OSIRIS: Tool for modeling plasma-based acceleration issues

Nearly 20 years old and it started as a tool for modeling E-157!

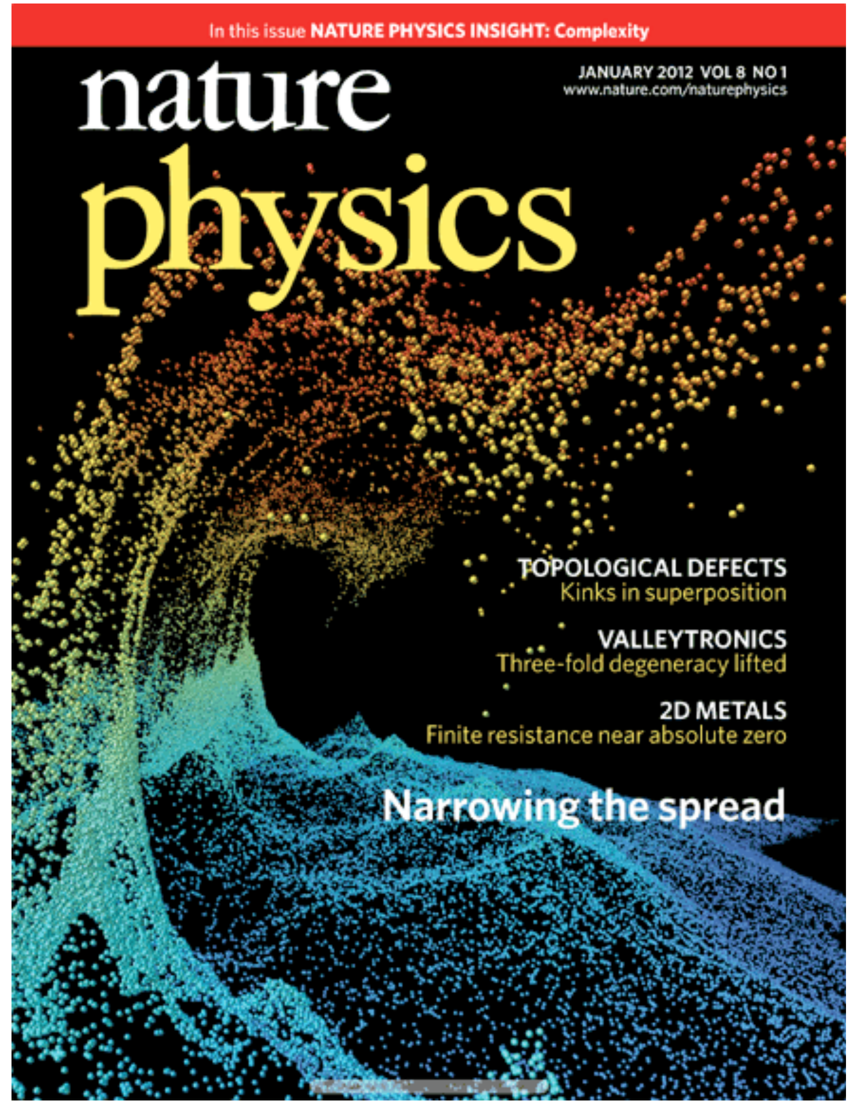
W.B.Mori
University of California
Los Angeles (UCLA)

Departments of Physics and
Astronomy and of Electrical Engineering

Institute of Digital Research and Education

mori@physics.ucla.edu

R. Fonseca, UCLA, PICKSC, IST, SLAC



UCLA

OSIRIS and QuickPIC have used to model FFTB and FACET for past 20 years: Design experiments, interpret experiments, study physics inaccessible to experiments

WarpX next talk

Simulations will be critical for FACET-II and PWFA linear collider research

- Need simulation tools that can support the design of experiments at FACET II.
- Need simulation tools that can aid in interpreting experiments at FACET II.
- Need simulation tools that can simulate new physics concepts, e.g., 3D down ramp injection and matching sections.
- Need simulation tools that can simulate physics of a PWFA-LC including the final focus and IP.
- Need simulation tools that aid in helping to design a self-consistent set of parameters for a PWFA-LC.

Simulations are critical for FACET-II and PWFA linear collider research

- Simulations tools need to be continually improved and validated.
- Simulation tools need to run on entire ecosystem of resources.
- Simulation and analysis tools need to be easy to use.
- Relationship between code developers and users is critical (best practices are not always easy to document).

1:Propose a major experiment that is consistent with DOE's one or more strategic goals

Proposal for an experiment at the FACET Science meeting at UCLA

QuickPIC and OSIRIS simulations have been essential in development of this experiment

- Deplete the drive beam of its energy
- 50% Energy extraction Efficiency
- 10 GeV energy gain for the trailing beam (TB)
- Minimize the energy spread of TB
- Demonstration of emittance preservation of TB
- (this is the first step towards eventually getting a collider quality beam)

- All at the same time

UCLA

2:Generation of ultra-low emittance beams

Proposal for an experiment at the FACET Science meeting at UCLA

- Need to produce electron bunches with brightness orders of magnitude larger than the brightest beams available today.

  - Localized ionization injection
  - Downramp injection
  - Colliding laser pulses inside the wake

UCLA

Collider specific research

- Develop beam loading scenarios for a single stage (need to be fully self-consistent):
  - High energy transfer efficiency
  - Large beam loading
  - Stable: no hosing
  - Develop approaches for transporting beams into and out of stages.
- Develop synchronized injection methods.
- Final focus
- Interaction point

Choices in software

- For PWFA-LC parameters use QuickPIC (quasi-static)
- For LWFA-LC parameters use full PIC (including boosted frame with and without quasi-3D) together with PGC (and perhaps quasi-static).
- For transport use QuickPIC (quasi-static) with other accelerator codes
- For synchronized injection methods use full PIC (including quasi-3D) together with PGC
- Final focus use quasi-static with QED
- Interaction point use full PIC with QED (OSIRIS)

There may be other points of view

Osiris and QuickPIC: Rough estimates

Total Number of Particle Pushes

	Osiris 3D (8ppc)	QuickPIC (8ppc)
FACET II	7×10^{15}	1×10^{13}
PWFA-LC	1×10^{21}	5.6×10^{16}

Total CPU-Hours: assuming no load imbalance

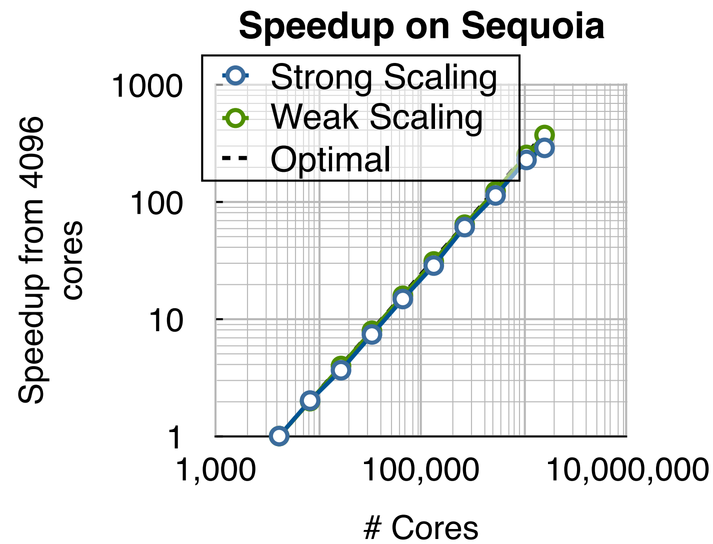
	Osiris 3D (8ppc)	QuickPIC (8ppc)
FACET II	5.9×10^{5}	2.8×10^{3}
PWFA-LC	8.7×10^{10}	1.5×10^{7}

Exascale is not needed for PWFA experiments at FACET II

OSIRIS 4.0: Open access through an MoU

osiris framework

- Massivelly Parallel, Fully Relativistic Particle-in-Cell (PIC) Code
- Visualization and Data Analysis Infrastructure
- Developed by the osiris.consortium
  ⇒ UCLA + IST

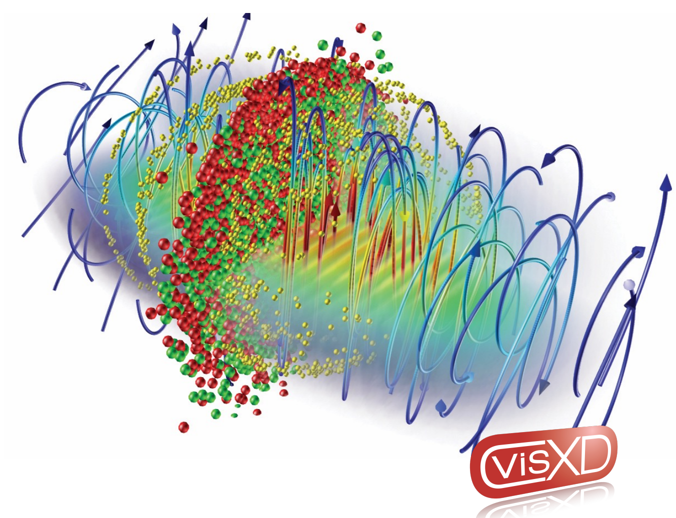


code features

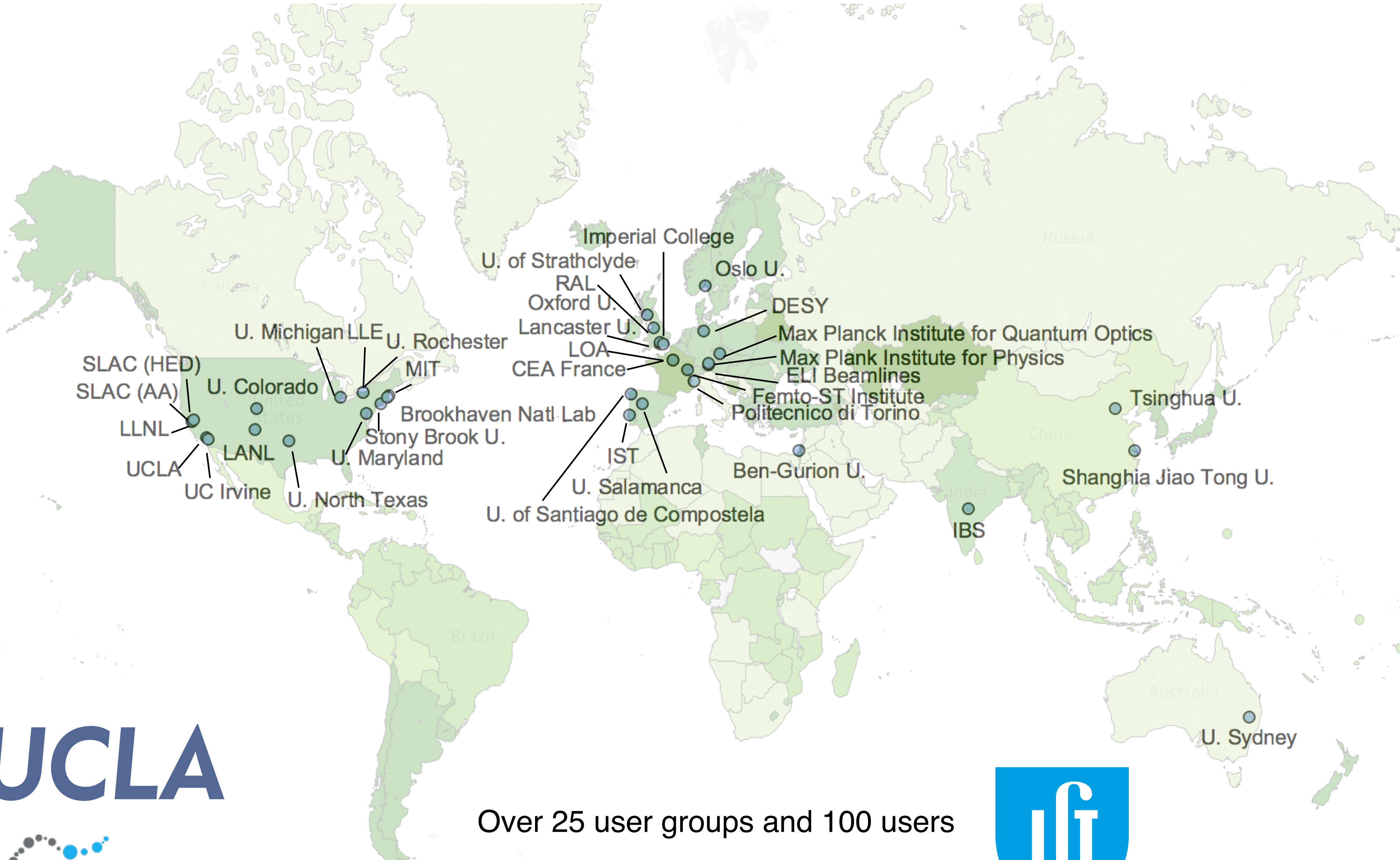
- Scalability to ~ 1.6 M cores
- SIMD hardware optimized
- Parallel I/O
- Dynamic Load Balancing
- Collisions
- Field ionization
- QED module
- Particle splitting/merging
- Quasi-3D
- Customized solver/Boosted frame
- GPGPU support
- Xeon Phi support

Ricardo Fonseca:
ricardo.fonseca@tecnico.ulisboa.pt
Adam Tableman:
tableman@physics.ucla.edu
Frank Tsung:
tsung@physics.ucla.edu

http://epp.tecnico.ulisboa.pt/
http://picks.idre.ucla.edu/



OSIRIS and QuickPIC access is international for HED and AA Science



UCLA

PICKSC

PLASMA SIMULATION UCLA

Over 25 user groups and 100 users

Welcome to the first annual OSIRIS Workshop
Sponsors: UCLA (PICKSC and IDRE) and IST

The goals of the workshop are:

1. To introduce users to the new features and design of OSIRIS 4.0.
2. To allow users of OSIRIS to share experiences and discuss best practices.
3. To identify useful test and demonstration problems.
4. To discuss how to transition from being a user to an active developer.
5. Identify areas for near term software improvements and a community strategy for carrying out the necessary development.

Currently there are over 100 users who have gained access through ~25 MoUs

NSF

PICKSC PLASMA SIMULATION UCLA

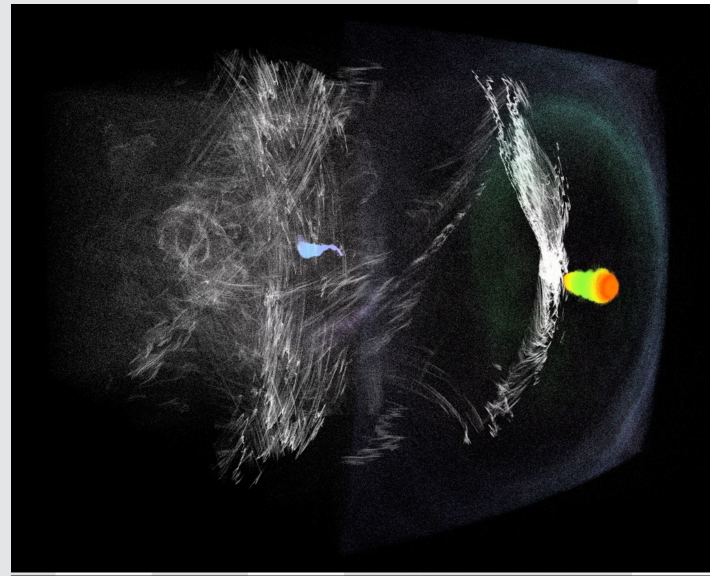
Big success
Over 60 atteneded
Agenda and talks at:
https://picksc.idre.ucla.edu/workshops/

PICKSC PLASMA SIMULATION UCLA

Message

TÉCNICO LISBOA

UCLA

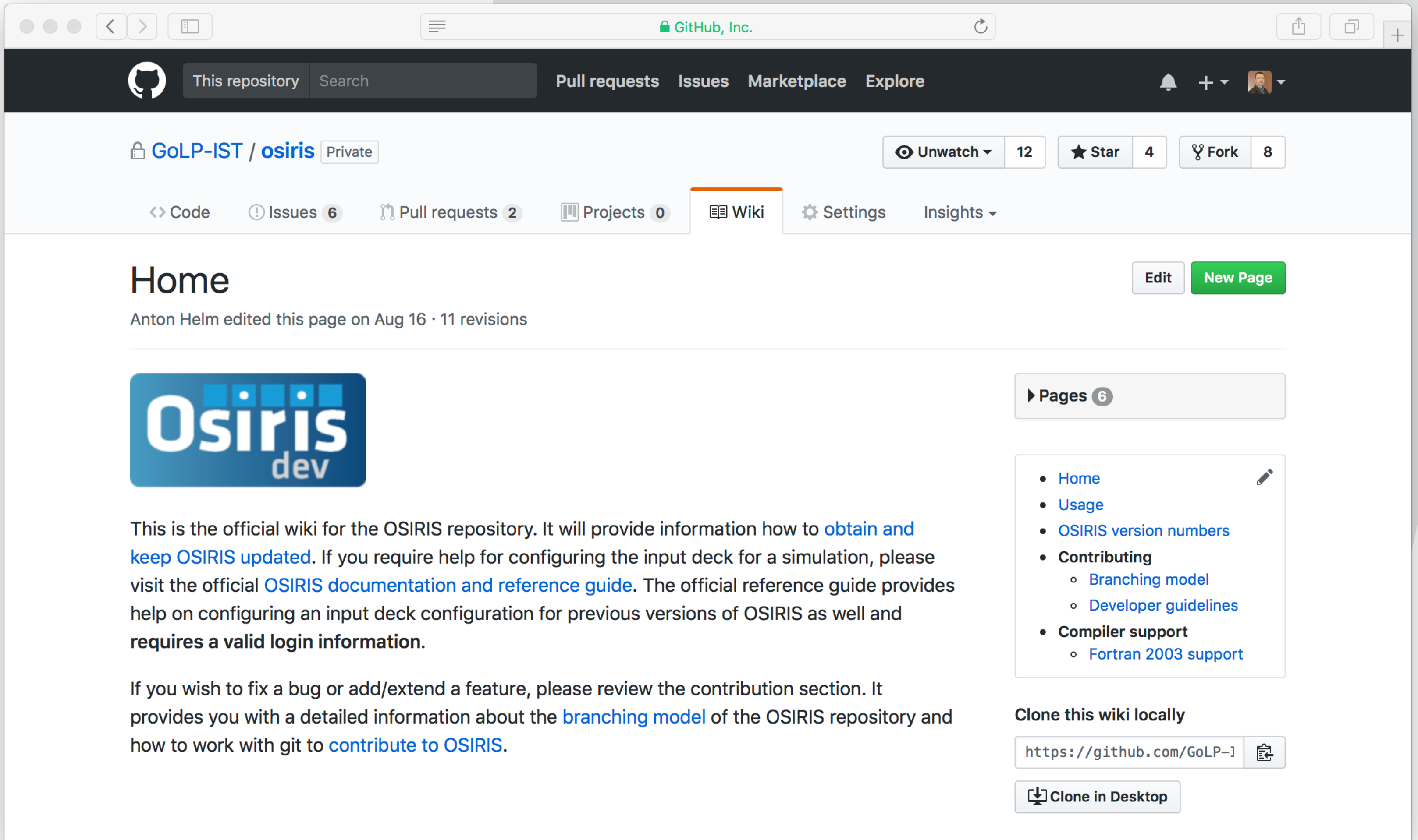


Osiris 4.0

- **OSIRIS 4.0 is a robust, extensible framework**
  - Fully object oriented using Fortran 2003
  - Supports many additional simulation modes and physical models
  - Can be safely and efficiently extended to include new features
- **Move to 4.0 now!**
  - The 4.x series is ready for production use
  - All new development must go into this version
  - Go check out the GitHub repository and start using it today!

All users and developers are moving to a single Github site

https://github.com/GoLP-IST/osiris/wiki

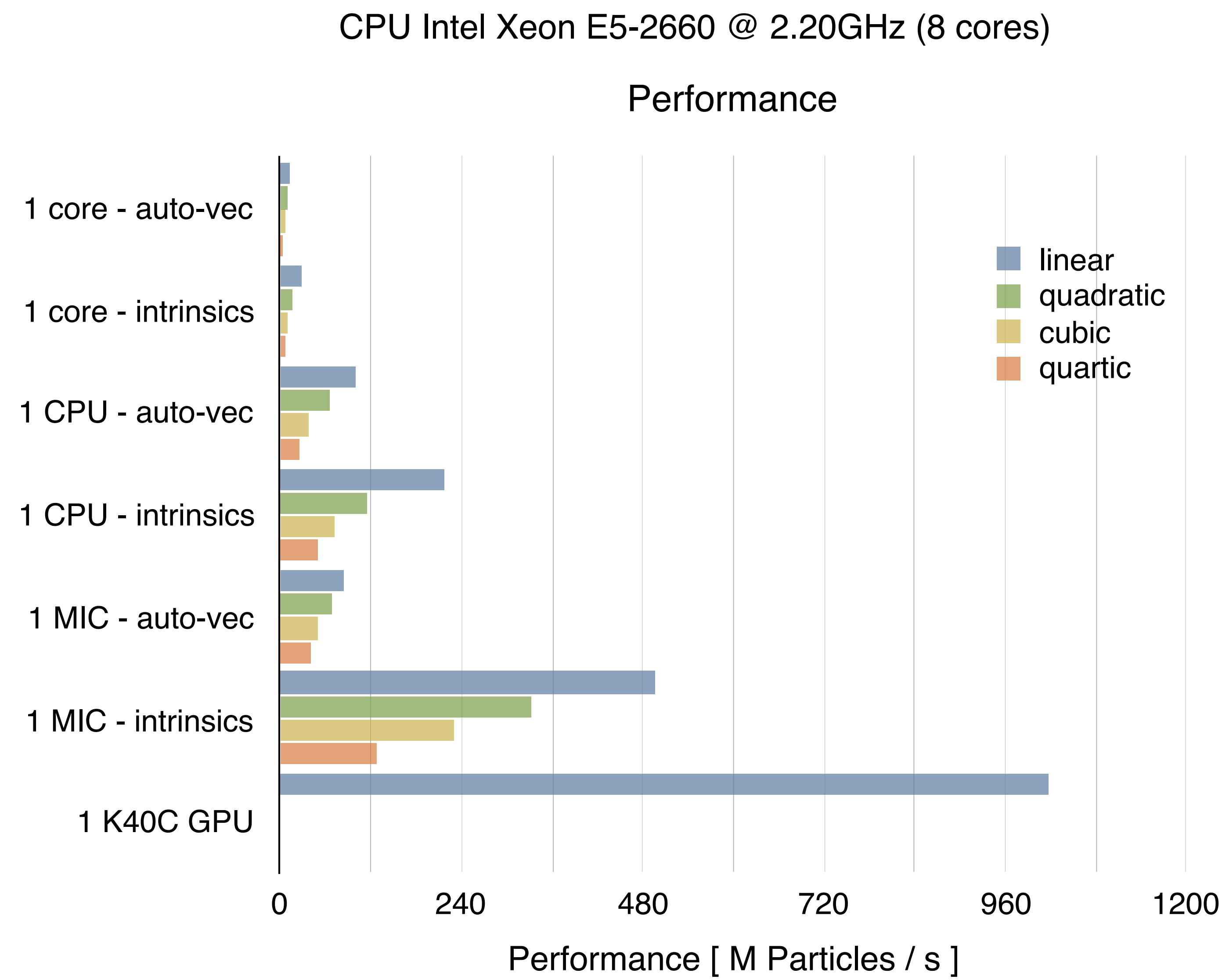


March towards exascale: many core

Ran	System	Cores	R_{max}	R_{peak}	Power	OSIRIS
1	Sunway	10649600	93014.6	125435.9	15371	No
2	Tianhe-2 (MilkyWay-2),	3120000	33862.7	54902.4	17808	Full (KNC)
3	Piz Daint, Switzerland	361760	19590.0	25326.3	2272	Full (CUDA)
4	Titan, United States	560640	17590.0	27112.5	8209	Full (CUDA)
5	Sequoia, United States	1572864	17173.2	20132.7	7890	Full (QPX)
6	Cori, United States	622336	14014.7	27880.7	3939	Full (KNL)
7	Oakforest-PACS, Japan	556104	13554.6	24913.5	2719	Full (KNL)
8	K computer, Japan	705024	10510.0	11280.4	12660	Standard
9	Mira, United States	786432	8586.6	10066.3	3945	Full (QPX)
10	Trinity, United States	301056	8100.9	11078.9	4233	Full (AVX2)

OSIRIS is Cuda and Intel PHI enabled

Knights Corner



Manual vectorization also plays a key role in CPU / core

Combining the 8 cores in the CPU yields over 200 M Particle pushes per second

Using automatic vectorization gives approximately the same performance for 1 MIC as for 1 CPU auto

Manual vectorization gives a significant boost from CPU version

OSIRIS runs across clusters of accelerators

R.A. Fonseca

Knights Landing vs. Knights Corner

KNL is the 2nd generation Intel MIC architecture

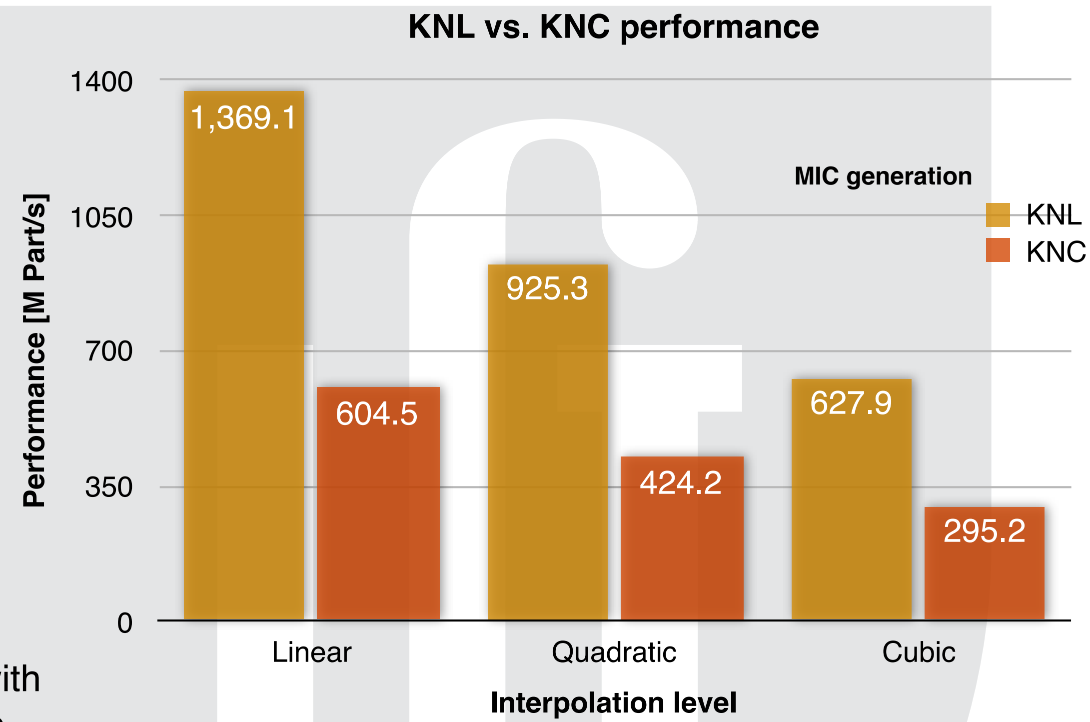
- First generation was the Knights Corner architecture
- Available only as coprocessor boards
- KNC board configuration
  - 60 x86_64 cores @ 1.053 GHz
  - 1× 512bit VPU/core
  - 8 GB GDDR5 RAM
  - Peak FP ~ 2 TFlop/s (SP)
- KNL main differences
  - More cores / higher clock speed
  - Twice the VPU units / core
  - 16 GB MCDRAM
  - Peak FP ~ 6 TFlop/s (SP)

Programming for KNL vs. KNC

- KNC intrinsics almost identical to AVX512 with a few exceptions, small changes required to vector code
- KNL has additional instructions for unaligned memory access
- Also additional AVX512 instructions (e.g. conflict detection), not explored yet.

KNL vs. KNC performance

- Avg. speedup was 2.2×
- Floating point efficiency lower on KNL
- Room for improvement on KNL code



Speedup KNL/KNC

Interpolation Level	Speedup
Linear	2,26 ×
Quadratic	2,18 ×
Cubic	2,13 ×

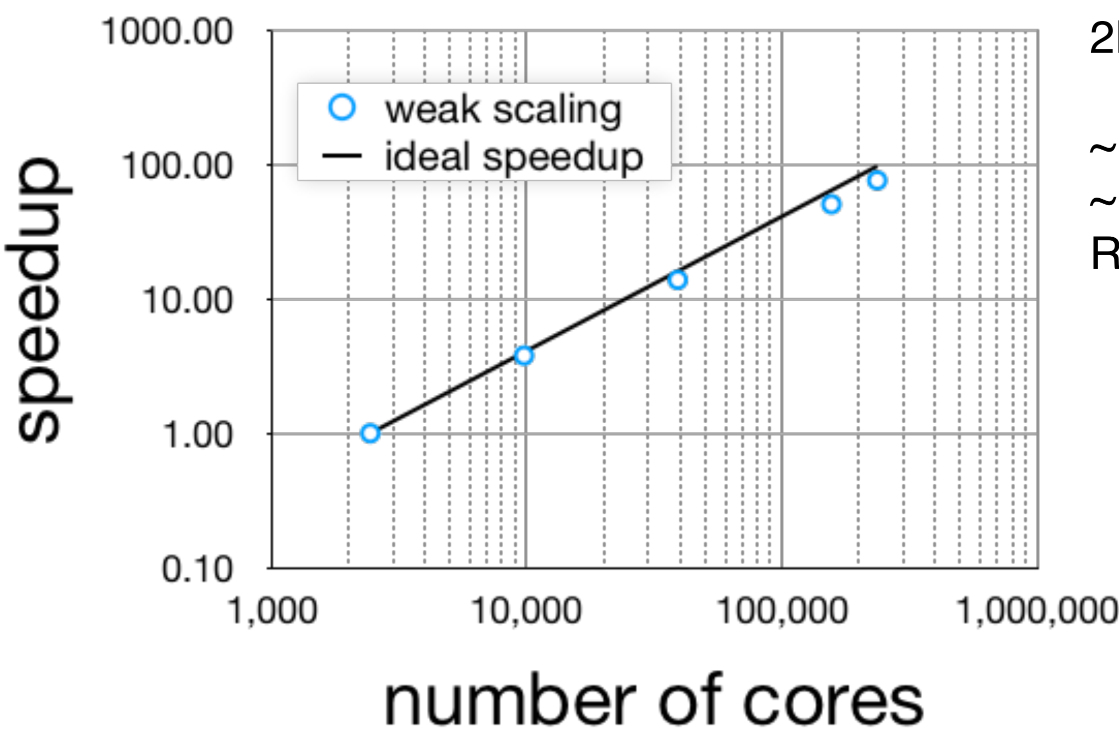
OSIRIS is running on Cori

number of cores	weak scaling speedup	ideal speedup	Weak Scaling Efficiency (%)	Timing (ns/(particle*step))
2,448	1.00	1	1.00	4.11000E-02
9,792	3.78	4.00	0.95	1.09000E-02
39,168	13.80	16.00	0.86	2.99E-03
156,672	50.40	64.00	0.79	8.16000E-04
236,708	75.80	96.69	0.78	5.42000E-04

TÉCNICO LISBOA

UCLA

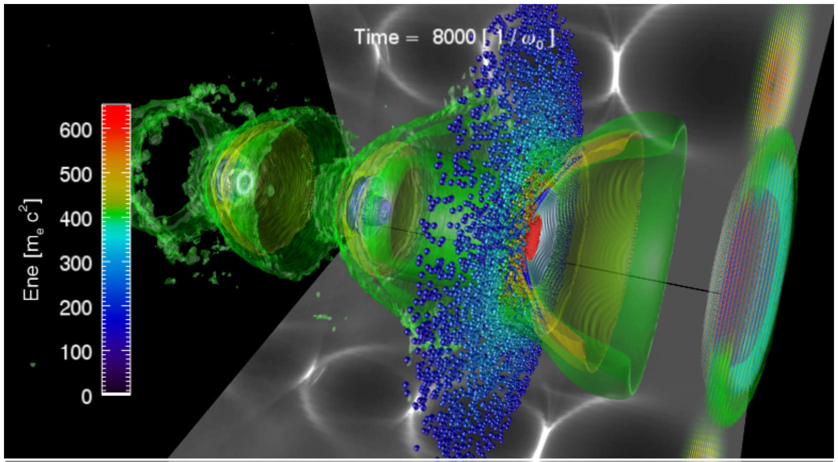
25ppc



2D second order particle shapes

~1 Billion pushes/s/ KNL node on one node
~.7 Billion pushes/s/KNL node on four nodes
Room for improvement

Maintaining parallel load balance can be crucial



LWFA Simulation

Parallel Partition

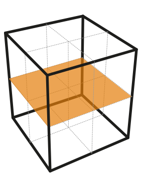
- 94×24×24 = 55k cores

Load Imbalance (max/avg load)

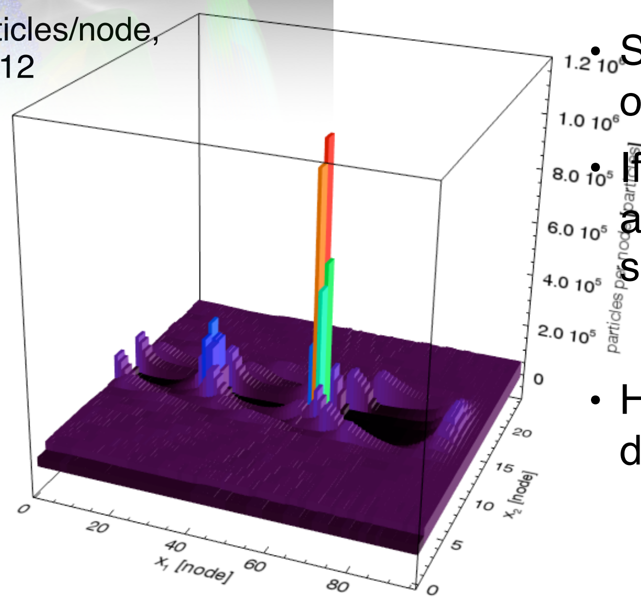
- 9.04×

Average Performance

- ~12% peak



Particles/node, $i_z = 12$

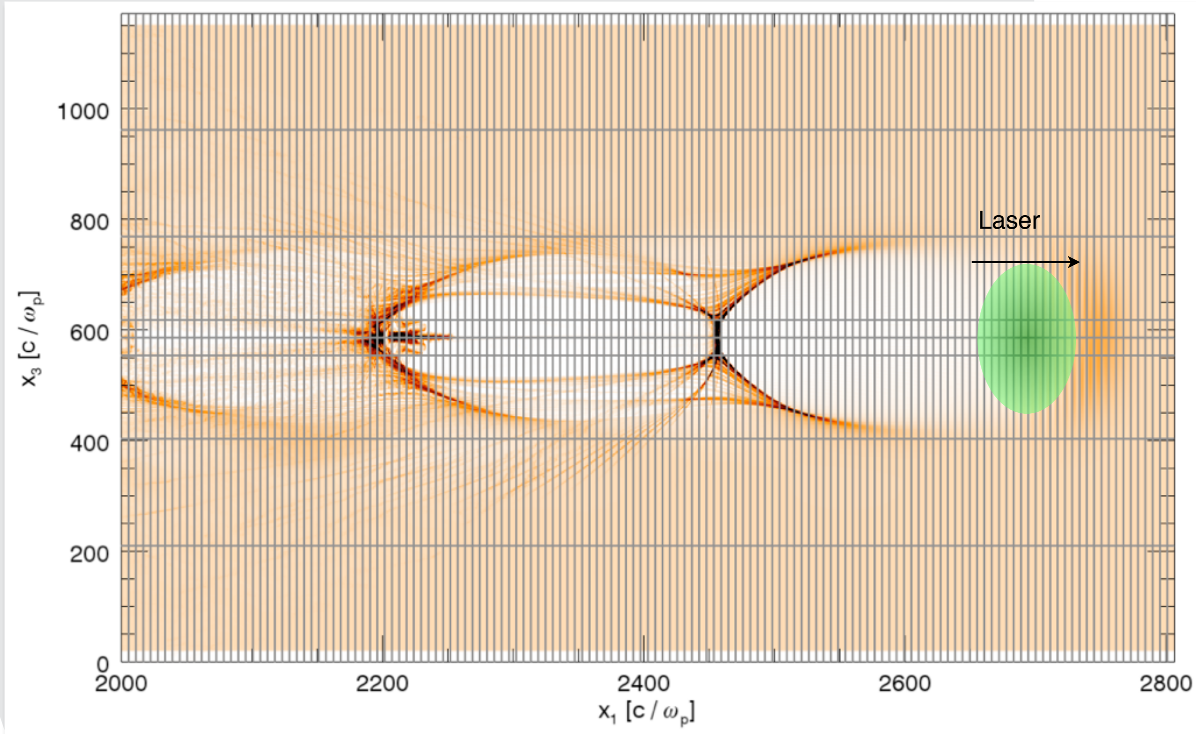


- **Full scale 3D modeling of relevant scenarios requires scalability to large number of cores**
  - Code performance must be sustained throughout the entire simulation
- **The overall performance will be limited by the slowest node**
  - Simulation time is dominated by particle calculations
  - Some nodes may have more particles than other
  - If the distribution of particles remains approximately constant throughout the simulation we could tackle this at initialization
    - Static load balancing
  - However this will usually depend on the dynamics of the simulation

Large scale LWFA run: Close but no cigar

Old result

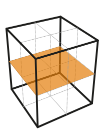
- **The ASCR problems are very difficult to load balance**
  - Very small problem size per node
  - When choosing partitions with less nodes along propagation direction imbalance degrades significantly
- **Not enough room to dynamic load balance along propagation direction**
- **Dynamic load balancing in the transverse directions does not yield big improvements**



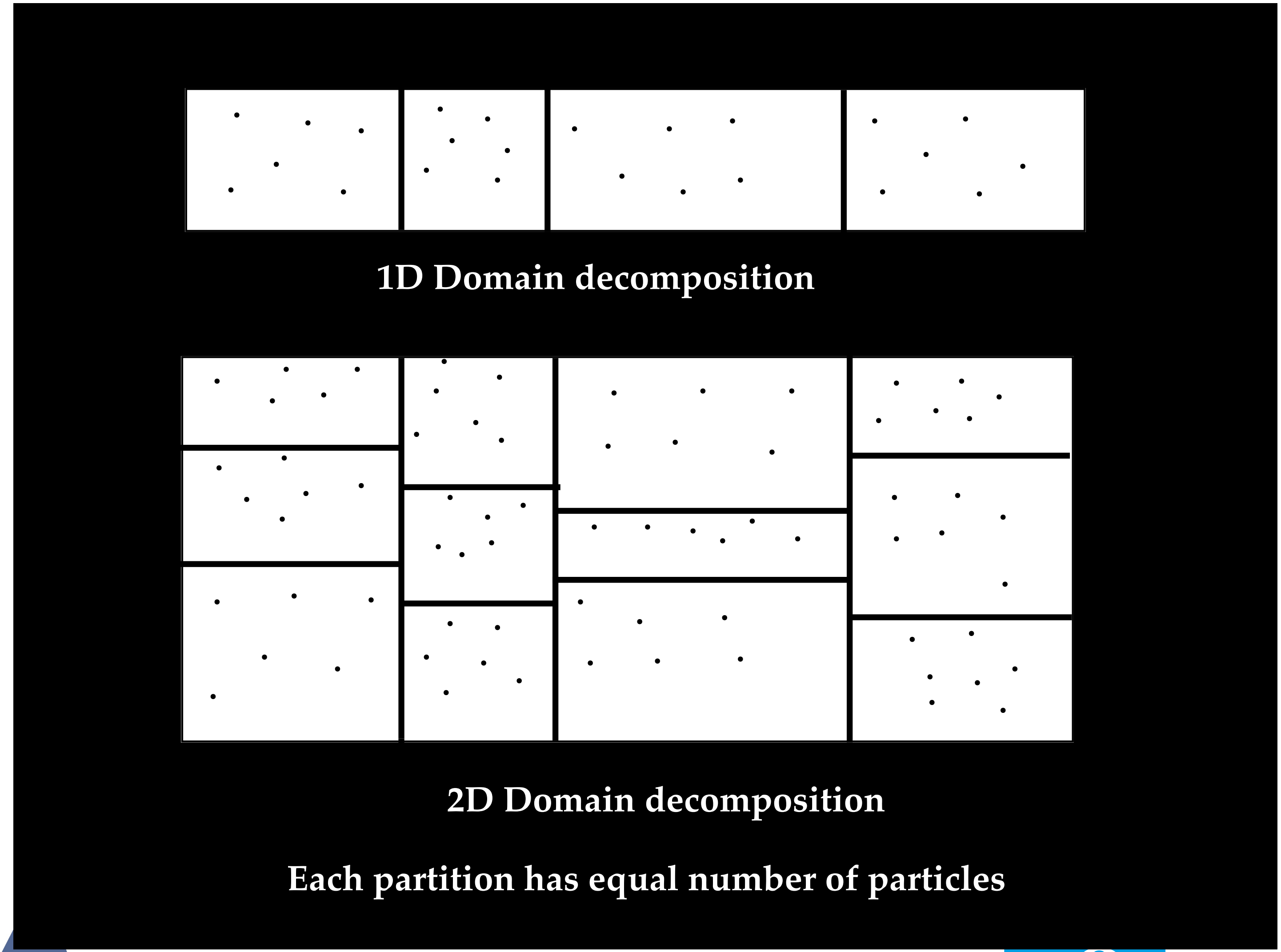
x_1-x_2 slice at box center
similar partition along x_3
> 30% improvement in inbalance

No overall speedup

- Best result:
  - Dynamic load balance along x_2 and x_3
  - Use max load



Use ideas from UPIC for OSIRIS: Non "four corners"
At workshop we started discussing this project.



Use simplest message passing sequence

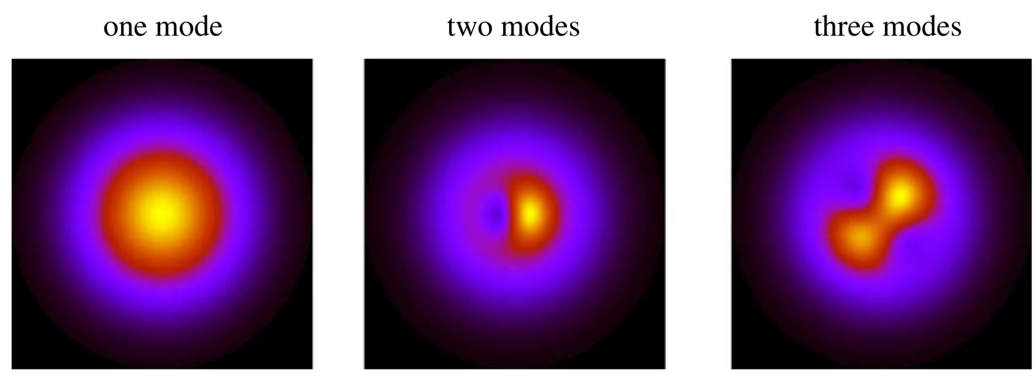
3D simulations of LWFA and PWFA (e and p) can be expensive, but “r-z” can be useful for parameter scans

- 2D cylindrical r-z simulations can get the geometric scaling correct: Used extensively for PWFA
- EM waves are radially polarized in r-z simulations, so cylindrical r-z simulations not used for LWFA studies.
- Expand in azimuthal mode number and truncate expansion! [1]: LASER is an m=1 mode. This is PIC in r-z and gridless in Φ.
- A charge conserving current deposit was developed and incorporated into OSIRIS [2].

[1] A.F. Lifshitz et al., JCP 228, pp.1803 (2009).

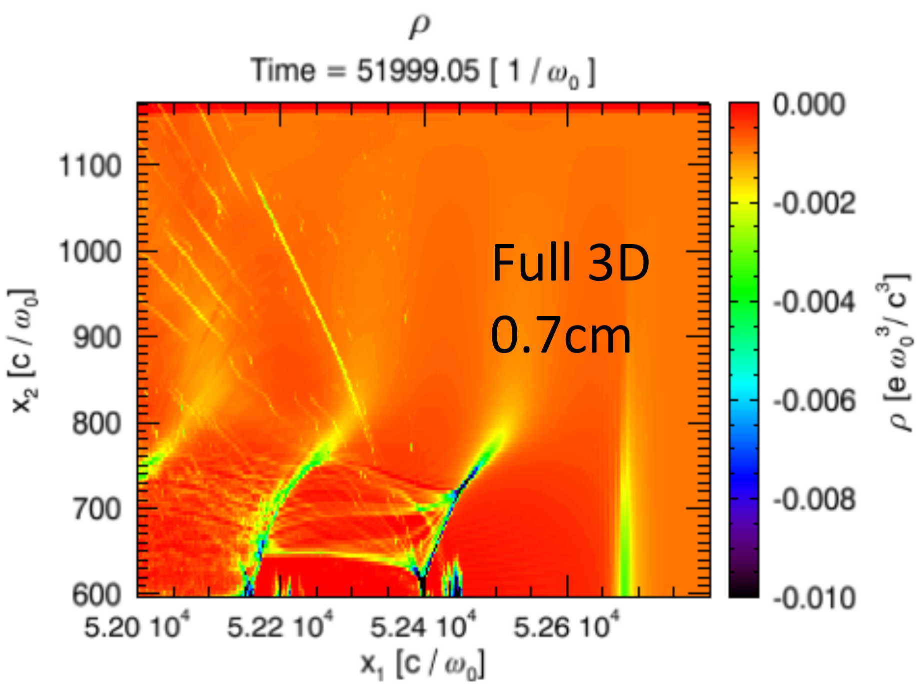
[2] A. Davidson et al., JCP 281, pp. 1063 (2014).

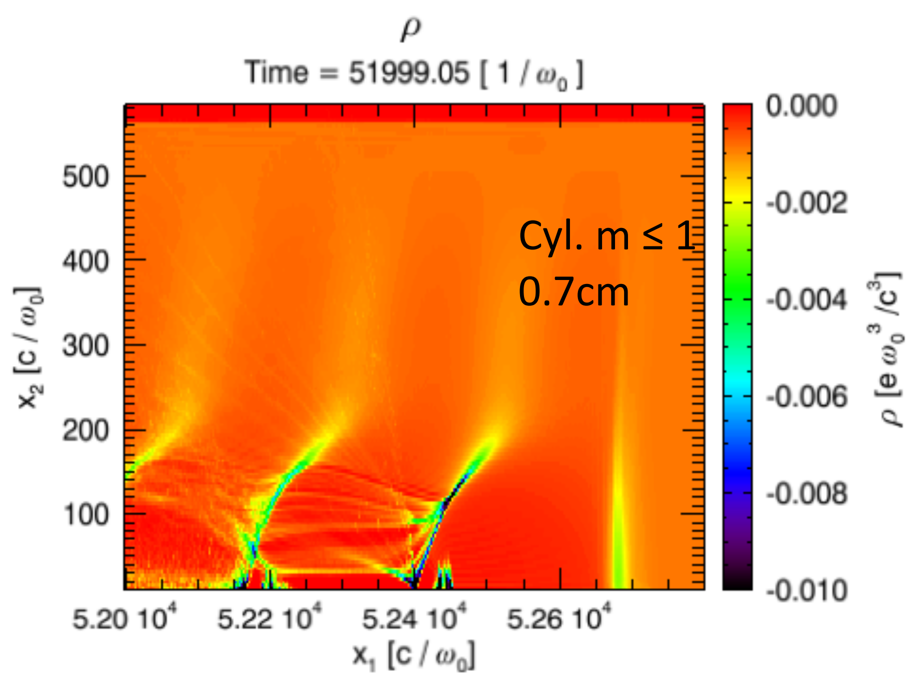
[3] R. Lehe et al., submitted (2015).



UCLA

OSIRIS quasi-3D agrees with full 3D for symmetric cases with CPU savings of ~100 or more: LWFA (but we use it for FACET II)

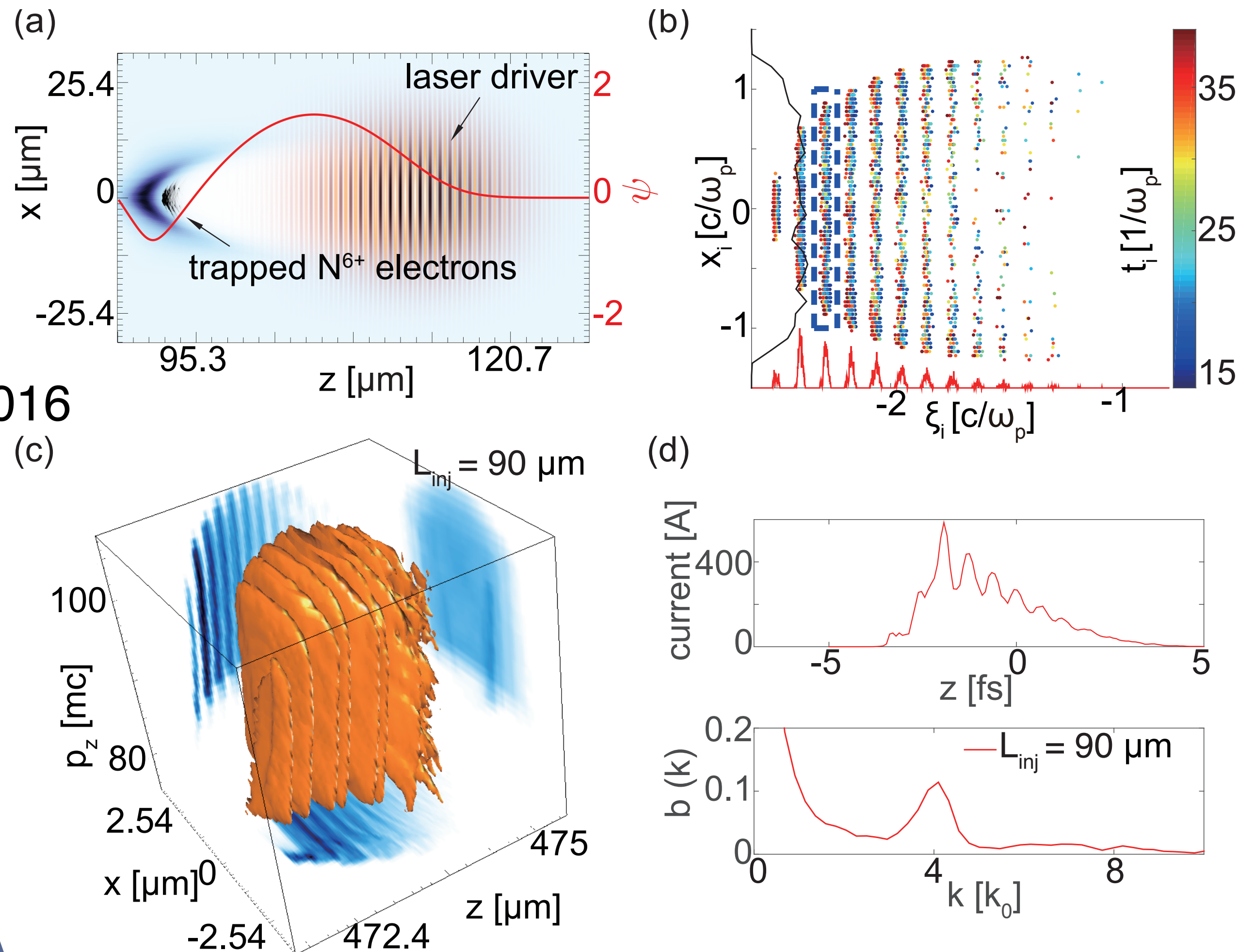




340pC 1.57 GeV

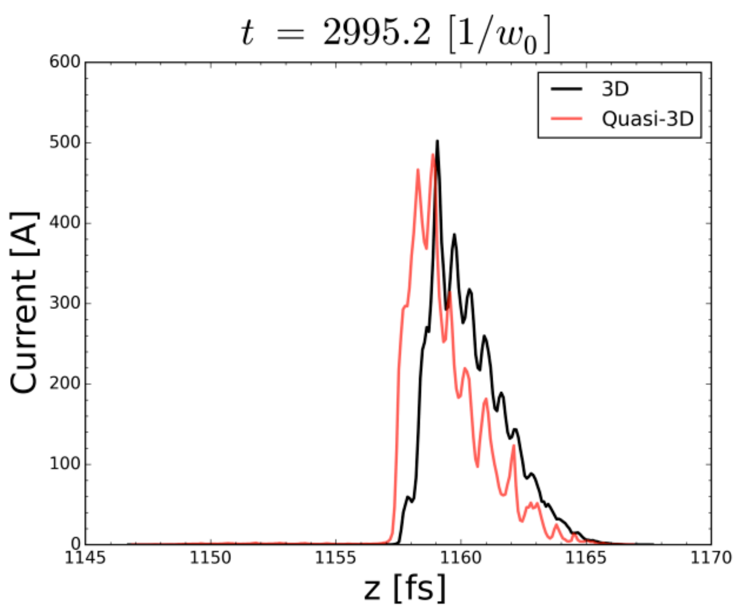
328pC 1.55 GeV

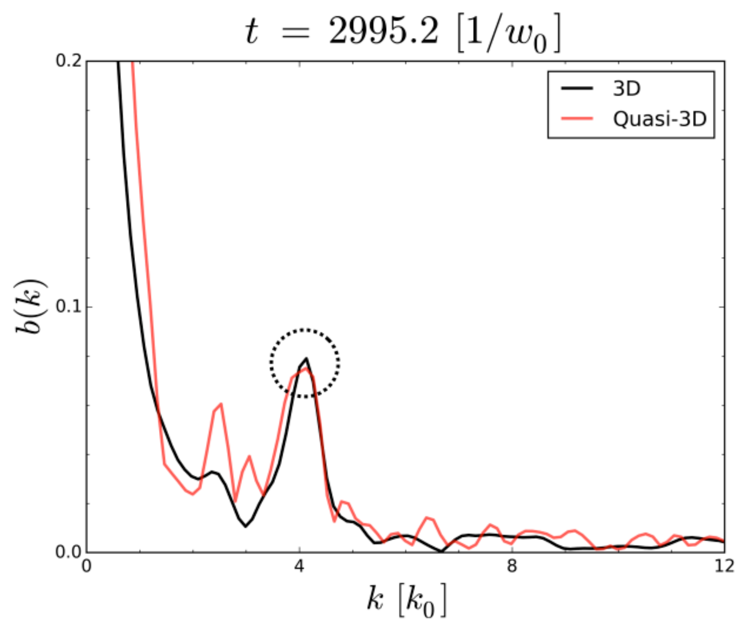
OSIRIS is used for studying unique beams from PBA: nano bunching: Still surprises

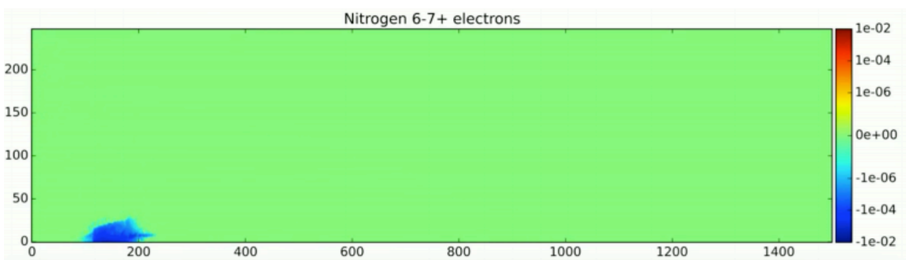


Xu et al. PRL 2016

quasi-3D OSIRIS and full OSIRIS agree well on complicated problems: nano bunching







- **Bunching factor:** $b(k) = \int dz\, n(z) exp(ikz)$
- **Nanoscale bunching at** $k = 4k_0$

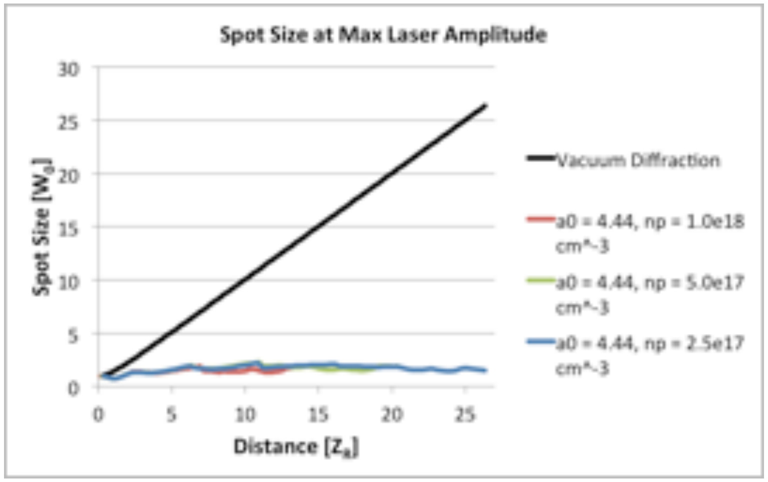
UCLA

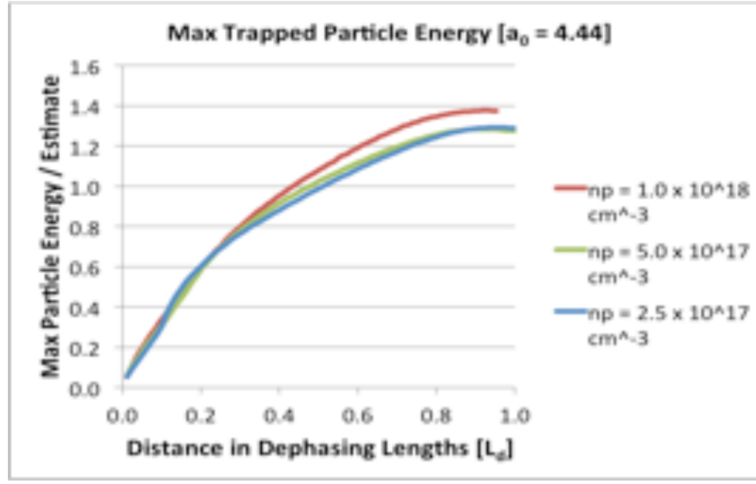
PICKSC

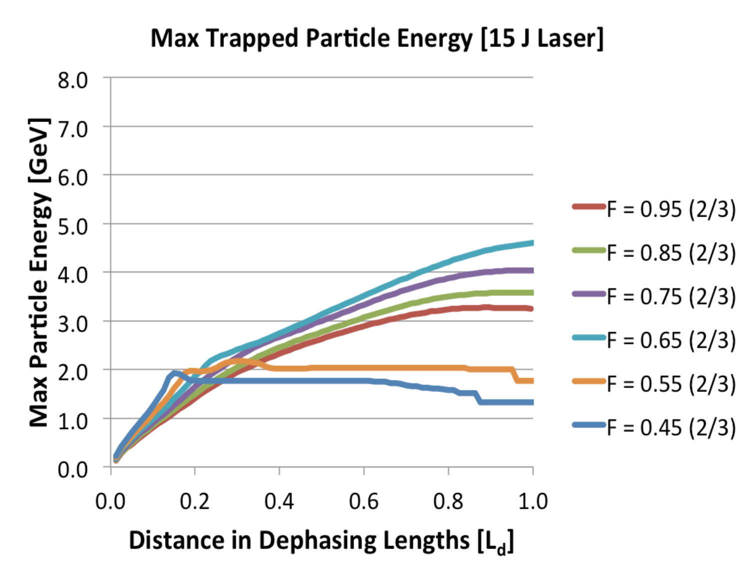
PLASMA SIMULATION UCLA

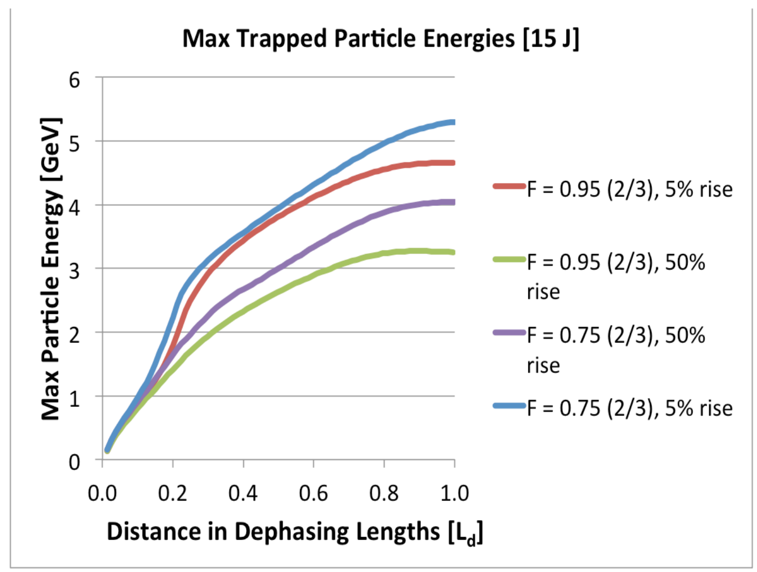
LWFA in self-guided nonlinear regime: 15J (30J) laser can produce 5+ GeV (8+) electrons

Lu et al. scaling and predictions is confirmed





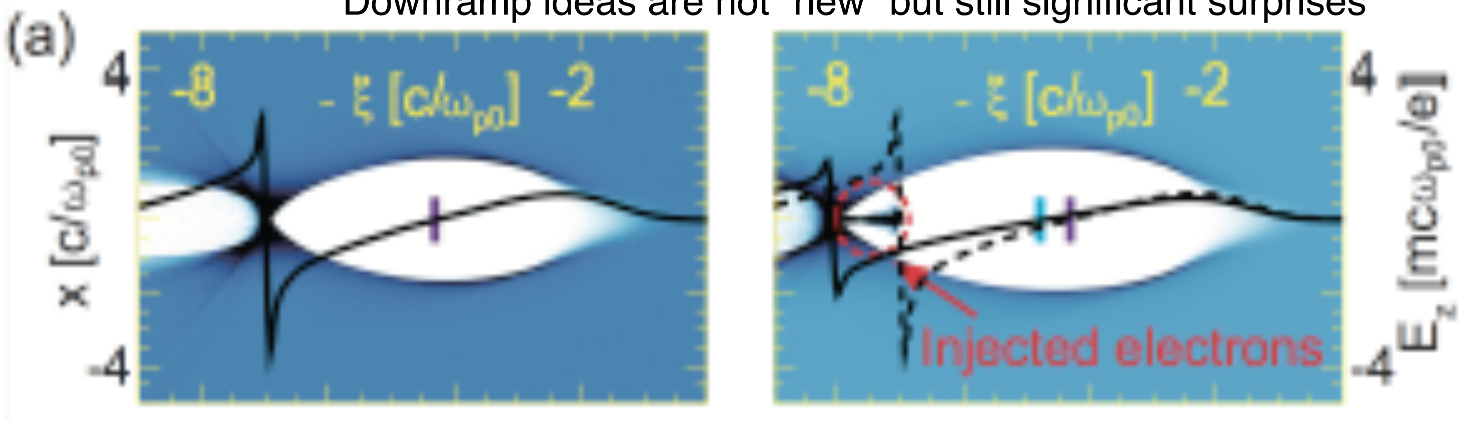




Optimizing lasers for fixed laser energy

OSIRIS has been used to study synchronized injection: downramp and ionization induced

"Downramp ideas are not "new" but still significant surprises



		I [kA]	ε [nm]	Energy Spread [MeV]	B [A/m²/rad²]
Ionization Injection[1]	Trojan horse (TH)	0.3	40	Several	7e17
	Downramp + TH	1	20	2.2	9e18
	Trans. colliding	0.4	8.5/6	0.2, 0.012 (slice)	1.7e19
	Two-color: Long.	0.3	50	1~2	2.5e17
	Two-color: Trans.	0.03	60	1,0.03 (slice)	2e16
Down-ramp Injection[2]	Laser (1e19 cm^{-3})	9	10	0.3	2e20
	Beam (3e18 cm^{-3})	10	30	0.5	2e19
	Beam (1e20 cm^{-3})	10	4	0.2	1.3e21

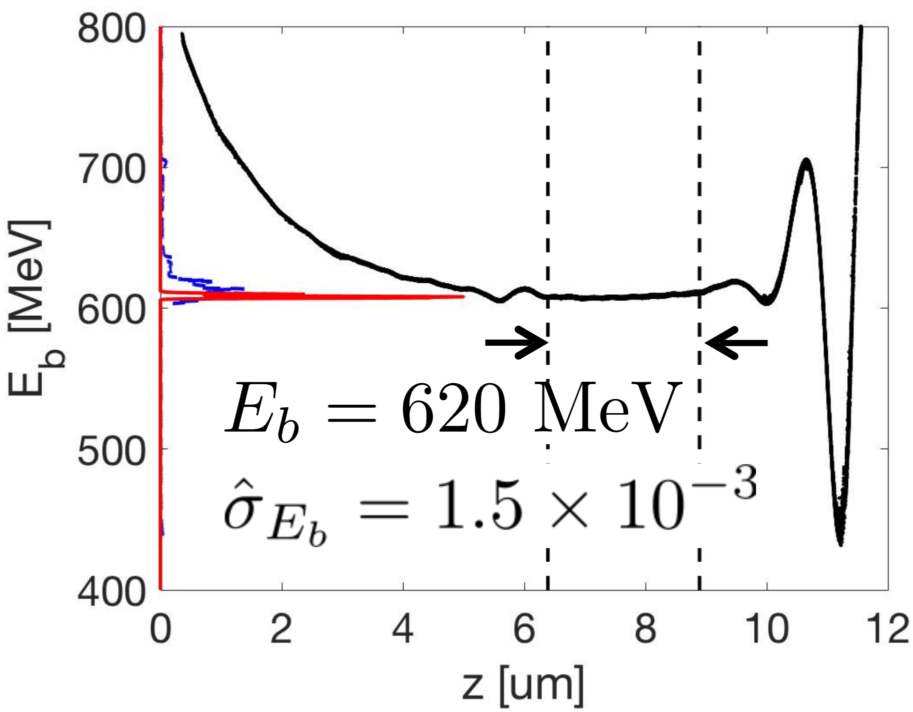
UCLA

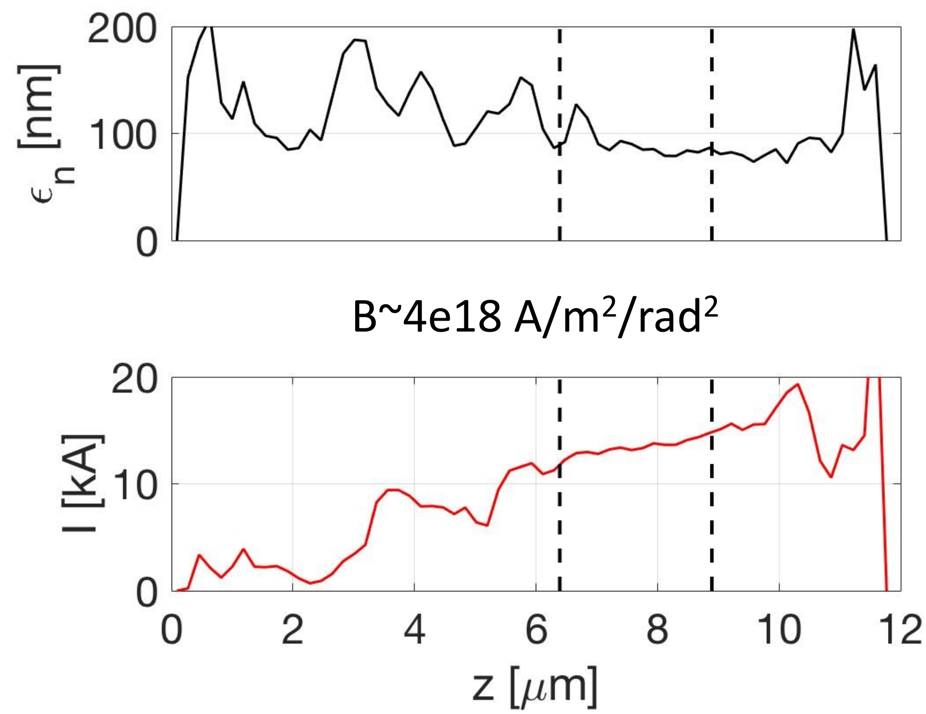
PICKSC PLASMA SIMULATION UCLA

Injector for an X-FEL
High brightness and low energy spread (need to migrate the NCI)

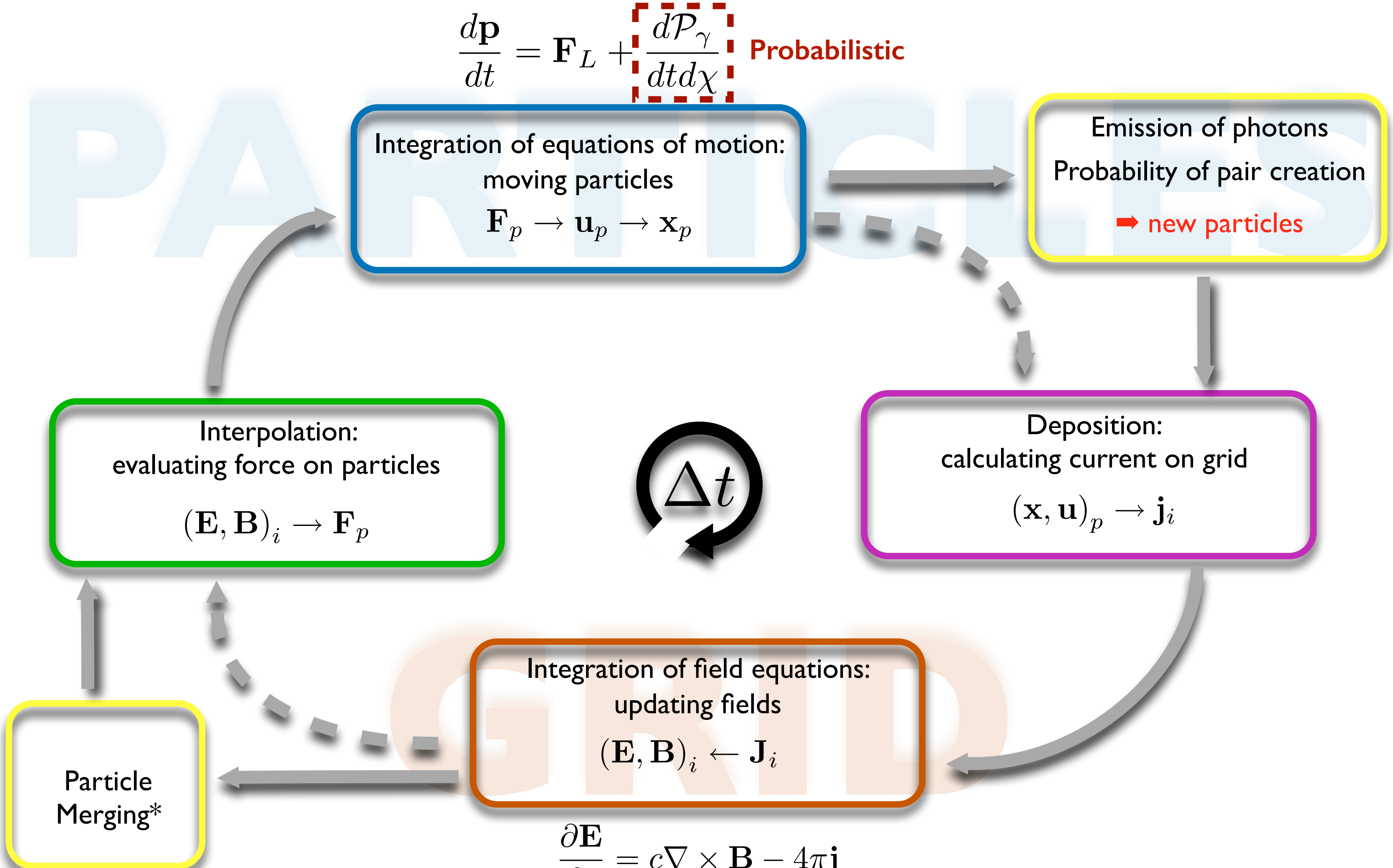
	I [kA]	σ_r [um]	σ_z [um]	ε_n [um]	Q [nC]	E_b [GeV]
Driver beam	34 (Λ=4)	5.3	5.3	5.3	1.5	10

	$n_{p,h}$ [cm^{-3}]	n_{p0} [cm^{-3}]	L_{ramp} [mm]	L_{acc} [mm]	Initial T [eV]
Plasma	1.5e18	1e18	1.33 (250 c/ω_{p0})	3.3	10





OSIRIS-QED PIC LOOP



$$\frac{\partial \mathbf{E}}{\partial t} = c\nabla \times \mathbf{B} - 4\pi \mathbf{j}$$

$$\frac{\partial \mathbf{B}}{\partial t} = -c\nabla \times \mathbf{E}$$

E.N Nerush et al., *PRL* 106, 035001 (**2011**)
C. P. Ridgers et al., *PRL*, 108, 165006 (**2012**)
M. Lobet et al., *PRL* 115, 215003 (**2015**)
A. Gonoskov et al., *PRE* 92, 023305 (**2015**)
*M.Vranic et al., *CPC* 191, 65-73 (**2015**)

The fundamental χ parameter

- **Schwinger field**

$$E_s = \frac{m^2 c^3}{e\hbar}$$

→ **Pair creation probability :**

$$W \propto \exp(-\pi E_s/E)$$

- **Let us introduce the parameter**

$$\chi = \frac{E}{E_s}$$

- **And generalized in any frame**

$$\chi = \frac{1}{E_s}\sqrt{(\gamma \mathbf{E} + \frac{\mathbf{p}}{mc} \times \mathbf{B})^2 - (\frac{\mathbf{p}}{mc} \cdot \mathbf{E})^2}$$

$$\chi \simeq \frac{\gamma E_\perp}{E_s}$$

Other configuration with lower E should allow pair creation ! ←

Disruption regimes for e⁻e⁺ colliding beams

Disruption parameter

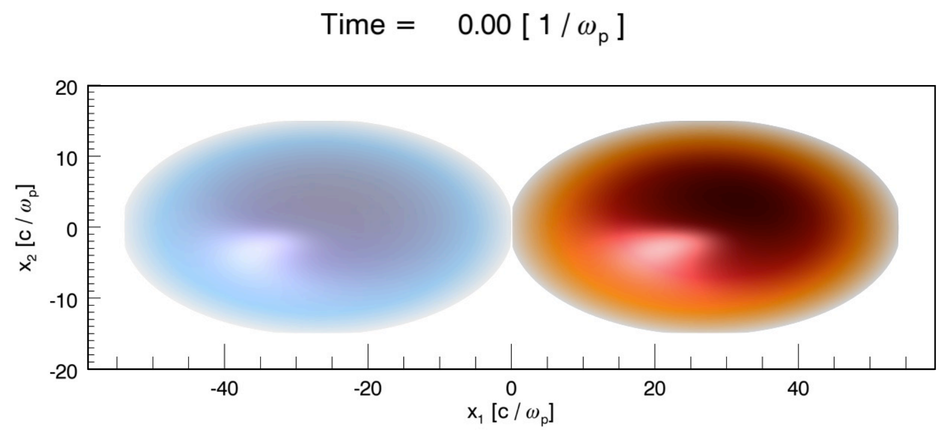
The disruption parameter relates to the number of pinching points of the beams during their interaction time

$$E_\perp \simeq B_\perp \sim e n_0 \sigma_0$$

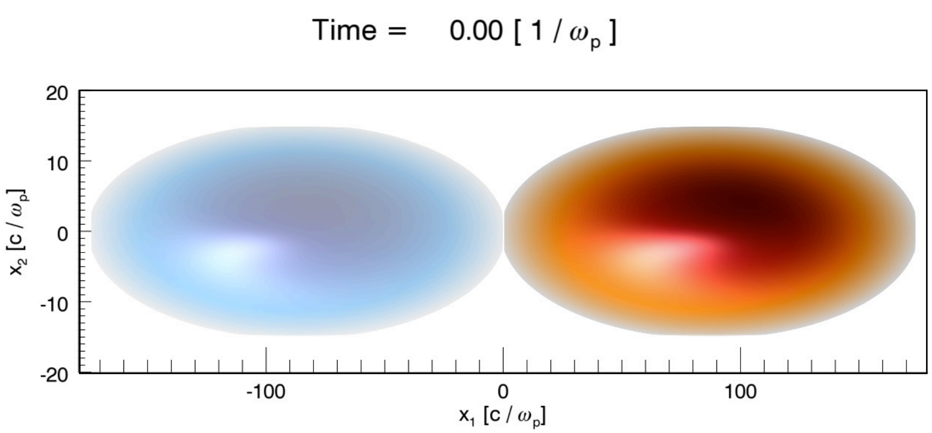
$$E_\parallel \sim E_\perp / \gamma$$

$$D = \frac{r_e N \sigma_z}{\gamma \sigma_0^2}$$

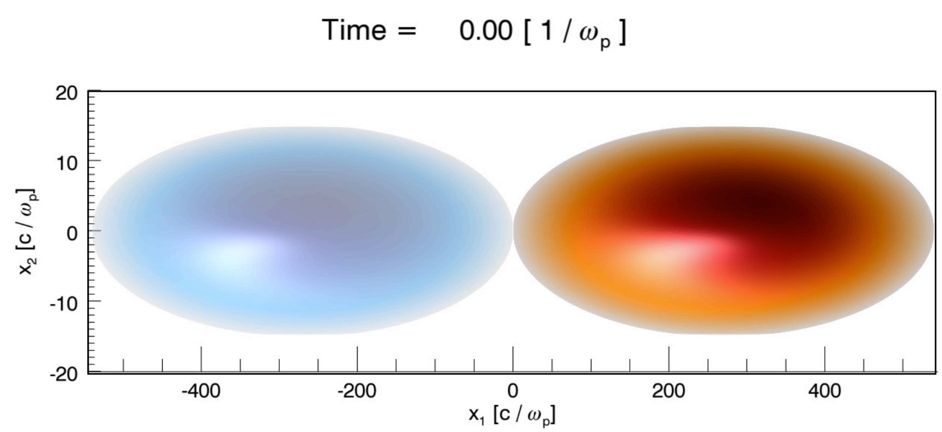
Low disruption regime D<1



Transition regime 1<D<10



Confinement regime D>10



Electron beam density

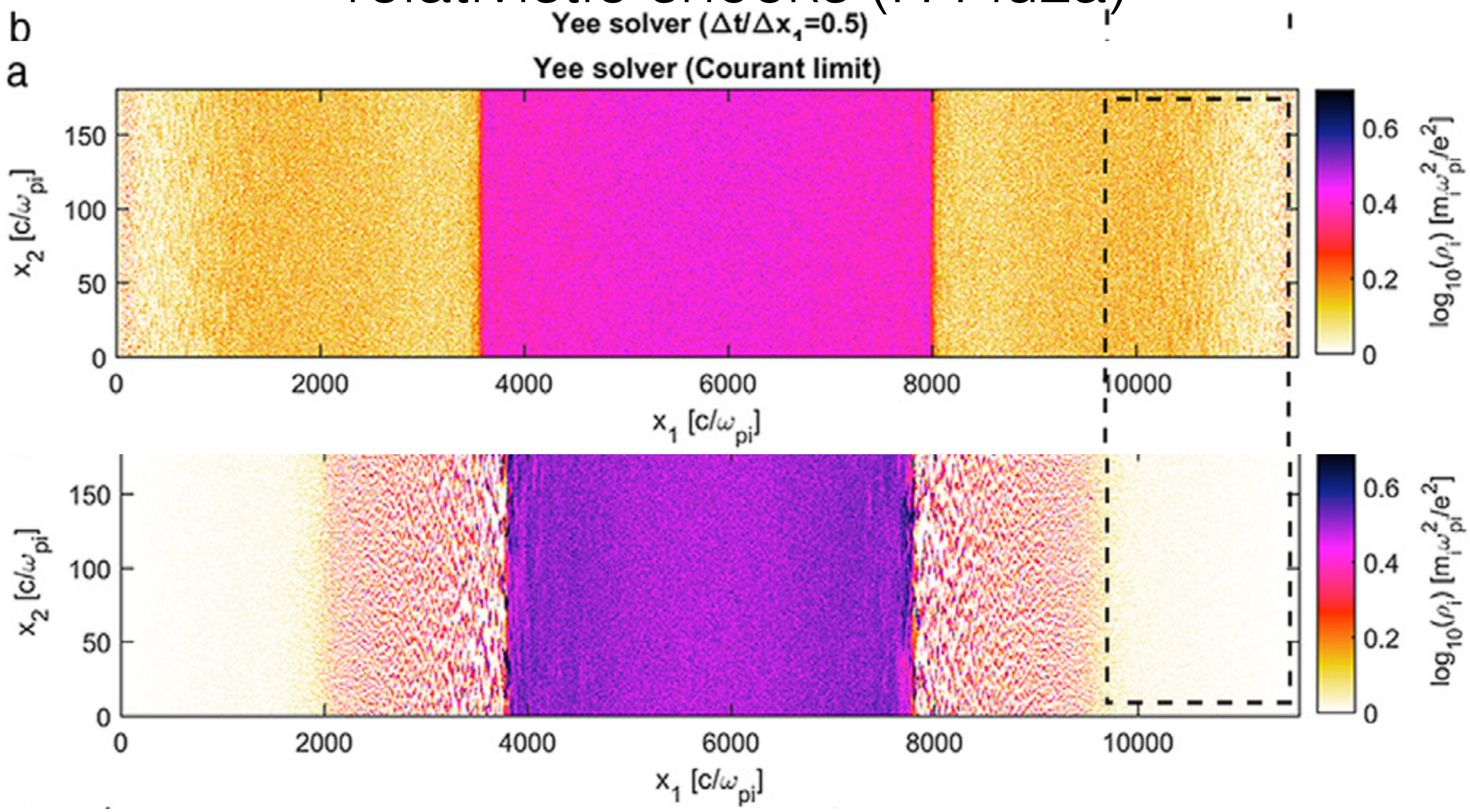
Positron beam density

Next disruption plus beamstrahlung: Copius pairs can be produced

UCLA

Need to remove numerical instabilities for high fidelity simulations: NCI is an issue

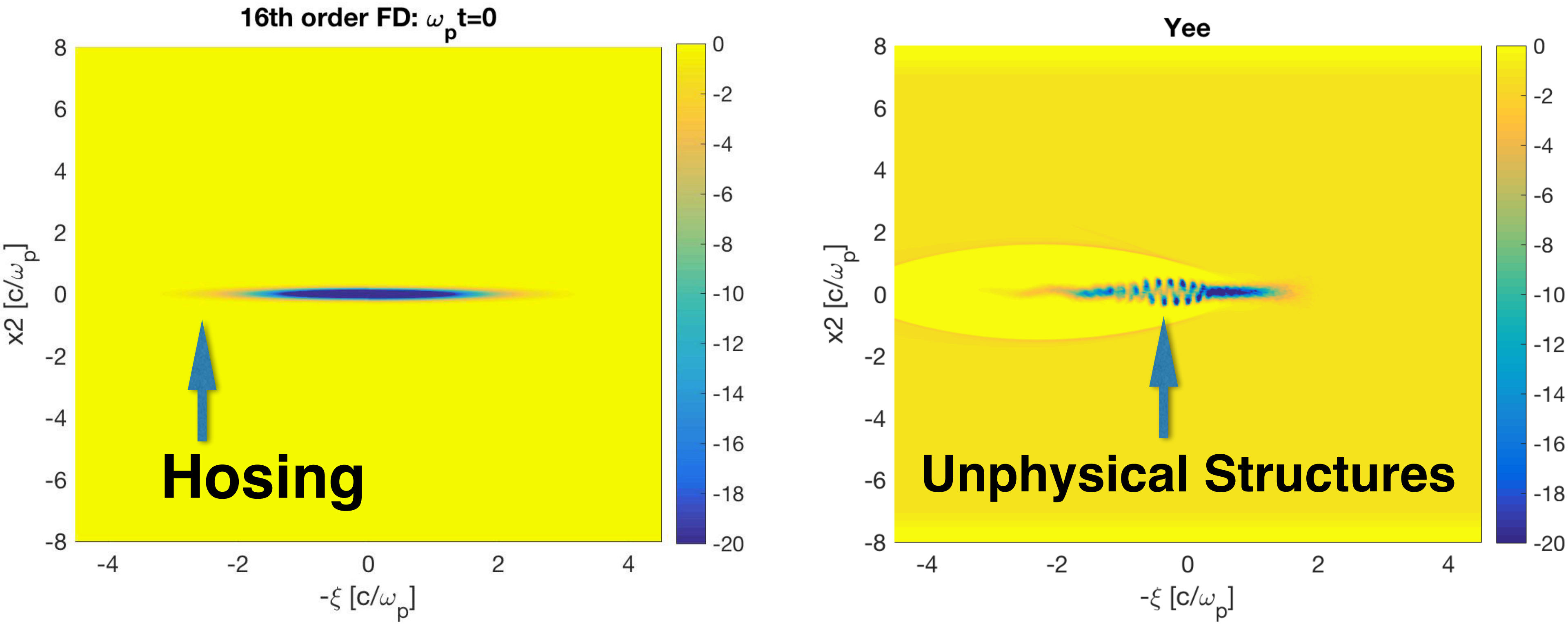
Two counter-streaming plasma flows ($\boldsymbol{\gamma}$=20): relativistic shocks (F. Fiuza)



Fei Li et al., Computer Physics Communications 214, 6 (2017).

NCI example
Hosing in Beam Plasma Wakefield Acceleration

A relativistic beam (1GeV) propagates into rest plasma.



UCLA

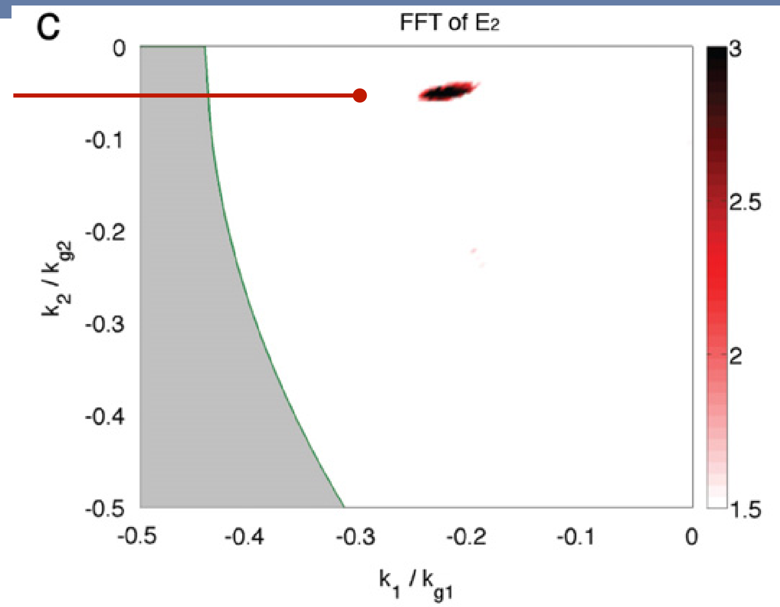
Elimination of (0,0) NCI in FFT based Solver

How to eliminate (0,0) modes?

$$k_1 \approx \pm 1.9 \left(\frac{\omega_p^2}{\gamma}\right)^{1/4} dt^{-1/2}$$

$$k_2 \approx \pm 1.4 \left(\frac{\omega_p^2}{\gamma}\right)^{1/2}$$

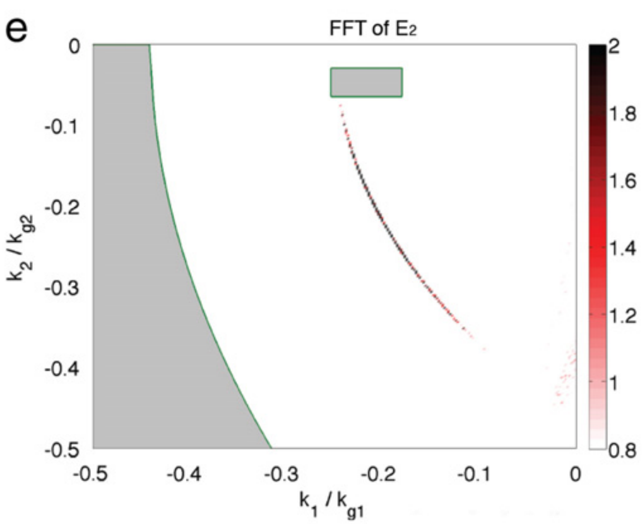
$$\Gamma \approx 0.12 \left(\frac{\omega_p^2}{\gamma}\right)^{3/4} dt^{1/2}$$

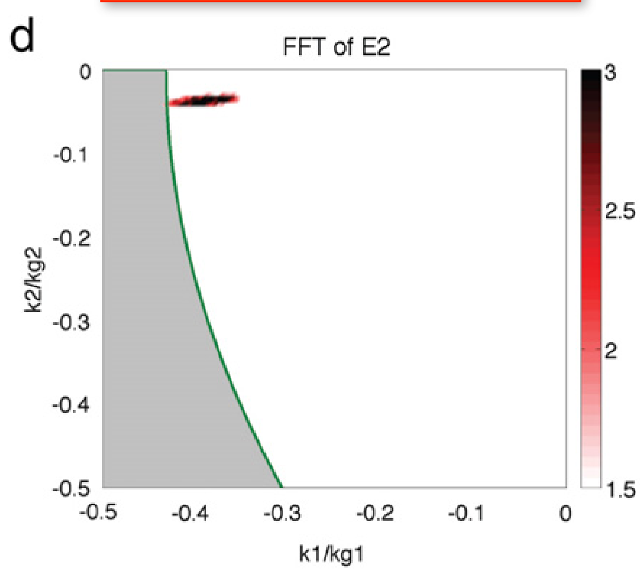


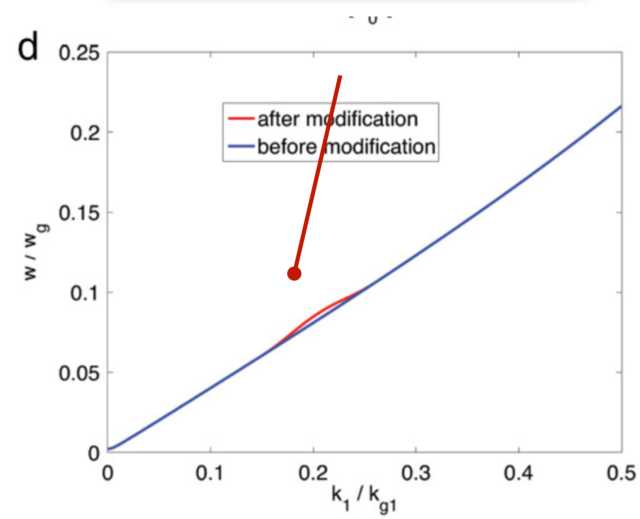
Direct filtering

Reduced time step

Modify EM dispersion



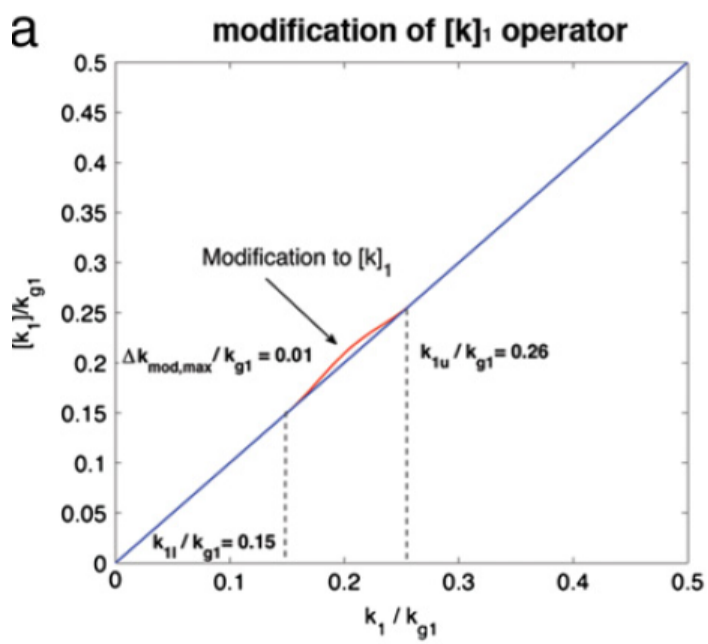




customized high-order FDTD solver

UCLA

Introducing a bump to the dispersion relation by using more coefficients



$$[k_1]_p = \sum_{l=1}^{p/2} C_l^p \frac{\sin[(2l-1)k_1\Delta x_1/2]}{\Delta x_1/2} \Longrightarrow [k_1]_{p*} = \sum_{l=1}^{M} \tilde{C}_l^p \frac{\sin[(2l-1)k_1\Delta x_1/2]}{\Delta x_1/2}$$

Minimize: $$F_1 = \int_0^{1/2} \left([k_1]_{p*} - [k_1]_p - \Delta k_{\text{mod}}\right)^2 dk_1$$

constrain: $$\mathcal{M}\vec{\tilde{C}}^p = \hat{\vec{e}}_1$$

where $\mathcal{M}_{ij} = (2j-1)^{2i-1}/(2i-1)!$ $(i = 1, ..., p/2)$ and $(j = 1, ..., M)$

$\vec{\tilde{C}}^p = (\tilde{C}_1^p, ..., \tilde{C}_M^p)^T$, $\hat{\vec{e}}_1 = (1, 0, ..., 0)^T$

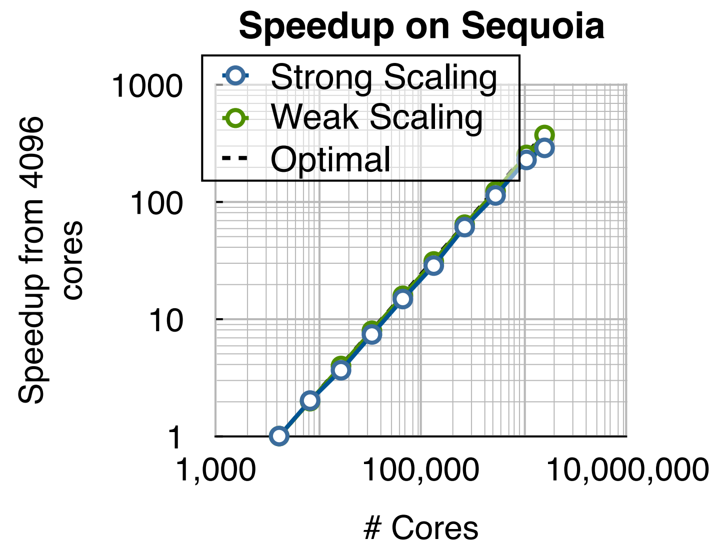
It's a constrained least-square minimization problem which can be solved using the Lagrange multipliers.

current correction

$$\tilde{J}_1^{n+\frac{1}{2}} = \frac{[k_1]_2}{[k_1]_{p*}} J_1^{n+\frac{1}{2}}$$

Fei Li et al., Computer Physics Communications 214, 6 (2017).

OSIRIS 4.0: Open access through an MoU



osiris framework

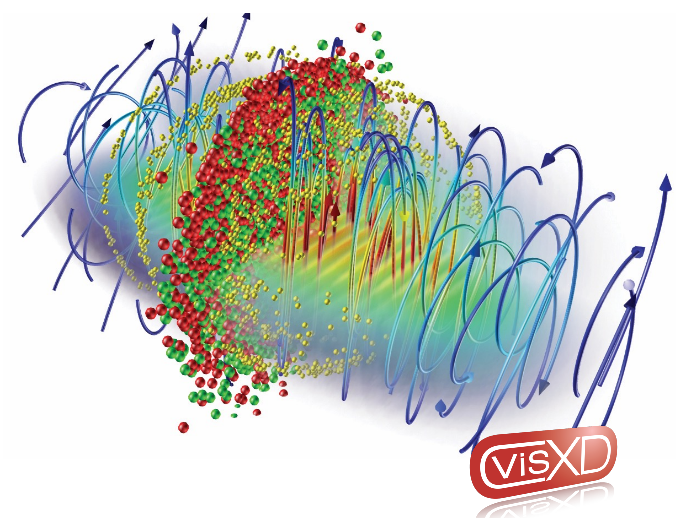
- Massivelly Parallel, Fully Relativistic Particle-in-Cell (PIC) Code
- Visualization and Data Analysis Infrastructure
- Developed by the osiris.consortium
  - ⇒ UCLA + IST

code features

- Scalability to ~ 1.6 M cores
- SIMD hardware optimized
- Parallel I/O
- Dynamic Load Balancing
- Collisions
- Field ionization
- QED module
- Particle splitting/merging
- Quasi-3D
- Customized solver/Boosted frame
- GPGPU support
- Xeon Phi support

Ricardo Fonseca: ricardo.fonseca@tecnico.ulisboa.pt
Adam Tableman: tableman@physics.ucla.edu
Frank Tsung: tsung@physics.ucla.edu

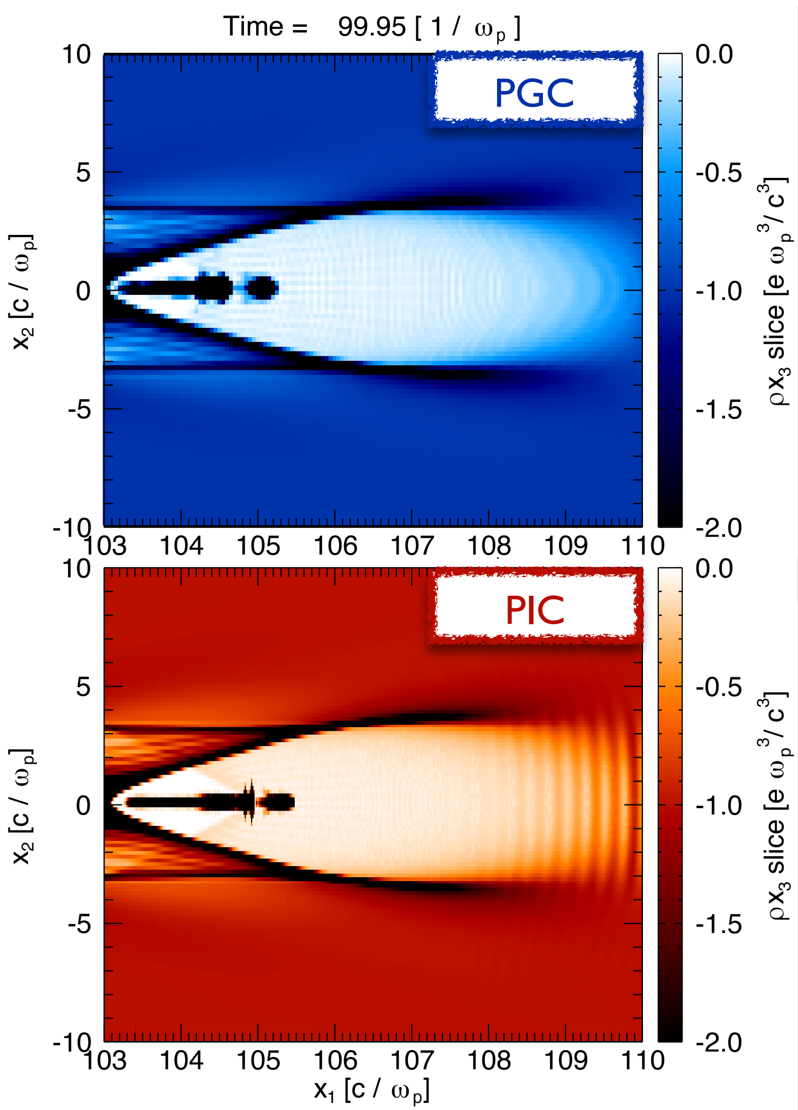
http://epp.tecnico.ulisboa.pt/
http://picks.idre.ucla.edu/



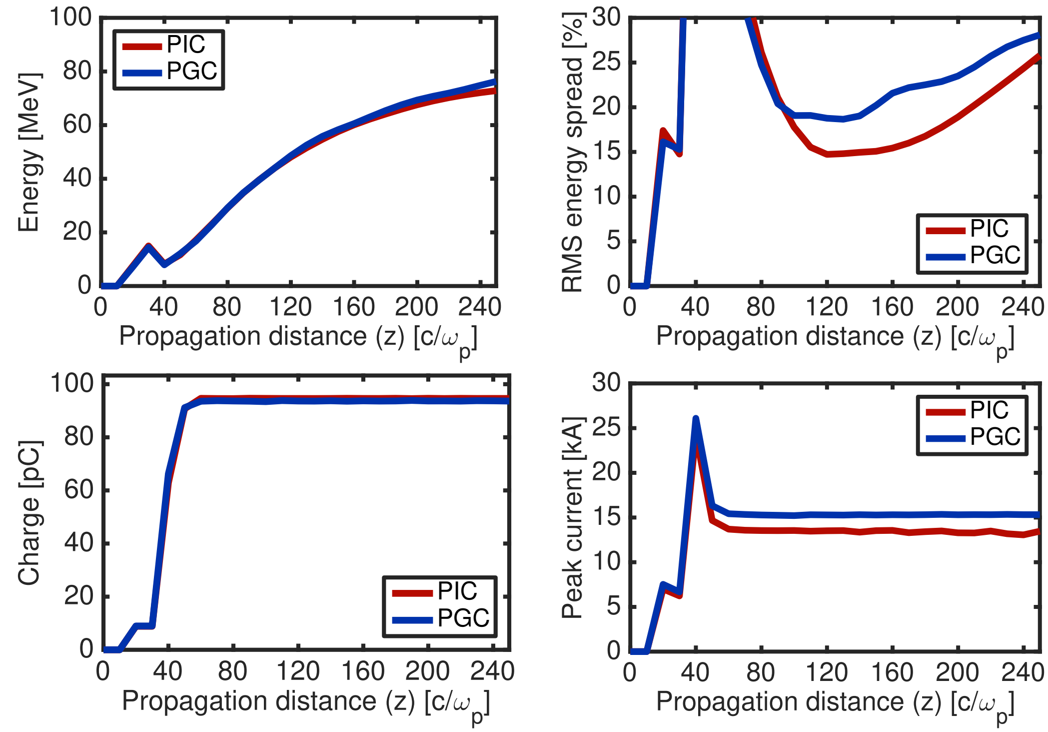
convergence of PGC algorithm for down-ramp injection

TÉCNICO LISBOA

plasma density (PGC/PIC)



beam parameters



- ✦ energy and charge of the injected bunch for PGC simulations agrees with full PIC simulations
- ✦ peak current and energy spread are overestimated by PGC
- ✦ due to coarser grid, macro particles in PGC have represent more electrons which leads to reduced statistics
- ✦ reduced statistics leads to more error prone variance
- ✦ it can be overcome by increasing the number of particles per cell