Title

From Compact Plasma Particle Sources to Advanced Accelerators with Modeling at Exascale

Authors

Huebl, Axel
Lehe, Remi
Zoni, Edoardo
et al.

Publication Date

2023-03-22

From Compact Plasma Particle Sources to Advanced Accelerators with Modeling at Exascale

Axel Huebl, Rémi Lehe, Edoardo Zoni, Olga Shapoval,
Ryan T. Sandberg, Marco Garten, Arianna Formenti,
Revathi Jambunathan, Prabhat Kumar, Kevin Gott,
Andrew Myers, Weiqun Zhang, Ann Almgren,
Chad E. Mitchell, Ji Qiang, Jean-Luc Vay
Lawrence Berkeley National Laboratory
Berkeley (CA), USA
{axelhuebl,jlvay}@lbl.gov

Alexander Sinn, Severin Diederichs,
Maxence Thévenet
Deutsches Elektronen-Synchrotron (DESY)
Hamburg, Germany

David Grote
Lawrence Livermore National Laboratory
Livermore (CA), USA

Luca Fedeli, Thomas Clark, Neil Zaïm, Henri Vincenti
LIDYL, CEA-Université Paris-Saclay, CEA Saclay
Gif-sur-Yvette, France

***Abstract*—Developing complex, reliable advanced accelerators requires a coordinated, extensible, and comprehensive approach in modeling, from source to the end of beam lifetime. We present highlights in Exascale Computing to scale accelerator modeling software to the requirements set for contemporary science drivers. In particular, we present the first laser-plasma modeling on an exaflop supercomputer using the US DOE Exascale Computing Project WarpX. Leveraging developments for Exascale, the new DOE SCIDAC-5 Consortium for Advanced Modeling of Particle Accelerators (CAMPA) will advance numerical algorithms and accelerate community modeling codes in a cohesive manner: from beam source, over energy boost, transport, injection, storage, to application or interaction. Such start-to-end modeling will enable the exploration of hybrid accelerators, with conventional and advanced elements, as the next step for advanced accelerator modeling. Following open community standards, we seed an open ecosystem of codes that can be readily combined with each other and machine learning frameworks. These will cover ultrafast to ultraprecise modeling for future hybrid accelerator design, even enabling virtual test stands and twins of accelerators that can be used in operations.**

This research was supported by the Exascale Computing Project (17-SC-20-SC), a joint project of the U.S. Department of Energy's Office of Science and National Nuclear Security Administration, responsible for delivering a capable exascale ecosystem, including software, applications, and hardware technology, to support the nation's exascale computing imperative. This material is based upon work supported by the CAMPA collaboration, a project of the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research and Office of High Energy Physics, Scientific Discovery through Advanced Computing (SciDAC) program. This work was supported by the Laboratory Directed Research and Development Program of Lawrence Berkeley National Laboratory under U.S. Department of Energy Contract No. DE-AC02-05CH11231 and by LLNL under Contract DE-AC52-07NA27344. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, 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, and the supercomputer Fugaku provided by RIKEN.

***Index Terms*—simulation, exascale, particle accelerator, particle-mesh, particle-in-cell, laser-plasma, modeling, HPC**

I. Introduction

Research of plasma-based accelerators has achieved significant milestones over the last decade. Highlights include achieving nearly 8 GeV electrons in a single-stage source [1], nC-class electron beams [2], demonstrating plasma-based FELs [3]–[5], reaching stable proton acceleration of ultrashort, nC-class pulses [6] and enabling studies into ultrahigh dose rate radiotherapy [7]–[9]. As the exploratory aspect of the field benefits significantly from the elucidation of fundamental processes through simulations, transitioning from intriguing sources to scalable accelerators requires universally integrated, quantitatively predictive capabilities for design and operations.

II. Next-Generation, Advanced Particle Accelerator Modeling

Computational modeling is an established method in particle accelerator research. Modeling fulfills tasks from exploratory research for new particle accelerator concepts (including beam physics, model building, validation) to design and optimization of accelerator components. For these tasks, computational model choices need to strike a balance between running with high fidelity (e.g., for detailed physics studies) or high speed (e.g., for ensemble runs). Figure 1 provides a schematic overview about possible model choices in a simulations workflow.

In advanced particle accelerator modeling, both extremes of modeling choices often benefit from using leadership-scale supercomputers. High fidelity modeling is a typical capability computing task in high-performance computing (HPC), requiring large portions of leadership-scale supercomputers at the

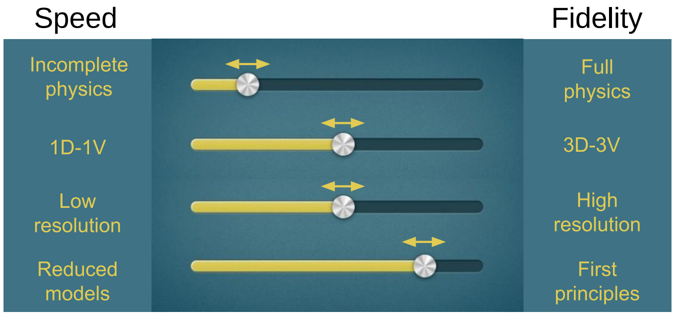

Fig. 1. Preference of speed (time-to-solution) or fidelity (accuracy) of results is determined by implemented algorithms and models in codes. Faster codes are preferred for design studies and optimization, more accurate codes for model building, validation and exploration.

same time to model a single simulation. On the other hand, fast modeling of ensembles is a typical capacity computing task in high-throughput computing (HTC).

As a consequence of these needs, the community developed more than one "type" of particle accelerator modeling code. Typical choices include fully-kinetic, electromagnetic particle-in-cell codes as the most accurate (and computationally expensive), electrostatic approximations, as well as (partially) fluid-based models. Most implemented methods these days are explicit algorithms, which iterate forward along either the independent variable of time t or the reference trajectory s of a beam. Implicit methods are more common in general plasma physics and their research is beneficial for applications with high accuracy requirements, e.g., for long-time stable modeling with high demands on energy conservation.

Scientific drivers for accelerator and beam physics research were recently described as "Grand Challenges" in the 2021 Snowmass HEP community exercise [10]: order-of-magnitude increases in beam intensity, beam quality and phase space density, beam control and beam prediction directly translate into significant needs for computational resolution (grids and no. of particles modeled) and predictive quality (simulate all the particles, conductors, dark currents, many turns, etc.). In advanced accelerator modeling, specific drivers that require Exascale-supercomputing resources include staged, wakefield-based acceleration for future particle colliders, compact (X)FEL sources, high-field physics experiments (QED) and novel, ultra-intense proton/ion sources. Closely related scientific domains that benefit from the same modeling capabilities are in high-energy density laser-plasma physics, fusion-energy science, extreme-field science, astrophysical plasmas, light-source modeling, and applications of compact acceleration sources to medical and industrial applications.

III. Advanced Particle Accelerator Modeling at Exascale

A. Exascale Computing in the US

The US DOE Exascale Computing Project (ECP) has set its goal to prepare scientists and computing facilities for supercomputers capable of 10^{18} double-precision floating point operations per second (1 ExaFlop/s). As part of ECP, developed scientific application software aims to reach a domain-specific figure-of-merit that is $50\times$ higher than at project start 7 years earlier.

Targeting the science case of staged wakefield acceleration, `WarpX`[1] [11]–[13] is developed in ECP as the successor to the successful `Warp` code [14]. As a particle-in-cell code, the figure-of-merit of `WarpX` in ECP is its parallel particle-and-cell update rate per second. Recently, `WarpX` completed the ECP goal and even reached a $500\times$ improvement to pre-ECP, due to significant improvements in both hardware and software [13].

B. First Runs on Exascale Machines Achieved

In 2022, the first reported Exascale machine was revealed at Oak Ridge National Lab, called Frontier.[2] Frontier is deployed as part of ECP and is powered by 9,408 compute nodes, each predominantly computing on four AMD MI250X GPUs per node, each with 2 Graphics Compute Dies. In July 2022, `WarpX` [13] ran for the first time on the full scale of this Exascale machine.[3]

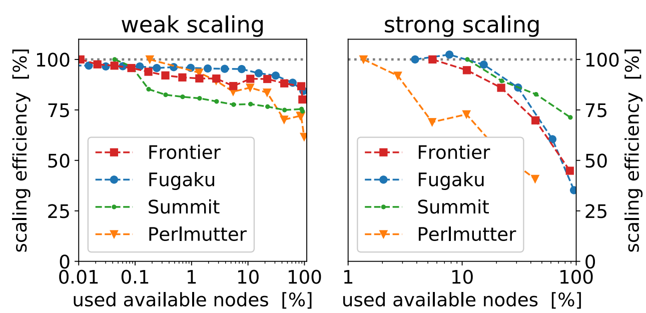

Fig. 2. Weak and strong scaling of WarpX relative to system size. Ideal scaling is the grey dotted line at 100 % in both graphs. Frontier and Perlmutter measurements were taken prior to their system acceptance dates. The network hardware of Perlmutter has since been updated from HPE Slingshot 10 to 11. Systems: Frontier (OLCF), Fugaku (Riken), Summit (OLCF), Perlmutter (NERSC). Published in [13].

Over the project years of ECP, `WarpX` has been developed with a performance-portable, single-source GPU-CPU programming model [16]. As a consequence, `WarpX` can run on traditional CPU machines and machines that use hardware accelerators, which are as of today GPUs by three different vendors (Nvidia, AMD, Intel).

Figure 2 from reference [13] shows the scalability of `WarpX` on some of the largest supercomputers available today. Highly relevant for the aforementioned challenges in accelerator modeling is the left plot, weak scaling, which shows the code performance when increasing both computational domain and provided parallel computing power on a supercomputer. `WarpX` achieves close to the ideal scaling over 4-5 orders

[1] Homepage: ecp-warpx.github.io

[2] https://www.olcf.ornl.gov/frontier/

[3] Also reported at AAC22: Few days later, the PIConGPU [15] team also measured full-scale OLCF Frontier performance results.

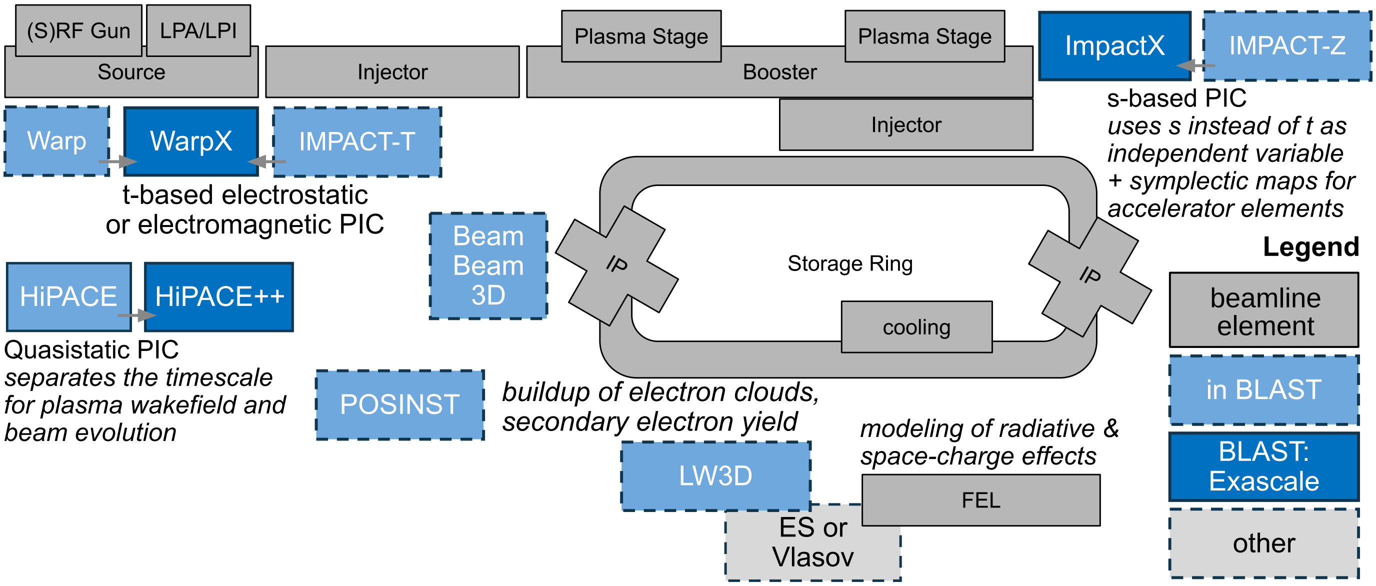

Fig. 3. Overview of a toy accelerator complex with beamline elements. As more BLAST codes are modernized for GPU-support, mesh-refinement and Exascale, projects evolve into the new Exascale code base [17]. Abbreviations: Interaction Point (IP), electrostatic (ES), Free Electron Laser (FEL).

of magnitude of system size increase. The plot to the right, strong scaling, shows the decrease in time-to-solution that can be gained by keeping the computational domain constant, but increasing provided parallel computing power. Due to communication needs in parallel machines, this plot cannot be arbitrarily scaled for complex applications. Nonetheless, `WarpX` still achieves a remarkable $> 50\,\%$ efficiency when scaled more than an order of magnitude, before being limited by data communication and GPU/CPU underutilization.

`WarpX` is developed fully in the open, following open science principles [18]–[21] and modern software engineering practices [22]. Besides running on supercomputers and cloud providers, `WarpX` can also be used on edge and personal computers with Linux, macOS and Windows operating system. The latter is particularly useful in designing simulation runs in lower dimension and resolution, as well as for development.

IV. Modernizing BLAST for Exascale

The `WarpX` code is, as was its predecessor `Warp`, part of the BLAST suite of codes.[4] BLAST, originally standing for "Berkeley Lab Accelerator Simulation Toolkit", was renamed in 2021 to "**B**eam, **Pla**sma & **A**ccelerator **S**imulation **T**oolkit". This reflects the grown community that spans maintainers, contributors and collaborators from many international institutions, and applies the BLAST codes to a growing number of applications. For instance, a code primarily developed outside of LBNL is the quasi-static `HiPACE++` code [23], which is maintained by DESY. Another code is `GEMPIX`, developed at IPP Garching for magnetic confinement fusion plasmas.

Leveraging the success and routines of `WarpX`, other existing BLAST codes are being transitioned to Exascale [17] as well. Figure 3 shows an overview of the new GPU-capable, mesh-refinement (MR)-enabled codes for accelerator modeling and prominent application. During this transition, existing codes are rewritten from Fortran to C++ for core compute routines and modern data structures. A common input layer is developed, which is standardized in the Python API `PICMI` (particle-in-cell modeling interface).[5] Time-based (t) implementations of the codes `Warp` and `IMPACT-T` [24] are combined in the new electromagnetic and electrostatic `WarpX` code. Beam-dynamics codes along a reference trajectory s as the independent variable, such as `IMPACT-Z` [25] and s-based `Warp` modules, are modernized in the new code `ImpactX` [17] and continue to include collective effects. Common particle-in-cell routines are shared via the Accelerated BLAST Recipes (`ABLASTR`) library, generalizing `WarpX` routines [17].

The Open Standard for Particle-Mesh Data (`openPMD`) is used in BLAST codes for data compatibility [26]. A high-performance, `openPMD` C++ and Python reference implementation, co-developed by LBNL and HZDR/CASUS, is used in massively parallel codes and in data analysis [27]. Recently, data streaming techniques were developed to transition traditional post-processing scripts to online, massively parallel data analysis workflows, which can be co-located with simulations and enable rapidly prototyping for scientific *in-transit* analysis [28].

V. Seeding a Community Ecosystem

From experience over the last years developing community standards such as `openPMD`, `PICMI`, the BLAST toolkit, or the `PIConGPU` software stack [15] - open standardization and modular, open source software collaborations emerge as the

[4] https://blast.lbl.gov

[5] https://picmi-standard.github.io

necessarily efficient way forward. From such seed projects, the US DOE SCIDAC-5 Collaboration for Advanced Modeling of Particle Accelerators (CAMPA) [29] will support advancement of numerical algorithms and accelerate community modeling codes in a cohesive manner: from beam source, over energy boost, transport, injection, storage, to application or interaction.

Making the existing particle accelerator modeling ecosystem compatible and interdependent will be beneficial towards the goal of predictive start-to-end modeling [30]. As a community, we should expect and ready models and codes for the exploration of "hybrid" accelerators, with conventional and advanced elements, as the next step for advanced accelerator modeling. Following open community standards, one can initiate an open ecosystem of codes that can be readily combined with each other and machine learning frameworks [17], towards enabling virtual test stands and twins of accelerators that can be used in operations.

Acknowledgment

This research used the open-source particle-in-cell code `WarpX` https://github.com/ECP-WarpX/WarpX, primarily funded by the US DOE Exascale Computing Project. Primary `WarpX` contributors are with LBNL, LLNL, CEA-LIDYL, SLAC, DESY, CERN, and TAE. We acknowledge all `WarpX`, `HiPACE++`, `ImpactX` and `openPMD` contributors. Slides of the plenary presentation are available under [31].

References

[1] A. J. Gonsalves, K. Nakamura, J. Daniels *et al.*, "Petawatt laser guiding and electron beam acceleration to 8 GeV in a laser-heated capillary discharge waveguide," *Phys. Rev. Lett.*, vol. 122, no. 8, p. 084801, Mar. 2019. DOI:10.1103/PhysRevLett.122.084801

[2] J. P. Couperus, R. Pausch, A. Köhler *et al.*, "Demonstration of a beam loaded nanocoulomb-class laser wakefield accelerator," *Nat. Commun.*, vol. 8, no. 1, p. 487, Sep. 2017. DOI:10.1038/s41467-017-00592-7

[3] W. Wang, K. Feng, L. Ke *et al.*, "Free-electron lasing at 27 nanometres based on a laser wakefield accelerator," *Nature*, vol. 595, no. 7868, pp. 516–520, Jul. 2021. DOI:10.1038/s41586-021-03678-x

[4] M. Galletti, D. Alesini, M. P. Anania *et al.*, "Stable operation of a free-electron laser driven by a plasma accelerator," *Phys. Rev. Lett.*, vol. 129, p. 234801, Nov 2022. DOI:10.1103/PhysRevLett.129.234801

[5] M. Labat, J. C. Cabadağ, A. Ghaith *et al.*, "Seeded free-electron laser driven by a compact laser plasma accelerator," *Nat. Photonics*, vol. 17, no. 2, pp. 150–156, Feb. 2023. DOI:10.1038/s41566-022-01104-w

[6] P. Hilz, T. M. Ostermayr, A. Huebl *et al.*, "Isolated proton bunch acceleration by a petawatt laser pulse," *Nat. Commun.*, vol. 9, no. 1, Dec. 2018. DOI:10.1038/s41467-017-02663-1

[7] F. Kroll, F.-E. Brack, C. Bernert *et al.*, "Tumour irradiation in mice with a laser-accelerated proton beam," *Nat. Phys.*, vol. 18, no. 3, pp. 316–322, Mar. 2022. DOI:10.1038/s41567-022-01520-3

[8] J. Bin, L. Obst-Huebl, J.-H. Mao *et al.*, "A new platform for ultra-high dose rate radiobiological research using the BELLA PW laser proton beamline," *Sci. Rep.*, vol. 12, no. 1, p. 1484, Jan. 2022. DOI:10.1038/s41598-022-05181-3

[9] L. D. Geulig, L. Obst-Huebl, K. Nakamura *et al.*, "Online charge measurement for petawatt laser-driven ion acceleration," *Review of Scientific Instruments*, vol. 93, no. 10, p. 103301, 2022. DOI:10.1063/5.0096423

[10] S. Nagaitsev, Z. Huang, J. Power *et al.*, "Accelerator and beam physics research goals and opportunities," 2021. DOI:10.48550/ARXIV.2101.04107

[11] J.-L. Vay, A. Almgren, J. Bell *et al.*, "Warp-X: A new exascale computing platform for beam–plasma simulations," *Nucl. Instrum. Methods Phys. Res. A*, vol. 909, pp. 476–479, Nov. 2018. DOI:10.1016/j.nima.2018.01.035

[12] J.-L. Vay, A. Huebl, A. Almgren *et al.*, "Modeling of a chain of three plasma accelerator stages with the WarpX electromagnetic PIC code on GPUs," *Phys. Plasmas*, vol. 28, no. 2, p. 023105, Feb. 2021. DOI:10.1063/5.0028512

[13] L. Fedeli, A. Huebl, F. Boillod-Cerneux *et al.*, "Pushing the frontier in the design of laser-based electron accelerators with groundbreaking mesh-refined particle-in-cell simulations on exascale-class supercomputers," in *2022 SC22: International Conference for High Performance Computing, Networking, Storage and Analysis (SC) (SC).* Los Alamitos, CA, USA: IEEE Computer Society, nov 2022, pp. 25–36. DOI:10.1109/SC41404.2022.00008

[14] J.-L. Vay, D. P. Grote, R. H. Cohen, and A. Friedman, "Novel methods in the particle-in-cell accelerator code-framework warp," *Computational Science & Discovery*, vol. 5, no. 1, p. 014019, dec 2012. DOI:10.1088/1749-4699/5/1/014019

[15] A. Huebl, "PIConGPU: Predictive Simulations of Laser-Particle Accelerators with Manycore Hardware," Ph.D. dissertation, Technische Universitaet Dresden, Jul. 2019. DOI:10.5281/zenodo.3266820

[16] A. Myers, A. Almgren, L. D. Amorim *et al.*, "Porting WarpX to GPU-accelerated platforms," *Parallel Comput.*, vol. 108, no. 102833, p. 102833, Dec. 2021. DOI:10.1016/j.parco.2021.102833

[17] A. Huebl, R. Lehé, C. Mitchell *et al.*, "Next generation computational tools for the modeling and design of particle accelerators at exascale," in *2022 North American Particle Accelerator Conference (NAPAC22).* JACoW Publishing, Geneva, Switzerland, Dec. 2022, invited Oral, TUYE2. DOI:10.18429/JACoW-NAPAC2022-TUYE2

[18] United Nations Educational, Scientific and Cultural Organization (UNESCO), "Open Science." https://www.unesco.org/en/open-science

[19] FOSTER Plus (European Union funded open project in Horizon 2020 and beyond), "Open Science Definition." https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

[20] Open Source Initiative, "The Open Source Definition," 2004. https://opensource.org/docs/osd

[21] Free Software Foundation, "The Free Software Definition," 1990. https://www.gnu.org/philosophy/free-sw.en.html

[22] R. Lehe, A. Huebl, J.-L. Vay *et al.*, "Embracing modern software tools and user-friendly practices, when distributing scientific codes," *Snowmass21 LOI*, 2020. https://www.snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF2_CompF0_Lehe-076.pdf

[23] S. Diederichs, C. Benedetti, A. Huebl *et al.*, "Hipace++: A portable, 3d quasi-static particle-in-cell code," *Computer Physics Communications*, vol. 278, p. 108421, 2022. DOI:10.1016/j.cpc.2022.108421

[24] J. Qiang, S. Lidia, R. D. Ryne, and C. Limborg-Deprey, "Three-dimensional quasistatic model for high brightness beam dynamics simulation," *Phys. Rev. ST Accel. Beams*, vol. 9, p. 044204, Apr 2006. DOI:10.1103/PhysRevSTAB.9.044204

[25] J. Qiang, R. Ryne, S. Habib, and V. Decyk, "An Object-Oriented Parallel Particle-in-Cell Code for Beam Dynamics Simulation in Linear Accelerators," *J. Comput. Phys.*, vol. 163, pp. 434–451, Jul. 2000. DOI:10.1006/jcph.2000.6570

[26] A. Huebl, R. Lehe, J.-L. Vay *et al.*, "openPMD: A meta data standard for particle and mesh based data," https://github.com/openPMD, 2015. https://www.openPMD.org. DOI:10.5281/zenodo.591699

[27] 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, 2018. https://openpmd-api.readthedocs.io. DOI:10.14278/rodare.27

[28] F. Poeschel, J. E, W. F. Godoy *et al.*, "Transitioning from file-based HPC workflows to streaming data pipelines with openPMD and ADIOS2," in *Driving Scientific and Engineering Discoveries Through the Integration of Experiment, Big Data, and Modeling and Simulation*, J. Nichols, A. B. Maccabe, J. Nutaro *et al.*, Eds. Cham: Springer International Publishing, 2022, pp. 99–118. DOI:10.1007/978-3-030-96498-6_6

[29] "Consortium for Advanced Modeling of Particle Accelerators (CAMPA)." https://campa.lbl.gov

[30] J.-L. Vay, A. Huebl, D. Sagan *et al.*, "A modular community ecosystem for multiphysics particle accelerator modeling and design," *Snowmass21 LOI*, 2020. https://snowmass21.org/docs/files/summaries/CompF/SNOWMASS21-CompF2_CompF0-AF1_AF0_Vay-070.pdf

[31] A. Huebl *et al.*, "From Compact Plasma Particle Sources to Advanced Accelerators with Modeling at Exascale," Nov. 2022, invited plenary, presentation slides. DOI:10.5281/zenodo.7697402