

RECEIVED: July 21, 2021

ACCEPTED: September 14, 2021

PUBLISHED: October 11, 2021

ICFA BEAM DYNAMICS NEWSLETTER#82 — ADVANCED ACCELERATOR MODELLING

Modeling of advanced accelerator concepts

J.-L. Vay,^{a,*} A. Huebl,^a R. Lehe,^a N.M. Cook,^b R.J. England,^c U. Niedermayer,^d P. Piot,^{e,f} F. Tsung^g and D. Winklehner^h

^a*Lawrence Berkeley National Laboratory,
Berkeley, CA 94720, U.S.A.*

^b*RadiaSoft LLC, RadiaSoft LLC,
Boulder, CO 80301 U.S.A.*

^c*SLAC National Accelerator Laboratory,
Menlo Park, CA, 94025, U.S.A.*

^d*Technical University Darmstadt,
64289 Darmstadt, Germany*

^e*Northern Illinois University,
DeKalb, IL 60115 U.S.A.*

^f*Argonne National Laboratory,
Lemont, IL 60439 U.S.A.*

^g*University of California Los Angeles,
Los Angeles, CA 90095, U.S.A.*

^h*Massachusetts Institute of Technology,
Cambridge, MA, 02139, U.S.A.*

E-mail: jlvay@lbl.gov

ABSTRACT: Computer modeling is essential to research on Advanced Accelerator Concepts (AAC), as well as to their design and operation. This paper summarizes the current status and future needs of AAC systems and reports on several key aspects of (i) high-performance computing (including performance, portability, scalability, advanced algorithms, scalable I/Os and In-Situ analysis), (ii) the benefits of ecosystems with integrated workflows based on standardized input and output and with integrated frameworks developed as a community, and (iii) sustainability and reliability (including code robustness and usability).

KEYWORDS: Accelerator modelling and simulations (multi-particle dynamics, single-particle dynamics); Simulation methods and programs; Software architectures (event data models, frameworks and databases); Wake-field acceleration (laser-driven, electron-driven)

*Corresponding author.

Contents

1	Introduction	1
2	Modeling advanced accelerator concepts	3
2.1	LWFA and PWFA	3
2.2	SWFA	5
2.3	DLA	6
2.4	Structured plasmas	7
2.5	From ultra-accurate to ultrafast	8
2.5.1	End-to-end Virtual Accelerators (EVA)	9
2.5.2	Optimization & Ultrafast AI/ML surrogate models	9
3	High-Performance Computing	10
3.1	Performance, portability and scalability	10
3.1.1	Performance	10
3.1.2	Portability	10
3.1.3	Parallel scalability	11
3.2	Advanced algorithms	11
3.3	Scalable I/O and in-situ analysis	12
4	Community ecosystem	14
4.1	Integrated workflows	14
4.1.1	Standardization of input & output	15
4.2	Integrated frameworks	16
4.2.1	Community development	16
5	Sustainability & reliability	18
5.1	Code robustness	18
5.2	Usability	19
6	Conclusion	19

1 Introduction

Advanced Accelerator Concepts (AAC) can produce accelerating or focusing gradients that are orders of magnitude higher than the gradients in conventional accelerator technologies. Thus they have the potential to enable much more compact — and in some cases cheaper — accelerator facilities. AAC encompass several technologies, including laser-driven plasma acceleration (a.k.a. LWFA or LPA), beam-driven plasma acceleration (a.k.a. PWFA), structure-based wakefield accelerators (SWFA), dielectric laser acceleration (DLA), laser-ion acceleration and novel plasma

devices (e.g., capillary discharge plasmas and laser-ionized plasma columns). Efficiently modeling these different concepts require different types of simulation tools. DLA, for instance, can be mostly modeled with simulation tools that are already in use for conventional accelerators (e.g. CST [1] or VSim [2]). On the other hand, plasma-based schemes require different simulation tools, and can be considerably more computationally expensive. In this paper, we describe the specific needs and challenges that apply when modeling advanced accelerator concepts. We note that some of the challenges of modeling conventional accelerators, which are described in another paper of this issue [3], also apply to advanced accelerator concepts. This includes for instance software sustainability, standardization, validation, verification, usability, integrated workflows and frameworks, toolkits, ecosystems, and future computing architectures.

Plasma acceleration involves a laser or charged particle beam “driver” that displaces electrons as it propagates through a narrow and long (in comparison to the dimensions of the beam) channel of pre-ionized plasma [4, 5]. The displaced electrons create a pocket of positively charged protons (which move very little thanks to their inertia) that attracts the electrons back on the axis of propagation of the driver, initiating plasma oscillations that form a “wake” of very intense, alternately positive and negative, electric fields that follow the driver beam. These fields can be used to accelerate and guide a “witness” electron or positron beam to very high energy in a much shorter distance than is possible using conventional accelerating techniques. Several aspects of plasma accelerators are particularly challenging to model. These include the detailed kinetic modeling (going beyond the fluid approximation by following plasma particle trajectories) of trapping of background plasma electrons in the wakefields, ion motion, the propagation of ultra-low emittance beams in the plasma, the long time (ps to ns) evolution of the perturbed plasma, and, because of the large number of 3D simulations required, evaluation of tolerances to non-ideal effects. It is also essential to integrate physics models beyond single-stage plasma physics in the current numerical tools, incorporating: plasma hydrodynamics for long term plasma/gas evolution and accurate plasma/gas profiles, collision, ionisation & recombination, coupling with conventional beamlines, production of secondary particles, spin polarization, QED physics at the interaction point.

Structure-based wakefield accelerators (SWFA) exploit the complex frequency dependence of dielectric and meta-material structures to selectively excite electromagnetic modes for the acceleration and control of energetic electron beams. As with plasma accelerators, the careful arrangement of a driver and witness beam can permit the acceleration of electron beams with GV/m-scale gradients [6]. Current research efforts aim to improve the efficiency and stability of these interactions through the manipulation of both the driver and structure properties. For example, the longitudinal shaping of the driver can enhance the resulting transformer ratio, improving the maximum energy gain of the witness beam for a given driver energy [7]. Alternatively, modifications to the structure composition and boundaries may improve the central frequency, bandwidth, and phase-velocity of the excited modes [8–11]. The modeling of these complex boundary conditions imposes additional requirements on core algorithms to capture the electromagnetic response at these boundaries without introducing numerical errors, while also demanding high level tools to describe and initialize representations of a structure within the simulation domain.

The dielectric laser-driven accelerator (DLA) uses a short-pulse infrared (IR) laser, typically in the 0.8 to 2 micron wavelength range, to drive an electromagnetic accelerating mode inside the narrow vacuum channel of a periodic dielectric device [12]. The combination of infrared frequen-

cies, femtosecond pulses, and dielectric materials possessing high damage thresholds enables GV/m accelerating fields [13]. Similar to conventional radio-frequency (RF) accelerators, acceleration in a DLA occurs in the vacuum region of a periodic material structure externally driven by an electromagnetic field. DLA is also in some ways similar to the structure-based wakefield accelerator (SWFA) approach, except that the fields in a DLA are directly excited by a laser rather than through wakefield excitation. Due to the approximately 4 orders of magnitude difference in wavelength from the RF to IR regimes, the corresponding feature sizes, cell-to-cell periodicity, and transverse dimensions in a DLA are reduced to the micron to sub-micron scale. The number of accelerating gaps increases accordingly, making a one-kick-per-period approach [14] very attractive. The similarities between DLA, conventional RF accelerators, and structure-based wakefield accelerators (SWFA), allow similar computational methods and codes to be used. However, since DLA does not rely upon metallic boundaries to confine electromagnetic fields, the boundary conditions and laser coupling methods required are in general different. These differences have been found to benefit from computational techniques more commonly used in the design of photonic devices, such as inverse design or adjoint variable method [15]. In addition, the smaller size of DLA devices and their use of sub-optical particle bunches requires particle dynamics modeling with high temporal and spatial resolution, and limits the accelerator length amenable to particle-in-cell (PIC) simulation to of order a few centimeters. Thus, transport and particle dynamics over longer (meter to kilometer) scales benefits from the development of more rapid particle tracking codes [14].

Lastly, novel plasma devices, including capillary discharge plasmas and laser-ionized plasma columns, show promise for application across a range of essential beamline components, including acceleration stages, focusing, energy compensators, and non-destructive beam diagnostics. These systems could enable fundamental advances in high quality electron and positron beams and support TeV-class high luminosity lepton collider concepts. However, modeling these systems with high fidelity requires integrated multi-physics capabilities at computational scales which exceed those of traditional electromagnetic simulation tools. Three-dimensional magnetohydrodynamic or hybrid kinetic solvers provide paths forward to resolving the difficulties associated with these systems.

In the following sections, we report on various key aspects of modeling of AAC systems, such as high-performance computing, the benefits of ecosystems, sustainability and reliability.

2 Modeling advanced accelerator concepts

2.1 LWFA and PWFA

The most widely used numerical tool to study plasma accelerators is the particle-in-cell (PIC) algorithm [16], where particle beams and plasmas follow a Lagrangian representation (based on particles trajectories) with electrically charged macroparticles while electromagnetic fields follow a Eulerian representation (based on field data on a mesh) on (usually Cartesian) grids. Exchanges between macroparticles and field quantities involve interpolations using B-splines at specified orders. Three-dimensional implementation of the PIC algorithm to model plasma accelerators can be computationally intensive, owing to the disparity of the spatial and time scales in the problem. Laser-driven plasma accelerators are typically (but not always) more computationally intense,

compared to beam-driven, since the laser wavelength λ must be resolved, and, in an underdense plasma, the laser wavelength is much less than the plasma wavelength λ_p , which is itself much less than the length of the plasma channel L_c , $\lambda \ll \lambda_p \ll L_c$.

Ab initio simulations of plasma accelerators with electromagnetic Particle-In-Cell codes are limited by (i) the number of time steps that are needed to resolve the crossing of the driver through the plasma channel that is orders of magnitude longer, (ii) the number of grid cells (and plasma macroparticles) that are needed to cover the wide disparity of spatial scales that span the physics of the driver beam, the wake and the accelerated witness beam. Several methods to reduce the computational cost, while maintaining physical fidelity, have been developed. Reduced dimensionality may be employed in many cases. Indeed, an ideal plasma accelerator typically has cylindrical symmetry and two-dimensional (r - z) modeling is often sufficient to describe the relevant physics. In addition, a truncated series of azimuthal modes may be used to describe nearly cylindrically symmetric cases [17]. Furthermore, fluid approximations (moment models) may be employed in situations where plasma kinetic effects are negligible (i.e., when plasma electrons flow are laminar, without crossing of electron trajectories).

One of the most successful method to speed up simulations is the *quasistatic* approximation [18, 19], which relies on separation of time scales to make approximations. The quasistatic approximation decouples the (long time scale) driver evolution and the (short time scale) plasma wave excitation. This approximation has been particularly successful in modeling PWFA driven by energetic beams, providing significant computational savings over full PIC (typically three to four orders of magnitude for a multi-GeV acceleration stage). It can also be used to efficiently model LWFA, provided that self-injection does not occur (this is typically the case for stages operating in linear or mildly nonlinear regimes) [19]. The massive speedup of the quasistatic approximation comes at the (comparatively small) cost of (i) an algorithm that is specialized and not as widely applicable as the standard PIC method, (ii) approximations that break when plasma electrons accelerate significantly in the wake and get trapped, (iii) more complicated schemes based on pipelines for computing and diagnostics on parallel computers.

For laser-driven plasma accelerators, the wakefield is driven by the ponderomotive force and averaging over the fast laser oscillations in the laser envelope may be employed to implement a ponderomotive guiding center model [19–23], eliminating the need to resolve the fast laser oscillation. Here the plasma responds to the averaged laser ponderomotive force and may be computed on a coarse grid, while the Maxwell equations determining the laser evolution may be solved on a localized fine grid, resulting in computational savings. The speedup offered by using the ponderomotive approximation with a laser envelope description comes at the (comparatively small) cost of a set of equations that is more somewhat complex to solve and parallelize than a standard electromagnetic PIC Maxwell solver.

An alternative to performing a quasistatic approximation to handle the disparity of scales is to shrink the range of scales using the *Lorentz-boosted frame method* [24]. With this method, the ultrahigh relativistic velocity of the driver is taken advantage of by using a frame of reference that is moving close to the speed of light in which the range of space and time scales of the key physics parameters is reduced by orders of magnitude, lowering the number of time steps — and thus speeding up the simulations — by the same amount. The tremendous speedup (typically three to four orders of magnitude for a multi-GeV acceleration stage) comes at the (comparatively small)

cost of additional complexity for the diagnostics that must reconstruct snapshots of the physics quantities (particles and fields) in the laboratory frame.

For LWFA and PWFA, the priorities for the modeling of the plasma physics for a single stage will be to (a) demonstrate that reduced models can be used to make accurate predictions for a single GeV-TeV stage (in the linear, mildly nonlinear and nonlinear regimes), (b) demonstrate that the PIC algorithm can systematically reproduce current experimental results (coordinated efforts with experiments) and (c) conduct tolerance studies and role of misalignments. There exist many codes that have been used to model plasma accelerators using the abovementioned simulation methods [25], with new codes emerging regularly.

2.2 SWFA

In SWFA, a “drive” bunch travels through a solid-state structure and excites electromagnetic wake-fields. The fields can directly accelerate a trailing “main” bunch [i.e. in collinear wakefield acceleration (CWA)] or can be extracted to power a separate accelerating structure [i.e. in two-beam acceleration (TBA)]. The simulation of SWFAs is often staged; the structure is first designed to provide the required electromagnetic properties before integrating the beam dynamics and interactions produced by the SWFA accelerator.

Designing the accelerating structure with the desired electromagnetic modes generally relies on the use of eigenmode solvers to model the structure or one of its unit cells to optimize the field distributions for acceleration while suppressing harmful modes. These tools can be used to efficiently refine the structure geometry independent of the drive beam dynamics. In the case of designing the power extraction and transfer structures (PETS) required in TBA, these solvers are also critical to optimizing the power-extraction coupler. Additionally, time-dependent simulations performed with an FDTD electromagnetic program such as, e.g., MEEP [26] or WARP [27], over a sufficiently large number of time steps, can provide an understanding of the mode excitation evolution. In such an approach, the beam is modeled by its current distribution. As a first step toward performing integrated simulations, the eigenmode analysis can be used to construct a Green’s function. Equivalently, FDTD simulations may be performed with a drive bunch of length much shorter than the mode wavelengths in order to directly extract the Green’s function. The obtained Green’s functions can be imported into available tracking programs to perform start-to-end simulation of the beam acceleration and guide the design of the SWFA linac (e.g. acceleration staging, understanding the trade-off between accelerating gradient and efficiency). Such an investigation can be performed with programs such as e.g., ELEGANT [28] or ASTRA [29]. Very often, they can be limited to the longitudinal beam dynamics using simple 1D-1V models [30, 31] to devise the optimum drive-beam shape and the accelerator architecture. The simulation model can be gradually refined by considering the transverse wakefield (modeled with its Green’s function approach) to any arbitrary order in complexity via the inclusion of successive multipole terms. This is especially important as most of the advanced structures are fully three-dimensional.

Ultimately, full-electromagnetic simulations coupled with PIC are performed, e.g., for a single stage of the acceleration, to verify the performance of the scheme. Over the years, several beam-dynamics tracking programs capable of handling collective effects have been altered to include a set of field “libraries” associated with some SWFA structures [29, 32]. This type of model has been used to simulate and guide experiments [33]. Likewise, a number of PIC codes have been

made available with some capability to model SWFAs (compared to L/PWFA, SWFA required matter-type boundaries with specified macroscopic parameters, e.g., electric permittivity, magnetic permeability, or conductivity including their possible dependence on frequency). Finally electromagnetic simulations can also be performed using wakefield models available within the `ECHO` [34] and `ACE3P` [35] suite of codes. Ultimately, self-consistent simulation coupling FDTD and PIC algorithm provide a complete understanding of the SWFA process, but their computational expense limits their practicality to the final stages of validation. Improvements to PIC models and performance discussed in following sections, including higher-order Maxwell solvers, moving window, and boosted frame techniques, have been applied with great success to plasma simulations, and could be similarly extended to dielectric structures [24, 36].

Precise tailoring of the wakefield will rely on a high degree of control over the electromagnetic properties of the SWFA so that the development of refined models capable of including frequency-dependent properties of the material are ultimately needed. Likewise, the material-science aspects of the SWFA and the fact that its properties can be altered in the strong-field regime are not captured by any of the programs currently available while recent experiments hinted that the material response to the established fields may ultimately limit the accelerating field [37].

2.3 DLA

The dielectric laser acceleration (DLA) scheme is analogous in many respects to a frequency-scaled version of a conventional accelerator. Consequently, modelling many aspects of the particle dynamics and transport in a DLA collider can rely largely upon well established accelerator codes and computational methods. This reduces the need for new code developments or large-scale computing infrastructure to a subset of DLA simulation tasks. For simulation purposes, a linear collider can be broken into three regimes which require different kinds of modeling: (1) Injector (source, emittance preparation, and acceleration up to 1 GeV); (2) Accelerator (1 GeV to 30 TeV) and (3) beam delivery, including final focus, crab kissing, IP design, and beamstrahlung. The proposed injection scheme outlined in Chapter 5, section 5 of [38] is based on superconducting RF technology, and thus is amenable to the same computational methods used to model such systems and will therefore not be addressed here. The main accelerator needs longitudinally coupled structures (traveling wave structures) for energy efficiency and mode stability reasons as described in Chapter 5, subsection 6.2 of [38]. The main accelerator can then be split hierarchically into separately addressed sections corresponding to their respective length scales:

1. One Optical Period (2 μ m)
2. Power coupling cell (10 cm)
3. Focusing cell (20 m at 3 TeV)
4. Betatron period (100 m at 3 TeV)
5. Bunch compression stage (1 km)
6. Full linac (4.2 km each side at 3 TeV)

An individual DLA cell (one optical period in length) can be straightforwardly simulated by any of a variety of FDTD, FDFD, and FEFD codes combined with various optimization techniques [15, 39, 40]. The resulting single-cell Fourier coefficients, combined with the corresponding phase and group velocities, can then be combined with 6D tracking codes to obtain long-distance phase space evolutions by symplectic one-kick-per-cell tracking [14]. The resulting code DLATRACK6D models the nonlinear fields both accurately and fast, but excluding boundary effects. Beam dynamics based DLA structure synthesis can be done by means of the alternating phase focusing (APF) scheme [41, 42]. Advanced codes and large-scale computing become necessary for the determination of the nonlinear particle trajectories through longer structure segments, since the small feature sizes of the structures need to be resolved. Particle-in-cell (PIC) simulations can be used to model full particle and field dynamics up to a few thousand or tens of thousands of structure periods but becomes computationally prohibitive at longer length scales. At optical-scale frequencies, a 10 cm interaction distance would be near the limit of what is possible with a modern supercomputer. A process of using simplified models to construct larger building blocks can be applied to successive levels of the design, using transfer maps to represent larger-scale components such as an entire power coupling cell or focusing cell.

Moreover, wake functions and beam coupling impedances must be included in the various structures, which can be done both in full 3D or by simplified 2D models in the frequency domain. The most crucial issues here are to find the beam loading and beam break up limits using the longitudinal and transverse wakes [43]. For the extremely short sub-optical bunches employed in a DLA scenario (with bunch length small compared to transverse size) the theory of beam instabilities might require extension by nonlinear parts of the transverse wakes and transverse position dependent parts of the longitudinal wakes. Such theories have to be validated by extensive PIC simulations. Moreover, the consequences of slight steering errors in the 10-nanometer range, leading to severe average beam power deposition in the structures needs to be studied. While much of the above simulation work can be accomplished by use of existing codes and/or well-established computational techniques, the beam dynamics in the final focus of a DLA collider needs to be completely redeveloped for DLA-type beams, unless, as assumed in Chapter 5, section 5 of [38], the microbunching is washed out prior to interaction at the IP. Direct collision of trains of extremely short, low intensity bunches with high repetition rate would behave very differently than conventional bunch crossings when it comes to beamstrahlung or crab kissing. Fast beam-beam interaction codes are available, but have yet to be adapted to attosecond bunches.

2.4 Structured plasmas

A structured plasma is a plasma system whose density, temperature, composition, and ionization state are tailored to enhance a desired interaction between a beam or laser with that plasma system. Technologies leveraging structured plasma are increasingly central to advanced accelerator concept schemes. Controlled plasma channels are integral to increasing the peak energy and quality of laser-accelerated electron beams [44–46]. For LPA schemes, generation of a plasma channel with specific radial density profiles enable matching of the drive laser to the plasma, reducing diffraction while maintaining peak on-axis intensities [18]. Longitudinal density control can be used to trigger injection via density downramp [47], and to modify dephasing through longitudinal tapering [48]. Similarly, PWFA schemes have benefited from the generation of meter-scale plasma

channels with near-uniform density [49–51]. The use of pre-ionized channels reduces head erosion caused by defocusing of the drive beam, thereby permitting longer acceleration lengths [52, 53]. Finally, recent demonstrations of positron wakefield acceleration have relied on pre-ionized hollow-channel plasmas [54]. Hollow-channel plasmas improve accelerating gradients and phase-stability compared with uniform plasma channels [55], and may enable the independent control of focusing and accelerating fields for emittance preservation [56]. Finite radius plasma columns have also been proposed as plasma targets to accelerate positrons in an electron driven PWFA [57].

Beyond sources and stages, structured plasmas have found application as flexible focusing elements. Discharge capillaries are capable of producing orders-of-magnitude larger magnetic field gradients than traditional quadrupoles or solenoids [58], subsequently enabling the compact staging of plasma accelerators [59]. Alternative approaches include passive plasma lenses, consisting of a narrow plasma jet outflow generated by laser pre-ionization [60, 61], which can provide comparable focusing with sufficiently high density. Structured plasmas have been employed as tunable dechirpers, capable of removing correlated energy spreads from GeV-scale electron beams [62]. Finally, a promising class of non-destructive diagnostics with high spatiotemporal resolution will rely on controllable plasma densities under interaction with electron and laser beams [63].

Modeling requirements for structured plasma systems vary significantly with the device type, scale, and application. For capillaries, proper modeling requires the characterization of the discharge current and resulting plasma transport properties, including electrical resistivity and thermal conductivity. Magnetohydrodynamic (MHD) codes are well suited to capturing the basic physics of these systems, while maintaining larger timesteps and reduced resolution requirements from kinetic approaches. Significant progress has been made in demonstrating their agreement with 1D analytical models and experimental results for waveguides and active plasma lenses [64, 65]. The coupling of such MHD codes with laser envelope models have contributed to record-breaking LPA acceleration of electrons to 8 GeV in 20 cm capillary [66]. Similarly, active plasma lens studies have benefited from the application of MHD to reproduce species-dependent nonlinearities in current flow and magnetic field [67–69].

Pre-ionized plasma sources, of the kind implemented for PWFA stages, passive plasma lenses, and hollow-channel plasmas, have traditionally leveraged high pressure gas jets or plasma ovens to produce the necessary neutral density profiles. Capturing the gas channel or sheet characteristics may necessitate hydrodynamic simulation on μs -timescales, while the subsequent laser ionization requires the computation of laser propagation and self-consistent field ionization profiles on fs-timescales. Furthermore, the pre-ionized plasma does not constitute a local thermal equilibrium (LTE), therefore the LTE dynamics implemented by most MHD codes is insufficient to capture the ionization and heating dynamics. Moreover, vacuum-plasma interfaces are of interest for matching incident drive beams [70], and multi-species plasmas have been employed for high-brightness injection scheme [71]. In these cases, kinetic codes, for example particle-in-cell techniques, may be coupled with hydrodynamic codes to obtain reasonable approximations, or alternatively, first principles models may be employed [60, 61, 72–74].

2.5 From ultra-accurate to ultrafast

The development of high-gradient particle accelerators demands for predictive modeling tools that range from (i) very detailed, full physics and dimensionality, first-principle kinetic simulation

tools that are needed for detailed runs for physics studies (typically based on the Particle-In-Cell method) [75], to (ii) ultrafast simulation tools that are needed for ensemble runs for design studies (using a combination of, e.g., reduced physics, low dimensionality, low resolution, artificial intelligence (AI)/machine learning (ML) surrogates [76, 77]).

2.5.1 End-to-end Virtual Accelerators (EVA)

The realization of software that are capable of end-to-end virtual accelerator (EVA) modeling (or “virtual twins”) has been identified as a Grand Challenge of Accelerator and Beam Physics [78].

Thorough modeling of the full particle accelerator in individual parts (injector, magnets, beam dynamics, etc.) is already an important aspect of particle accelerator development; one, without which no project nowadays can proceed to the building stage. However, simulating the system in many small units poses two significant challenges: (i) The various codes used for the individual parts are often from different eras, written in different languages (C/C++, Fortran, Python, etc.), and use different standards for data I/O (particles, fields, etc.), slowing down the design and simulation process; (ii) Without multiphysics couplings, subtle, but important effects (collective effects, halo, coherent synchrotron radiation, etc.) might not be considered in the design, ultimately leading to issues during commissioning and the need to modify the machine (shimming, additional steering, shielding, etc.) and longer downtimes due to added radiation. Furthermore, opportunities for more efficient working points that can be achieved only from system optimization of the entire accelerator may be missed entirely.

Thus, there is a clear need for full physics 6-D computer simulations of the entire accelerator system that incorporate all components (including both conventional and AAC sections), all pertinent physical effects, and that execute fast and reliably. Furthermore, these EVAs should be able to leverage modern computing infrastructure like HPC clusters and GPU computing, and fully integrate AI/ML tools to maximize efficiency for practical applications. The development of such tools requires continuous advances in fundamental beam theory and applied mathematics, improvements in mathematical formulations and algorithms, and their optimized implementation on the latest computer architectures.

2.5.2 Optimization & Ultrafast AI/ML surrogate models

Surrogate models can be built by training a neural network on a smaller (compared to other optimization techniques) set of high fidelity simulations (often particle-in-cell) that coarsely maps out the hyperspace of possible input parameters. Non-simulated sets of input parameters can then be approximated by the surrogate model. This can be highly useful for optimization and online feedback about the accelerator. Some examples of successful use of surrogate models in particle accelerator optimization are [79–82]. A speedup of one to several orders of magnitude compared to conventional techniques has been observed in these cases.

Other examples, specifically using Bayesian Optimization using Gaussian Process models, are the tuning of SwissFEL [83, 84] and the Linac Coherent Light Source (LCLS) [85].

These surrogate models will have strong impact on accelerator modelling and optimization through the ultrafast execution of the trained models. They can be trained for full start-to-end virtual accelerators or individual components. The ultrafast execution enables quick turnarounds

for design optimization and as an on-line feedback tool during commissioning, tuning, and operation of an accelerator system.

Important steps forward to make surrogate models widely usable have been identified in ref. [76]. The main paths forward are *finding the best practices* for various different accelerator systems and modeling needs thereof, and *developing a robust and flexible framework* for surrogate modeling.

3 High-Performance Computing

High-Performance Computing (HPC) continues to be a tightly connected domain for advanced accelerator modeling research. One of the main reasons for this is the large computational model size for full-fidelity PIC simulations, especially when describing kinetic effects in relativistic laser-matter interaction.

Over the last decade, HPC machines established so-called hardware accelerator components, most notably general purpose programmable graphics processing units (GPUs), in a majority of leadership-class machines. Such novel compute hardware provides an order of magnitude higher computational density compared to traditional computing hardware under reduced energy costs. GPUs are highly parallel computing devices and need algorithms that are 10- to 100-thousandfold parallelized compared the conventional parallelism in CPUs (10- to 100-fold parallel). When redesigning algorithms for modern compute architectures, PIC codes for accelerator modeling can usually achieve an order-of-magnitude speedup for time-to-solution [86, 87].

3.1 Performance, portability and scalability

3.1.1 Performance

Performance increase for modeling software is generally desirable, even outside of large-scale HPC runs, and improves overall scientific productivity when studying particle accelerators at all scales. For instance, most optimization and machine-learning workflows are GPU-accelerated as well to increase computational throughput for studies, while ultraprecise numerical schemes can often use GPU-accelerated linear algebra and FFT routines; thus GPU acceleration also benefits small model sizes.

Developing software for multiple GPU-platforms in parallel to existing CPU architectures is an undertaking that requires the redesign of many existing algorithms and code-bases. Programming models and languages evolve rapidly and the need to port algorithms to significantly more fine-grained parallelism is a good incentive to rewrite and adopt modern software practices and popular programming languages.

3.1.2 Portability

Portability of implementations is important as many scientists work with three major platforms (Linux, macOS, Windows), although the former two (Unix-like) variants are certainly dominating with HPC developers and machines. An additional challenge arises with the diversity in GPU and other compute hardware, which ideally should be addressed with *performance-portable* software design. In such a design, an algorithm is ideally implemented only in a single, abstract form and then specialized/optimized for the specific hardware needs at compilation-time.

Performance portability relies on using and contributing to standard, industry-supported programming languages to allow such a single-source approach. Performance portability layers (e.g., Kokkos [88], Alpaka [89], Raja [90]) are currently implemented predominantly as modern C++ libraries with other languages lacking significantly in release time, compiler variety and robustness. Common numerical and data management libraries (e.g., AMReX [91], CoPA [92]) build on such performance portability layers and only as the next steps, domain-specific libraries and applications are implemented to benefit from all aforementioned developments.

3.1.3 Parallel scalability

Parallel scalability on leadership-scale machines continues to be only achievable with additional parallelization over multiple coupled compute nodes (each powered with CPUs or GPUs). Methods that (a) split the computational domain into local sub-problems, and (b) ideally overlap communications (usually of data in guard cell regions) between collaborating nodes with computation within each node, are essential for such capability simulation runs. Furthermore, at large-scale, load balancing of particle-based methods as well as refinement of fidelity needs (e.g., adaptive mesh-refinement) of the simulation domain become important to make optimal use of available resources.

Looking at the rapidly evolving HPC landscape, many themes in advancing accelerator and beam modeling software are common, e.g. Poisson solvers, robust particle pushers, need to model hybrid particle accelerators combining conventional and advanced accelerator elements, etc. It is thus evident that collaborative efforts between conventional and advanced accelerator development can benefit the community, relying for instance on similar or compatible low-level parallelization layers, but also coordination when implementing and sharing domain-specific numerics can ensure that scalable solution can be readily extended for research needs.

3.2 Advanced algorithms

In addition to developing codes that use the latest architectures as efficiently as possible, it is essential to constantly reassess the capabilities and limitations of existing algorithms, and develop new ones that improve stability, accuracy and/or performance. The extreme computational intensity required for full PIC modeling of LWFA has fostered many innovations in Particle-In-Cell methodology. In particular, the impetus to complete these simulations in a reasonable time using a Lorentz boosted frame has led to active research to mitigate some numerical instabilities that are otherwise particularly violent.

In a Lorentz boosted frame PIC simulation, the plasma column, which is at rest in the laboratory, is moving near the speed of light in the frame of the simulation. This results in coupling between the fields on the (static) computational grid and the plasma modes and its aliases that lead to resonances and the so-called *numerical Cherenkov Instability* [93, 94]. Research that was spurred by the drive to perform Lorentz boosted frame LWFA simulations has led to tremendous progress in the understanding of the instability and its mitigation [95–111]. One of the most notable of these advances is the *Galilean Boosted Frame approach*, which consists in using the ability to integrate Maxwell’s equation analytically in time over one time step of the *Pseudo-Spectral Analytical Time-Domain* (PSATD) Maxwell solver [112, 113] to synchronize all the fields and plasma modes with a Galilean frame that follows the plasma in the boosted frame, hereby removing the resonances between all the corresponding modes and leading to stability [107, 108].

This research has also spurred advances in the PSATD solver itself, enabling its efficient usage in parallel using domain decomposition (where one FFT is performed per MPI subdomain) with arbitrary order of accuracy (and locality) of the spatial derivatives [36, 109, 113–115]. This research is ongoing and very active. Most recently, a novel algorithm was proposed that takes again advantage of the PSATD ability to integrate analytical the field over a time step to average the field that is used to push particles over one time step [116]. This overcomes a longstanding barrier of boosted frame simulation that were limited to small time steps by the Courant-Friedrich-Lowy (CFL) condition, when the transverse cell size of the grid Δx is smaller than the longitudinal cell size Δz , $c\Delta t \leq \Delta x < \Delta z$, where c is the speed of light in vacuum. With the novel algorithm, the simulation is still stable even when $c\Delta t > \Delta x < \Delta z$. In another recent advance, a novel *hybrid* PSATD PIC scheme was proposed that combines the advantages of stability of nodal PIC methods and efficiency of staggered PIC methods, demonstrating speedups of boosted frame full PIC LWFA and PWFA simulations by factors of two [117].

In addition to reducing the number of time steps, techniques to reduce the number grid cells that are needed are being developed. Mesh refinement, a technique that is well established in the modeling of fluids, has been used with success in the modeling of conventional accelerators with electrostatic PIC solvers [118], and has been explored more recently for the modeling of plasma accelerators with electromagnetic PIC solvers [119–122]. Reduction of dimensionality using 2D planar or axisymmetric geometries is a very common way to reduce the computational workload. Fourier decomposition in azimuthal modes, a technique that has been borrowed from other areas (including conventional accelerators), has been introduced in this community in 2009 [17] and been adopted by more codes since, using full PIC with boosted frame [123] or quasistatic PIC [124]. The rapid adoption by the community of the code FBPIC [125] demonstrates the effectiveness of combining the full PIC approach with reduced dimensionality using azimuthal Fourier decomposition in a software that is easy to install (as a Python module) and yet efficient on both CPUs and GPUs.

It is important to note that the constant development of novel algorithms, which has been fostered by the extraordinary numerical and computational challenges associated with the modeling of high-gradient accelerator components, has found application to other related fields. For example, the development of distributed parallel PSATD PIC [36, 113, 114, 126] has spurred research into the numerical properties of Maxwell solvers when modeling plasma mirrors with electromagnetic PIC codes. These types of simulations were shown to benefit greatly from ultrahigh-order pseudospectral solvers of order 50 or more [127], leading to simulations of plasma mirror experiments and applications [128–131] that are otherwise out of reach to standard PIC methods with finite-difference Maxwell solvers. Another example is the development of the novel hybrid PSATD scheme that was mentioned above, which has found application in the modeling of electron-positron pair creation in PIC simulations of high-field physics [117], with speedups by an order of magnitude or more over conventional methods, exemplifying the cross-fertilization of the progress in new algorithms across applications.

3.3 Scalable I/O and in-situ analysis

While modern programming models, software design and advanced algorithms ensure that simulations can execute efficiently, one further challenge emerges to scientific productivity on HPC hardware: when comparing the computational peak performance evolution to the parallel system

storage bandwidth of HPC machines as in table 1, the gap between potential simulation fidelity that can be computed and data that can be stored for analysis widens with every new machine generation.

Table 1. Evolution of leadership-scale machines at the Oak Ridge Leadership Computing Facility with respect to peak performance and storage bandwidth, according to <https://web.archive.org/web/20210715000401/> <https://www.olcf.ornl.gov/frontier/>.

System Specs	Titan	Summit	Frontier
Peak Performance	27 PFLOP	200 PFLOP	>1.5 EFLOP
Storage Bandwidth	1 TB/s	2.5 TB/s	2-4x Summit
Ratio	27:1	80:1	150-300:1

Following this trend, modeling applications that run at a system’s capability and follow the traditional workflow of simulation + post-processing will spend more of their time in large-scale data input and output, which eventually can dominate the time-to-solution. Driven by this trend, several steps can be taken by simulation developers.

As a first step, adopting scalable, parallel I/O libraries with adequate data layouts [132] can extent the applicability to truly make optimal use of parallel filesystems [133]. Implementing these libraries in both data producing as well as data consuming codes opens another possibility to standardize on and share the community solutions, as demonstrated in the openPMD project [134].

As a second step and long-term solution, data analysis and processing need to be transitioned to in situ-processing, running online with the modeling simulation and avoiding traditional, file-based long-term storage for raw data. This approach essentially moves the design of data analysis to the setup and design phase of a simulation that, similar to an experimental setup, needs to plan the locations, acceptance and dynamic ranges of “virtual diagnostics” before even starting a high-fidelity modeling run. Virtual diagnostics can reduce data sizes and thus data throughput needs by orders of magnitudes, e.g., by calculating beam momenta, phase space histograms, spectra and even 3D videos as a simulation runs instead of storing realistically trillions of particles in regular intervals to disk for later analysis. While in-situ data processing and analysis are not new, and have even been mainstream in some accelerator codes or frameworks (e.g., Warp [27], PIConGPU [86]), the discrepancy between peak performance and storage bandwidth reported in table 1 triggers the development of community libraries that enable a more systematic description of new diagnostics in an in-situ approach, ensuring high efficiency and scalability.

A challenge arises in the rapid setup of in situ processing pipelines, since typical approaches of implementing a virtual detector in the simulation code itself can be more complicated compared to scripted (serial) analysis workflows on files. Innovative software design can potentially alleviate this challenge, with approaches such as streaming data with close-to-file-like analysis syntax [135] or embedding of interpreters into running simulations to execute scripts on in-memory data. Continued research in this area and close collaboration with computer science and engineering teams will be essential to ensure that flexible, science-case specific analysis can be designed by accelerator and beam physicists.

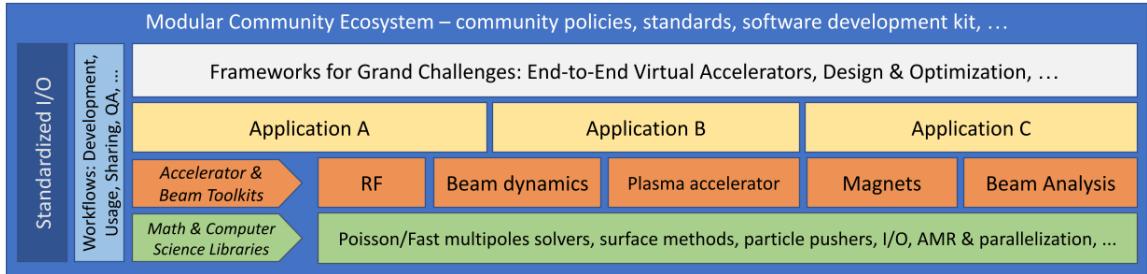


Figure 1. Diagram of a possible *Ecosystem* for particle beam and accelerator modeling.

4 Community ecosystem

Simulations are critical to the design, development and operation of all accelerator facilities, some costing in the billions of U.S. Dollars. The simulation of all the components and physical processes that are involved is a very wide-ranging and complex endeavor. While a lot of efforts have been aimed at the development and maintenance of very useful and successful simulation tools, they were largely uncoordinated. This resulted in a rich collection of (often relatively small, with some large and sophisticated) codes that are not interoperable, use different I/O formats and quite often duplicate some physics functionalities using the exact same underlying algorithms (e.g., many codes have standard Particle-in-Cell cores that are distinct but algorithmically identical), impoverishing the effective diversity of the whole.

Accelerator and beam physics software, like many other types of software, are subject to a ‘natural cycle’, driven by, e.g., developers retiring or moving to other projects, evolution of programming languages or computer hardware, leading to regular retirement or loss of codes that may need to be rewritten from scratch. The rewriting can be particularly costly when the previous code is poorly documented (if at all) and the original developer(s) is/are not available.

A concrete community ecosystem for accelerator and beam modelling could be structured as in figure 1: individual components are represented as boxes with dependencies building on top of each other. The fundamental building blocks are implementations of solvers/numerical schemes, efficient I/O, parallelization and a performance portability layer (*Math & Computer Science Libraries*). Depending on this functionality, domain-specific libraries for, e.g., RF, beam or plasma modeling can be implemented (*Accelerator & Beam Toolkits*) that then act as toolkits to be combined into concrete *Applications*. Compatibility as well as development productivity, usage, and quality assurance of an ecosystem would be coordinated with standardization (e.g., for I/O and data layouts) as well as common workflows, e.g., shared continuous benchmarks and practices. Overarching frameworks then implement concrete workflows that combine applications or toolkit components in an innovative way for the study of complex, integrated research questions.

4.1 Integrated workflows

Integrated workflows are often needed to answer complex contemporary research questions, such as the start-to-end description of a hybrid accelerator with conventional and plasma elements. Furthermore, all optimization, AI/ML training as well as exploratory studies benefit from establishing and maintenance of workflows. These workflows might (a) link several codes together to solve a

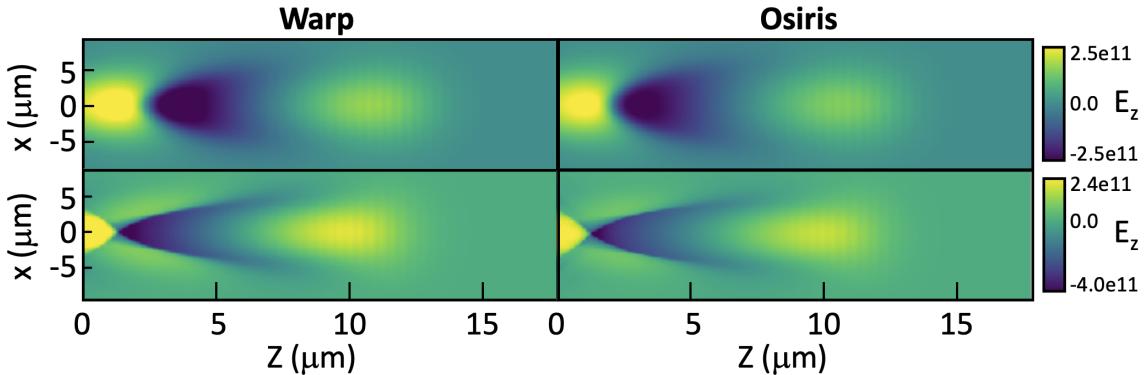


Figure 2. Longitudinal electric field (in V/m) in a laser-driven plasma acceleration stage at two times (top: $t \approx 300$ fs, bottom: $t \approx 600$ fs) along the laser propagation from 2-D PIC simulations with: (left) Warp; (right) Osiris. Plots are based on rendering from the openPMD-viewer.

larger problem or solve a problem with more physical processes, (b) benchmark and validate the various codes against each other and known solutions, and (c) establish reproducibility for new scientific results.

4.1.1 Standardization of input & output

At the moment, many simulation codes are developed independently with little coordination between various groups. In order to establish workflows that span multiple applications, a typical approach so far is to rely on file-based data exchange and code-specific high-level wrapping of input options. Along these lines, the Consortium for Advanced Modeling of Particle Accelerators (CAMPA [136]) supports (among other activities) the development of standardized input and output formats through the openPMD [134] and PICMI [137] projects.

The Open Standard for Particle-Mesh Data (openPMD) is a meta-data standard for research data, adding scientific self-description on top of modern, portable, scalable file-formats such as ADIOS [138] and HDF5 [139]. openPMD achieves this by collaboratively defining an open standard document that evolves in versions, similar to software. Community members then implement openPMD data handling in their simulation codes and analysis types either directly against supported file formats or use an open source library layer, openPMD-api [140].

The standard input format for Particle-In-Cell codes (PICMI) strives to unify the usage of accelerator modeling codes by standardization on an API for simulation design. Conceptually, PICMI defines a common, explicit PIC modeling interface (API) for a wide range of numerical and initialization options. A PICMI script can be run with multiple codes, given that both implement the requested features, and simplifies comparability of physics cases tremendously. Code “backend” choices can be picked based on performance or numerical needs, without breaking context for the user and relearning the simulation design for each specific application.

As part of the CAMPA efforts to foster compatibility and coordination between particle accelerator modeling applications, a code-validation workflow between the Warp [27] and Osiris [141] code has been carried out. In the study on the two codes in figure 2, both independent code bases modeled the same physical setup of an LWFA (with a laser driving a wake in a plasma column)

and used standardized openPMD output and the same, community-developed post-processing tool (openPMD-viewer) to analyze the data. A one-to-one comparison to machine precision between the two codes was beyond the scope of the study because: (a) while using the same physical parameters to initialize the simulations with each code, some secondary parameters were left unspecified such as, e.g., the exact phase of the laser oscillations, (b) some numerical aspects are different such as, e.g., the method of initialization of the laser in the boosted frame. Nonetheless, it is rewarding to observe on figure 2 that the two codes predict very similar wake evolutions and dephasing of the laser as it propagates through the plasma column.

Another series of comparisons is underway that also incorporates a common Python input script using PICMI, and including more codes, in the workflow, to validate the workflow concept for further adoption and development.

4.2 Integrated frameworks

Integrated developments of frameworks could implement more tightly coupled workflows on a library level instead of relying on high-level application coupling. As presented at the beginning of the section in the concept of an modular software ecosystem in figure 1, modular, cross-cutting software designs can provide reusability, efficient development investments, and compatibility. Built-in features of lower-level software such as GPU-support, multi-node parallelism, translation to AI/ML frameworks and multi-physics support can, if evolved together, be inherited by a wide community. In particular, common or compatible in-memory data structures have not been attempted in accelerator and beam modeling so far, but would provide significant performance and software maintenance benefits for tightly-coupled, co-developed physics modules for integrated modeling frameworks that address grand challenges.

Several community examples exist to demonstrate modular development. For instance, the codes WarpX [142] and PICConGPU [86] as well as several analysis frameworks share the development of the I/O library openPMD-api [140], which provides access to high-performant, portable low-level I/O libraries. WarpX, FBPIC implement PICMI assisted by a shared, central interface library, with implementation underway in other codes (e.g., OSIRIS). WarpX, HiPACE++ [143] and related codes share the data structures from AMReX [91] and parts of their code bases.

4.2.1 Community development

This transition to a more coordinated and collaborative approach is also driven by the need to transition a large body of software from CPUs to GPUs, a transition that is more disruptive than most past transitions, including the transition from serial to parallel codes (and it is anticipated to be one of many transitions needed to adopt to the rapidly evolving computer hardware). Three dimensional simulations of advanced accelerator concepts have been around since the early 2000's (for wakefield accelerators). Ever since then, more realisms (e.g., ionization, QED effects), better numerics, and advanced algorithms have been added to almost all code bases, and code developers have developed best practices for adding new modules while maintaining the performance needed to run on state-of-the-art HPC centers around the world. Thus, instead of porting solely to GPUs, accelerator modeling teams already teamed up with computer scientists and embrace libraries that abstract computer hardware specific details for performance and sustainability. Establishing more

community standards, funding reliable software dependencies and reusable physics modules are the logical steps to speed up development of code bases further.

An efficient mean for guiding community development is the collaborative setting of and adherence to common policies. Rather than starting from scratch, one can adopt an existing libraries and codes and build a beam and accelerator modeling ecosystem on it. As a blueprint, one could build on community development concepts of the Interoperable Design of Extreme-scale Application Software (IDEAS) project [144] and its Extreme-scale Scientific Software Development Kit (xSDK) [145].

The main goal of the IDEAS project is to “help move scientific software development toward an approach of building new applications as a composition of reusable, robust, and scalable software components and libraries, using the best available practices and tools.” [144], while the xSDK project is being developed to “provide a coordinated infrastructure for independent mathematical libraries to support the productive and efficient development of high-quality applications” [145]:

Rapid, efficient production of high-quality, sustainable extreme-scale scientific applications is best accomplished using a rich ecosystem of state-of-the art reusable libraries, tools, lightweight frameworks, and defined software methodologies, developed by a community of scientists who are striving to identify, adapt, and adopt best practices in software engineering. The vision of the xSDK is to provide infrastructure for and interoperability of a collection of related and complementary software elements — developed by diverse, independent teams throughout the high-performance computing (HPC) community — that provide the building blocks, tools, models, processes, and related artifacts for rapid and efficient development of high-quality applications.

Although xSDK targets “extreme-scale scientific applications” of relevance to exascale supercomputing, its derived community policies apply well to all scales of computing, from laptops to clusters or Cloud computing. Hence, the paradigm and tools are readily applicable to the full set of modeling needs of the particle accelerator community.

Individual packages in such an ecosystem are composable libraries, frameworks, and domain components developed by individual groups in the community. Each package publishes technical design documents, documentation (reference, tutorials, how-to guides), API definitions, and a concrete implementation [146]. In order to solve a specific physics case, a typical application uses functionalities from several packages, often in an innovative way.

There are many advantages to adopting policies and tools such as the IDEAS/xSDK:

- Community policies have been established over years by teams of specialists in scientific software development and computing, and they include best practices in software development.
- For reusable components, the policies require the use of *permissive open source licenses* (“Non-critical optional dependencies can use any OSI-approved license.”) that allow reuse of source and binary code, including for commercial and proprietary applications, fostering collaborations across laboratories, academia and industrial partners.
- A wide-ranging set of open source, interoperable tools from a variety of backgrounds (including numerical solvers, e.g., Hypre, multi-parameter optimizer, e.g. LibEnsemble, and mesh-refinement frameworks, e.g. AMReX) can be combined and used as foundation of accelerator and beam physics toolkits and codes.

- A wide community of developers that can help improve software capability and sustainability across projects.
- Partial overlap in functionality is not problematic and in fact improves the diversity of the whole.
- Time of development and maintenance of domain specific software is greatly reduced, as the domain scientists can concentrate on the domain-specific functionalities while lower-level, cross-cutting, numerical packages are maintained by dedicated specialists.
- Portability across platforms (CPUs, GPUs, etc.) of low-level libraries ensures portability of the software that is built upon them. Building new tools on these low-level libraries greatly reduces the overall burden on the community for porting codes to new platforms.
- The domain-specific packages that are built, following established policies are portable across platforms and interoperable, and can be further combined to form larger toolkits and codes.
- Since packages are reused across software and duplication is minimized, bugs and inefficiencies are spotted earlier.
- Many codes, libraries and packages exist in the community that can be reused as building blocks, and progressively modernized and blended in a module that is portable and runs efficiently on modern CPUs and GPUs. Hence, not all functionalities have to be rewritten from scratch at once, offering a progressive path from the current status to an ecosystem of accelerator modeling tools.

It is to be emphasized that the goal is not to propose that only one code be developed to study, e.g., RF accelerators or plasma-based accelerators. On the contrary, the goal is to provide a coordinated infrastructure that enables inclusive and collaborative development of modeling tools for the community, and for these tools, to follow modern practices for scientific software: be portable, efficient, robust and leading to consistently accurate and reproducible results.

5 Sustainability & reliability

An important aspect of the development of scientific modeling tools is to ensure that the corresponding software is robust, easy to use and can be extended for new science cases. This is made more difficult by the fact that these tools are oftentimes constantly evolving, with new capabilities being continuously added to a given software. Here we mention a few important practices [146] that facilitate this process.

5.1 Code robustness

Because modeling tools are being continuously improved and extended by developers, there is always a risk that a given change to the source code may introduce bugs and produce erroneous results in some specific cases. It is therefore paramount that the developers track the changes to the source code (with a version control tool such as `git`), and that they simultaneously maintain a comprehensive suite of tests that the code should always successfully pass. In the case of scientific

modeling codes, these tests can consist of a number of physical setups for which there are well-known theoretical predictions. It can thus be checked that the results of the code conform to these predictions. In order for these tests to be most effective, they need to be run *automatically* whenever the source code is modified — a process known as *continuous integration*.

In addition, the development of large-scale scientific tools nowadays involve a whole community of developers and users rather than a single, isolated developer. Therefore, the robustness and quality of the software is generally greatly improved by the adoption of collaborative development platforms. These online platforms allow the developers to easily and efficiently review each other's changes to the source code, before these changes are incorporated in the mainline version of the code. They also provide a central venue for users to raise issues and questions, and generally communicate their needs with the developers.

Finally, another aspect of a code's robustness is its interface with the user. As the code evolves, the interface with the user often needs to be modified so as to enable functionalities or generalize existing ones. It is not uncommon that these changes break some of the users' scientific workflows. Thus, it is important for the developers to regularly publish releases of the source code, where these changes are clearly documented. The different releases of a given software can be made citable and available through an online archival service, so that users can roll back to a previous release if needed and scrutinize changes in a reproducible manner.

5.2 Usability

In order to maximize the impact of a given software tool on its scientific community, it is also key that this tool be user-friendly. For example, a common obstacle to the adoption of some scientific tools is their complex installation procedure, especially if many dependencies are involved. The developers can drastically simplify the installation procedure of a given software tool by making it available through modern package managers [147–149]. These package managers can automatically download and compile the source code as well as its set of dependencies. They can also ensure that compatible versions of the dependencies are being used. Importantly, some of these package managers can allow users to have the exact same installation workflow on large-scale HPC resources as on their desktop workstation.

Another key aspect of the user experience is the code documentation. For scientific modeling tools, the documentation usually includes sections on how to install and use the code, as well as a description of algorithm being used along with relevant references. Since, again, scientific tools are often continuously evolving, it can be beneficial that the documentation be generated automatically, whenever the source code is modified or whenever a new release is published. Just as the modeling tools themselves, documentation should be easily accessible for the whole community and easy to improve from developers and users alike.

6 Conclusion

Computer modeling will remain an essential component of advanced accelerator research for design, operation and interpretation, even more so as the accelerators mature to increasingly complex production tools with stringent operational constraints. Important grand challenges of the coming

years include the accurate, end-to-end description of multi-stage collider designs with AAC components, the development of hybrid accelerators that arises with upgrades of beamlines with, e.g., plasma focusing elements or boosters.

Fast turn-around and scalable modeling capabilities are indispensable, yet still often far from providing real-time feedback that could benefit interactive design and guide experiments, especially with sufficient fidelity. Thus, training of AI/ML surrogate models adds to the need to accelerate modeling capabilities with modern computing hardware and algorithms. HPC computing trends based on new hardware, most notably GPUs, can contribute to improving the time-to-solution and accuracy of simulations. With a higher level of parallelism and software complexity, deeper software stacks are needed than in traditional modeling approaches and community libraries emerge around themes such as performance-portability layers, accelerated numerics libraries, multi-node data management, dynamic load balancing, scalable I/O, parallel and *in situ* data analysis as well as visualization, to name a few.

Ultimately, the realization of such ambitious goals as the realization of start-to-end virtual twins of particle accelerators that can provide feedback in real time calls for (a) a more effective coordination of the codes development and maintenance across the community - which should be encouraged among laboratories, academia, and industrial partners - together with (b) the development of ecosystems of modeling tools for multiphysics particle accelerator modeling and design, standards and end-to-end workflows. To maximize efficiency and benefits to the community, these should be developed using modern software best practices and adopt open science principles to the largest extent possible, with renewed attention on usability, reliability and sustainability.

Acknowledgments

We acknowledge the support of the Director, Office of Science, Office of High Energy Physics, of the U.S. Department of Energy under Contract No. DEAC02-05CH11231, including support to the Consortium for Advanced Modeling of Particle Accelerators (CAMP) project, and of the Exascale Computing Project (No. 17-SC-20-SC), a collaborative effort of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration. We also acknowledge support by the U.S. Department of Energy, Office of Science, Office of High Energy Physics under Award Number DE-SC0018719. We acknowledge funding by the Gordon and Betty Moore Foundation under grant GMBF4744 (ACHIP) and the German Ministry of Education and Research (Grant No. 05K19RDE). We acknowledge support by the U.S. Department of Energy, Office of Science under contracts DE-SC0018656 with Northern Illinois University and DE-AC02-06CH11357 with Argonne National Laboratory. This research used resources of the National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231. We acknowledge the support by the US National Science Foundation under award number PHY-1505858, as well as the Bose Foundation.

References

[1] The CST Team, *CST*, <https://www.3ds.com/products-services/simulia/products/cst-studio-suite/>.

- [2] The VSIM Team, *VSIM*, <https://txcorp.com/vsim/>.
- [3] D. Sagan, N.M. Cook, Y. Hao, G. Hoffstaetter, A. Huebl, C.-K. Huang et al., *Simulations of Future Particle Accelerators: Issues and Mitigations*, in *ICFA Beam Dynamics Newsletter#82 — Advanced Accelerator Modelling* (2021), **2021 JINST** **16** T10002 [[arXiv:2108.11027](https://arxiv.org/abs/2108.11027)]
- [4] P. Chen, J.M. Dawson, R.W. Huff and T.C. Katsouleas, *Acceleration of Electrons by the Interaction of a Bunched Electron Beam With a Plasma*, *Phys. Rev. Lett.* **54** (1985) 693 [*Erratum ibid.* **55** (1985) 1537]
- [5] E. Esarey, C.B. Schroeder and W.P. Leemans, *Physics of laser-driven plasma-based electron accelerators*, *Rev. Mod. Phys.* **81** (2009) 1229.
- [6] B.D. O’Shea, G. Andonian, S.S. Baturin, C.I. Clarke, P.D. Hoang, M.J. Hogan et al., *Suppression of Deflecting Forces in Planar-Symmetric Dielectric Wakefield Accelerating Structures with Elliptical Bunches*, *Phys. Rev. Lett.* **124** (2020) 104801.
- [7] Q. Gao et al., *Observation of High Transformer Ratio of Shaped Bunch Generated by an Emittance-Exchange Beam Line*, *Phys. Rev. Lett.* **120** (2018) 114801.
- [8] G. Andonian et al., *Planar-Dielectric-Wakefield Accelerator Structure Using Bragg-Reflector Boundaries*, *Phys. Rev. Lett.* **113** (2014) 264801.
- [9] P.D. Hoang et al., *Experimental characterization of electron beam driven wakefield modes in a dielectric woodpile Cartesian symmetric structure*, *Phys. Rev. Lett.* **120** (2018) 164801 [[arXiv:1708.00108](https://arxiv.org/abs/1708.00108)].
- [10] F. Lemery, K. Floettmann, P. Piot, F.X. Kärtner and R. Aßmann, *Synchronous acceleration with tapered dielectric-lined waveguides*, *Phys. Rev. Accel. Beams* **21** (2018) 051302 [[arXiv:1712.08403](https://arxiv.org/abs/1712.08403)].
- [11] X. Lu, M.A. Shapiro, I. Mastovsky, R.J. Temkin, M. Conde, J.G. Power et al., *Generation of High-Power, Reversed-Cherenkov Wakefield Radiation in a Metamaterial Structure*, *Phys. Rev. Lett.* **122** (2019) 014801.
- [12] R.J. England, R.J. Noble, Z. Wu and M. Qi, *Dielectric Laser Acceleration*, *Rev. Mod. Phys.* **86** (2014) 1337 [[arXiv:1309.7637](https://arxiv.org/abs/1309.7637)].
- [13] D. Cesar, S. Custodio, J. Maxson, P. Musumeci, X. Shen, E. Threlkeld et al., *High-field nonlinear optical response and phase control in a dielectric laser accelerator*, *Commun. Phys.* **1** (2018) 46.
- [14] U. Niedermayer, T. Egenolf and O. Boine-Frankenheim, *Beam Dynamics Analysis of Dielectric Laser Acceleration using a Fast 6D Tracking Scheme*, *Phys. Rev. Accel. Beams* **20** (2017) 111302 [[arXiv:1707.09815](https://arxiv.org/abs/1707.09815)].
- [15] T. Hughes, G. Veronis, K.P. Wootton, R.J. England and S. Fan, *Method for computationally efficient design of dielectric laser accelerator structures*, *Opt. Exp.* **25** (2017) 15414.
- [16] C.K. Birdsall and A.B. Langdon, *Plasma Physics Via Computer Simulation*, Adam-Hilger (1991).
- [17] A. Lifschitz, X. Davoine, E. Lefebvre, J. Faure, C. Rechatin and V. Malka, *Particle-in-cell modelling of laser-plasma interaction using fourier decomposition*, *J. Comput. Phys.* **228** (2009) 1803.
- [18] P. Sprangle, E. Esarey and A. Ting, *Nonlinear theory of intense laser-plasma interactions*, *Phys. Rev. Lett.* **64** (1990) 2011.
- [19] T.M. Antonsen and P. Mora, *Self-focusing and Raman scattering of laser pulses in tenuous plasmas*, *Phys. Rev. Lett.* **69** (1992) 2204.

- [20] D. Gordon, W. Mori and T. Antonsen, *A ponderomotive guiding center particle-in-cell code for efficient modeling of laser-plasma interactions*, *IEEE Trans. Plasma Sci.* **28** (2000) 1135.
- [21] C. Benedetti, C.B. Schroeder, E. Esarey, C.G.R. Geddes and W.P. Leemans, *Efficient Modeling Of Laser-Plasma Accelerators With Inf&Rno*, *AIP Conf. Proc.* **1299** (2010) 250.
- [22] B.M. Cowan, D.L. Bruhwiler, E. Cormier-Michel, E. Esarey, C.G.R. Geddes, P. Messmer et al., *Characteristics Of An Envelope Model For Laser-Plasma Accelerator Simulation*, *J. Computat. Phys.* **230** (2011) 61.
- [23] D. Terzani, C. Benedetti, C.B. Schroeder and E. Esarey, *Accuracy of the time-averaged ponderomotive approximation for laser-plasma accelerator modeling*, *Phys. Plasmas* **28** (2021) 063105.
- [24] J.L. Vay, *Noninvariance of Space- and Time-Scale Ranges under a Lorentz Transformation and the Implications for the Study of Relativistic Interactions*, *Phys. Rev. Lett.* **98** (2007) 130405.
- [25] J.-L. Vay and R. Lehe, *Simulations for Plasma and Laser Acceleration*, *Rev. Accel. Sci. Technol.* **9** (2017) 165.
- [26] A.F. Oskooi, D. Roundy, M. Ibanescu, P. Bermel, J. Joannopoulos and S.G. Johnson, *Meep: A flexible free-software package for electromagnetic simulations by the fDTD method*, *Comp. Phys. Commun.* **181** (2010) 687.
- [27] J.-L. Vay, D.P. Grote, R.H. Cohen and A. Friedman, *Novel methods in the particle-in-cell accelerator code-framework warp*, *Comp. Sci. Discovery* **5** (2012) 014019.
- [28] M. Borland, *elegant: A Flexible SDDS-Compliant Code for Accelerator Simulation*, in *Proceedings of the 6th International Computational Accelerator Physics Conference*, Darmstadt, Germany, 11–14 September 2000 [DOI: [10.2172/761286](https://doi.org/10.2172/761286)].
- [29] S.M. Lidia, K. Floettmann and P.R.-G. Piot, *Recent Improvements to the ASTRA Particle Tracking Code*, *Conf. Proc. C* **030512** (2003) 3500.
- [30] K.L.F. Bane and P. Emma, *LiTrack: A Fast longitudinal phase space tracking code with graphical user interface*, *Conf. Proc. C* **0505161** (2005) 4266.
- [31] W.-H. Tan, P. Piot and A. Zholents, *Formation of Temporally Shaped Electron Bunches for Beam-Driven Collinear Wakefield Accelerators*, *Phys. Rev. Accel. Beams* **24** (2021) 051303 [[arXiv:2101.07414](https://arxiv.org/abs/2101.07414)].
- [32] D. Mihalcea, P. Piot and P. Stoltz, *Three-Dimensional Analysis of Wakefields Generated by Flat Electron Beams in Planar Dielectric-Loaded Structures*, *Phys. Rev. ST Accel. Beams* **15** (2012) 081304 [[arXiv:1204.6724](https://arxiv.org/abs/1204.6724)].
- [33] T.H. Pacey, Y. Saveliev, G. Xia and J. Smith, *Simulation studies for dielectric wakefield programme at CLARA facility*, *Nucl. Instrum. Meth. A* **909** (2018) 261 [[arXiv:1712.03716](https://arxiv.org/abs/1712.03716)].
- [34] I. Zagorodnov, *Wakefield Code ECHO 2(3)D*, in *ICFA mini-Workshop on “Electromagnetic wake fields and impedances in particle accelerators*, Erice, Italy, 24–28 April 2014 [<https://bib-pubdb1.desy.de/record/206299>.]
- [35] ACE3P, https://portal.slac.stanford.edu/sites/ard_public/acd/Pages/Default.aspx
- [36] H. Vincenti and J.-L. Vay, *Ultrahigh-order Maxwell solver with extreme scalability for electromagnetic PIC simulations of plasmas*, *Comp. Phys. Commun.* **228** (2018) 22.
- [37] B.D. O’Shea, G. Andonian, S.K. Barber, C.I. Clarke, P.D. Hoang, M.J. Hogan et al., *Conductivity Induced by High-Field Terahertz Waves in Dielectric Material*, *Phys. Rev. Lett.* **123** (2019) 134801.

- [38] ALEGRO collaboration, *Towards an Advanced Linear International Collider*, [arXiv:1901.10370](https://arxiv.org/abs/1901.10370).
- [39] W. Shin and S. Fan, *Accelerated solution of the frequency-domain maxwell's equations by engineering the eigenvalue distribution of the operator*, *Opt. Exp.* **21** (2013) 22578.
- [40] T. Egenolf, O. Boine-Frankenheim and U. Niedermayer, *Simulation of DLA grating structures in the frequency domain*, *J. Phys. Conf. Ser.* **874** (2017) 012040.
- [41] U. Niedermayer, T. Egenolf, O. Boine-Frankenheim and P. Hommelhoff, *Alternating Phase Focusing for Dielectric Laser Acceleration*, *Phys. Rev. Lett.* **121** (2018) 214801 [[arXiv:1806.07287](https://arxiv.org/abs/1806.07287)].
- [42] U. Niedermayer, T. Egenolf and O. Boine-Frankenheim, *Three Dimensional Alternating-Phase Focusing for Dielectric-Laser Electron Accelerators*, *Phys. Rev. Lett.* **125** (2020) 164801 [[arXiv:2004.05458](https://arxiv.org/abs/2004.05458)].
- [43] T. Egenolf, U. Niedermayer and O. Boine-Frankenheim, *Tracking with wakefields in dielectric laser acceleration grating structures*, *Phys. Rev. Accel. Beams* **23** (2020) 054402 [[arXiv:1911.03337](https://arxiv.org/abs/1911.03337)].
- [44] C.G.R. Geddes, C. Toth, J. van Tilborg, E. Esarey, C.B. Schroeder, D. Bruhwiler et al., *High-quality electron beams from a laser wakefield accelerator using plasma-channel guiding*, *Nature* **431** (2004) 538.
- [45] W.P. Leemans, B. Nagler, A.J. Gonsalves, C. Toth, K. Nakamura, C.G.R. Geddes et al., *GeV electron beams from a cm-scale accelerator*, *Nature Phys.* **2** (2006) 696.
- [46] T.P.A. Ibotson et al., *Laser-wakefield acceleration of electron beams in a low density plasma channel*, *Phys. Rev. ST Accel. Beams* **13** (2010) 031301.
- [47] H. Suk, N. Barov, J.B. Rosenzweig and E. Esarey, *Plasma electron trapping and acceleration in a plasma wake field using a density transition*, *Phys. Rev. Lett.* **86** (2001) 1011.
- [48] P. Sprangle, B. Hafizi, J.R. Penano, R.F. Hubbard, A. Ting, A. Zigler et al., *Stable laser-pulse propagation in plasma channels for GeV electron acceleration*, *Phys. Rev. Lett.* **85** (2000) 5110.
- [49] B.E. Blue et al., *Plasma wake field acceleration of an intense positron beam*, *Phys. Rev. Lett.* **90** (2003) 214801.
- [50] I. Blumenfeld et al., *Energy doubling of 42 GeV electrons in a metre-scale plasma wakefield accelerator*, *Nature* **445** (2007) 741.
- [51] M. Litos et al., *High-efficiency acceleration of an electron beam in a plasma wakefield accelerator*, *Nature* **515** (2014) 92.
- [52] M.J. Hogan et al., *Plasma wakefield acceleration experiments at FACET*, *New J. Phys.* **12** (2010) 055030.
- [53] S.Z. Green, E. Adli, C.I. Clarke, S. Corde, S.A. Edstrom, A.S. Fisher et al., *Laser ionized preformed plasma at FACET*, *Plasma Phys. Control. Fusion* **56** (2014) 084011.
- [54] S. Gessner, E. Adli, J.M. Allen, W. An, C.I. Clarke, C.E. Clayton et al., *Demonstration of a positron beam-driven hollow channel plasma wakefield accelerator*, *Nature Commun.* **7** (2016) .
- [55] S. Lee, T.C. Katsouleas, R.G. Hemker, E.S. Dodd and W.B. Mori, *Plasma-wakefield acceleration of a positron beam*, *Phys. Rev. E* **64** (2001) 045501.
- [56] C.B. Schroeder, E. Esarey, C. Benedetti and W.P. Leemans, *Control of focusing forces and emittances in plasma-based accelerators using near-hollow plasma channels*, *Phys. Plasmas* **20** (2013) 080701.

[57] S. Diederichs, T.J. Mehrling, C. Benedetti, C.B. Schroeder, A. Knetsch, E. Esarey et al., *Positron transport and acceleration in beam-driven plasma wakefield accelerators using plasma columns*, *Phys. Rev. Accel. Beams* **22** (2019) 081301.

[58] J. van Tilborg et al., *Active Plasma Lensing for Relativistic Laser-Plasma-Accelerated Electron Beams*, *Phys. Rev. Lett.* **115** (2015) 184802.

[59] S. Steinke et al., *Multistage coupling of independent laser-plasma accelerators*, *Nature* **530** (2016) 190.

[60] R. Lehe, C. Thaury, E. Guillaume, A. Lifschitz and V. Malka, *Laser-plasma lens for laser-wakefield accelerators*, *Phys. Rev. ST Accel. Beams* **17** (2014) 121301.

[61] C.E. Doss et al., *Laser-ionized, beam-driven, underdense, passive thin plasma lens*, *Phys. Rev. Accel. Beams* **22** (2019) 111001.

[62] R. D'Arcy et al., *Tunable Plasma-Based Energy Dechirper*, *Phys. Rev. Lett.* **122** (2019) 034801 [[arXiv:1810.06307](https://arxiv.org/abs/1810.06307)].

[63] P. Scherkl et al., *Plasma-photonic spatiotemporal synchronization of relativistic electron and laser beams*, [arXiv:1908.09263](https://arxiv.org/abs/1908.09263).

[64] G.A. Bagdasarov, P.V. Sasorov, V.A. Gasilov, A.S. Boldarev, O.G. Olkhovskaya, C. Benedetti et al., *Laser beam coupling with capillary discharge plasma for laser wakefield acceleration applications*, *Phys. Plasmas* **24** (2017) 083109.

[65] A.J. Gonsalves et al., *Petawatt Laser Guiding and Electron Beam Acceleration to 8 GeV in a Laser-Heated Capillary Discharge Waveguide*, *Phys. Rev. Lett.* **122** (2019) 084801.

[66] G.A. Bagdasarov, N.A. Bobrova, O.G. Olkhovskaya, V.A. Gasilov, C. Benedetti, S.S. Bulanov et al., *Creation of an axially uniform plasma channel in a laser-assisted capillary discharge*, *Phys. Plasmas* **28** (2021) 053104.

[67] J. van Tilborg et al., *Nonuniform discharge currents in active plasma lenses*, *Phys. Rev. Accel. Beams* **20** (2017) 032803.

[68] C.A. Lindstrøm et al., *Emittance Preservation in an Aberration-Free Active Plasma Lens*, *Phys. Rev. Lett.* **121** (2018) 194801 [*Erratum ibid.* **122** (2019) 129901] [[arXiv:1808.03691](https://arxiv.org/abs/1808.03691)].

[69] B.E. Carlsten, D. Schulte and W. Lu, *Summary of Working Group 8: Advanced and Novel Accelerators for High Energy Physics*, *J. Phys. Conf. Ser.* **1596** (2020) 012063.

[70] X.L. Xu et al., *Physics of Phase Space Matching for Staging Plasma and Traditional Accelerator Components Using Longitudinally Tailored Plasma Profiles*, *Phys. Rev. Lett.* **116** (2016) 124801 [[arXiv:1411.4386](https://arxiv.org/abs/1411.4386)].

[71] A. Deng et al., *Generation and acceleration of electron bunches from a plasma photocathode*, *Nature Phys.* **15** (2019) 1156.

[72] Y. Xi, B. Hidding, D. Bruhwiler, G. Pretzler and J.B. Rosenzweig, *Hybrid modeling of relativistic underdense plasma photocathode injectors*, *Phys. Rev. ST Accel. Beams* **16** (2013) 031303.

[73] G.G. Manahan, A.F. Habib, P. Scherkl, D. Ullmann, A. Beaton, A. Sutherland et al., *Advanced schemes for underdense plasma photocathode wakefield accelerators: pathways towards ultrahigh brightness electron beams*, *Philos. Trans. Roy. Soc. A* **377** (2019) 20180182.

[74] AWAKE collaboration, *Evolution of a plasma column measured through modulation of a high-energy proton beam*, [arXiv:2006.09991](https://arxiv.org/abs/2006.09991).

[75] J.-L. Vay, D. Sagan, A. Huebl, M. Thévenet, R. Lehe, C.-K. Ng et al., *End-to-End Virtual Accelerators (EVA)*, submitted to *Snowmass21* [https://snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF2_CompF0-AF1_AF0_Vay-067.pdf].

[76] R. Lehe, A. Hanuka, A. Edelen, X. Huang, C. Mayes, N. Cook et al., *Machine learning and surrogates models for simulation-based optimization of accelerator design*, submitted to *Snowmass21* [https://snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF2_CompF3-AF1_AF6_Lehe-075.pdf].

[77] D. Winklehner and A. Adelmann, *Application of Machine Learning to Particle Accelerator Simulations*, submitted to *Snowmass21* [https://snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF3_CompF0-AF1_AF0_Winklehner-108.pdf].

[78] S. Nagaitsev et al., *Accelerator and Beam Physics Research Goals and Opportunities*, submitted to *Snowmass21* [https://snowmass21.org/docs/files/summaries/AF/SNOWMASS21-AF1_AF7_S_Nagaitsev-056.pdf], [[arXiv:2101.04107](https://arxiv.org/abs/2101.04107)].

[79] A. Adelmann, *On Nonintrusive Uncertainty Quantification and Surrogate Model Construction in Particle Accelerator Modeling*, *SIAM/ASA J. Uncertain. Quantif.* **7** (2019) 383.

[80] F. Van der Veken et al., *Machine learning in accelerator physics: Applications at the CERN Large Hadron Collider*, *PoS AISIS2019* (2020) 044.

[81] A. Edelen, S. Biedron, J. Edelen and S. Milton, *First Steps Toward Incorporating Image Based Diagnostics into Particle Accelerator Control Systems Using Convolutional Neural Networks*, in *Proceedings of the 2nd North American Particle Accelerator Conference*, Chicago, IL, U.S.A., 9–14 October 2016, TUPOA51 [[arXiv:1612.05662](https://arxiv.org/abs/1612.05662)].

[82] A. Edelen, N. Neveu, Y. Huber, M. Frey, C. Mayes and A. Adelmann, *Machine learning for orders of magnitude speedup in multiobjective optimization of particle accelerator systems*, *Phys. Rev. Accel. Beams* **23** (2020) 044601 [[arXiv:1903.07759](https://arxiv.org/abs/1903.07759)].

[83] J. Kirschner, M. Mutny, N. Hiller, R. Ischebeck and A. Krause, *Adaptive and Safe Bayesian Optimization in High Dimensions via One-Dimensional Subspaces*, [arXiv:1902.03229](https://arxiv.org/abs/1902.03229).

[84] J. Kirschner, M. Nonnenmacher, M. Mutny, A. Krause, N. Hiller, R. Ischebeck et al., *Bayesian Optimisation for Fast and Safe Parameter Tuning of SwissFEL*, in *Proceedings of the 39th International Free-Electron Laser Conference*, Hamburg, Germany, 26–30 August 2019, pp. 707–710.

[85] J. Duris et al., *Bayesian optimization of a free-electron laser*, *Phys. Rev. Lett.* **124** (2020) 124801 [[arXiv:1909.05963](https://arxiv.org/abs/1909.05963)].

[86] M. Bussmann, H. Burau, T.E. Cowan, A. Debus, A. Huebl, G. Juckeland et al., *Radiative signatures of the relativistic kelvin-helmholtz instability*, in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*, Denver, CO, U.S.A., 17–21 November 2013, pp. 5:1–5:12.

[87] A. Myers et al., *Porting WarpX to GPU-accelerated platforms*, [arXiv:2101.12149](https://arxiv.org/abs/2101.12149).

[88] *Kokkos*, <https://github.com/kokkos>.

[89] E. Zenker, B. Worpitz, R. Widera, A. Huebl, G. Juckeland, A. Knüpfer et al., *Alpaka — An Abstraction Library for Parallel Kernel Acceleration*, *Proceedings of the IEEE 30th International Parallel and Distributed Processing Symposium*, Chicago, IL, U.S.A., 23–27 May 2016, pp. 631–640.

[90] *LLNL/RAJA: RAJA Performance Portability Layer (C++)*, <https://github.com/LLNL/RAJA>.

- [91] *AMReX*, <https://amrex-codes.github.io/>.
- [92] The *Copa* Team, *Copa*, <https://github.com/ECP-copa>.
- [93] B.B. Godfrey, *Numerical cherenkov instabilities in electromagnetic particle codes*, *J. Comput. Phys.* **15** (1974) 504.
- [94] B.B. Godfrey, *Canonical momenta and numerical instabilities in particle codes*, *J. Comput. Phys.* **19** (1975) 58 .
- [95] J.-L. Vay et al., *Application Of The Reduction Of Scale Range In A Lorentz Boosted Frame To The Numerical Simulation Of Particle Acceleration Devices*, in *Proceedings of the 23rd Particle Accelerator Conference*, Vancouver, Canada, 4–8 May 2009.
- [96] S.F. Martins, R.A. Fonseca, L.O. Silva, W. Lu and W.B. Mori, *Numerical Simulations Of Laser Wakefield Accelerators In Optimal Lorentz Frames*, *Comp. Phys. Commun.* **181** (2010) 869.
- [97] J.L. Vay, C.G.R. Geddes, E. Cormier-Michel and D.P. Grote, *Numerical Methods For Instability Mitigation In The Modeling Of Laser Wakefield Accelerators In A Lorentz-Boosted Frame*, *J. Comput. Phys.* **230** (2011) 5908.
- [98] J. Vay, C.G.R. Geddes, E. Cormier-Michel and D.P. Grote, *Effects Of Hyperbolic Rotation In Minkowski Space On The Modeling Of Plasma Accelerators In A Lorentz Boosted Frame*, *Phys. Plasmas* **18** (2011) 30701.
- [99] B.B. Godfrey and J.-L. Vay, *Numerical stability of relativistic beam multidimensional PIC simulations employing the Esirkepov algorithm*, *J. Comput. Phys.* **248** (2013) 33.
- [100] X. Xu, P. Yu, S.F. Martins, F.S. Tsung, V.K. Decyk, J. Vieira et al., *Numerical instability due to relativistic plasma drift in EM-PIC simulations*, *Computer Physics Communications* **184** (2013) 2503.
- [101] B.B. Godfrey, J.-L. Vay and I. Haber, *Numerical stability analysis of the pseudo-spectral analytical time-domain PIC algorithm*, *J. Comput. Phys.* **258** (2014) 689.
- [102] B.B. Godfrey and J.-L. Vay, *Suppressing the numerical cherenkov instability in FDTD PIC codes*, *J. Comput. Phys.* **267** (2014) 1.
- [103] B.B. Godfrey, J.-L. Vay and I. Haber, *Numerical Stability Improvements for the Pseudo-Spectral EM PIC Algorithm*, *IEEE Trans. Plasma Sci.* **42** (2014) 1339 [[arXiv:1309.0116](https://arxiv.org/abs/1309.0116)].
- [104] B.B. Godfrey and J.L. Vay, *Improved numerical Cherenkov instability suppression in the generalized PSTD PIC algorithm*, *Comput. Phys. Commun.* **196** (2015) 221.
- [105] P. Yu, X. Xu, V.K. Decyk, F. Fiuza, J. Vieira, F.S. Tsung et al., *Elimination of the numerical Cerenkov instability for spectral EM-PIC codes*, *Comput. Phys. Commun.* **192** (2015) 32.
- [106] P. Yu, X. Xu, A. Tableman, V.K. Decyk, F.S. Tsung, F. Fiuza et al., *Mitigation of numerical Cerenkov radiation and instability using a hybrid finite difference-FFT Maxwell solver and a local charge conserving current deposit*, *Comput. Phys. Commun.* **197** (2015) 144.
- [107] R. Lehe, M. Kirchen, B.B. Godfrey, A.R. Maier and J.-L. Vay, *Elimination of numerical Cherenkov instability in flowing-plasma particle-in-cell simulations by using Galilean coordinates*, *Phys. Rev. E* **94** (2016) 053305.
- [108] M. Kirchen, R. Lehe, B.B. Godfrey, I. Dornmair, S. Jalas, K. Peters et al., *Stable discrete representation of relativistically drifting plasmas*, *Phys. Plasmas* **23** (2016) 100704.

[109] M. Kirchen, R. Lehe, S. Jalas, O. Shapoval, J.L. Vay and A.R. Maier, *Scalable spectral solver in Galilean coordinates for eliminating the numerical Cherenkov instability in particle-in-cell simulations of streaming plasmas*, *Phys. Rev. E* **102** (2020) 13202.

[110] F. Li, P. Yu, X. Xu, F. Fiuza, V.K. Decyk, T. Dalichaouch et al., *Controlling the numerical cerenkov instability in pic simulations using a customized finite difference maxwell solver and a local fft based current correction*, *Comput. Phys. Commun.* **214** (2017) 6.

[111] A. Pukhov, *X-dispersionless Maxwell solver for plasma-based particle acceleration*, *J. Comput. Phys.* **418** (2020) 109622.

[112] I. Haber, R. Lee, H. Klein and J. Boris, *Advances In Electromagnetic Simulation Techniques*, in *Proceedings of 6th Conference on Numerical Simulation of Plasmas*, 1973, Berkeley, CA, U.S.A., pp. 46–48.

[113] J.L. Vay, I. Haber and B.B. Godfrey, *A domain decomposition method for pseudo-spectral electromagnetic simulations of plasmas*, *J. Comput. Phys.* **243** (2013) 260.

[114] H. Vincenti and J.-L. Vay, *Detailed analysis of the effects of stencil spatial variations with arbitrary high-order finite-difference Maxwell solver*, *Comput. Phys. Commun.* **200** (2016) 147.

[115] S. Jalas, I. Dornmair, R. Lehe, H. Vincenti, J.-L. Vay, M. Kirchen et al., *Accurate modeling of plasma acceleration with arbitrary order pseudo-spectral particle-in-cell methods*, *Phys. Plasmas* **24** (2017) 033115.

[116] O. Shapoval, R. Lehe, M. Thévenet, E. Zoni, Y. Zhao and J.-L. Vay, *Overcoming timestep limitations in boosted-frame Particle-In-Cell simulations of plasma-based acceleration*, [arXiv:2104.13995](https://arxiv.org/abs/2104.13995).

[117] E. Zoni, R. Lehe, O. Shapoval, D. Belkin, N. Zaïm, L. Fedeli et al., *A Hybrid Nodal-Staggered Pseudo-Spectral Electromagnetic Particle-In-Cell Method with Finite-Order Centering*, [arXiv:2106.12919](https://arxiv.org/abs/2106.12919).

[118] J.-L. Vay, P. Colella, J.W. Kwan, P. McCorquodale, D.B. Serafini, A. Friedman et al., *Application of adaptive mesh refinement to particle-in-cell simulations of plasmas and beams*, *Phys. Plasmas* **11** (2004) 2928.

[119] J.-L. Vay, J.-C. Adam and A. Héron, *Asymmetric PML for the absorption of waves. application to mesh refinement in electromagnetic particle-in-cell plasma simulations*, *Comput. Phys. Commun.* **164** (2004) 171.

[120] J.-L. Vay, D.P. Grote, R.H. Cohen and A. Friedman, *Novel methods in the particle-in-cell accelerator code-framework warp*, *Comput. Sci. Discovery* **5** (2012) 014019.

[121] J.L. Vay et al., *Warp-X: A new exascale computing platform for beam–plasma simulations*, *Nucl. Instrum. Meth. A* **909** (2018) 476 [[arXiv:1801.02568](https://arxiv.org/abs/1801.02568)].

[122] J.L. Vay, A. Almgren, J. Bell, R. Lehe, A. Myers, J. Park et al., *Toward plasma wakefield simulations at exascale*, in *Proceedings of the IEEE Advanced Accelerator Concepts Workshop*, Breckenridge, CO, U.S.A., 12–17 August 2018, pp. 1–5.

[123] R. Lehe, M. Kirchen, I.A. Andriyash, B.B. Godfrey and J.-L. Vay, *A spectral, quasi-cylindrical and dispersion-free Particle-In-Cell algorithm*, *Comput. Phys. Commun.* **203** (2016) 66.

[124] F. Li, W. An, V.K. Decyk, X. Xu, M.J. Hogan and W.B. Mori, *A quasi-static particle-in-cell algorithm based on an azimuthal Fourier decomposition for highly efficient simulations of plasma-based acceleration: QPAD*, *Comput. Phys. Commun.* **261** (2021) 107784.

[125] *FBPIC*, <https://fbpic.github.io>.

[126] H. Kallala, J.L. Vay and H. Vincenti, *A generalized massively parallel ultra-high order FFT-based Maxwell solver*, *Comput. Phys. Commun.* **244** (2019) 25.

[127] G. Blalard, H. Vincenti, R. Lehe and J.L. Vay, *Pseudospectral Maxwell solvers for an accurate modeling of Doppler harmonic generation on plasma mirrors with particle-in-cell codes*, *Phys. Rev. E* **96** (2017) 033305.

[128] A. Leblanc, S. Monchocé, H. Vincenti, S. Kahaly, J.-L. Vay and F. Quéré, *Spatial Properties of High-Order Harmonic Beams from Plasma Mirrors: A Ptychographic Study*, *Phys. Rev. Lett.* **119** (2017) 155001.

[129] H. Vincenti, *Achieving Extreme Light Intensities using Optically Curved Relativistic Plasma Mirrors*, *Phys. Rev. Lett.* **123** (2019) 105001 [[arXiv:1812.05357](https://arxiv.org/abs/1812.05357)].

[130] L. Chopineau, A. Leblanc, G. Blalard, A. Denoeud, M. Thévenet, J.L. Vay et al., *Identification of Coupling Mechanisms between Ultraintense Laser Light and Dense Plasmas*, *Phys. Rev. X* **9** (2019) 011050.

[131] L. Fedeli, A. Sainte-Marie, N. Zaïm, M. Thévenet, J.-L. Vay, A. Myers et al., *Probing Strong-Field QED with Doppler-Boosted Petawatt-Class Lasers*, *Phys. Rev. Lett.* **127** (2021) 114801 [[arXiv:2012.07696](https://arxiv.org/abs/2012.07696)].

[132] L. Wan, A. Huebl, J. Gu, F. Poeschel, A. Gainaru, R. Wang et al., *Improving I/O Performance for Exascale Applications through Online Data Layout Reorganization*, accepted in *IEEE Trans. Parallel Distrib. Syst.* (2021) [[DOI: 10.1109/TPDS.2021.3100784](https://doi.org/10.1109/TPDS.2021.3100784)].

[133] A. Huebl, R. Widera, F. Schmitt, A. Matthes, N. Podhorszki, J.Y. Choi et al., *On the scalability of data reduction techniques in current and upcoming hpc systems from an application perspective*, in *High Performance Computing*, J.M. Kunkel, R. Yokota, M. Taufer and J. Shalf, eds., Springer International Publishing (2017), pp. 15–29.

[134] A. Huebl, R. Lehe, J.-L. Vay, D.P. Grote, I. Sbalzarini, S. Kuschel et al., *openPMD: A meta data standard for particle and mesh based data*, <https://www.openpmd.org> [[DOI: 10.5281/zenodo.591699](https://doi.org/10.5281/zenodo.591699)].

[135] F. Poeschel, J. E, W.F. Godoy, N. Podhorszki, S. Klasky, G. Eisenhauer et al., *Transitioning from file-based HPC workflows to streaming data pipelines with openPMD and ADIOS2*, submitted to *Smoky Mountains Computational Sciences & Engineering Conference*, 2021.

[136] *CAMPA: Consortium for Advanced Modeling of Particle Accelerators*, <http://campa.lbl.gov>.

[137] *PICMI*, <https://github.com/picmi-standard>.

[138] W.F. Godoy, N. Podhorszki, R. Wang, C. Atkins, G. Eisenhauer, J. Gu et al., *Adios 2: The adaptable input output system. a framework for high-performance data management*, *SoftwareX* **12** (2020) 100561.

[139] The HDF Group, *Hierarchical data format version 5*, <http://www.hdfgroup.org/HDF5>.

[140] A. Huebl, F. Poeschel, F. Koller and J. Gu, *openPMD-api: C++ & Python API for Scientific I/O with openPMD*, <https://github.com/openPMD/openPMD-api> [[DOI: 10.14278/rodare.27](https://doi.org/10.14278/rodare.27)].

[141] R.A. Fonseca et al., *OSIRIS: A Three-Dimensional, Fully Relativistic Particle in Cell Code for Modeling Plasma Based Accelerators*, in *Computational Science — ICCS 2002. Part III*, Lecture Notes in Computer Science, vol. 2331, Springer (2002), pp. 342–351.

- [142] J.L. Vay, A. Huebl, A. Almgren, L.D. Amorim, J. Bell, L. Fedeli et al., *Modeling of a chain of three plasma accelerator stages with the WarpX electromagnetic PIC code on GPUs*, *Phys. Plasmas* **28** (2021) 23105.
- [143] M. Thévenet, S. Diederichs, A. Sinn, R. Lehe, A. Huebl, A. Myers et al., *HiPACE++: Highly efficient Plasma Accelerator Emulation, quasistatic particle-in-cell code*, 2021, <https://github.com/Hi-PACE/hipace>.
- [144] *IDEAS: Interoperable Design of Extreme-scale Application Software*, <https://ideas-productivity.org>.
- [145] *xSDK: Extreme-scale Scientific Software Development Kit*, <http://xsdk.info>.
- [146] R. Lehe, A. Huebl, J.-L. Vay, A. Friedman, M. Thevenet, C. Mitchell et al., *Embracing modern software tools and user-friendly practices, when distributing scientific codes*, submitted to *Snowmass21* [https://snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF2_CompF0_Lehe-076.pdf].
- [147] *Spack*, <https://spack.io/>.
- [148] *Conda*, <https://docs.conda.io/en/latest/>.
- [149] *The python package installer*, <https://pip.pypa.io/en/stable/>.