On the road toward real-time virtual prototyping of particle accelerators

Jean-Luc Vay


March 26, 2016
Particle accelerators are essential tools in modern life that power scientific discovery, cure cancer, secure our borders, and help create a wide range of products.

**Medicine**
- ~9000 medical accelerators in operation worldwide
- 10’s of millions of patients treated/yr
- 50 medical isotopes, routinely produced with accelerators

**Industry**
- ~20,000 industrial accelerators in use
  - Semiconductor manufacturing
  - Cross-linking/polymerization
  - Sterilization/irradiation
  - Welding/cutting
- Annual value of all products that use accel. Tech.: $500B

**National Security**
- Cargo scanning
- Active interrogation
- Stockpile stewardship: materials characterization, radiography, support of non-proliferation

**Discovery Science**
- ~30% of Nobel Prizes in Physics since 1939 enabled by accelerators
- 4 of last 14 Nobel Prizes in Chemistry for research utilizing accelerator facilities

There are 30,000 Particle Accelerators Making an Impact on Our Lives
Problem: size & cost often a limiting factor

Example 1: Proton Therapy Center

New Rochester Mayo Clinic
Proton Therapy Center

- 4 chambers
- $188M

120-ton gantry directs proton beam
to appropriate spot on patient
by rotating around a three-story chamber.

http://finance-commerce.com/2014/03/status-report-mayo-proton-therapy-facility/#ixzz43DJgnIlA
http://blogs.mprnews.org/statewide/2014/03/mayos-proton-beam-facility-on-track-for-2015-opening/
Problem: size & cost often a limiting factor

Example 2: Carbon Therapy Center

Heidelberg Proton & Carbon Therapy Center
• 2 scans chambers
• one $4\pi$ chamber
• €119M

http://medicalphysicsweb.org/cws/article/research/51684
https://www.klinikum.uni-heidelberg.de/About-us.124447.0.html?&L=1

670-ton gantry
~1/10 Eiffel Tower
Problem: size & cost often a limiting factor

Example 3: High-Energy Physics collider

CERN LHC

Cost: $10B
Cons.: 150MW

CLIC

Cost: $?B
Cons.: 415MW

Future colliders?

ILC

Cost: $8B-$20B?
Cons.: 230MW

FCC

Cost: $?B – Cons.: ?MW
CERN’s next director-general on the LHC and her hopes for international particle physics

Fabiola Gianotti talks to *Nature* ahead of taking the helm at Europe’s particle-physics laboratory on 1 January.

Elizabeth Gibney

22 December 2015

Some people think that future governments will be unwilling to fund larger and more expensive facilities. Do you think a collider bigger than the LHC will ever be built? And will it depend on the LHC finding something new?

The outstanding questions in physics are important and complex and difficult, and they require the deployment of all the approaches the discipline has developed, from high-energy colliders to precision experiments and cosmic surveys. High-energy accelerators have been our most powerful tools of exploration in particle physics, so we cannot abandon them. What we have to do is push the research and development in accelerator technology, so that we will be able to reach higher energy with compact accelerators.
Most applications of accelerators make use of conventional technology (with roots in post-WWII era)

This technology has been refined and extended...

1956: Stanford Medical Center
1940: Cockcroft Walton generator Cavendish Lab
Today: Varian Medical
Today: Superconducting resonators

Today, we are on the brink of technological breakthroughs...

Two-beam acceleration
— Power transfer from one beam to another with RF structures

Dielectric Wakefield Accelerator
— Strong fields induced in dielectric structure

Beam-Plasma Wakefield Accelerator
— Beam generated plasma wake with very high fields

Laser-Plasma Accelerator
— Laser generated plasma wake with very high fields
Laser plasma acceleration enables development of compact accelerators

State-of-the-art

Multi-GeV electron beam from 9 cm laser plasma accelerator
Let’s go from this…

...to this...

...and make it ready for application.

Can this technology be developed for light sources, medical and security applications, and even high energy colliders?

*High-performance computing is key to answering these questions and reaching that goal!*
Modeling of particle accelerators
All accelerators in the world rely on modeling

- CERN (HL-)LHC
- FNAL PIP(-II/III)
- SLAC LCLS-(II)
- LBNL (k-)BELLA
- SLAC FACET(-II)
- LBNL ALS(-U)
Advanced simulations have key impact on the support & analysis of experiments

Start-to-end simulations of LCLS match measurements

Supported world record BELLA 4.25 GeV beam over 9 cm

\[ \text{LH } \sigma_E = 9.1 \text{ keV} \]
\[ \text{LH } \sigma_E = 19 \text{ keV} \]

Simulations with IMPACT

Advanced simulations also key to the design of future accelerators

Parallel genetic optimization improves design of ALS upgrade.

3 Objectives:

- Horizontal emittance
- Momentum aperture
- Total diffusion rate

Predict amplification of instability in LCLS-II.

and validate novel mitigation scheme

Simulations with IMPACT

M. Venturini and J. Qiang, PRST-AB 18, 054401 (2015).

Advanced simulations are essential tools for the exploration of new concepts

Demonstrate concept of two-color injection of ultra-high quality beam

Next generation of accelerators needs next generation of modeling tools

Our vision

Fast – runs in seconds to minutes

Hi-Fi – full & accurate physics

Link – integrated ecosystem

Real-time virtual prototyping of entire accelerator

with intuitive interface, dissemination & user support.

Simulations take too long!

- X-FEL start-to-end: 6 Hrs
- 2-color injection: 3 days
- Beam-beam LHC: 1 day
- BELLA: 7 days

Combine best algorithms

- Speed
  - 2025
- Port codes to fastest hardware
  - NERSC-8 Cori
  - 2016

Need to speedup by $\times 10^n$
## Investment in new algorithms pushes the frontier

<table>
<thead>
<tr>
<th>Algorithm/method</th>
<th>Reference</th>
<th>Originated</th>
<th>Adopted by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Integrated Maps for rf cavity dynamics</td>
<td>Ryne, LANL Report 1995</td>
<td>ML/IMPACT</td>
<td>D. Abell nonlinear model</td>
</tr>
<tr>
<td>Stochastic Leap-Frog for Brownian motion</td>
<td>Qiang &amp; Habib, PRE 2000</td>
<td>IMPACT</td>
<td></td>
</tr>
<tr>
<td>Spectral-finite difference multigrid solver</td>
<td>Qiang &amp; Ryne, CPC 2001/2006</td>
<td>IMPACT</td>
<td></td>
</tr>
<tr>
<td>Improved Perfectly Matched Layers</td>
<td>Vay, JCP 2000/JCP 2002</td>
<td>Warp</td>
<td>Osiris</td>
</tr>
<tr>
<td>AMR-PIC electrostatic</td>
<td>Vay et al, LPB 2002/PoP 2004</td>
<td>Warp</td>
<td></td>
</tr>
<tr>
<td>PIC w/ shift-Green function method</td>
<td>Qiang et al, PRSTAB 2002/CPC 2004</td>
<td>BBeam3D</td>
<td></td>
</tr>
<tr>
<td>Secondary emission of electrons algorithm</td>
<td>Furman &amp; Pivi, PRST-AB 2003</td>
<td>Posinst</td>
<td>TxPhysics, Warp, spacecraft charging codes</td>
</tr>
<tr>
<td>AMR-PIC electromagnetic</td>
<td>Vay et al, CPC 2004</td>
<td>Emi2D</td>
<td>Warp</td>
</tr>
<tr>
<td>3D Poisson solver with large aspect ratio</td>
<td>Qiang &amp; Gluckstern, CPC 2004</td>
<td>IMPACT</td>
<td></td>
</tr>
<tr>
<td>PIC w/ integrated Green function</td>
<td>Qiang et al, PRSTAB 2006</td>
<td>ML/IMPACT</td>
<td>BB3D, Pyheadtail, Opal</td>
</tr>
<tr>
<td>Hybrid Lorentz particle pusher</td>
<td>Cohen et al, NIMA 2007</td>
<td>Warp</td>
<td></td>
</tr>
<tr>
<td>Lorentz boosted frame</td>
<td>Vay, PRL 2007</td>
<td>Warp</td>
<td>Osiris, Vorpal, PIConGPU, INF&amp;RNO, JPIC, ...</td>
</tr>
</tbody>
</table>
and it also benefits the entire community

<table>
<thead>
<tr>
<th>Algorithm/method (cont.)</th>
<th>Reference</th>
<th>Originated</th>
<th>Adopted by</th>
</tr>
</thead>
<tbody>
<tr>
<td>Explicit Lorentz invariant particle pusher</td>
<td>Vay, <em>PoP</em> 2008</td>
<td>Warp</td>
<td>Tristan, Osiris, PIConGPU, Photon-Plasma, QED, etc.</td>
</tr>
<tr>
<td>New convolution integral w/ smooth kernel</td>
<td>Qiang, <em>CPC</em> 2010</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>Mixed Particle-Field decomposition method</td>
<td>Qiang &amp; Li, <em>CPC</em> 2010</td>
<td>BBBeam3D</td>
<td></td>
</tr>
<tr>
<td>PIC with tunable electromagnetic solver</td>
<td>Vay <em>et al</em>, <em>JCP</em> 2011</td>
<td>Warp</td>
<td>Osiris, Vorpal</td>
</tr>
<tr>
<td>Efficient digital filter for PIC</td>
<td>Vay <em>et al</em>, <em>JCP</em> 2011</td>
<td>Warp</td>
<td>Osiris, Vorpal</td>
</tr>
<tr>
<td>Laser launcher from moving antenna</td>
<td>Vay <em>et al</em>, <em>PoP</em> 2011</td>
<td>Warp</td>
<td>Osiris, Vorpal</td>
</tr>
<tr>
<td>High-precision laser envelope model</td>
<td>Benedetti <em>et al</em>, 2011</td>
<td>Inf&amp;rno</td>
<td></td>
</tr>
<tr>
<td>Domain decomposition for EM spectral solver</td>
<td>Vay <em>et al</em>, <em>JCP</em> 2013</td>
<td>Warp</td>
<td></td>
</tr>
<tr>
<td>Spectral solver with azimuthal decomposition</td>
<td>Lehe <em>et al</em>, <em>CPC</em>, <em>in press</em> 2016</td>
<td>FBPIC/Warp</td>
<td></td>
</tr>
<tr>
<td>Novel algorithm for vectorization</td>
<td>Vincenti <em>et al</em>, <em>submitted</em> 2016</td>
<td>Warp/PICSAR</td>
<td></td>
</tr>
<tr>
<td>Generalized spectral solver</td>
<td>Vay <em>et al</em>, <em>in preparation</em> 2016</td>
<td>Warp</td>
<td></td>
</tr>
</tbody>
</table>

*Latest with more in preparation*
It also has relevance beyond original application.

Secondary e- yield (SEY) model (Furman-Pivi, PRST-AB 2003)

Prediction of spacecraft charging from charged particles impact.

Lorentz boosted frame (Vay, PRL 2007)

Speed-up first principles simulations $x \ 10^n$.

Relativistic particle pusher (Vay, PoP 2008)

More precision integration of relativistic particle motion.
We are now combining algorithms for maximum efficiency.

**Lower # time steps:**
- optimal Lorentz boosted frame

**Higher accuracy:**
- Lorentz invariant particle pusher
- Pseudo-spectral Maxwell solvers
- AMR

**Lower dimensionality**
- FFT+Hankel Transform Maxwell solver for quasi-RZ geom

**Higher scalability**
- FFT Maxwell solvers with domain decomposition

**Higher stability**
- Analysis & mitigation of Numerical Cherenkov Instability
Preparing Accelerator Codes for Exascale Supercomputing
Our proposal was selected as part of the NERSC Exascale Applications Program (NESAP)

NESAP Codes

<table>
<thead>
<tr>
<th>Advanced Scientific Computing Research</th>
<th>Basic Energy Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Almgren (LBNL)</td>
<td>BoxLib</td>
</tr>
<tr>
<td>AMR Framework</td>
<td>Chombo-crunch</td>
</tr>
<tr>
<td>Trebotich (LBNL)</td>
<td>Chombo-crunch</td>
</tr>
<tr>
<td>Vay (LBNL)</td>
<td>WARP &amp; IMPACT</td>
</tr>
<tr>
<td>IMPACT</td>
<td>IMPACT</td>
</tr>
<tr>
<td>Toussaint (Arizona)</td>
<td>MILC</td>
</tr>
<tr>
<td>Habib (ANL)</td>
<td>HACC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nuclear Physics</th>
<th>Biological and Environmental Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maris (Iowa St.)</td>
<td>MFDn</td>
</tr>
<tr>
<td>Joo (JLAB)</td>
<td>Chroma</td>
</tr>
<tr>
<td>Christ/Karsch (Columbia/BNL)</td>
<td>DWF/HISQ</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fusion Energy Sciences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jardin (PPPL)</td>
</tr>
<tr>
<td>Chang (PPPL)</td>
</tr>
</tbody>
</table>

Mathieu Lobet
NESAP postdoc
Started on 2/1/16

Courtesy R. Gerber, NUGEX 2016
LBNL home of unique simulation toolset
for conventional & advanced concepts accelerators

State-of-the-art codes:
• WARP, IMPACT, BEAMBEAM3D, INF&RNO, POSINST, FBPIC.

For detailed modeling of the largest set of physics and components:
• beams, plasmas, lasers, structures, etc in linacs, rings, injectors, traps, ...

Supporting accelerator modeling:
• across DOE (HEP, BES, NP, FES, DNN) and beyond (CERN, DESY, KEK, ...).

All codes share Particle-In-Cell loop at core
(as do other accelerator & plasma codes elsewhere)
Strategy relies then on small kernel

- NESAP work should benefit other BLAST codes, and beyond (OSIRIS, Synergia, ...)

- Warp is a complex code:
  - ~150k lines FORTRAN + ~100k lines Python

Particle-In-Cell Scalable Applications Resources (PICSAR)

- initiated with kernel of Warp’s Particle-In-Cell main loop (all FORTRAN)
  - optimization independently of Warp complexity and legacy
- open source repository for collaborative development and distribution with other codes/groups
  - licensing is underway
  (derived class of PIC solver enables usage of new optimized routines from Warp)
Novel pre-exascale supercomputers require restructuration with “multi-level parallelism”

To run effectively on future systems

- **Manage Domain Parallelism**
  - independent program units; explicit

- **Increase Thread Parallelism**
  - independent execution units within the program; generally explicit

- **Exploit Data Parallelism**
  - Same operation on multiple elements

- **Improve data locality**
  - Cache blocking;
  - Use on-package memory

---

**Particle-In-Cell**

- Domain decomposition
  - 1 MPI task/domain

- Subdomain tiling
  - 1 OpenMP thread/tile

---

```
| --> DO I = 1, N
|    R(I) = B(I) + A(I)
| --> ENDDO
```
Initial OpenMP scheme gave poor scaling

One MPI SUBDOMAIN

One big loop with one thread/tile

- Two adjacent tiles may write to the same location (conflicts)
- Adding $OMP REDUCTION clause not sufficient
- Poorly scales with # OpenMP threads:
  5x on Edison on one socket (12cores)
Novel OpenMP scheme to improve scaling

Main loop split in 4:

- Loop 1: Each thread writes on its center part
- Loop 2: Each thread writes in +/-X direction
- Loop 3: Each thread writes in +/-Y direction
- Loop 4: Each thread writes in +/-Z direction

OpenMP scaling: 5x -> 11x! (on Edison)
Novel vectorization algorithm leads to >2x speedup on charge/current deposition routines

• Previous algorithms developed on vector supercomputers in 70s-90s do not work

• Novel algorithm developed* (implemented in Warp/PICSAR)

Tests demonstrate speedups >2x (max. theoretical=4)

Load imbalance develops during simulations

- Distribution of particles, initially regular, becomes very irregular
- E.g.: laser-driven ion simulation with 12 cores

\[
\text{Imbalance (\%) = } 100 \times \frac{\text{MAXTIME(Pi)} - \text{MINTIME(Pi)}}{\text{MINTIME(Pi)}}
\]

Irregularity also affects balance between tiles

needs for intra- & inter-node dynamic load balancing (DLB)
• DLB implemented using $OMP$ SCHEDULE(runtime) with runtime="guided" or "dynamic"

• Initial testing show better results with "dynamic"
Internode DLB by moving domain boundaries

Simple strategy:
(i) Project work load (field solvers+particle routines) along X,Y,Z
(ii) Compute new CPU boundaries along X,Y,Z
(iii) Exchange particles and fields between CPUs

Testing of efficiency on large simulations is underway
More in talk by A. Bhagatwala tomorrow
I/O, data analysis and visualization
Large scale HPC simulations need efficient I/O, analysis and visualization

- I/O, analysis and visualization can be bottleneck

- collaborations on HDF5 and ADIOS parallel I/O

- collaboration on new data layout specifications for PIC
  - https://github.com/openPMD
  - http://www.openpmd.org

- collaboration on in situ data analysis and visualization
  - https://bitbucket.org/berkeleyleab/warpiv

Presentation by B. Loring on Monday
In situ analysis & viz. complements raw data dumps

- **Dumps of particles and fields are needed for post-processing**
  - but limited by parallel I/O efficiency & amount of data generated

- **In situ analysis** enables increased frequency
  - down-samples, derived quantities, reduced geometry etc
  - focused analysis, write exactly what you (think you) need

- **In situ visualization** enables maximum temporal resolution and I/O reduction of multi-dimensional datasets
  - reduces problem sized data to image sized data
  - fast, small, serial I/O, doesn’t impact file system.
simple idea: use a laser to create a plasma. charge separation accelerates protons

- active area of research with many applications!
  - including cancer therapy
    - precise control over depth at which energy is deposited
    - less tissue damage
    - relatively compact and cheap

- HPC simulations are essential
  - used to develop and prove theory
  - when mature, will be used to model experimental apparatus

Example: laser-driven ion beam acceleration

- active area of research with many applications!
  - including cancer therapy
    - precise control over depth at which energy is deposited
      ➔ less tissue damage
    - relatively compact and cheap

- HPC simulations are essential
  - used to develop and prove theory
  - when mature, will be used to model experimental apparatus

**simple idea:** use a laser to create a plasma. Charge separation accelerates protons

3D simulations are needed to capture correct physics

- 2D is qualitatively similar, but in 3D:
  - higher energies are reached
  - the beam propagates faster and further over the same time period.
- In-situ data analysis & visualization enables high frequency projections & histories of derived quantities
3D simulations are needed to capture correct physics

- **2D** is qualitatively similar, but in **3D**:
  - higher energies are reached
  - the beam propagates faster and further over the same time period.

- **In-situ data analysis & visualization** enables high frequency projections & histories of derived quantities
In situ enables high-frequency visualization of isosurfaces

**Particle Density**

**Kinetic Energy**
In situ analysis and visualization: small portion of runtime and small I/O footprint

- Raw data for same analysis ~ **3.4 TB**.
- In situ $\Rightarrow 4033\times$ reduction in I/O.
Summary

• Particle accelerators are essential tools of science, medicine, industry and security
• Plasma-based methods on the brink to deliver much smaller, cheaper accelerators, with profound impacts
• HPC already essential and can play key role in the development of new technologies
• Ultimate goal is real-time virtual prototyping of entire accelerators
• Preparation toward exascale is vital and underway (NESAP)
• Novel in situ analysis and visualization tools will complement standard parallel I/O for maximum utilization
NERSC is an essential partner of our research program and we look forward to our continued partnership and future co-discoveries!