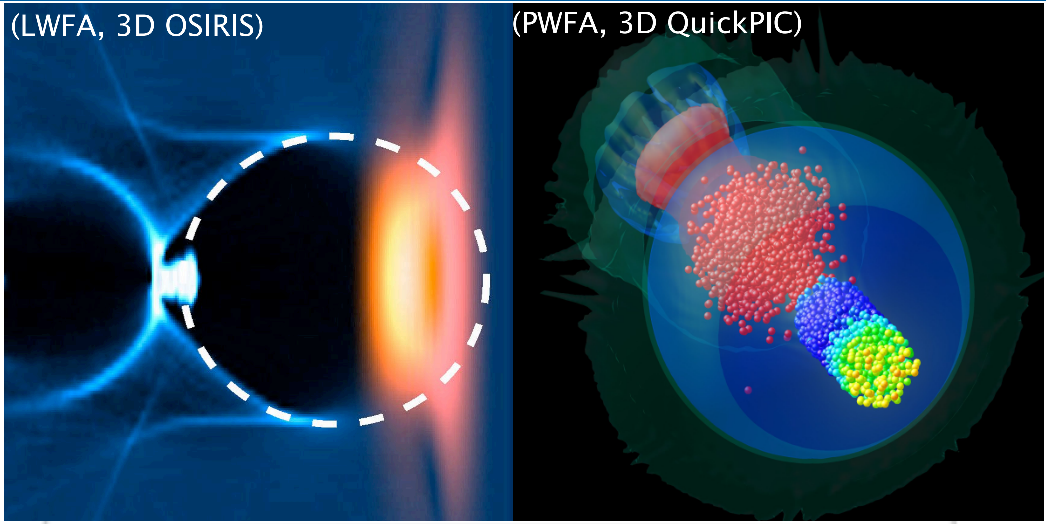
Project Summary -- Simulations of Plasma Based Accelerator Experiments Around the World.

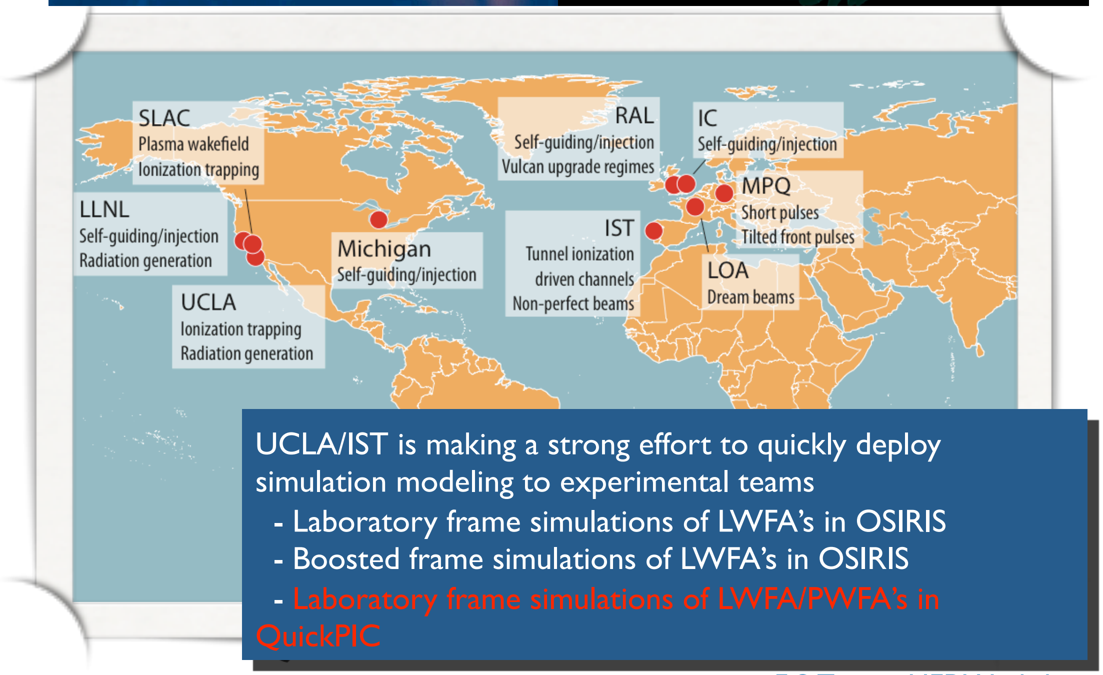
- Plasma based accelerators can achieve accelerating gradients 1,000 x that of those created by conventional accelerators.

- Recently, 2 plasma-based accelerator facilities have been approved.
  - BELLA -- LBNL (ref. C. Geddes)
  - FACET -- 25GeV e-/e+ beams for single-stage PWFA demonstration.

- Simulations have played an important role in the understanding of current experiments. Our objective in the next 3-5 years is to help the design of current LWFA & PWFA experiments, and explore parameters which are not currently accessible.

- This talk will concentrate on PWFA's





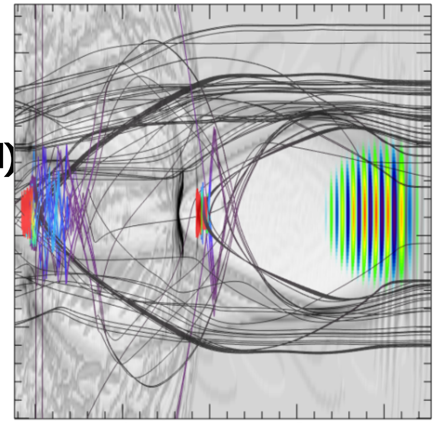
UCLA/IST is making a strong effort to quickly deploy simulation modeling to experimental teams
- Laboratory frame simulations of LWFA's in OSIRIS
- Boosted frame simulations of LWFA's in OSIRIS
- Laboratory frame simulations of LWFA/PWFA's in QuickPIC

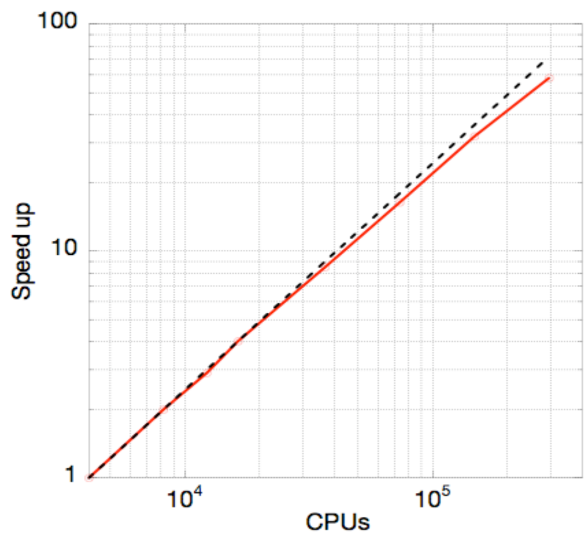
PIC simulations: OSIRIS framework

osiris framework

- Massivelly Parallel, Fully Relativistic Particle-in-Cell (PIC) Code
- Local FDTD field solver
- Visualization and Data Analysis Infrastructure (viz_xd)
- Strong scaling to at least 5000 processors
- Developed by the osiris.consortium

⇒ UCLA + IST + USC





In Recent Strong Scaling Studies, OSIRIS is shown to be >80% efficient on ~300k cores on the BlueGene Supercomputer Jugene

New Features in v2.0

- Bessel Beams
- Binary Collision Module
- Tunnel (ADK) and Impact Ionization
- Dynamic Load Balancing
- Higher Order Shape Functions
- Perfectly Matched Layer (PML)
- Parallel I/O

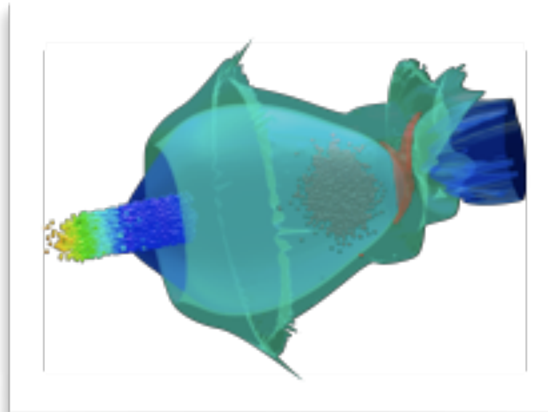
UPIC: UCLA Particle-in-Cell Framework

Features of UPIC:

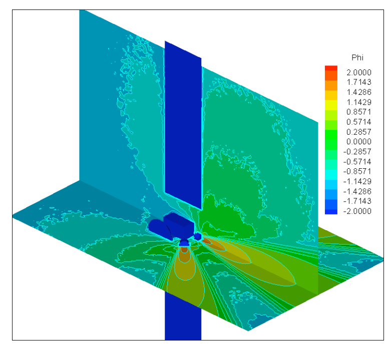
- Provides trusted components for rapid construction of new parallel PIC codes (You-PICK)
- Support multiple physics models, levels of accuracy, optimizations, computer architectures.
- Supports both MPI and threaded programming models.
- Hides parallel processing by reusing communication patterns: Physicists only need to know the data layout.
- Components used in wide variety of applications: Magnetic Fusion, Space Physics, Plasma Accelerators (QuickPIC), Cosmology, Quantum Plasmas, Ion Propulsion (DRACO).

(V. K. Decyk, Comp. Phys. Comm. 17, 95 (2007).)

Recently UPIC has been ported to the GPU, and we will show some preliminary results and discuss the move to new multi-core architectures.

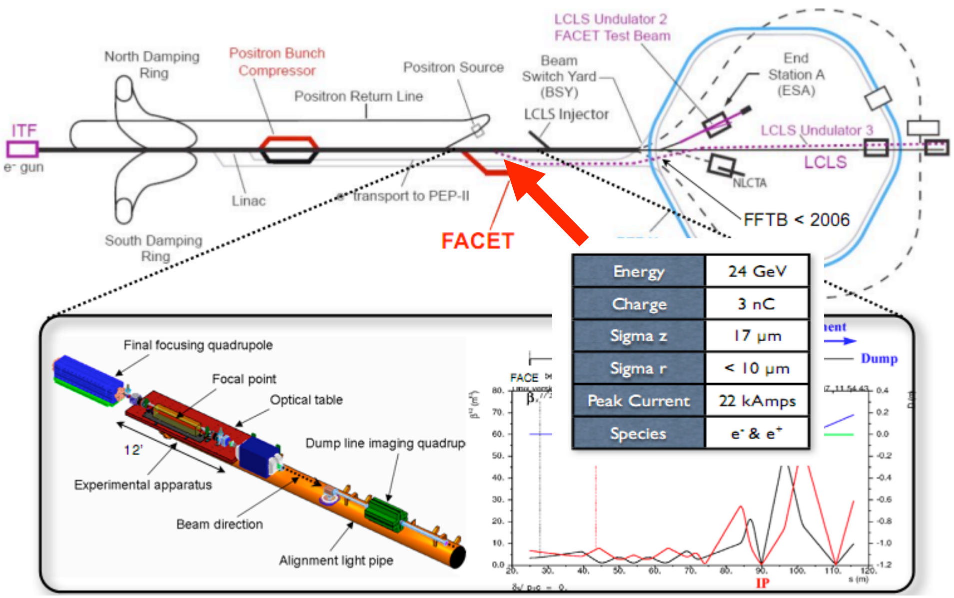


QuickPIC: Plasma Accelerators
(C. K. Huang, *et al*)



DRACO: Ion Propulsion
(J. Wang, *et al*)

Facilities for ACcelerator science and Experimental Test Beams



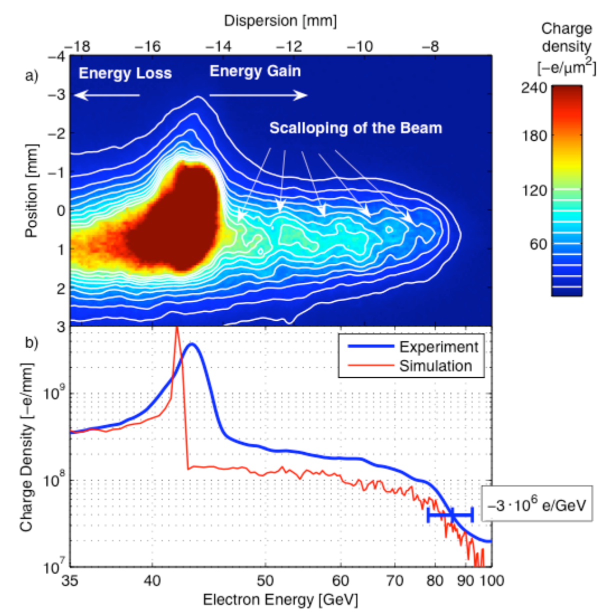
Energy	24 GeV
Charge	3 nC
Sigma z	17 μm
Sigma r	< 10 μm
Peak Current	22 kAmps
Species	e^- & e^+

The PWFA-LC illustrates the key questions that must be answered:

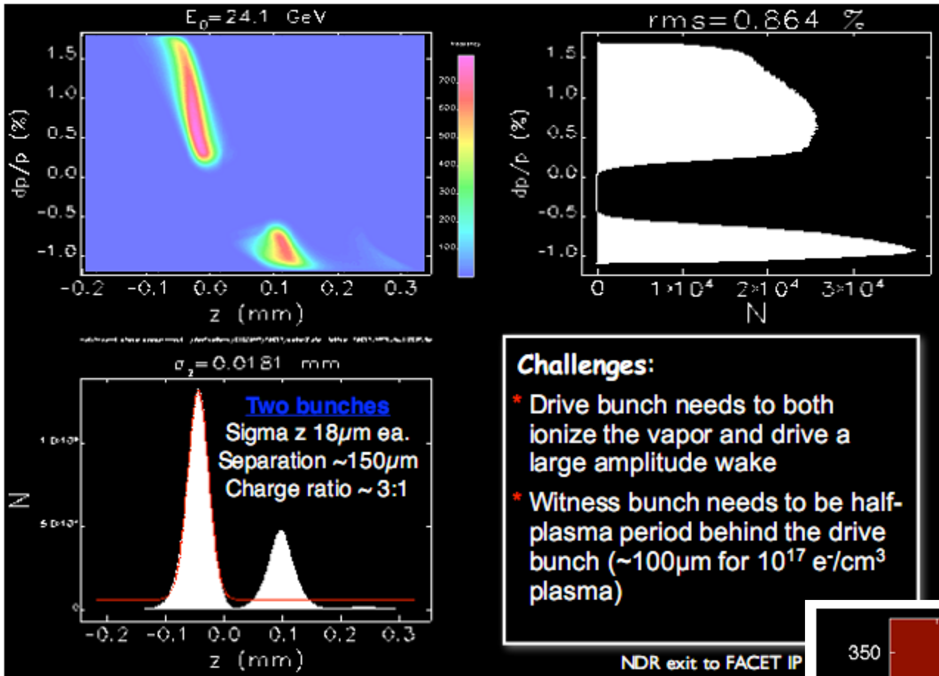
- High beam loading efficiency with both e- and e+ trailing bunches.
- Small energy spread (required to achieve luminosity and luminosity spectrum)
- Small emittance and small emittance dilution (required to achieve luminosity).
- Average bunch repetition rates in the 10's of kHz's.
- Multiple stages.

Plasma Acceleration Research Program at FACET will focus on the first three issues.

FACET is a new facility to provide high-energy, high peak current e^- & e^+ beams for PWFA experiments at SLAC, the goal is to achieve high efficiency, with low energy spread and low emittance. In particular, up to now, although PWFAs have achieved much higher energy gains (>40GeV vs 1GeV in LWFA's), the energy spread of the accelerated particle is not good because the same beam is used to create the wake (and lose energy) and for acceleration (and gain energy), creating spectrum shown on the right. (Ref: Nature, **445**, p. 741)

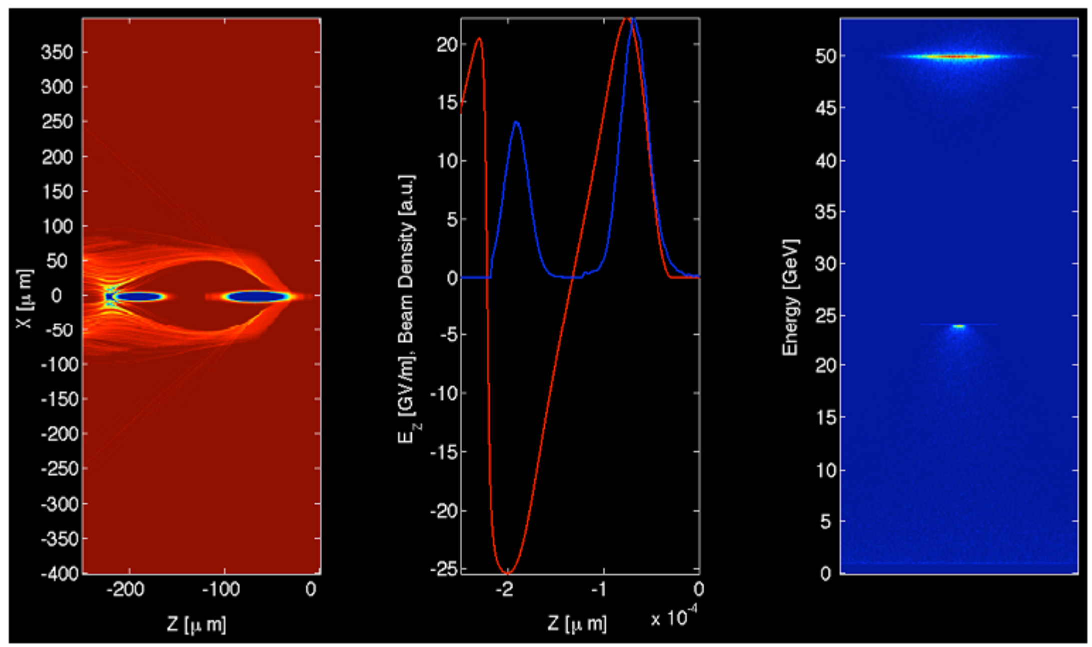


Simulating 2-bunch experiment @ FACET



Two-bunch generation

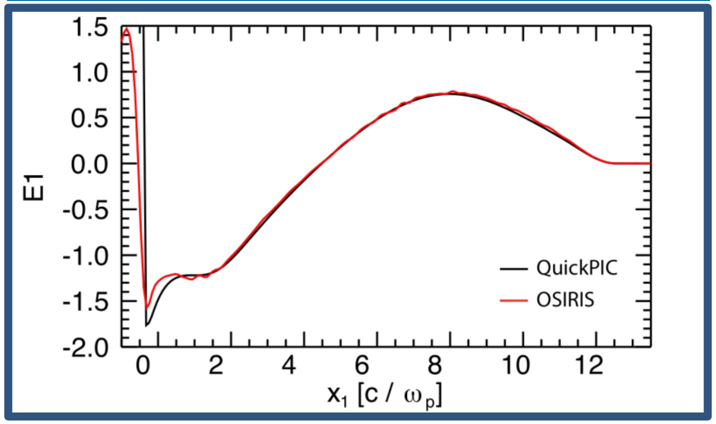
Possible FACET experimental parameters simulated in QuickPIC



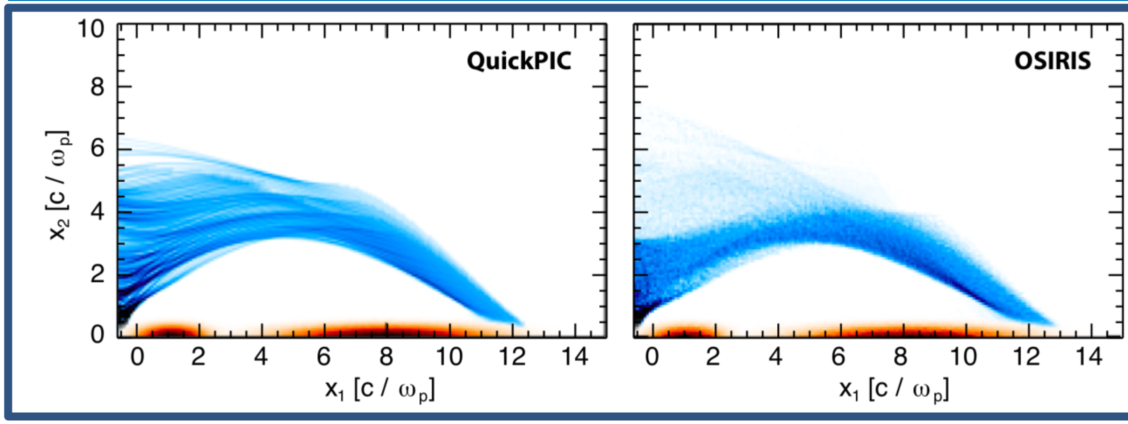
Comparison of OSIRIS 2D Cylindrical vs. QuickPIC

Simulation of a nominal FACET stage with ionization over meter distances

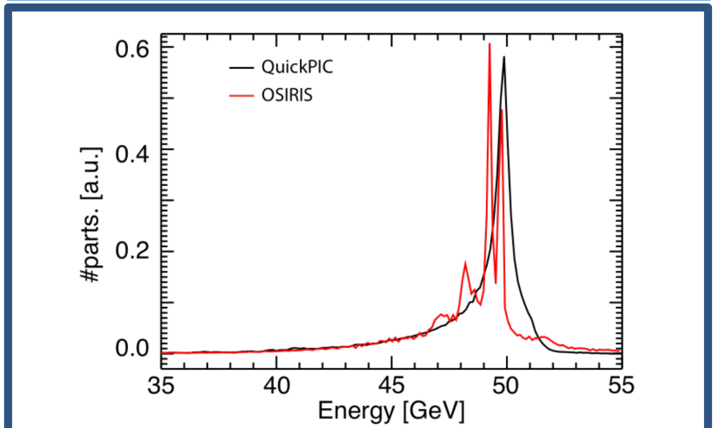
Accelerating field



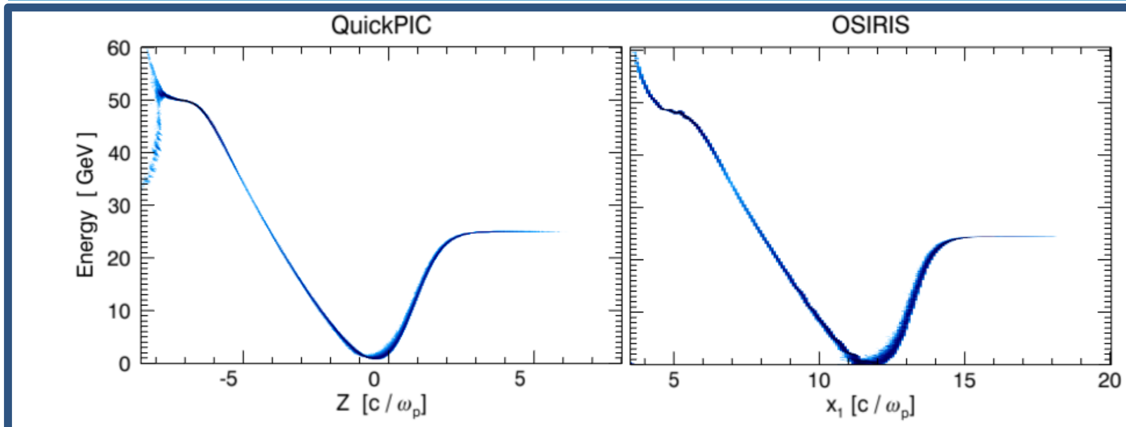
Charge density :: Lithium plasma + beams



Energy spectrum :: Trailing beam



Phasespace :: Driving & Trailing beams



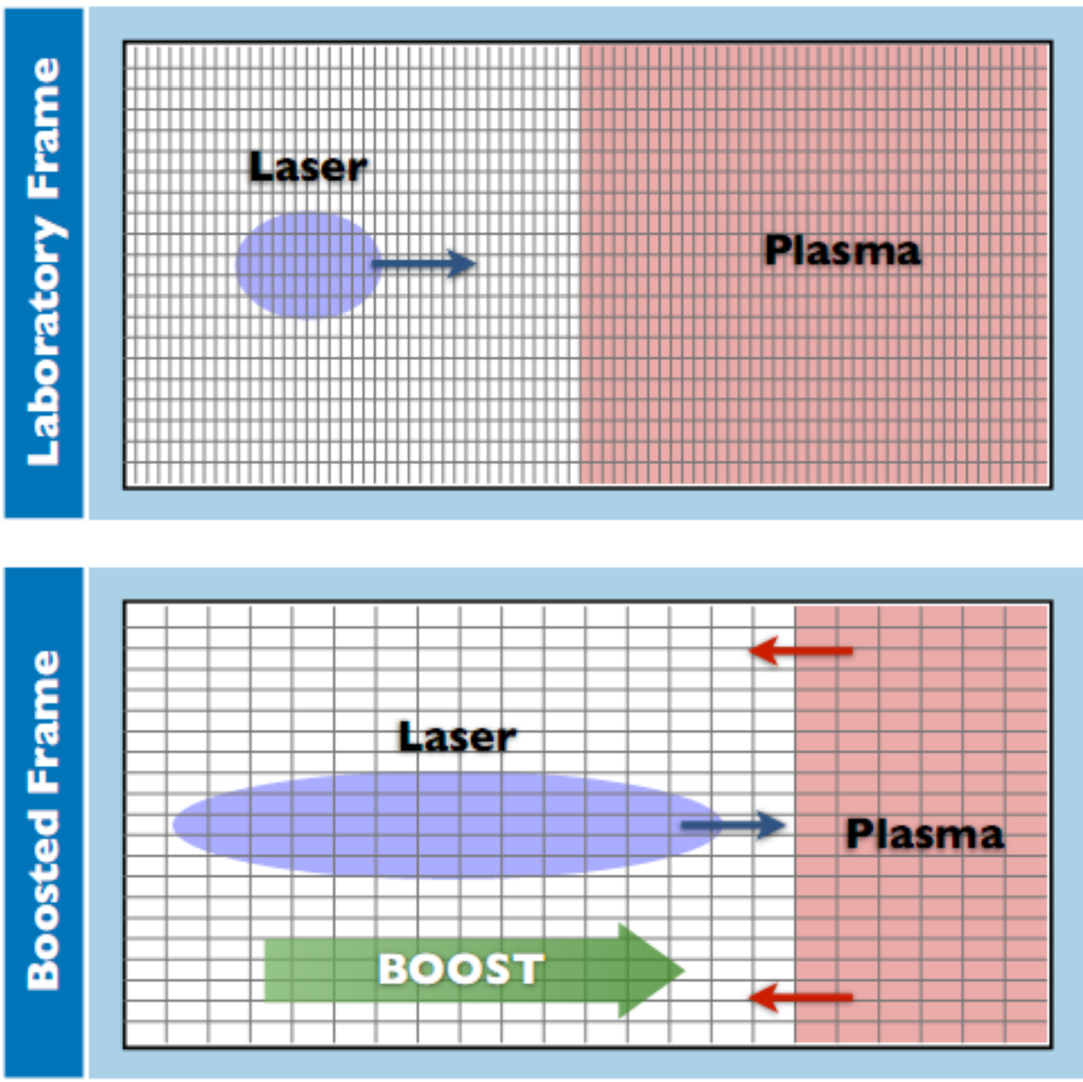
Computational Time [CPU.h]	**QuickPIC**	**OSIRS 2D Cyl**	**OSIRIS 3D** (estimate)
	~250	200	~50000

Current HPC Requirements (3D LWFA Simulations Using OSIRIS)

- Architectures
  - Franklin/Cray XT4
- Compute/memory load
  - 2,048 nodes, 0.5TB data (needs 1GB/core for temp data), 50,000 (cpu*hours) per run
- Data read/written
  - 200GB data (total), 512GB checkpoint (each) 200GB moved in/out of NERSC per month.
- Necessary software, services or infrastructure
  - HDF5/MPI
- Current primary codes and their methods or algorithm
  - OSIRIS
- Known limitations/obstacles/bottlenecks
  - None (recall earlier scaling data which showed good scaling for up to 300,000 cores)
- Anything else?

Boost frame

Grid resolution in Laboratory and Boosted frame



Resolution gains

Particles

- Resolutio… → …$\beta)$
- Plas… cont…on → γ

Moving window in Lab Frame

Time steps

- Time step → $\gamma(1+\beta)$
- Total time → $\gamma(1+\beta)$

Total → $\gamma^2(1+\beta)^2$

Current HPC Requirements (3D High Resolution PWFA Simulations using QuickPIC of a PWFA LC design)

- Architectures
  - Franklin/Cray XT4
- Compute/memory load
  - 8192 cores, 5TB total memory, 100 hours total (8 restarts), roughly 1GB/core
- Data read/written
  - 10GB simulation data per simulation
- Necessary software, services or infrastructure
  - HDF/MPI
- Current primary codes and their methods or algorithm
  - QuickPIC
- Known limitations/obstacles/bottlenecks
  - QuickPIC uses FFT to solve for its fields, this limits the scalability of basic QuickPIC to 100 cores. Pipelining (explained later) is used to improve scaling.
- Anything else?

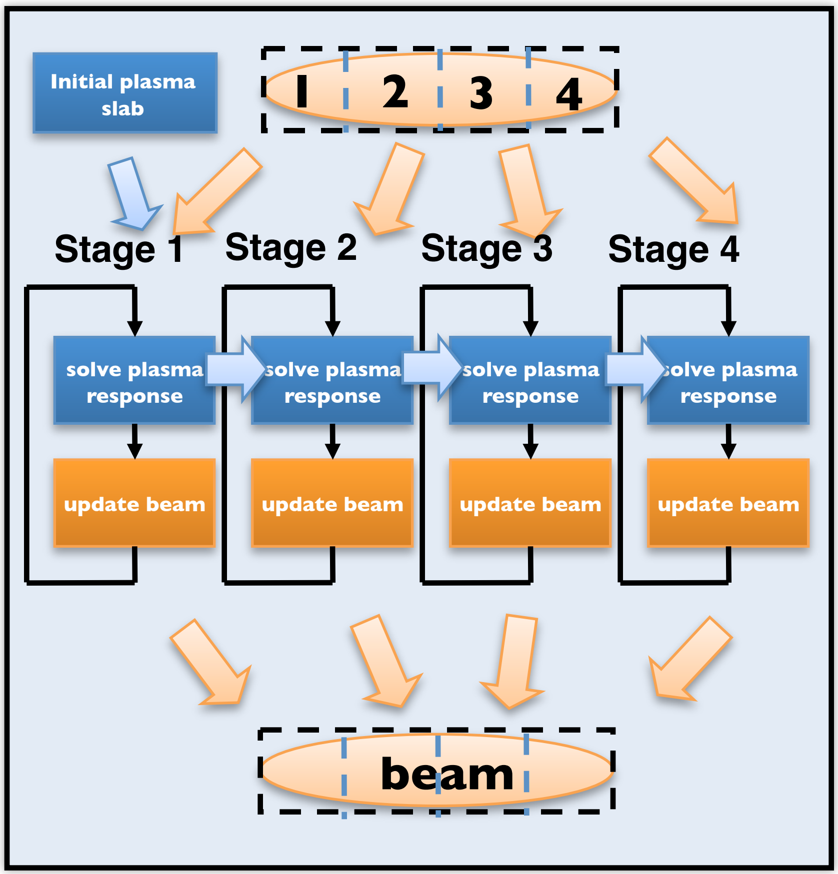
HPC Usage and Methods for the Next 3-5 Years

- Upcoming changes to codes/methods/approaches
  - We have begun to port UPIC to multi-core architectures like GPU'S (in upcoming slides)
- Changes to Compute/memory load
  - The resolution will go up by 10x for asymmetric beam in the collider scenario, we expect future simulations to use 100,000 CPU's. Also, as the transverse size of the beam decreases, the resolution will also need to increase.
- Changes to Data read/written
  - OSIRIS now uses parallel HDF5, and QuickPIC will follow suit soon.
- Changes to necessary software, services or infrastructure
  - (see GPU slides)
- Anticipated limitations/obstacles/bottlenecks on 10K-1000K PE system.
  - OSIRIS has shown good (>80%) strong scaling for 300k processors, QuickPIC scales to > 10k CPU's with pipelining (see next 2 slides) and should scale up to 100's of thousands of CPU's.
- Strategy for dealing with multi-core/many-core architectures
  - (see GPU slides)

Scaling to 100,000+ processors and enabling high resolution capability: Pipelining

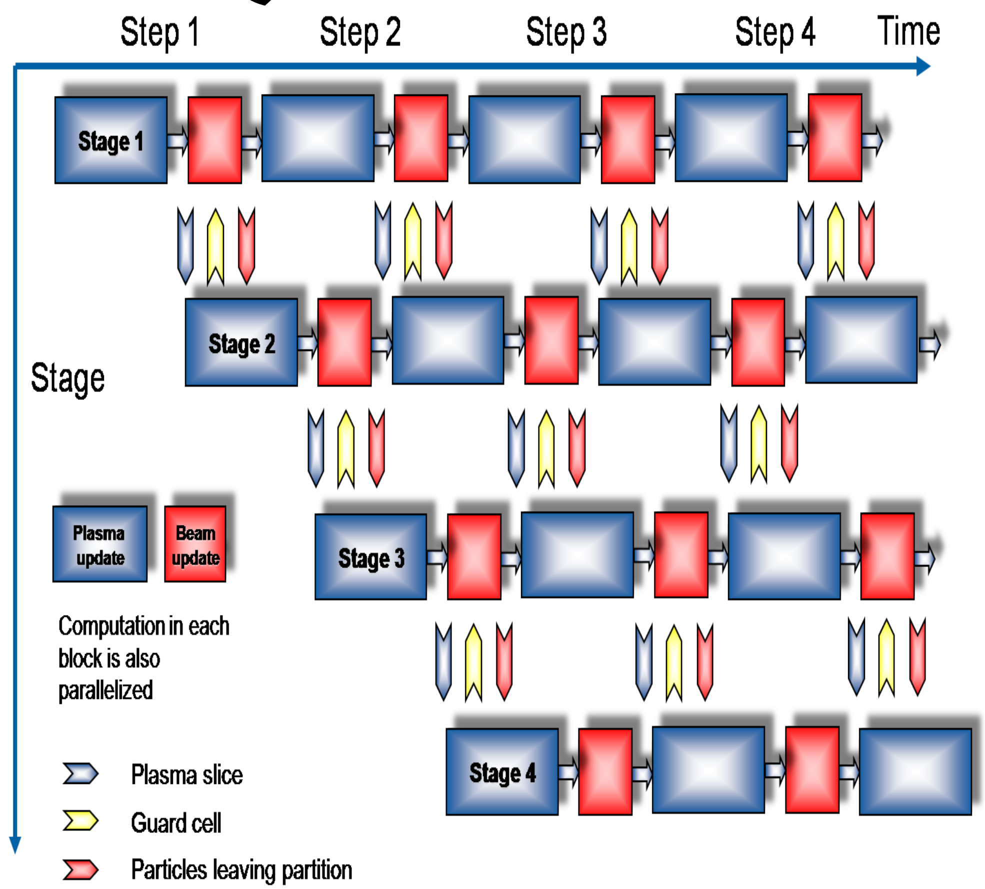
UCLA

Schematic of Pipelining



- Increase throughput through more execution units (similar to CPU)
- Pipeline stages separate in time/space
- Can work with arbitrary domain decomposition

Implementation in QuickPIC

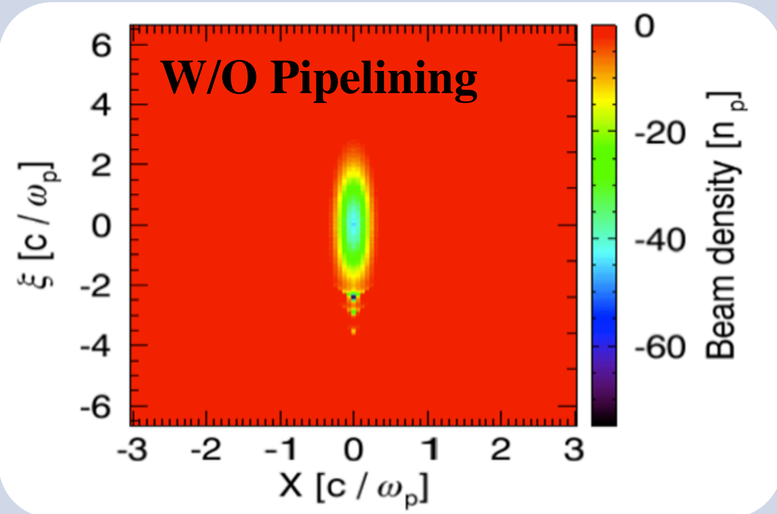


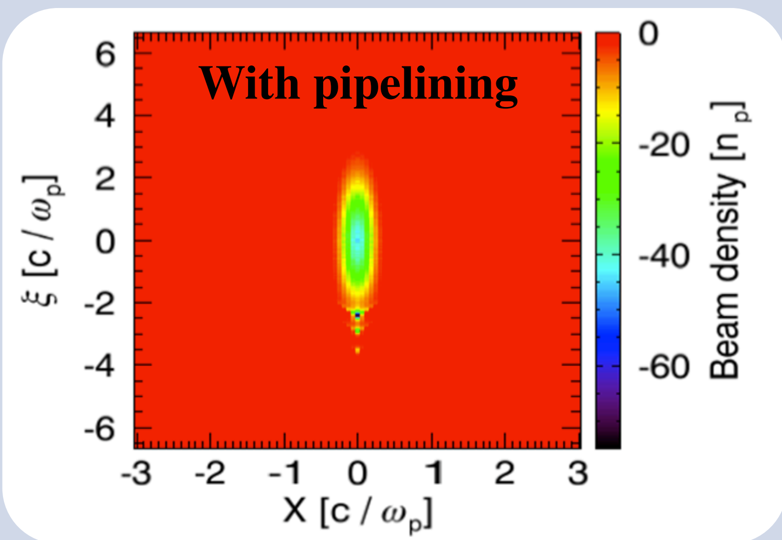
- Communication overlap with computation
- Particles leaving pipeline stage are buffered
- Overall efficiency as high as 85% (2048 processors in 64 pipeline stages), should scale to >100k processors

Pipeline algorithm verification and scaling (>60% efficient on >8k cores)

UCLA

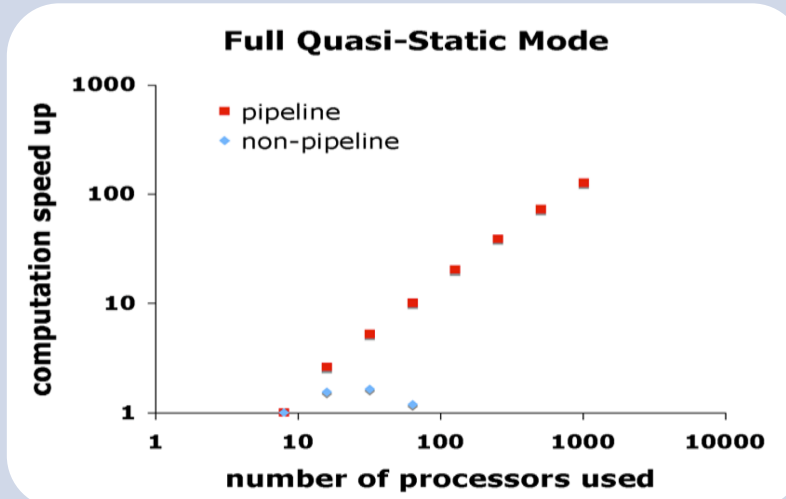
Verification

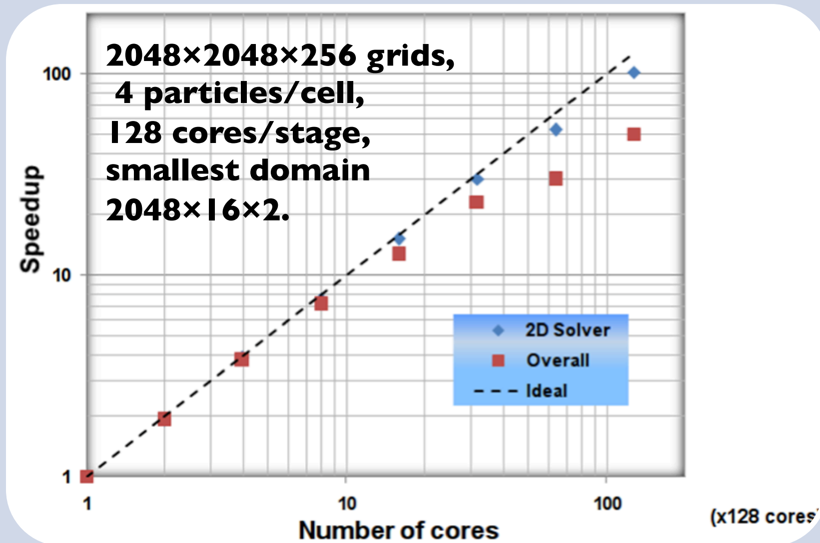




Feng et al, JCP (2009)

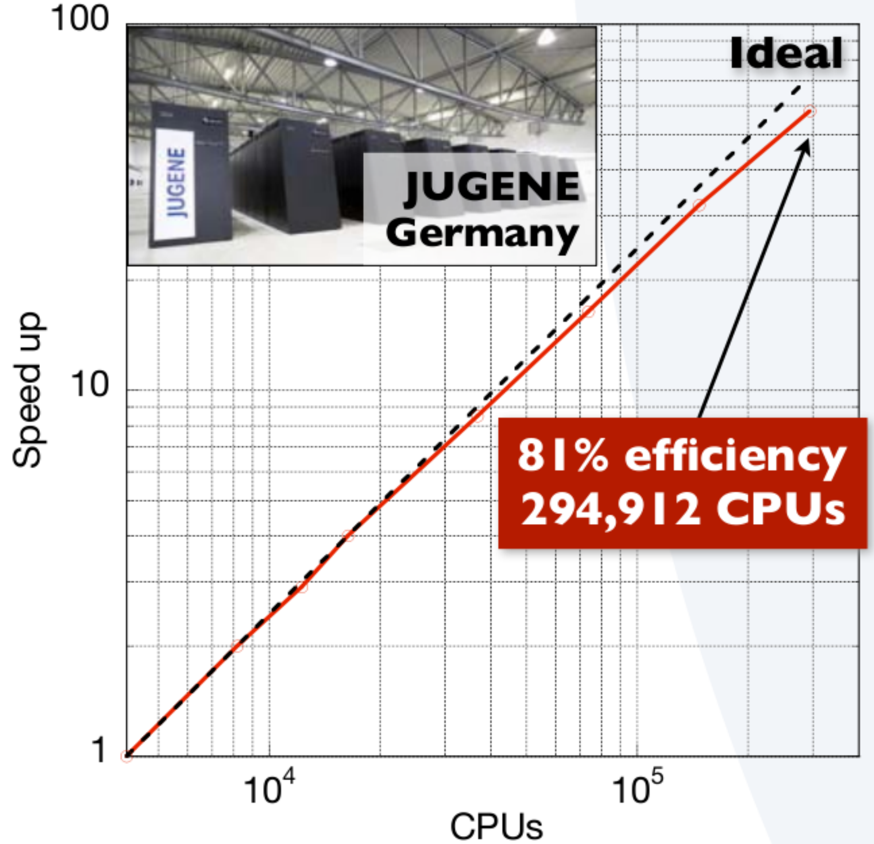
Parallel scaling





Optimize scalability and tap new hardwares

OSIRIS strong scaling up to ~300k CPUs



* Spatial domain decomposition
* Local field solver
* Minimal communication
* Dynamic Load Balancing

New hardware features

SIMD units

tailored code already in production

GPUs

CUDA development (test PIC code)

PowerXCell

PIC algorithm and strategies for multi-core

- Particle-in-Cell (PIC) Algorithms
  - PIC codes have low computational intensity (few FLOPs/memory access), and due to statistical reasons, there are many particles per each cell and particle tasks dominate the timing.
  - 2D Electrostatic code has 45 FLOPs/particle update (11 for deposit, 34 for push)
  - Memory access is still the bottleneck. FLOPs are cheap, therefore, moving data from memory to the processing units efficiently will give the best results on both CPU's and GPU's.
- PIC codes can implement a streaming algorithm by keeping particles constantly sorted by their grid positions. In the process, only the particles' position relative to the grid point is needed and single precision is sufficient.
  - Minimizes global memory access since field elements need to be read only once.
  - Cache is not needed, no gather/scatter.
  - Deposit and updating particles can have optimal stride 1 access.
  - Single precision can be used for particles' positions.
- Challenge: optimizing particle sort

(red texts represents points which also applies to current CPU's, i.e., what we learned on the GPU's have improved the performance of our codes on more traditional architectures)

Timing Results on GPU's

2D ES Benchmark with 256x512 grid, 4,718,592 particles, 36 particles/cell, dt = .025

NVIDA GTX 280 compared to the 3.0 GHz Intel Core 2 X9650 Host:

NVIDA Tesla (C1060) compared to the 3.2 GHz Intel Nehalem Host:

Deposit

- **GTX 280**: 0.18 nsec/particle/time step, a speedup of 40. This is about 35% of memory bandwidth limit.
- **Tesla**: 0.19 nsec/particle/time step, a speedup of 36. This is about 45% of memory bandwidth limit.

Push+Sort

- **GTX 280**: 1.13 nsec/particle/time step, a speedup of 16. This is about 20% of memory bandwidth limit.
- **Tesla**: 1.21 nsec/particle/time step, a speedup of 14. This is about 25% of memory bandwidth limit.

Entire Code

- **GTX 280**: 1.49 nsec/particle/time step, a speedup of 18.
- **Tesla**: 1.63 nsec/particle/time step, a speedup of 15.

Last words on GPU's

- Problem areas:
  - Very difficult to debug, emulator not very faithful.
  - Occasional incorrect result (no ECC yet)
- To debug, we run a Fortran code on the host simultaneously.
  - We can run either the CUDA or Fortran routine at any point
  - Copy out from CUDA and compare

Future looks very promising.

- Software development should improve in future
  - Emerging standards should help: OpenCL , co-Array Fortran.
  - More libraries becoming available: BLAS, FFT, CUDPP
  - Non-standard features and extra manual labor should disappear.

Summary

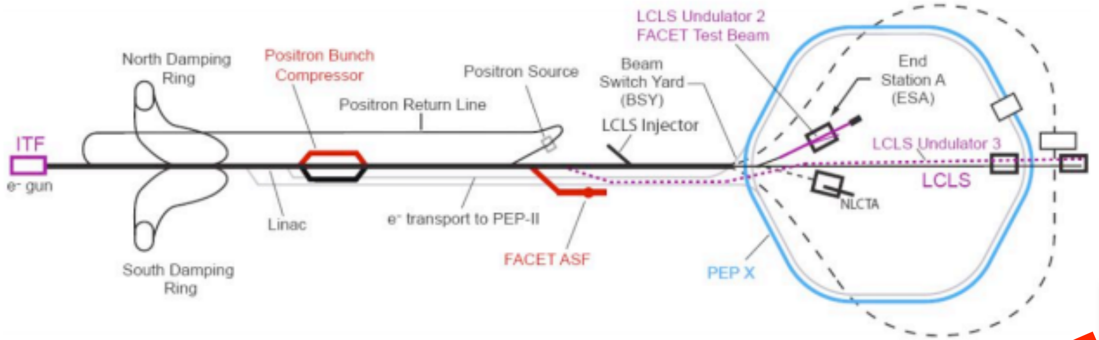
UCLA

- Recommendations on NERSC architecture, system configuration and associated service requirements needed for your science:
  - What to do with restart on future supercomputers? A 10TB simulation (using >10,000 cpu's) would take ~1 day to transfer from the compute nodes to servers over Gigabit network. Although this is a worse-case estimate, it shows that restarts can represent a large fraction of the NERSC allocation if not done efficiently
- What significant scientific progress could you achieve over the next 5 years with access to ~50X NERSC resources?
  - Better understanding of the injection process, which requires better diagnostics and higher resolutions of the "plasma sheet"
  - Simulations of longer/larger future experiments, adding more realistic models into our codes, such as radiative loss, to QuickPIC.
  - Ion Motion (which is needed for future, more energetic drivers)
  - Using the radiation gauge in QuickPIC to include self-injection in LWFA/PWFA's
- Any other special needs or NERSC wish lists?
  - Historically NERSC has been very responsive to its users and it has been our main resource for high performance computing!

PWFA research @ FACET

The PWFA-LC concept illustrates the key questions that must be answered:

* High beam loading with both electrons and positrons (required for high efficiency)
* Small energy spreads (required to achieve luminosity and luminosity spectrum),
* Small emittances and small emittance dilution (required to achieve luminosity),
* Average bunch repetition rates in the 10's of kHz (required to achieve luminosity)
* Multiple plasma stages to achieve the desired energy.



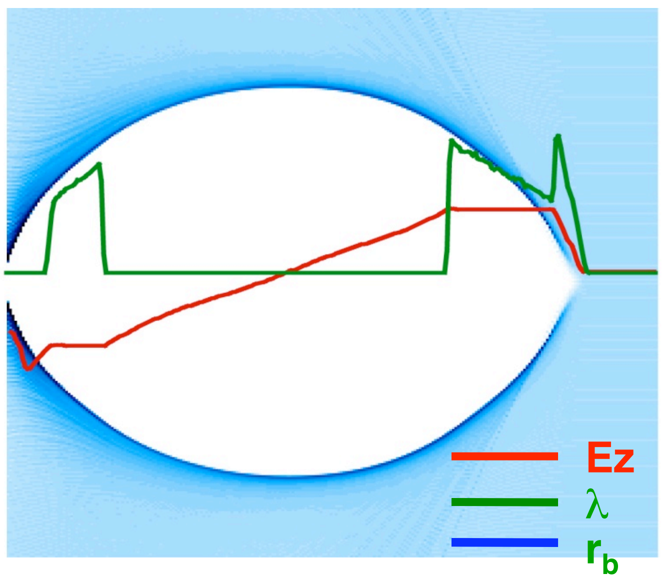
Plasma Acceleration Research Program at FACET will focus on the first thi

Energy	24 GeV
Charge	3 nC
Sigma z	17 μm
Sigma r	< 10 μm
Peak Current	22 kAmps
Species	e^- & e^+

Designing modules for PWFA-LC

Formulas for designing flat wakefield in blow-out regime:

- **Wake structure in blow-out regime: Lu et al., PRL 2006.**
- **Beam-loading: Tzoufras et al., PRL 2008.**



Simulation of the first and the last stages of a 19 stages 0.5TeV PWFA

Physical Parameters		
	Drive beam	Trailing beam
Beam Charge (1E10e-)	0.82 + 3.6	1.73
Beam Length (micron)	13.4 + 44.7	22.4
Emittance (mm mrad)	10 / 62.9	62.9
Plasma density (1E16 cm^{-3})	5.66	
Plasma Length (m)	0.59	
Transformer ratio	1.22	
Loaded wake (GeV/m)	42.7 GeV/m	

Numerical Parameters	
Box size	1000x1000x247
Grids	1024x1024x256
Plasma particles	4 part./cell
Beam particles	8.4 E6 x 3
Time step	60 ω_p^{-1}
Total steps	440

- Summarize your projects and its scientific objectives for the next 3-5 years
  - For the past 80 years, the tool of choice in experimental high energy physics has been particle accelerators. The Large Hadron Collider (LHC) at CERN came online in 2008. The construction cost alone for the LHC machine is nearly 10 billion dollars and it is clear that if the same technology is used that the world's next "atom smasher" will cost at least several times that in today's dollars. The long-term future of experimental high-energy physics research using accelerators depends on the successful development of novel ultra high-gradient acceleration methods. New acceleration techniques using lasers and plasmas have already been shown to exhibit gradients and focusing forces more than 1000 times greater than conventional technology, raising the possibility of ultra-compact accelerators for applications in science, industry, and medicine.
  - In plasma based acceleration the coulomb force of a particle beam or the radiation pressure of a laser beam pushes (or pulls) to create a plasma wake that moves near the speed of light. The accelerating gradients in plasma wakefields are more than 1000 times higher than in conventional accelerators. Properly placed particles surf these wakes to ultra-high energies. Plasma-based accelerators has been a fast growing field due to a combination of breakthrough experiments, parallel code developments, and a deeper understanding of the underlying physics of the nonlinear wake excitation in the so-called blowout regime. In a recent PWFA experiment at SLAC, electrons in the tail of a 42 GeV electron beam were accelerated out to ~80 GeV in only 80 cm. This corresponds to greater than 40 GeV/m energy gain for nearly one meter! In recent LWFA experiments at LBNL monoenergetic electron beams at 1GeV have been reported (in a recent experiment by scientists from UCLA and LLNL 1 GeV beams have also been observed). In each case the wakefield was excited in the nonlinear regime in which plasma electrons are radially expelled. Additionally, in the past few years, parallel simulation tools for plasma based acceleration have been verified against each other, against experiment, and against theory.
  - Based on this progress in experiment, theory, and simulation linear collider concepts using wakefields have been developed and two facilities have been approved. One facility is FACET (at SLAC). This facility will provide 25 GeV electron and positron beams. The other facility is BELLA (at LBNL). It will provide a 30 Joule/30 fs laser. The goal for each facility is to experimentally test key aspects of a single cell within the collider concepts. Furthermore, there are other lasers both within the US, and in Europe and Asia that are currently or will be able to experimentally study LWFA in nonlinear regimes.
  - While some simulations will be conducted to help design and interpret near term experiments, another goal of this proposal is to use these our advanced simulation tools to study parameters that are in regimes that

- Summarize your projects and its scientific objectives for the next 3-5 years
  - For the past 80 years, the tool of choice in experimental high energy physics has been particle accelerators. The Large Hadron Collider (LHC) at CERN came online in 2008. The construction cost alone for the LHC machine is nearly 10 billion dollars and it is clear that if the same technology is used that the world's next "atom smasher" will cost at least several times that in today's dollars. The long-term future of experimental high-energy physics research using accelerators depends on the successful development of novel ultra high-gradient acceleration methods. New acceleration techniques using lasers and plasmas have already been shown to exhibit gradients and focusing forces more than 1000 times greater than conventional technology, raising the possibility of ultra-compact accelerators for applications in science, industry, and medicine.
  - In plasma based acceleration the coulomb force of a particle beam or the radiation pressure of a laser beam pushes (or pulls) to create a plasma wake that moves near the speed of light. The accelerating gradients in plasma wakefields are more than 1000 times higher than in conventional accelerators. Properly placed particles surf these wakes to ultra-high energies. Plasma-based accelerators has been a fast growing field due to a combination of breakthrough experiments, parallel code developments, and a deeper understanding of the underlying physics of the nonlinear wake excitation in the so-called blowout regime. In a recent PWFA experiment at SLAC, electrons in the tail of a 42 GeV electron beam were accelerated out to ~80 GeV in only 80 cm. This corresponds to greater than 40 GeV/m energy gain for nearly one meter! In recent LWFA experiments at LBNL monoenergetic electron beams at 1GeV have been reported (in a recent experiment by scientists from UCLA and LLNL 1 GeV beams have also been observed). In each case the wakefield was excited in the nonlinear regime in which plasma electrons are radially expelled. Additionally, in the past few years, parallel simulation tools for plasma based acceleration have been verified against each other, against experiment, and against theory.
  - Based on this progress in experiment, theory, and simulation linear collider concepts using wakefields have been developed and two facilities have been approved. One facility is FACET (at SLAC). This facility will provide 25 GeV electron and positron beams. The other facility is BELLA (at LBNL). It will provide a 30 Joule/30 fs laser. The goal for each facility is to experimentally test key aspects of a single cell within the collider concepts. Furthermore, there are other lasers both within the US, and in Europe and Asia that are currently or will be able to experimentally study LWFA in nonlinear regimes.
  - While some simulations will be conducted to help design and interpret near term experiments, another goal of this proposal is to use these our advanced simulation tools to study parameters that are in regimes that will not be accessible. We will therefore dramatically advance the rate of discovery and progress in plasma-based accelerator research. We are in a unique position as we are the only group in the world with three-dimensional full (OSIRIS) and quasi-static (QuickPIC) particle-in-cell (PIC) codes. The quasi-static algorithm provides a savings of 100-10000 in computer time without loss of accuracy. Because much of the physics