy app (<https://github.com/ECP-copa/ExaMiniMD>) and then ported to LAMMPS. At the time ECP began, the fraction-of-peak performance for SNAP for this baseline version was very low on GPUs. To address this concern, a collaboration between EXAALT, CoPA, NERSC/NESAP, Cray, and NVIDIA was formed. A new proxy app version of the SNAP model, called TestSNAP, was created (<https://github.com/FitSNAP/TestSNAP>).

TestSNAP is a serial code derived from the parallel CPU version of SNAP in LAMMPS. It is a good proxy in terms of memory and computational costs. It computes step 3 of the MD sub-motif in [Figure 1](#), which dominates all other parts of the time step for a simulation using SNAP. Importantly, the isolation of the SNAP algorithm in the proxy code made it possible to rapidly experiment with different formulations of the high-level algorithm as well as low-level optimizations such as data structure alterations or loop reordering. The proxy also includes a correctness check which was very helpful to insure changes did not alter the numerical results. The team used the proxy to explore a variety of GPU strategies, first using the OpenACC and CUDA programming models, and then Kokkos. Improvements made in TestSNAP were ported back to the Kokkos version of SNAP in the production LAMMPS code. Further improvements were also implemented directly in LAMMPS ([Gayatri et al., 2020](#)).

The following optimizations improved both CPU and GPU performance of the SNAP potential in LAMMPS:

1. Altered the structure of the SNAP equations to avoid duplicate computations in different terms as well as the order of summations by using an adjoint refactorization. This enabled a dramatic reduction in the flop count, as well as reduced memory footprint and memory access count.
2. Flattened jagged multidimensional arrays which further reduced memory use.
3. Symmetrized data layouts of certain matrices, which reduced memory overhead and use of thread atomics on GPUs.

These optimizations were GPU specific:

4. Broke up one large kernel into multiple kernels. This reduced register pressure, but also greatly increased memory use as intermediate quantities needed to be stored between kernel launches. However, with other optimizations, the net effect was a large reduction in memory use with reduced register pressure.
5. Reversed the order of per-atom and per-neighbor loops.
6. Optimized the memory data layout for the chosen access patterns (e.g., column-major vs row-major).
7. Changed the memory data layout of an array between kernels via transpose operations.
8. Refactored loop indices and data structures to use complex numbers and multidimensional arrays instead of arrays of structs.
9. Refactored some of the kernels to avoid thread atomics and use of global memory.
10. Judiciously used Kokkos hierarchical parallelism and GPU shared memory.
11. Fused a few selected kernels, which helped eliminate intermediate data structures and reduced memory use.
12. Added a new memory data layout inspired by Cabana, which enforced perfect coalescing and load balancing in one of the kernels.
13. Pre-computation of certain parameters.

Figure 17 shows the effect of these optimizations on the SNAP potential performance over time for the EXAALT benchmark problem running on a single NVIDIA V100 GPU on OLCF Summit. For the original Kokkos version of SNAP in LAMMPS, the performance was 5.09 Katom-steps/s per GPU (Katom-steps = 1000s of atom-steps). With the improvements listed above, the new performance is now 110.7 Katom-steps/s per GPU, which is a $\sim 22\times$ speedup.

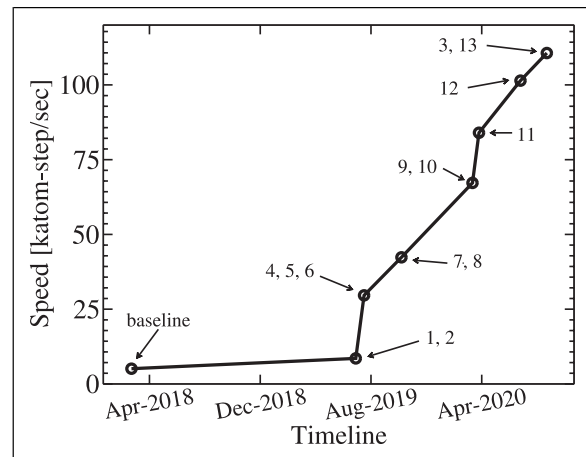


Figure 17. $22\times$ improvement over time of SNAP performance in LAMMPS on an NVIDIA V100 GPU. The numbered data points refer to specific optimizations in the CPU/GPU and GPU lists.

The algorithmic improvements have also been implemented in the CPU (non-Kokkos) version of SNAP in LAMMPS, with the exception of the third item in the CPU/GPU list. Running on 36 MPI ranks of a dual-socket Intel Broadwell CPU, these changes increased the performance of the CPU version of SNAP by a factor of ~ 3 for the same benchmark.

6. Summary

Library efforts, algorithm development, and interactions with particle applications represented within CoPA all contribute to our co-design process and strategy. The anatomy of a time step for particle applications (Figure 1) provides a window into the scope of the CoPA Co-design Center. The computational kernels requiring optimization for exascale computing are associated with the nature of particle interactions. Applications with short-ranged, long-ranged, and particle-grid interactions are addressed within the Cabana library. While applications requiring a quantum mechanical description of interactions are addressed within the PROGRESS/BML libraries. Inclusion of expertise and application partners representing all the sub-motifs has allowed us to understand and create these libraries as well as proxy apps of interest for short-range MD, long-range MD, PIC, and QMD applications. Success is measured by the use of these products within both ECP and non-ECP projects. We close by highlighting some lessons learned, followed by impacts within ECP and the broader community.

Important lessons learned include:

1. Many times over we have discovered the benefits of proxy apps for rapid prototyping of different ideas

and speedup of the performance optimization process.

2. Improving performance on GPUs requires multiple approaches including optimizing data layout, coalescing memory accesses, increasing arithmetic intensity, and using profiling to guide optimizations. Gains can come from both improving the algorithm as well as improving the implementation.
3. Co-design teams of domain scientists, computational scientists, and expert programmers in hardware-specific languages and programming models, working together, proved beneficial to design and optimization efforts.
4. Focused hackathon sessions proved highly productive for small teams over short timeframes, collaborating on algorithms, implementations, and benchmarking.

Impacts as successes with our application partners across all sub-motifs include:

1. WDMApp/XGC is transitioning from Fortran to C++ using Kokkos/Cabana, replacing much of their code with Cabana kernels. The result will be a single flexible codebase with performance portability across relevant architectures.
2. EXAALT/LAMMPS, as part of a co-design team, was able to improve the performance of their SNAP ML model by $\sim 22\times$.
3. Integration of the PROGRESS/BML QMD capability, the LATTE electronic structure code, and the NAMD MD code, has enabled hybrid QM/MM simulations of proteins. This capability will extend the impact of our ECP work to biomedical research including studies of SARS-CoV-2 proteins.

Library efforts have influenced improvements in a number of the ECP ST libraries, such as Kokkos, heFFTe, ArborX, and others. CoPA's library co-design capability allows for integration into existing particle applications, as well as creation of new applications as we continue on the road to exascale.

Acknowledgments

This article describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the article do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was performed as part of the CoPA, supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the US DOE Office of Science and the NNSA. Assigned: Los Alamos Unclassified Report (LA-UR) 20-26599. This work was performed at Argonne National Laboratory under the U.S. Department of Energy contract DE-AC02-06CH11357, Lawrence Livermore National Laboratory under U.S. Government Contract DE-AC52-07NA27344, Oak Ridge National Laboratory under U.S. Government Contract DE-AC05-00OR22725, Princeton Plasma Physics Laboratory under U.S. Department of Energy contract DE-AC02-06CH11357 with Princeton University, Los Alamos National Laboratory, and at Sandia National Laboratories. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of the U.S. Department of Energy (Contract No. 89233218NCA000001). Sandia National Laboratories is a multitechnology 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 Security Administration under contract number DE-NA-0003525. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF), the Argonne Leadership Computing Facility (ALCF), and the National Energy Research Scientific Computing Center (NERSC), supported by DOE under the contract numbers DE-AC05-00OR22725, DE-AC02-06CH11357, and DEAC02-05CH11231, respectively.

ORCID iD

Susan M Mniszewski  <https://orcid.org/0000-0002-0077-0537>

References

- Adedoyin AA, Negre CFA, Bock N, et al. (2019) Performance optimizations of recursive electronic structure solvers targeting multi-core architectures (LA-UR, 20-26665). arXiv: 2102.08505.
- Alexander F, Almgren A, Bell J, et al. (2020) Exascale applications: skin in the game. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378(2166): 20190056. DOI: [10.1098/rsta.2019.0056](https://doi.org/10.1098/rsta.2019.0056).
- Almgren AS, Bell JB, Lijewski MJ, et al. (2013) Nyx: a massively parallel AMR code for computational cosmology. *The Astrophysical Journal* 765(1): 39. DOI: [10.1088/0004-637X/765/1/39](https://doi.org/10.1088/0004-637X/765/1/39).
- Ayala A, Tomov S, Luo X, et al. (2019) Impacts of multi-GPU MPI collective communications on large FFT computation. In: Workshop on Exascale MPI (ExaMPI) at SC19, Denver, CO, 17 November 2019. New York: IEEE. DOI: [10.1109/ExaMPI49596.2019.00007](https://doi.org/10.1109/ExaMPI49596.2019.00007).

- Behler J and Parrinello M (2007) Generalized neural-network representation of high-dimensional potential-energy surfaces. *Physical Review Letters* 98(14): 146401. DOI: [10.1103/PhysRevLett.98.146401](https://doi.org/10.1103/PhysRevLett.98.146401).
- Bock N, Cawkwell MJ, Coe JD, et al. (2008) Latte. Available at: <https://github.com/lan/LATTE>.
- Bock N, Negre CFA, Mniszewski SM, et al. (2018) The basic matrix library (BML) for quantum chemistry. *The Journal of Supercomputing* 74(11): 6201–6219.
- Bowers KJ, Albright BJ, Yin L, et al. (2009) Advances in petascale kinetic plasma simulation with VPIC and Roadrunner. *Journal of Physics: Conference Series* 180: 012055.
- Brackbill JU and Forslund DW (1985) Simulation of low-frequency, electromagnetic phenomena in plasmas. In: Brackbill JU and Cohen BI (eds) *Multiple Time Scales*. Cambridge Mass: Academic Press.
- Chen G and Chacón L (2015) A multi-dimensional, energy- and charge-conserving, nonlinearly implicit, electromagnetic Vlasov-Darwin particle-in-cell algorithm. *Computer Physics Communications* 197: 73–87.
- Chen G, Chacón L and Barnes DC (2011) An energy- and charge-conserving, implicit, electrostatic particle-in-cell algorithm. *Journal of Computational Physics* 230(18): 7018–7036.
- Chen G, Chacón L and Barnes DC (2012) An efficient mixed-precision, hybrid CPU-GPU implementation of a nonlinearly implicit one-dimensional particle-in-cell algorithm. *Journal of Computational Physics* 231(16): 5374–5388.
- Chen G, Chacón L, Yin L, et al. (2020) A semi-implicit, energy- and charge-conserving particle-in-cell algorithm for the relativistic Vlasov-Maxwell equations. *Journal of Computational Physics* 407: 109228.
- Cui Q and Elstner M (2014) Density functional tight binding: values of semi-empirical methods in an *ab initio* era. *Physical Chemistry Chemical Physics* 16(28): 14368–14377. DOI: [10.1039/c4cp00908h](https://doi.org/10.1039/c4cp00908h).
- Davidson RC (2001) *Physics of Nonneutral Plasmas*. London, England: Imperial College Press London.
- Dawson W and Nakajima T (2018) Massively parallel sparse matrix function calculations with NTPoly. *Computer Physics Communications* 225: 154–165. DOI: [10.1016/j.cpc.2017.12.010](https://doi.org/10.1016/j.cpc.2017.12.010).
- Desai S, Reeve ST and Belak JF (2020) Implementing a neural network interatomic model with performance portability for emerging exascale architectures. arXiv:2002.00054.
- Dongarra J, Gates M, Haidar A, et al. (2014) Accelerating numerical dense linear algebra calculations with GPUs. In: Kindratenko V (ed) *Numerical Computations with GPUs*. New York: Springer.
- Edwards HC, Trott CR and Sunderland D (2014) Kokkos: enabling manycore performance portability through polymorphic memory access patterns. *Journal of Parallel and Distributed Computing* 74(12): 3202–3216. DOI: [10.1016/j.jpdc.2014.07.003](https://doi.org/10.1016/j.jpdc.2014.07.003).
- Essmann U, Perera L, Berkowitz ML, et al. (1995) A smooth particle mesh Ewald method. *The Journal of Chemical Physics* 103(19): 8577–8593. DOI: [10.1063/1.470117](https://doi.org/10.1063/1.470117).
- Falgout RD and Yang UM (2002) hypre: a library of high performance preconditioners. In: Sliotta PMA, Hoekstra AG, Tan CJK, et al. (eds) *Computational Science—ICCS 2002, Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer, 632–641. ISBN 978-3-540-47789-1. DOI: [10.1007/3-540-47789-6_66](https://doi.org/10.1007/3-540-47789-6_66).
- Fattebert JL, Osei-Kuffuor D, Draeger EW, et al. (2016) Modeling dilute solutions using first-principles molecular dynamics: computing more than a million atoms with over a million cores. In: SC'16: proceedings of the international conference for high performance computing, networking, storage and analysis, Salt Lake City, UT, 13–18 November 2016, pp. 12–22. New York: IEEE.
- Franchetti F, Spampinato DG, Kulkarni A, et al. (2020) FFT and solver libraries for exascale: FFTx and spectralpack. In: Exascale computing project (ECP) annual meeting poster, Houston, 3–7 February 2020.
- Frontiere N, Raskin CD and Owen JM (2017) CRKSPH—A conservative reproducing kernel smoothed particle hydrodynamics scheme. *Journal of Computational Physics* 332: 160–209. DOI: [10.1016/j.jcp.2016.12.004](https://doi.org/10.1016/j.jcp.2016.12.004).
- Gayatri R, Moore S, Weinberg E, et al. (2020) Rapid exploration of optimization strategies on advanced architectures using TestSNAP and LAMMPS. arXiv e-prints: arXiv:2011.12875.
- Genoni T, Clark R and Welch D (2010) A fast implicit algorithm for highly magnetized charged particle motion. *The Open Plasma Physics Journal* 3(1): 36–41.
- Germann TC, McPherson AL, Belak JF, et al. (2013) Exascale co-design center for Materials in Extreme environments (ExMatEx) annual report—year 2. Lawrence Livermore National Laboratory, Technical Report LLNL-SR-647437.
- Griebel M, Schneider M and Zenger C (1992) A combination technique for the solution of sparse grid problems. In: De Groen P and Beauwens R (eds) *Iterative Methods in Linear Algebra*. North Holland, Amsterdam.
- Habib S, Morozov V, Finkel H, et al. (2012) The universe at extreme scale: multi-petaflop sky simulation on the BG/Q. arXiv e-prints: arXiv:1211.4864.
- Habib S, Pope A, Finkel H, et al. (2016) HACC: simulating sky surveys on state-of-the-art supercomputing architectures. *New Astronomy* 42: 49–65. DOI: [10.1016/j.newast.2015.06.003](https://doi.org/10.1016/j.newast.2015.06.003).
- Hockney RW and Eastwood JW (1989) *Computer Simulation Using Particles*. 1st edition. Bristol England, PA: CRC Press. ISBN 978-0-85274-392-8.
- Hourahine B, Aradi B, Blum V, et al. (2020) DFTB+, a software package for efficient approximate density functional theory based atomistic simulations. *The Journal of Chemical Physics* 152(12): 124101.
- Ku S, Chang CS, Hager R, et al. (2018) A fast low-to-high confinement mode bifurcation dynamics in the boundary-plasma gyrokinetic code XGC1. *Physics of Plasmas* 25: 056107. DOI: [10.1063/1.5020792](https://doi.org/10.1063/1.5020792).
- Lau Y-T and Finn JM (1990) Three-dimensional kinematic reconnection in the presence of field nulls and closed field lines. *The Astrophysical Journal* 350: 672–691.

- Lebrun-Grandié D, Prokopenko A, Turcksin B, et al. (2019) ArborX: a performance portable search library. arXiv: 1908.11807.
- Liu GR and Liu MB (2003) *Smoothed Particle Hydrodynamics: A Meshfree Particle Method*. Singapore: World Scientific.
- Marx D and Hutter J (2009) *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*. Cambridge, England: Cambridge University Press.
- Melo MCR, Bernardi RC, Rudack T, et al. (2018) NAMD goes quantum: an integrative suite for hybrid simulations. *Nature Methods* 15(5): 351–354. DOI: [10.1038/nmeth.4638](https://doi.org/10.1038/nmeth.4638).
- Mniszewski SM, Cawkwell MJ, Wall ME, et al. (2015) Efficient parallel linear scaling construction of the density matrix for Born-Oppenheimer molecular dynamics. *Journal of Chemical Theory and Computation* 11(10): 4644–4654.
- Mniszewski SM, Perriot R, Rubensson EH, et al. (2019) Linear scaling pseudo Fermi-operator expansion for fractional occupation. *Journal of Chemical Theory and Computation* 15(1): 190–200.
- Mohd-Yusof J, Sakharmykh N, Mniszewski SM, et al. (2015) Fast sparse matrix multiplication for QMD using parallel merge. In: GPU Technology Conference, San Jose, CA, 17–20 March 2015: NVIDIA.
- Negre CFA, Mniszewski SM, Cawkwell MJ, et al. (2016) Recursive factorization of the inverse overlap matrix in linear-scaling quantum molecular dynamics simulations. *Journal of Chemical Theory and Computation* 12(7): 3063–3073.
- Niklasson AMN, Mniszewski SM, Negre CFA, et al. (2016) Graph-based linear scaling electronic structure theory. *The Journal of Chemical Physics* 144: 234101.
- Niklasson AMN, Tymczak CJ and Challacombe M (2003) Trace resetting density matrix purification in $O(N)$ self-consistent-field theory. *The Journal of Chemical Physics* 118(19): 8611–8620. DOI: [10.1063/1.1559913](https://doi.org/10.1063/1.1559913).
- Parker SE and Birdsall CK (1991) Numerical error in electron orbits with large $\omega\epsilon\Delta t$. *Journal of Computational Physics* 97(1): 91–102.
- Peebles PJE (1980) *The Large-Scale Structure of the Universe*. Princeton, NJ: Princeton Univ Press.
- Perez D., Cubuk E. D., Waterland A., et al. (2016) Long-time dynamics through parallel trajectory splicing. *Journal of Chemical Theory and Computation* 12(1): 18–28.
- Plimpton S (1995) Fast parallel algorithms for short-range molecular dynamics. *Journal of Computational Physics* 117(1): 1–19. Available at: <http://lammps.sandia.gov>.
- Plimpton S, Kohlmeyer A, Coffman P, et al. (2018) fftMPI, a library for performing 2d and 3d FFTs in parallel, version 00. Available at: <https://www.osti.gov/servlets/purl/1457552>.
- Pope A, Daniel D and Frontiere N (2017) SWFFT: a stand-alone version of HACC's distributed-memory, pencil-decomposed, parallel 3D FFT. Available at: <https://xgitlab.cels.anl.gov/hacc/SWFFT>.
- Ricketson LF and Cerfon AJ (2016) Sparse grid techniques for particle-in-cell schemes. *Plasma Physics and Controlled Fusion* 59(2): 024002.
- Ricketson LF and Cerfon AJ (2018) Sparse grid particle-in-cell scheme for noise reduction in beam simulations. In: Proceeding of the 13th international computational accelerator physics conference, Key West, FL, 20–24 October 2018.
- Ricketson LF and Chacón L (2020) An energy-conserving and asymptotic-preserving charged-particle orbit implicit time integrator for arbitrary electromagnetic fields. *Journal of Computational Physics* 418: 109639
- Saad Y (2003) *Iterative Methods for Sparse Linear Systems*. 2nd edition. Philadelphia: Society for Industrial and Applied Mathematics. DOI: [10.1137/1.9780898718003](https://doi.org/10.1137/1.9780898718003).
- Scheinberg A, Chen G, Ethier S, et al. (2019) Kokkos and Fortran in the exascale computing project plasma physics code XGC. In: Proceedings of SC19 conference, Denver, CO, 17–22 November 2019. New York: IEEE.
- Stanier A, Chacón L and Chen G (2019) A fully implicit, conservative, non-linear, electromagnetic hybrid particle-ion/fluid-electron algorithm. *Journal of Computational Physics* 376: 597–616.
- The CP2K Developers Group (2020) DBCSR: distributed block compressed sparse row matrix library. Available at: <https://github.com/cp2k/dbcsr>.
- Thompson AP, Swiler LP, Trott CR, et al. (2015) Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics* 285: 316–330.
- Tuszewski M (1988) Field reversed configurations. *Nuclear Fusion* 28(11): 2033–2092.
- Vay J-L, Almgren A, Bell J, et al. (2018) Warp-X: a new exascale computing platform for beam-plasma simulations. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 909: 476–479.
- Vay J-L, Haber I and Godfrey BB (2013) A domain decomposition method for pseudo-spectral electromagnetic simulations of plasmas. *Journal of Computational Physics* 243: 260–268.
- Vu HX and Brackbill JU (1995) Accurate numerical solution of charged particle motion in a magnetic field. *Journal of Computational Physics* 116(2): 384–387.

Author biographies

Susan M Mniszewski is a Senior Scientist at Los Alamos National Laboratory and PI for the Co-design Center for Particle Applications (CoPA). She is also a Co-PI for a Quantum Computing LDRD Project. She has contributed to the High Performance Computing PROGRESS/BML libraries for quantum molecular dynamics (QMD). Her research interests include new algorithm approaches for material science applications, machine learning, and novel computing (quantum and neuromorphic).

James Belak is a Senior Scientist at Lawrence Livermore National Laboratory and Co-PI for the CoPA co-design center. His career has centered around the application of

High Performance Computing to equilibrium and non-equilibrium problems in Materials Physics.

Jean-Luc Fattebert is a Staff Scientist at Oak Ridge National Laboratory, working on the BML and PROGRESS libraries. His research interest is on computational algorithms to solve problems in materials sciences, chemistry and biology, from the atomistic scale to the mesoscale.

Christian FA Negre is a Scientist at Los Alamos National Laboratory. He has spent most of his career working as a computational chemist studying optical properties of metallic nanoparticles, absorption spectra of organic molecules, interfacial electron and energy transfer between molecules and semiconductors, and molecular electronics. Dr Negre is now developing techniques to improve the performance of quantum-based molecular dynamics simulations (QMD) focusing on methods to solve problems in the field of applied theoretical chemistry.

Stuart R Slattery is a Computational Scientist at Oak Ridge National Laboratory where he is Team Lead for Scalable Algorithms and Applications. His work focuses on scalable algorithms and performance portable software for applications in advanced manufacturing and nuclear engineering.

Adetokunbo A Adedoyin is a Scientist at Los Alamos National Laboratory specializing in scientific application performance on state-of-the-art and future computer architectures. Prior to LANL, he served as a Computational Physicist at the University of Notre Dame specializing in constitutive modeling of advanced reactive materials at the macro- and meso-scopic scale.

Robert F Bird is a Scientist at Los Alamos National Laboratory, who specializes in the development of performance-portable code and algorithms for next generation compute platforms. His work focuses primarily on particle methods, but also extends to other areas. Within CoPA, he is both a core Cabana developer and a plasma-PIC specialist.

Choongseok Chang is a Managing Principal Physicist at Princeton Plasma Physics Laboratory. He is the head of the SciDAC Partnership Center for High-fidelity Boundary Plasma Simulation and the Co-Lead for Science of the ECP WDMApp project. He is also the leader of the international XGC particle-in-cell code development team. His interest is focused around the extreme-scale HPC study of the non-local, nonlinear, multi-scale plasma turbulence and transport.

Guangye Chen is a Scientist at Los Alamos National Laboratory. His research interests include computational plasma physics, novel algorithm development, scientific high-performance computing, and software development.

Stéphane Ethier is a Principal Computational Scientist at the Princeton Plasma Physics Laboratory (PPPL) and co-head

of the Advanced Computing Group. His work focuses on high performance computing on large-scale systems, particle-in-cell methods for magnetic fusion research, GPU programming, data management, and other related fields. He is a member of the ECP Whole Device Modeling Application project, as well as CoPA.

Shane Fogerty is a Scientist at Los Alamos National Laboratory. His research spans topics related to performance-portable computational methods for scientific simulation software. He is particularly interested in performance opportunities from mixed-precision algorithms for multiphysics simulations on modern computer architectures.

Salman Habib is the Director of the Computational Science Division at Argonne National Laboratory with joint positions at The University of Chicago and Northwestern University. His research interests cover a wide range of problems in physics, ranging from cosmology to quantum information, with a major interest in supercomputing applications and algorithms. Habib leads the ExaSky project within the ECP.

Christoph Junghans is the Deputy Group Leader of the applied computer science group at Los Alamos National Laboratory. His research interests span from scientific software development and engineering over molecular dynamics methods to multi-scale simulation techniques.

Damien Lebrun-Grandié is a Computational Scientist at Oak Ridge National Laboratory. He is co-maintainer of the Kokkos core library which provides performance portability to hundreds of scientific HPC applications, as well as the lead developer of the ArborX geometric search library. Within CoPA, Damien is primarily involved with the development of Cabana.

Jamaludin Mohd-Yusof is a Scientist at Los Alamos National Laboratory. His interests include materials science, machine learning and fluid mechanics, where he develops novel algorithms and applications for High Performance Computing and emerging architectures. Within CoPA he primarily contributes to the BML effort.

Stan G Moore is a Staff Member at Sandia National Laboratories. He specializes in using Kokkos to extend particle-based simulation methods such as molecular dynamics to run efficiently on HPC platforms, and running particle-based simulations at large scale. He is a core software developer of the LAMMPS molecular dynamics code.

Daniel Osei-Kuffuor is a Staff Scientist in the Center for Applied Scientific Computing (CASC) at Lawrence Livermore National Laboratory. His research interests include numerical linear algebra, sparse matrix computations, and scalable numerical solver and algorithm development for HPC applications, including electronic structure calculations.

His work on CoPA supports the BML and PROGRESS libraries.

Steven J Plimpton is a Staff Member at Sandia National Laboratories. He has worked on a variety of particle-based methods and open-source simulation software, mostly for materials modeling. He is a developer for the LAMMPS molecular dynamics package.

Adrian Pope is a Staff Scientist at Argonne National Laboratory. His research focuses on cosmological n-body simulations and statistical inference from astronomical surveys. He is a core developer of the HACC cosmological simulation code, maintains the stand-alone version of HACC's 3D FFT called SWFFT, and works with CoPA on potential technology transfer from HACC to other particle-based codes and solvers.

Samuel Temple Reeve is a Computational Scientist at Oak Ridge National Laboratory, formerly a postdoctoral researcher at Lawrence Livermore National Laboratory, working on the Cabana library and CabanaMD proxy app. His research interests span atomistic and microstructural simulation methods for problems in materials science.

Lee Ricketson is a Staff Scientist at Lawrence Livermore National Laboratory. His research focuses on numerical methods for the kinetic equations governing plasma dynamics. He is particularly interested in the advancement of particle-in-cell methods.

Aaron Scheinberg is a Computational Scientist focusing on exascale computing, scientific application performance, particle-based methods, magnetic fusion simulations, and GPU programming. Formerly at the Princeton Plasma Physics Laboratory, he is now a consultant at Jubilee Development.

Amil Y Sharma is an Associate Research Physicist at the Princeton Plasma Physics Laboratory. He is a developer of the magnetic fusion simulation code XGC, which is part of the ECP WDMApp project.

Michael E Wall is a Scientist at Los Alamos National Laboratory. His main expertise is in data processing and simulations for macromolecular X-ray diffraction studies. His recent focus has been on molecular-dynamics simulations for protein crystallography, parallel processing of diffuse X-ray scattering data, and quantum molecular dynamics simulations of proteins.