HiPACE++: GPU-accelerated modeling of plasma wakefield accelerators

S. Diederichs,[1,2,3] * C. Benedetti,[2] A. Huebl,[2] R. Lehe,[2] A. Myers,[2] A. Sinn,[1] J.-L. Vay,[2] W. Zhang,[2] and M. Thévenet[1]*

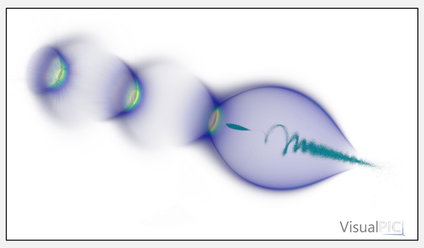
[1]Deutsches Elektronen-Synchrotron DESY, Notkestraße 85, 22607 Hamburg, Germany
[2]Lawrence Berkeley National Laboratory, 1 Cyclotron Rd, Berkeley, California 94720, USA
[3]University of Hamburg, Institute of Experimental Physics, Luruper Chaussee 149, 22607 Hamburg, Germany

* Contacts: severin.diederichs@desy.de
maxence.thevenet@desy.de

Motivation

Quasi-static Particle-in-Cell

- Modeling plasma-based accelerators can be notoriously challenging
- Particle-in-Cell (PIC) codes are the method of choice, but can be computationally expensive
- Quasi-static (QS) PIC is ideal for modeling plasma accelerators due to its large time steps ($\omega_p \gg \omega_\beta$)



Graphics Processing Units (GPUs)

- GPUs are **highly performant**
  → 14 of the top 20 of the Top500[1] are GPU-based
- GPUs are **energy-efficient** and reduce environmental impact
  → 19 of the top 20 of the Green500[1] are GPU-based
- With thousands of cores operating on a fast shared memory pool, GPUs are inherently parallel
  → ideally suited for PIC

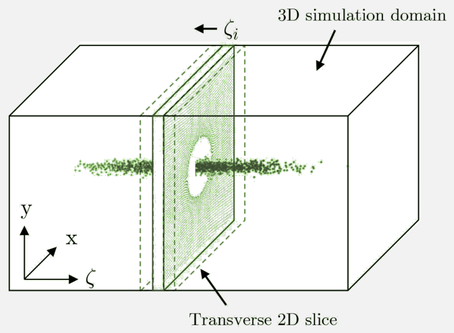
Top 20 of the Green500[1]

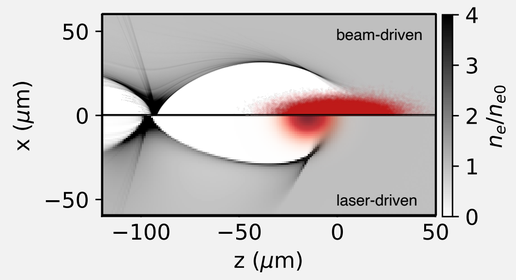
Rank	Name	Hardware
1	Frontier TDS	AMD Instinct MI250X
2	Frontier	AMD Instinct MI250X
3	LUMI	AMD Instinct MI250X
4	Adastra	AMD Instinct MI250X
5	ATOS THX.A.B	NVIDIA A100
6	MN-3	Xeon Platinum 8260M
7	SSC-21 Scalable Module	NVIDIA A100
8	Tethys	NVIDIA A100
9	Wilkes-3	NVIDIA A100
10	Athena	NVIDIA A100
11	Phoenix - 2022	NVIDIA A100
12	HiPerGator AI	NVIDIA A100
13	Snellius Phase 1 GPU	NVIDIA A100
14	Perlmutter	NVIDIA A100
15	Karolina, GPU partition	NVIDIA A100
16	MeluXina - Accelerator Module	NVIDIA A100
17	Alex	NVIDIA A100
18	NVIDIA DGX SuperPOD	NVIDIA A100
19	JUWELS Booster Module	NVIDIA A100
20	JURECA Data Centric Module	NVIDIA A100

Quasi-static PIC suitable for GPU computing

- QS PIC calculates a 3D problem from head to tail in n_z 2D slices
- All beam and laser operations can be integrated in the loop over slices
- A modest amount of data (beams, plasma, and field slices) is stored on GPU memory

  → high-resolution simulations fit on a single GPU
  → suitable for GPU-equipped laptops and supercomputers





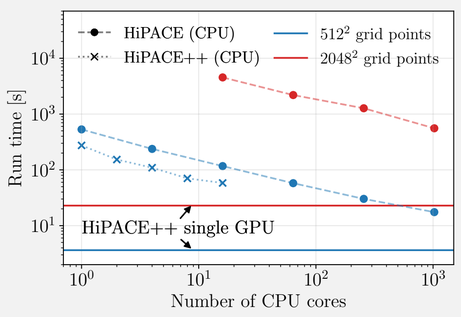
GPU-accelerated, quasi-static PIC code HiPACE++ provides high-resolution 3D-simulations with ease

- HiPACE++ is the first portable QS PIC code, optimized for GPU
  Diederichs et al., CPC 278, 108421 (2022)
- **Open-source**, join the community![3]
- Based on the AMReX library[4], providing portability (currently to NVIDIA and AMD GPUs)
- Uses common openPMD standard[5] for I/O

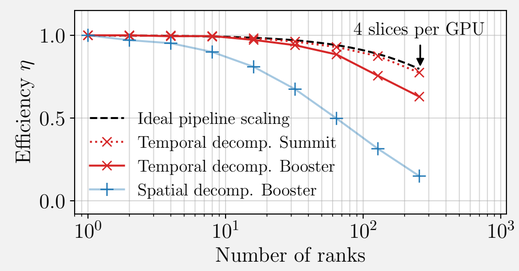
Features:
field ionization (ADK), normalized and SI units, two field solvers, temperature effects, static mesh refinement, laser envelope solver (soon)

SCAN ME

1. Production runs **10x faster** and **1000x lower costs** on NVIDIA A100 GPU than transversely-parallelized CPU implementation[2]



2. GPU-optimized pipeline enables **strong scaling to hundreds of GPUs** on the JUWELS Booster[2]

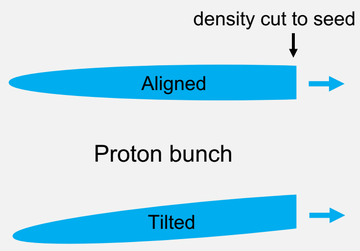


3. GPUs estimated to be **~10x more energy-efficient** in a node-to-node comparison with CPUs only

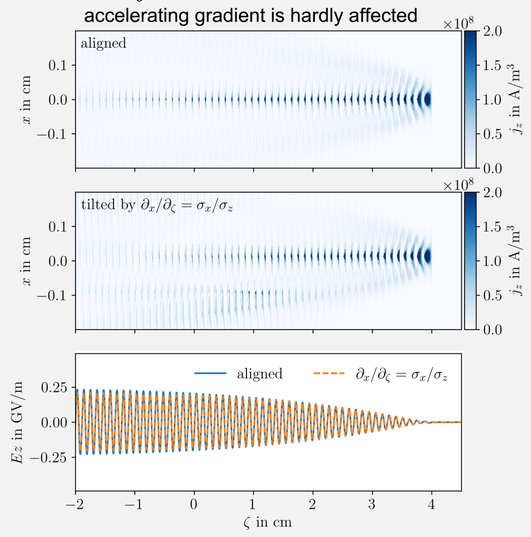
Challenging physics scenarios become feasible and affordable

Physics case 1: AWAKE

- AWAKE[6] uses a long, self-modulating (SM) proton bunch
- Tilted proton bunch may induce hosing instability



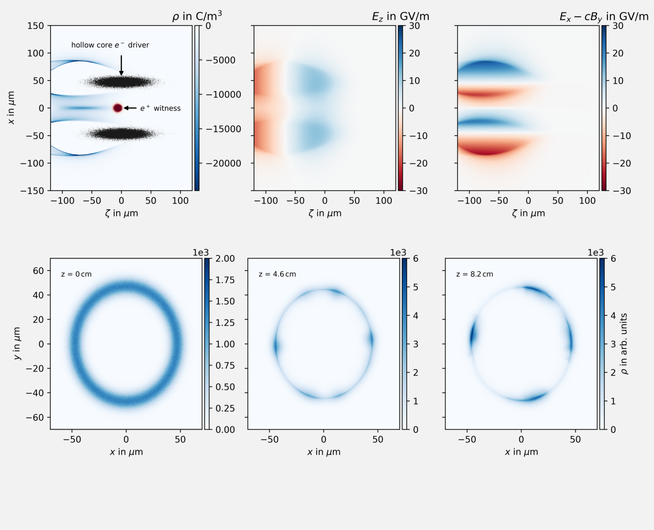
After 10 m propagation in a plasma with $n_0 = 9{\times}10^{14}\ cm^{-3}$ accelerating gradient is hardly affected



- **Costs**: **1 node hour** on the JUWELS Booster (NVIDIA A100) using 16 nodes, in double precision, $512{\times}512{\times}2048$ grid points, $120{\times}10^6$ beam particles, 400 time steps

Physics case 2: positron acceleration with hollow-core drive beam[7,8]

- 2015: full 3D simulations were too expensive[7]
- Today: 3D simulations with HiPACE++ reveal Weibel instability of the drive beam[9]



- **Costs**: **1 hour** on a laptop (NVIDIA RTX2070) in single precision, $1024{\times}1024{\times}1024$ grid points, 10^7 beam particles, 300 time steps

Conclusion

- GPUs are both powerful and environmental-friendly – use them!
- Quasi-static PIC is a prime candidate for GPU computing due to its low memory requirement
- HiPACE++ is open-source tool and ideally suited for modeling most of the challenging 3D wakefield scenarios

References

[1] https://www.top500.org/lists/green500/list/2022/06/
[2] Diederichs et al., Comput. Phys. Commun. 278, 108421 (2022)
[3] https://github.com/Hi-PACE/hipace/
[4] Zhang et al., J. Open Sour. Softw. 4 (37) 1370 (2019)
[5] Huebl et al., DOI:10.5281/zenodo.591699 (2015)
[6] E. Gschwendtner, Nucl. Instrum. Methods Phys. Res., Sect. A 829, 76 (2016)
[7] Jain et al., Phys. Rev. Lett. 115, 1950001 (2015)
[8] Hue et al., Phys. Rev. Research 3, 043063 (2021)
[9] Su et al., IEEE Trans. Plasma Sci. 15, 192 (1987)

Acknowledgements

This poster presentation has received support from the European Union's Horizon 2020 Research and Innovation program under Grant Agreement No 101004730. This work was supported by the Director, Office of Science, Office of High Energy Physics, of the U.S. Department of Energy, under Contract No. DE-AC02-05CH11231 and used the computational facilities at the National Energy Research Scientific Computing Center (NERSC). We gratefully acknowledge the Gauss Centre for Supercomputing e.V. (www.gauss-centre.eu) for funding this project by providing computing time through the John von Neumann Institute for Computing (NIC) on the GCS Supercomputer JUWELS at Jülich Supercomputing Centre (JSC). This research was supported in part through the Maxwell computational resources operated at Deutsches Elektronen-Synchrotron DESY, Hamburg, Germany. We acknowledge the Funding by the Helmholtz Matter and Technologies Accelerator Research and Development Program.