Dream beams: Extreme-scale computing enabling new accelerator technologies for the energy and intensity frontiers

W.B. Mori,[1] W. An,[1] V.K. Decyk,[1] W. Lu,[1] F.S. Tsung,[1] R.A. Fonseca,[2] S.F. Martins,[2] J. Vieira,[2] L.O. Silva,[2] M. Chen,[3] E. Esarey,[3] C.G.R. Geddes,[3] W.P. Leemans,[3] C.B. Schroeder,[3] J.-L. Vay,[3] K. Amyx,[4] D.L. Bruhwiler,[4] J.R. Cary,[4] E. Cormier-Michel,[4] B. Cowan,[4] P. Messmer,[4] C. Huang,[5] and T. Antonsen[6]

[1] University of California, Los Angeles, CA 90095, USA
[2] Instituto Superior Tecnico, Lisbon 1049-001, Portugal
[3] Lawrence Berkeley National Lab, Berkeley, CA 94720, USA,
[4] Tech-X Corporation, Boulder, CO 80303, USA
[5] Los Alamos National Laboratory, NM 87545, USA
[6] University of Maryland, College Park, MD 20742, USA

Email: mori@physics.ucla.edu

Abstract: For the past 80 years, the instrument of choice in experimental high-energy physics has been particle accelerators, with the newest being the Large Hadron Collider (LHC) at CERN. The construction cost alone for the LHC machine is nearly 10 billion dollars. It is clear that if the same technology is used then the world's next atom smasher" at the energy frontier will cost at least several time that in today's dollars, making it unlikely to be built. One might ask what to do next? One way is to improve the intensity of the particle beams to be collided against each other and keep their energy constant, while another is to develop new accelerator technologies that could reduce the size and cost of a future higher energy machine. In this article we discuss how large scale simulations using state-of-art codes have played and will continue to play a central role in the development of plasma-based acceleration which is currently the leading candidate for a "new" accelerator technology. We will describe the current status of algorithms and codes including recent developments, how these codes have led to scientific discovery, how they were used to approve to new experimental facilities (FACET and BELLA), how they are being used to develop collider concepts based on plasma-based acceleration, and what advances in simulation capability can be expected in the near future.

1. Introduction

Particle accelerators are critical to scientific discovery in the United States Department of Energy (DOE) program and indeed the world. Of the 28 facilities listed in the 2003 DOE report Facilities for the Future of Science: A Twenty-Year Outlook, 14 involve accelerators [Abraham03]. The development and optimization of accelerators are essential for advancing our understanding of the fundamental properties of matter, energy, space, and time, and for enabling research in aspects of materials science, chemistry, geosciences, and bioscience. Modeling of accelerator components and simulation of beam dynamics are necessary for understanding and optimizing the performance of existing accelerators and for optimizing the design and cost effectiveness of future accelerators.

In the next decade, the High Energy Physics (HEP) community will explore the energy frontier by operating and upgrading the Large Hadron Collider (LHC), by developing novel concepts and technologies necessary for the design of the next lepton collider, and through newly approved operating facilities such as BELLA and FACET at which R&D for these new acceleration technologies will be undertaken. It will also be exploring the intensity frontier by designing high intensity proton sources for neutrino physics and rare process program uses accelerators to study the properties of nuclear matter and the structure of the nucleus, understand the mechanism of quark confinement, and create and study the quark-gluon plasma state. The flagship DOE Nuclear Physics (NP) accelerators are CEBAF at Jefferson Lab and RHIC at Brookhaven National Laboratory (BNL). The worldwide nuclear physics community has identified the construction of a polarized electron-ion collider as a long-term objective in the ongoing effort to understand and answer questions in these areas. DOE/NP is considering two approaches for such a facility: e-RHIC and ELIC. DOE/NP is also constructing a facility for rare isotope beams (FRIB) that will permit studies of nuclei far from stability that promise to radically improve our understanding of atomic nuclei.

Today's accelerators are complex, large, and expensive. For example, the LHC is 30 kilometers in circumference and cost over 10 billion dollars to build. Designing and developing a new accelerator or upgrading an existing one is not only time consuming but is also expensive. Computer simulations have the potential to greatly reduce the time and cost it takes from the conceptual design of an accelerator to operation.

In this article, we will concentrate on the role that simulations have played and will continue to play in the development of a new accelerator technology for a lepton collider at the energy frontier. This technology is plasma-based acceleration [Joshi06, Esarey96]. In plasma based acceleration a short and intense laser or relativistic particle beam (the driver) propagates through a plasma near the speed of light. The light pressure of the laser or the space charge forces from the particle beam displaces plasma electrons. The ions then pull the electrons back towards where they started thereby creating a plasma wave wake with a phase velocity near the speed of light. Schematics of both are shown in Figure 1. The accelerating (electric) field in these wakes are more than 1000 times higher than those in existing accelerators. Properly shaped and phased electrons or positron beams (witness beams) are loaded onto the wake and they surf to ultra-high energies in very short distances.

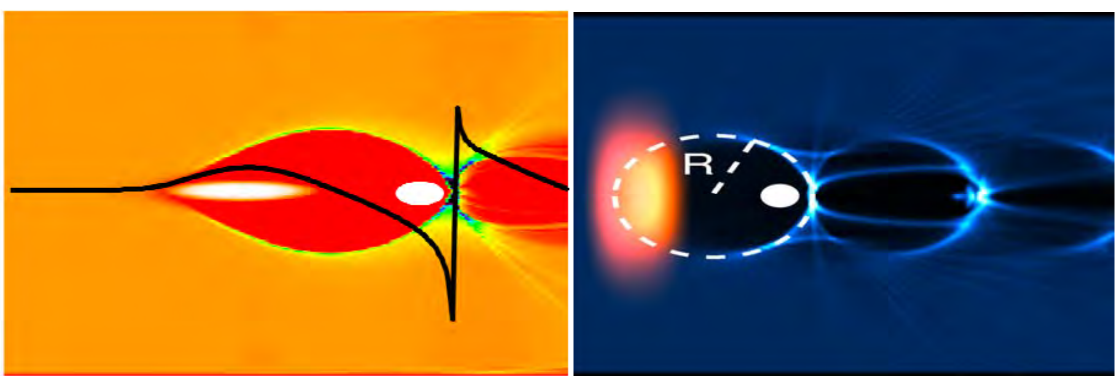

Figure 1. On the left is an electron beam (white) moving from right to left. It forms a wakefield (density of plasma is shown. A lineout of the accelerating field is shown in black. A trailing bunch is shown in white in the back of the wakefield. On the right a laser (orange) is moving from right to left. It also creates a wakefield. The wakefield in both cases is a moving bubble of a radius R. A trailing beam is shown in white as well.

2. Hierarchy of Algorithms and Codes

From the very beginning computer simulation has played a central role in the advancement of plasma based acceleration. The very first paper on the subject (it was using a laser as the driver) [Tajima79] was completely based on simulations (one dimensional), as the laser technology did not exist (in fact one could argue that the proper type of laser did not exist until a few years ago). Since that time, simulations have moved from being used to test new ideas, to modeling the full scale of experiments in three dimensions, to being used to design new facilities, and to being used to test key parts of a collider concept for parameters not yet accessible in a laboratory. As in most fields this was achieved through a combination of improved and more sophisticated code, parallelization of code, new algorithms including reduced models (that were developed through an improved understanding of the theory), and faster and larger computers. Much of this work was done under the SciDAC program including the current COMPASS project. As we will argue later, new hardware and new software developed for the hardware, together with improved algorithms, and workflows will lead to simulations being used for real-time steering of experiments, being used with genetic algorithms to find optimal laser and particle beam profiles etc, as well as simulations which integrate plasma stages with other components in a complex accelerator.

To accurately simulate this process one needs to model how a short and intense driver evolves over "large" distances, how the wake is excited and how it evolves, and how the properties of the witness beams evolve as they are accelerated. In many cases the excitation of the wake is nonlinear so a particle or kinetic description is needed. The leading kinetic description is the particle-in-cell (PIC) method. In some cases fluid models for the plasma response may be useful, however PIC algorithms are now getting very fast and reduced PIC models can have very little noise. It is worth noting that currently ~12% of INCITE awards are for applications using the PIC algorithm. Much of the national expertise for optimizing the basic PIC algorithm on current and future extreme computing platforms resides within the COMPASS project.

There is now a well established hierarchy of methods for modeling plasma based acceleration (many developed within SciDAC). Here we quickly review the hierarchy of methods. More details can be found in several references. The most "complete" method is the fully explicit, full Maxwell solver, fully relativistic, PIC method [Birdsall91]. In this method, a large number of macro-particles are initialized with positions within a grid and momenta. Each timestep, the currents (and in some cases the charge density) are "deposited" onto the appropriate corners or segments of the grid. Maxwell's equations are then advanced forward to calculate new electric and magnetic fields. This is most commonly done using the finite difference time domain (FDTD) method for which the fields need to be appropriately centered with respect to the cell (some are located at corners, some at the midplane of the cell boundaries, while others are defined in the middle of the cell) to achieve second order accuracy. This is commonly done using the Yee mesh. The fields can then be used to calculate the force on each particle so that their momentum and position can be advanced forward to the next time step after which the current can be determined and the whole sequence can be repeated. This "loop" can then be repeated the desired number of timesteps. The cell size is chosen to resolve the smallest scale length of physical relevance and the time step is then determined from the Courant condition. The shortest length of relevance is usually the laser wavelength for a laser driver and the wake's wavelength for a particle beam driver. In some cases the charged particles can radiate x-rays as they are accelerated. The wavelength of the x-rays is far too short to be resolved, and can be handled by secondary algorithms. While this algorithm is straightforward to explain, production codes are now very complex due to parallelization, single core optimization, various choices for how to deposit the current, higher order particle shapes, complex boundary conditions, diagnostics, and physics packages required to realistically model phenomena of interest.

The next method within the hierarchy takes advantage of the fact that a short pulse laser does not backscatter. In this limit, the laser pulse can be separated out into a central frequency (corresponding to a forward moving wave) and an "envelope" which is a complex number describing the pulse shape and length. Only the evolution of the envelope (both amplitude and phase) for the laser is calculated and it is explicitly assumed that the time scale that the laser evolves over is much longer than a laser period (in fact longer even than a plasma wake period). Therefore, the time steps and grids used to advance the laser envelope can now be larger than the laser period and wavelength. The particles still oscillate at the laser frequency and wavenumber. However, one can further separate out the fields and forces into plasma wakefields and laser fields and then average over the laser frequency and wavenumber (assuming there is no backscatter). The result is that the force on the particles has the usual Lorentz force from the plasma wakefields and a ponderomotive or light pressure force from the laser. This is idea is called the ponderomotive guiding center (PGC) approximation [Mora97, Gordon00]. . The potential savings is very large. The cell size can be $\sim\omega_o/\omega_p$ larger since now the wake and not the laser wavelength needs to be resolved. The time step is also larger by the same factor (the Courant condition sets the time step to be the cell size divided by the speed of light). Since one wants to run the simulation for the same physical distance then the number of required time steps is therefore reduced by $\sim\omega_o/\omega_p$. Assuming the same number of particles per cell are used and the speed of the algorithm per particle step remains unchanged then this technique can lead to a speed up of $\sim(\omega_o/\omega_p)^2$. As we will show later such speed ups are being obtained. One potential issue is that as the laser propagates deeply into the plasma it pump depletes. As this occurs the central frequency decreases. Therefore over time the envelope approximation breaks down and methods to get around this issue are currently under development [Cowan10]. The actual speed up can be lower since one might need more particles per cell in the PGC because the cells are large, and because as the longitudinal cell size is reduced it becomes similar to the transverse cell size such that the actual time step is reduced by factors of $2^{1/2}$ and $3^{1/2}$ because of the difference in the Courant condition in 2d and 3d respectively. In addition, the PGC averaging procedure breaks down for very intense pump strengths.

Another method that takes advantage of the fact that there is no backscatter was recently proposed [Vay07]. In this method, one does the simulation is a Lorentz boosted frame (Figure 2). In a Lorentz frame moving at a speed near the speed of light with the laser, the laser will appear Lorentz expanded by a factor $(1+v_f/c)\gamma_f$ where $\gamma_f=(1-v_f^2/c^2)^{-1/2}$ and v_f is the velocity of the frame and c is the speed of light. The plasma (now moving towards you) is now Lorentz contracted (by a factor γ_f). In this frame the shortest wavelength is still the laser wavelength (so long as there is not light reflected as this light will have a very short wavelength in this frame) which has also been Lorentz expanded, and the number of wavelengths within the laser is a Lorentz invariant, i.e., it is the same in every reference frame. So although the laser is longer the number of cells required to resolve it does not change and the number of cells used for the simulation does not change. In physical units the time step is now larger [by $\sim(1+v_f/c)\gamma_f$]. The required number of time steps is therefore smaller by $(1+v_f/c)^2\gamma_f^2$ [a factor (1+v/c) arises because the pulse and plasma are moving towards each other]. The optimal frame to transform into is one moving near the group velocity of light within the plasma, $v_g=(1-\omega_p^2/\omega_o^2)^{1/2}$. In this frame $\gamma_f=\gamma_g=\omega_o/\omega_p$ such that the speed up once again scales as $\sim(\omega_o/\omega_p)^2$. And just like in the PGC method, the speed up is not quite as high as the above estimate since in the group velocity frame, the cells become nearly squares or cubes in 2d and 3d so that the time step needs to be slightly smaller. During the past year, substantial progress on this technique has been made [Martins10a, Martins10b, Huang09, Vay10]. This progress has involved working around a numerical instability that is still not completely understood. A sweet spot has been found that requires substantial smoothing which, based on convergence tests, appears to be fine for large γ_f . This technique is attractive because in principal it should work for full pump depletion distances and for arbitrary pump strengths. However, one must use smaller γ_f when investigating self-trapping, that is when the wake gets so large that background plasma electrons become self-injected into the wake and are accelerated. This is because the plasma is represented by γ_f fewer particles which can lead to a statistical issue. The PGC method has also been used for self-trapping but with its own limitations mentioned above. In addition, thinking about the physics in different inertial frames can help one better understand the physics.

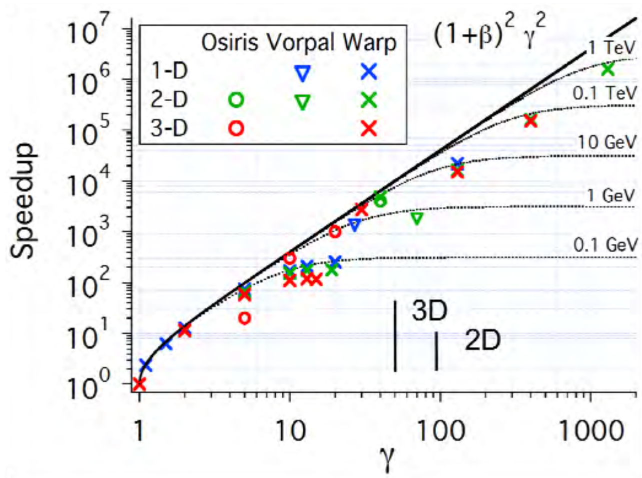

Figure 2. Plot of speed ups obtained do date using the boosted frame technique. Note that the OSIRIS simulations were for self-trapping and beam loading simulations using high intensity lasers. The Warp simulations were for test particles in a quasi-linear regime.

The last method in the hierarchy is the quasi-static approximation (which is used together with the PGC when using a laser driver). The quasi-static approximation [Mora97, Huang06] exploits the fact that during the time it takes the laser or particle beam driver to pass by a plasma electron the laser and particle beam appear static, i.e., neither their shape nor, their energy distribution (for a laser, this the frequency content) change. Under this approximation, one can calculate the wake for a given driver shape and then use the potentials and fields within this wake to get source terms to advance the laser or particle beam forward with time steps that resolve the time scale that the driver evolves over. These time scales are much larger than a plasma period. A laser driver typically evolves over a Rayleigh length, $k_0\ w_0^2/2$ while for a particle beam evolves over the betatron period $k_p/(2\gamma)^{1/2}$ (the period that particles oscillate transversely in the wakefield). For a laser driver the speed up on paper could therefore be the product of a factor of $\sim(\omega_o/\omega_p)^2$ from the use of PGC and an additional factor of $\sim(\omega_o/\omega_p)k_p^2w_0^2/2$, i.e., $\sim(\omega_o/\omega_p)^3k_p^2w_0^2/2$ from the quasi-static approximation. However, accuracy and numerical stability issues keep the time step for the laser solver to $\sim\sim(\omega_o/\omega_p)k_p\Delta$ where Δ is the transverse cell size. For a particle beam driver the speed up scales as $(2\gamma)^{1/2}$. Furthermore, the algorithm requires a predictor corrector loop so the field solve and particle push has additional overhead. This technique has already been used to verify some scaling laws out to 100 GeV [Tzoufras08b, Geddes08. Huang09]. An advantage of the quasi-static model is that it is not susceptible to numerical Cerenkov radiation so smaller transverse cell sizes can be used for narrow trailing bunches. In addition, mesh refinement should be easier to implement in principle. Calculating radiation reaction is also more straightforward. A disadvantage is that no prescription for handling self-trapping is presently available.

As the above indicates there is no one code and algorithm that fits all for modeling plasma-based acceleration. Currently, there are a suite of production codes that have been supported under the advanced accelerator part of SciDAC, OSIRIS [Fonseca02], VORPAL [Nieter04], and QuickPIC [Huang06]. In addition, another code Warp has been supported for electron cloud work. It has recently been used to develop the boosted frame technique and is starting to be used by some in SciDAC for LWFA simulations. OSIRIS is a full PIC code that also can run using the boosted frame technique primarily to study self-injection. VORPAL is a full PIC code that also can operate with the PGC, with cut cells for complicated boundaries, and is currently being updated to run using the boosted frame. Warp has been used to run laser wakefield simulations using the boosted frame technique (for very high γ in the external injection and weak beam loading regime). QuickPIC is a 3d quasi-static code which is used for both laser and particle beam drivers.

Software verification is an important part of the SciDAC mission. Previously we have reported comparison between all of these codes on a simple set of laser wakefield simulations [Paul08]. In these simulations the plasma wakefields were compared early in time before the laser had evolved.

The laser intensity was varied. The codes OSIRIS, VORPAL (with and without the PGC), Warp, and QuickPIC, and VORPAL all give very similar results. Comparison for longer run times will be undertaken in the following year.

3. Parallel Scalability

For extreme computing, it is not only necessary to have a code or algorithm, it is also important that the code scales well on parallel platforms. There are two types of parallel scaling studies. In strong scaling the problem size remains fixed as one increases the number or cores. In weak scaling the problem size increase in proportion the number of cores so that the size per core remains fixed. For many situations the weak scaling is most relevant since one typically uses extreme computing for large problems. This is generally true for plasma based acceleration. However, the strong scaling study is useful to understand the limitations of ones algorithm. The codes used under this project scale almost perfectly for weak scaling. Each of the codes also scales well for strong scaling so long as the number of particles per core is greater than ~100,000 and the domain size on a node is greater than ~4000 cells. The scaling will depend on how optimized the code is on a single core.

Examples of some recent results include strong scaling studies of OSIRIS on the Jugene Bluegene machine in Germany and the Jaguar computer at Oak Ridge. On Jugene OSIRIS had 81% efficiency out to 294912 cores (entire machine) with 1892 cells and 56000 particles per core at the end. On Jaguar OSIRIS had 60% parallel efficiency for 128000 cores with 4000 cells and 64000 particles per core at the end.

4. Results

In this section we describe some recent results and discoveries obtained from using OSIRIS, VORPAL, Warp, and QuickPIC. This includes helping to provide the scientific basis for two recently approved facilities dedicated to researching issues related to extending plasma-based acceleration towards a collider at the energy frontier.

4.1. Science Discoveries

An important issue for plasma based accelerators is how to simultaneously inject large amounts of charge into a plasma accelerator stage while maintaining high beam quality (emittance and energy spread). One issue with self-injection is that electrons which are injected also have large transverse momentum [Tsung04, Geddes05, Lu2007]. Several of the discoveries are related to understanding how to control injection. Such an understanding may ultimately lead to the injection and acceleration of high quality beams with significant charge. In addition, in the near future there may be 250 Joule short pulse lasers available. Another discovery is related to understanding how choices in the pulse length and spot size could affect the possible output energy and charge from a single stage.

4.1.1. Understanding the acceleration of self-trapped electrons. Experiments and simulations demonstrated in 2004 the formation of the first narrow energy spread and narrow divergence beams from laser plasma accelerators [Geddes04, Mangles04, Faure 04, Tsung04, Pukhov02]. Initial simulations described the essential physics of beam formation from self-trapped electrons, and the physics and optimization of these beams continues to be an important topic. Additional work has led to improved numerical techniques which control unphysical momentum spread [Cormier-Michel08], and quantitative agreement with the experiments [Tsung06, Geddes09] which together with advanced visualization [Ruebel10] allow detailed analysis of the physics of these stages (Figure 3). Tracking particles through the simulations shows that they were injected from the sides through the strong fields of the nonlinear wake, increasing their transverse momentum and hence the beam divergence undesirably [Tsung04, Geddes05, Lu07, Ruebel10]. Simulations also modeled GeV self trapping experiments in cm-scale capillary plasmas [Leemans06; Geddes09], showing that this process

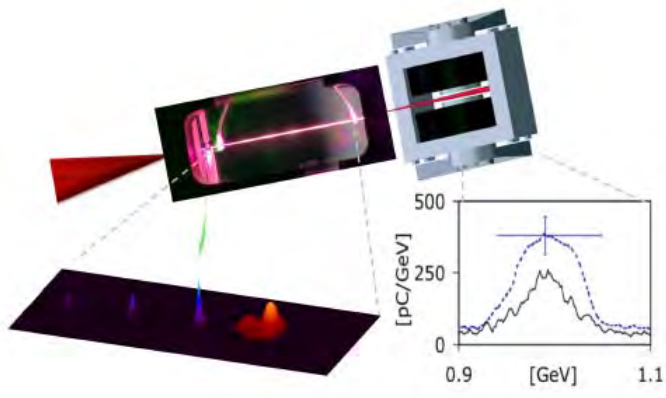

Figure 3. VORPAL simulations model formation of GeV electron beams from LBNL capillary experiments (top) showing the laser-driven wake and particles (left) and reproducing the experimental beam (right) [figure from Geddes09].

scales in a well understood way with plasma density. Such results are motivating next-generation experiments and simulations to control injection of electrons to further stabilize and improve beam quality.

4.1.2. Ionization trapping. One possibility for reducing the transverse momentum of self-injected electrons is to use ionization induced injection [Oz07, Pak10]. For a particle to be injected the change the wake must give it a change in potential that approaches ~mc^2/e. This is challenging if the particle originates outside the wake where the initial potential is 0. However, if the particle can be born inside the wake at a phase (location) where the potential is near a maximum (positive value) then it is easy for the particle to get trapped. It is easier to trap electrons near the axis and one can control where injection occurs by varying the gas mixture along the propagation direction. This allows one to control the emittance and energy spread. This was originally investigated for particle beam drivers [Oz07]. Recently, this idea was extended to laser drivers where the laser field also leaves an imprint on the particles transverse momentum [Pak10]. In the experiment of Pak et al. a mixture of He and N was used. Electrons from the K shell of nitrogen were tunnel ionized near the peak of the laser pulse and were injected into and trapped by the wake created by electrons from majority helium atoms and the L shell of nitrogen. The spectrum of the accelerated electrons, the threshold intensity at which trapping occurs, the forward transmitted laser spectrum, and the beam divergence are all consistent with this injection process. In the experiment it is not possible to conclusively show which shell an electron came from. However, this is possible in simulations. Full-scale 3D OSIRIS simulations showed that the trapped electrons came from the K shell of Nitrogen and they verified the theoretical threshold. Results from OSIRIS for both a beam or laser driver are shown in Figure 4.

4.1.3. Controlling injection to improve beam quality and stability. Further improvement of bunch momentum spread and day-to-day accelerator stability are required for applications such as light sources and high energy physics. Simulations are supporting experiments to develop controlled injection of particles decoupled from the guiding and acceleration process (in contrast to past self trapped experiments) to improve beam quality using tailored plasmas and multiple laser pulses.

By focusing the laser in a decreasing plasma density gradient, beams with longitudinal and transverse momentum spreads an order of magnitude lower than previously observed were produced [Geddes08]. Simulations showed these bunches result from modulation of the plasma wave phase velocity and can function as injectors to improve stability and reduce energy spread of high energy LWFAs [Geddes08, Geddes09], and initial results on density tailoring in a capillary waveguide have

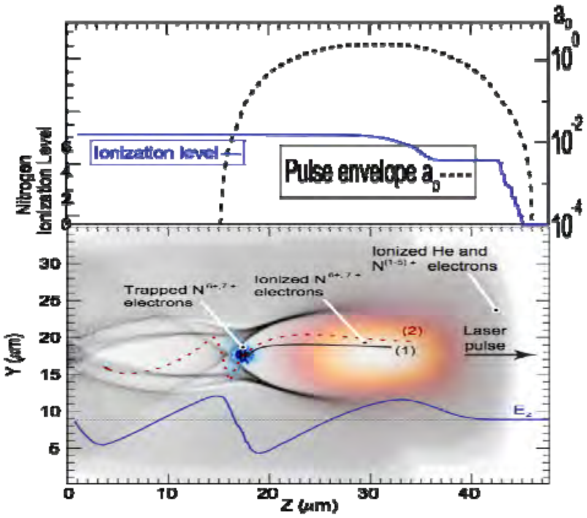

Figure 4a. In the top, a plot of the laser intensity envelope (log scale) showing the ionization level of the plasma mixture. In the bottom is a plot of the laser (orange) and wake (gray made up of mostly He electrons), the trajectory of a trapped (born near the axis) and untrapped K shell electrons as well as the axial wakefield (blue).

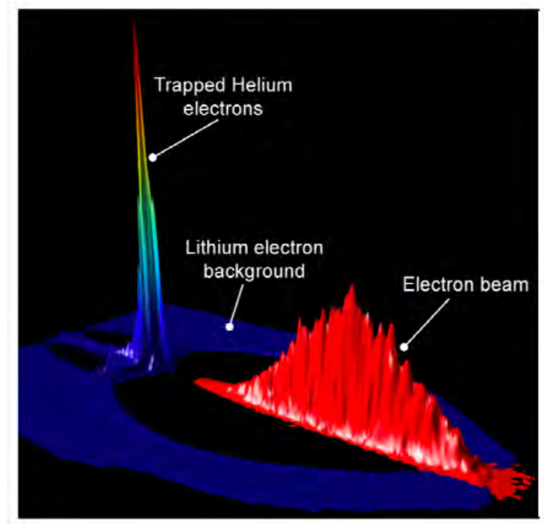

Figure 4b. A plot of the beam (red), the electrons make the wake (blue) and trapped electrons from Helium buffer gas. This OSIRIS simulation is for the E-167 experiment at SLAC.

demonstrated the increased stability predicted [Gonsalves07] and are continuing to tune energy spread and emittance, which is being supported by VORPAL envelope simulations which resolve trapping while allowing rapid 3D simulation required for these experiments.

Colliding laser pulses [Esarey97, Faure06] have been proposed and explored as an all-optical approach to producing stable, tunable, high-quality electron beams. The collision of two laser pulses creates a low velocity beat wave, which can give electrons an initial kick in velocity and position after putting them in phase to be accelerated by the plasma wake of the drive pulse. Self-consistent PIC simulations are important to characterize the dynamics of plasma electrons trapping into the plasma wake, in order to help maximize the trapped charge and minimize beam emittance and energy spread.

Simulations carried out with the parallel VORPAL framework are being used to guide and interpret experiments at the LOASIS Program of LBNL [E. Cormier-Michel10).]. Effects beam timing, angle, and intensity as well as plasma guiding were explored, demonstrating parameters that produce high quality beams. 2D simulations showed comparable results to 3D allowing parameter scans to be performed in 2D, reducing computational time and allowing for more thorough exploration of parameter space. Scaling laws were developed to understand scaling of laser and beam parameters with the plasma density and to allow rapid design of higher energy laser-plasma accelerators which use longer interaction lengths/lower densities.

The simulations are guiding parameters for colliding pulse experiments using the LBNL (0.6 J driver and <0.25 J collider) at 19 angle and show a beam of the order of 20 pC can be injected for a plasma density of ~4.10^{18} cm^{-3}. Recent experimental data have shown good agreement with simulation results including density and timing behavior (Figure 5), demonstrating the ability to quantitatively design such an experiment [Cormier-Michel08]. With optimized guiding of the laser pulse simulations indicate the beam could be accelerated to ~400 MeV with percent level energy spread, and effects of guiding and realistic laser mode have been characterized. Predictive verification against experiments provides confidence in the simulation scans being used to design future experiments to provide beam quality needed for applications.

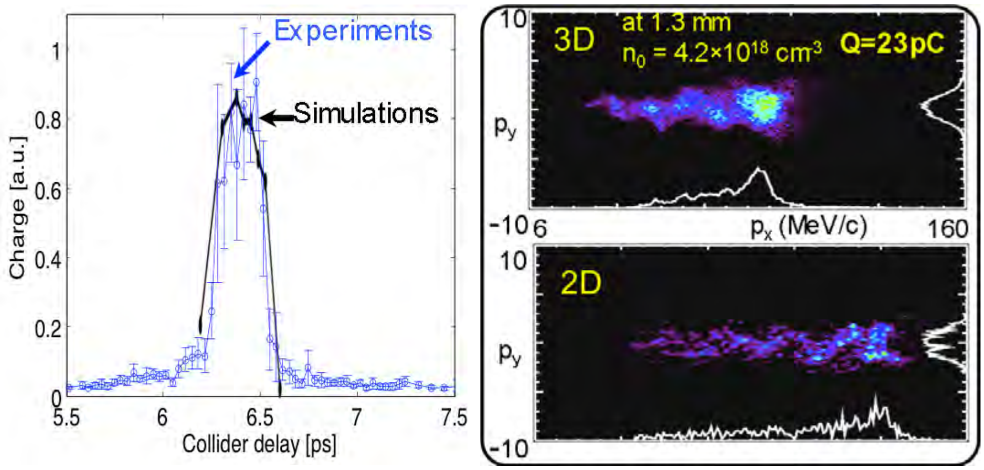

Figure 5. Simulations predicted parameters for colliding pulse experiments, including timing behavior (left), and demonstrate the formation of high quality beams (right, showing agreement of 2d and 3d results). Figure from Cormier-Michel et al, AAC 2008.

4.1.4. Next generation experiments. Since the results of Geddes04, Mangles04, Fuare04, and Tsung04, the past five years have seen an explosion of experimental results on generating mono-energetic electron beams from 100 MeV to 1 GeV [Leeman06, Clayton10]using a plasma wave accelerating structure driven by an intense laser (in fact there are too many to reference them all). The next decade will see tremendous increases in laser power and energy permitting parameter spaces to be accessed that could lead to bunches at the 10 GeV with .1 to 1nC of without external guiding. These new lasers will provide powers above a petawatt and energies at the 100s of Joule levels. They will also permit accessing the fully nonlinear regimes described in the recent theories.

We have been investigating through fully nonlinear particle-in-cell simulations the effects that choices in laser intensity and pulse length as well as plasma density can make for fixed laser energy in the output electron beam energy, charge, and quality [Martins10a]. In Figure 6 we show results from simulations where the laser energy was between 150 and 250 J. The simulation on the left was done in the lab frame (it simulated a 150Joule/30fs laser propagating in a 1.5 10^{19}cm^{-3} plasma, 2.3 GeV electrons were produced in a mm) while the simulation on the right (it simulated a 250J/220fs laser propagating in a 1.5 10^{19}cm^{-3} plasma, 40 GeV electrons were produced in 22cm) was only possible with the use of calculations in a Lorentz boosted frames We have also simulated several experiments including one at LLNL that recently observed 1 GeV quasi mono-energetic electron beams. The results are consistent with the predictions of Lu et al [Lu:07].

4.2. Supporting next generation facilities: BELLA and FACET

4.2.1. BELLA. Next-generation experiments at the BELLA laser facility under construction at LBNL will use controlled injection coupled with meter-scale plasmas to increase bunch quality and increase energy to 10 GeV. A future collider could use many modules in series, each increasing the beam energy to reach TeV energies [Leemans09]. These stages must preserve low energy spread and good emittance for electrons, and also for positrons motivating study of less strongly driven, quasilinear wakes, where dynamics are similar for electrons and positrons and where the shape of the wake is controlled by thelaser profile. With simulation of present cm-scale experiments pushing state of the art computing, new techniques are required. Scaling of the design with plasma density, envelope simulations and Lorentz boosted simulations have been combined to model these stages.

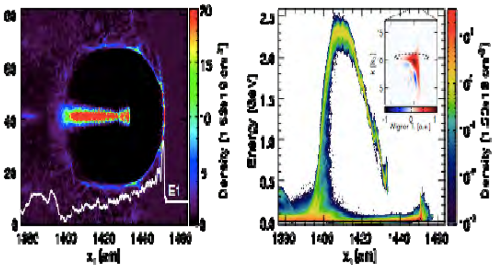

Figure 6a. Results for using a tightly focused 150 J laser at high density. On the left is the plasma density and the trapped particle density as well as a lineout of the accelerating field. On the right is the energy vs. position of the trapped particles. Note that the integrated energy spread is large while the slice energy spread is small.

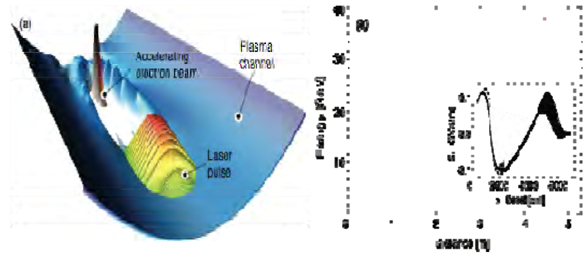

Figure 6b. Results for a wide focus and long pulse regime for a 250J laser propagating in a channel. On the left is laser pulse (yellow and red) propagating in a channel. The blue is the plasma density and the fin in the back is the trailing beam. On the right is the energy vs. time for the beam and the lineout of the acceleration field.

Theory predicts that wake structure will scale with plasma density but does not predict the shape of the wake or how the plasma will focus the laser. A series of VORPAL PIC simulations characterized wake structure and evolution and verified predicted scaling with density, allowing scalable design [Cormier-Michel08]. Hence a short simulation at high density predicts performance for high energy stages and also provides understanding of how stage performance will scale with laser and plasma performance. Variation of laser pulse amplitude, width and length together with plasma longitudinal taper (for beam phase control) and transverse profile (for guiding) established parameters that best deplete the laser energy into the accelerating field of the wake while maintaining quasilinear structure. Electron beam charge, length, and width were next adjusted to obtain efficient acceleration—that is, high transfer of laser energy deposited in the wake to the particles resulting in a design for 10 GeV stages with the Petawatt BELLA laser.

Some parameters such as electron bunch oscillation in the focusing field do not scale with density, requiring simulation of the full meter-long stage at 10^{17}/cc densities, and this in turn requires advanced models which reduce computational load. Over 4 orders of magnitude speedup has been obtained by modeling the laser as a complex scalar envelope modulating a plane wave, enabling unscaled, 3D simulations of full meter-scale stages (PGC code). Recently, the ability to track separately a local laser phase has allowed this envelope model to correctly resolve the laser field into significant depletion. This development is being continued to allow the envelope model to propagate into deep depletion [Cowan10].

Lorentz boosted frame simulations using WARP have been used to verify the scaling of energy gain from 100 MeV to up to 1 TeV [Vay10] on simulations of externally injected low charge beams, thanks to speedups ranging respectively from 100 to over 1 million times (compared to explicit PIC in the laboratory frame). At 10 GeV, the maximum attainable speedup of 10,000 times was achieved, speeding up the modeling of one BELLA stage with explicit PIC from years to hours (assuming a few thousands of CPUs on present day supercomputers) (Figure 7).

Combined simulation with scaled, envelope and boost techniques have been used to detail understanding of 10 GeV stages with few hundred pC charge and low energy spread for both electrons and positrons using a petawatt laser, and how such stages scale for other application such as GeV light sources stages. This checks the models against one another for accuracy, and derives physical

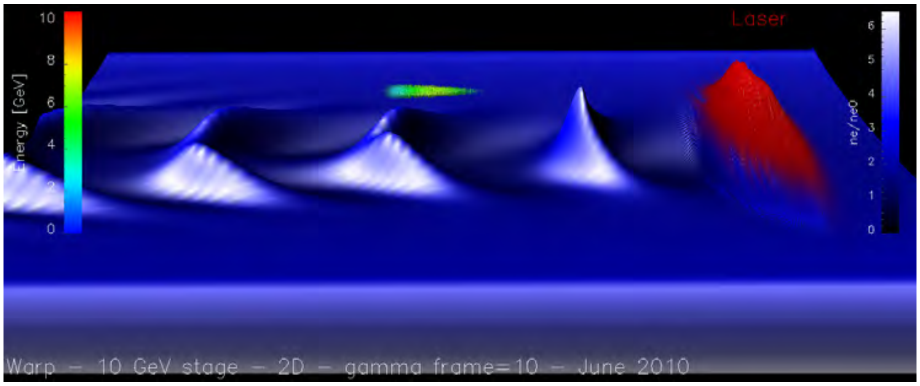

Figure 7. Image from a WARP simulation of a strongly depleted 10 GeV BELLA stage in a Lorentz boosted simulation frame showing the wake (blue) driven by the laser (red) and the accelerated particle bunch.

understanding from the differences each approximation reveals. Future simulations will use the improved computational efficiency demonstrated together with new computing resources to increase resolution to capture and tune the stages to preserve the very low emittances required for applications, and staging to the energy frontier.

4.2.2. FACET. Next generation experiments at the FACET facility which is under construction at SLAC will be aimed at showing that a trailing bunch of electrons can be accelerated with high transfer efficiency while maintaining its emittance and energy spread. The FACET facility will have electron and positron bunches at 23 GeV with 3nC of charge. In addition, the beams can have bunch lengths as short as 14 m which corresponds to peak currents of 22 kA. Designs of these experiments indicate that two bunches with ~1nC in the drive beam and ~.5nC in the trailing beam and a separation ~100 m are possible. These conditions have been simulated with QuickPIC to find the optimum density. Other beam parameters have also been simulated. In Figure 8 we show that for the expected conditions, that the trailing beam can be accelerated from 23 GeV to 50 GeV with a few % energy spread in less than 1 meter. The QuickPIC simulations are an essential tool in the design of these experiments.

4.3. Issues for Colliders

The high fidelity SciDAC codes are critical for assessing critical issues related to a future collider. Issues related to emittance preservation for collider parameters, the use of asymmetric beams, staging, optimal beam and laser profiles, and the possibility of higher power lasers and beams cannot all be addressed at FACET or BELLA. They can be addressed using trusted codes and extreme computing. In this section we give two examples of how computing is addressing issues beyond BELLA and FACET.

The ultra-low emittance beams required for next-generation lepton colliders and the large focusing forces in typical plasma accelerators imply that the matched spot size of the electron beam is much smaller than the plasma wake width. For linear wakes this can limit bunch charge and hence the efficiency of a plasma accelerator, because a small bunch with high charge modifies the plasma fields causing nonlinear emittance growth. This problem can however be mitigated by using nonlinear wakes [Tzoufras:09] (which can lead to other issues) or by tailoring the focusing fields in the linear regime through the use higher order laser modes. Simulations with the VORPAL framework have shown that it is possible to co-propagate different laser modes in a plasma channel up to the electron beam dephasing length, keeping the focusing forces small during the whole acceleration process, and that

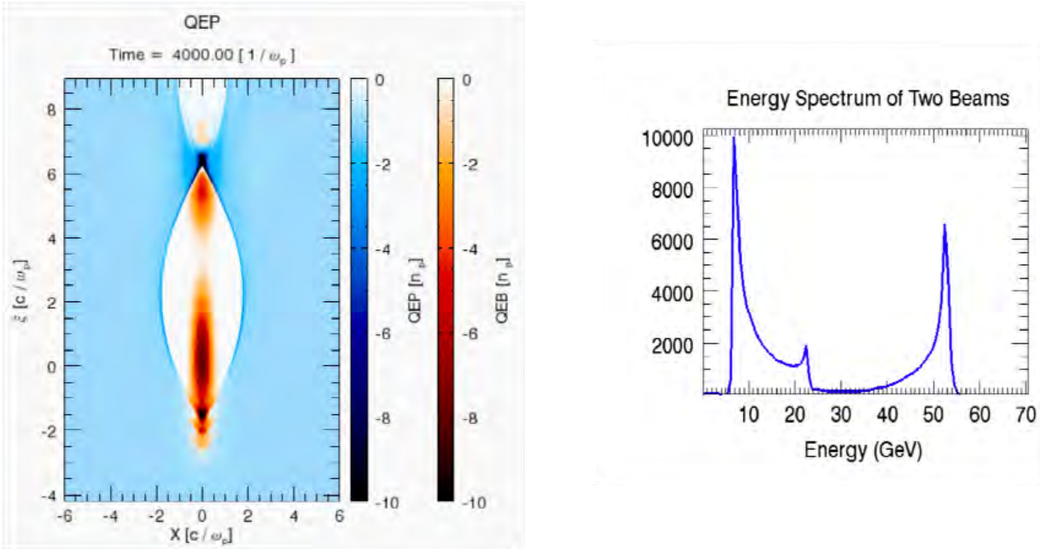

Figure 8. QuickPIC simulations of possible two bunch experiments at FACET. On the left is the beam and plasma density. The plasma density is chosen so that the trailing bunch loads the rear of the wake. On the right is the energy spectra of both bunches. The drive bunch is been decelerated from 23 GeV to 6 GeV while the trailing bunch has been accelerated to 50 GeV.

phasing between the modes can be used to control the focusing phase and field gradients. Simulations with test particles have shown that the matched beam spot size can be increased by almost a factor of 3 using this method, while keeping the low emittance of the beam (Figure 9) [E. Cormier-Michel10].

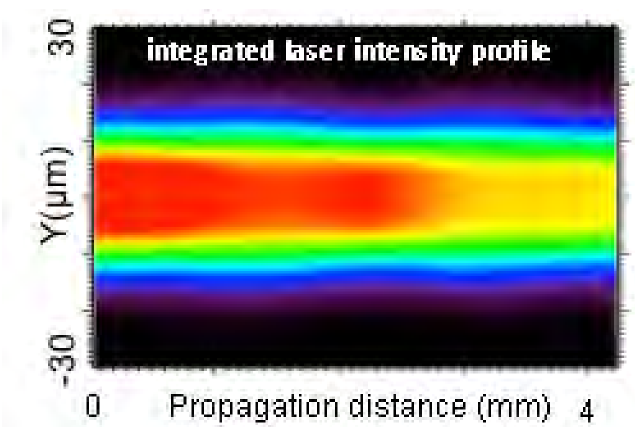

Figure 9. VORPAL simulation shows that multiple laser modes can co-propagate up to the depletion & dephasing distance required for an efficient accelerator while maintaining a flat transverse (Y) profile to tailor focusing forces [Figure from Cormier-Michel et al, sub PRSTAB].

A PWFA-LC design [Seryi09] uses a conventional 25 GeV electron drive beam accelerator, to produce trains of drive bunches distributed in counter-propagating directions to 20 PWFA cells for both the electron and the positron arms of the collider to reach energy of 500 GeV for each beam. Each cell provides 25 GeV of energy to the main beam in about a meter of plasma. The layout and parameters were chosen to optimize PWFA performance while also providing feasible parameters at the interaction point and a practical design for the main beam injector and the drive beam acceleration and distribution system. The wakes are to be excited in a nonlinear or weakly nonlinear blow-out regime where if the plasma ions remain stationary a uniform ion column results which provides a linear focusing force and a radially independent focusing force for both the driving and accelerated electron beams. There is a small region in a weakly nonlinear wake that can both focus and accelerate positrons.

Properties of the drive and main beam bunches have been optimized by particle-in-cell simulations using the code QUICKPIC [Huang:06]. Simulation results are shown in Figure 10. The main beam bunch charge is 1.0×10^{10} particles with a Gaussian distribution. A plasma density of $10^{17}cm^{-3}$ and a Gaussian drive bunch charge of 2.9×10^{10} were chosen, the simulations showed 48% power transfer efficiency from the drive beam to the main beam with a gradient of roughly 25 GV/m. The drive beam bunch length is 30 m while the main beam bunch length is 10 m and the drive-main beam bunch separation is 115 m. The separation between the two bunches must be approximately equal to the plasma wavelength. The energy spread was maintained at less than 2%. We have also developed theories for how to tailor the current profile of the drive and main beams to achieve higher transfer efficiencies and smaller energy spreads [Tzoufras:08a]. QuickPIC simulations show that lower energy spreads are possible (less than 1%) but that hosing could potentially limit the energy transfer efficiency for the longer bunches. Hosing is also an issue for LWFA.

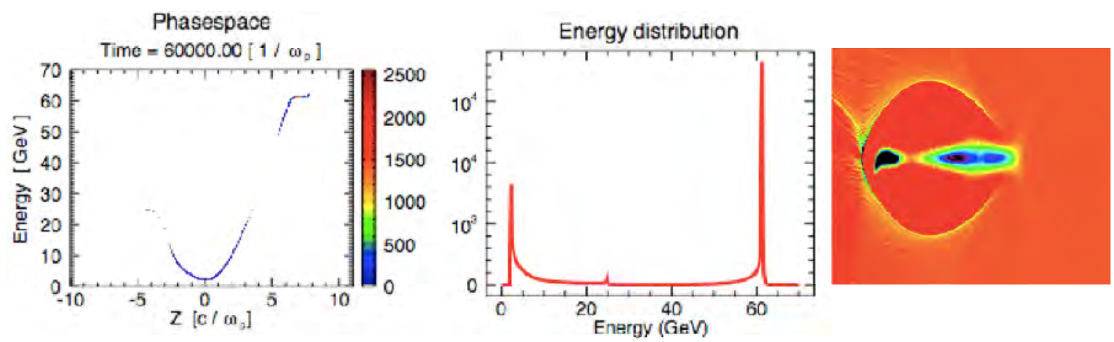

Figure10. QuickPIC results for a single stage in a PWFA-LC design. Using non-optimized Gaussian bunches gives 48% transfer efficiency while maintaining ~1% energy spread. On the left is the energy vs. z. In the middle is the energy spectrum. On the right is the beam and plasma densities. In each case both the drive and trailing beams are shown.

5. Summary and Some Thoughts on What is Next

Simulations have played a critical role in the development of plasma based accelerators. They are now routinely used to design and interpret state-of-the-art experiments. They were an essential component for making the science case for two recently approved experimental facilities, BELLA and FACET. They have helped make science discoveries and are being used to study what is possible for laser facilities being considered at the 250 Joule energy level. They are also currently being used to address issues for colliders based on plasma accelerator stages under consideration as next steps beyond BELLA and FACET.

Simulation tools continue to improve. Within the COMPASS project there is a hierarchy of codes and methods. These methods include full PIC, full PIC in a boosted frame, full PIC with ponderomotive guiding center (PGC), quasi-static PIC with PGC, and fluid models. Previously, the quasi-static technique had allowed the study of 1-100 GeV LWFA stages [Tzoufras08b, Huang09] and upto 1 TeV PWFA stages. In the past year, there has been tremendous progress in the use of each. For self-injected cases boosted frame simulations give speed ups ~300 and for external injection speedups $\sim(\omega_o/\omega_p)^2$ are possible making it possible to model TeV LWFA external injected stages. PGC simulations are also giving speed ups $\sim(\omega_o/\omega_p)^2$. The next year will see further comparison between all techniques.

The future is even more exciting, as we are about to enter the era of exascale computing. The new systems are likely to share some of the following characteristics: a heterogeneous hierarchy of powerful shared memory multi-core nodes, with some element of SIMD (vector) processors as part of their design, coupled together with traditional message passing networks. To prevent memory starvation, they will likely have lightweight threads which can hide memory latency. To meet this challenge for PIC codes, new data structures and algorithms will be required. These algorithms will need to be parameterized so they can work on any future architecture that has these elements.

New kernels have already been implemented for key elements of the PIC algorithm [Decyk09, Decyk10]. For example, a kernel for a 2D electrostatic code that contains 6 subroutines (including a spectral field solver) for a 2D electrostatic PIC code has been tested on a NVIDIA GTX 280 GPU. For an optimum case where the particles never leave a cell the speed up was 44, while for a realistic case (typical plasma temperature) the speed up was 28 (1 ns per particle advance) over an Intel i7 CPU. From this performance it can be estimated that a 3D electromagnetic code could achieve speed ups in excess of 100 (<3ns per full particle advance) for currently used algorithms. A 2D EM kernel for a rigorous charge conserving algorithm based on a different data structure and sorting routine has achieved speed ups from 27-81 depending on the initial plasma temperature [Kong10].

While current GPUs appear to offer the greatest acceleration, one should expect that not only will they evolve significantly but future many core platforms will also evolve rapidly. As a result algorithms will need to be retuned regularly. Therefore, as noted above new algorithms should be parameterized as much as possible. It should also be noted that on current systems there is usually one GPU per 4 cores so speed ups of only 10 are not competitive with SIMD units for which there are usually one per core.

If one assumes speed ups ~100 for next generation machines per core and similar numbers of cores and that additional speed ups of ~100 from existing algorithms are possible (for example the quasi-static could produce this) then the turnaround time of simulations for next generation experiments could be ~minutes or less. This could lead to real time steering of experiments as well as the ability of using simulations to design and optimize laser, beam, and density profiles including energy chirps.

Acknowledgments

Work supported by the U.S. Department of Energy (DOE) Office of Science, Office of High Energy Physics SciDAC program and under grants including DE-AC02-05CH11231, DE-FG03-95ER40926, DE-FG02-01ER41178, DE-FG02-03ER83857, DE-FC02-07ER41500; and by the National Science Foundation under the grants grants PHY-0904039 and PHY-0936266**.** This work used the resources of the National Energy Research Scientific Computing Center, the Hoffman 2 cluster at UCLA, and at the National Center for Computational Sciences.

References

[Abraham03] S. Abraham, Spencer Abraham announces department of energy 20-year science facility plan, text available at http://www.er.doe.gov/Sub/Newsroom/News_Releases/DOE-SC/2003/Energy%2020-Year%20Science20Facility%20Plan.htm, 2003.

[Bethel09] E.W. Bethel, et al. "Occam's Razor and Petascale Visual Data Analysis," Journal of Physics: Conference Series 180, 012084 (2009).

[Birdsall91] C. K. Birdsall, A. B. Langdon, V. Vehedi, and J. P. Verboncoeur, Plasma Physics via Computer Simulations, Adam Hilger, Bristol, 1991.

[Blumenfeld07a] I. Blumenfeld et al., Nature 445, 741 (2007).

[Clayton10] D. Froula et al., Self-guided laser wakef eld acceleration beyond 1 GeV using ionization-induced injection, to appear in Phys. Rev. Lett.

[Cormier-Michel08] E. Cormier-Michel, et al., "Unphysical kinetic effects in laser wakefield accelerators modeled with particle-in-cell codes," Phys. Rev. E. 78, 016404 (2008).

[Cormier-Michel10]E. Cormier-Michel, et al. "Control of focusing fields in laser-plasma accelerators using higher-order modes," PRSTAB submitted.

[Cowan08] B.M. Cowan, et al, “Characteristics of an envelope model for laser plasma accelerator simulation”, J. Comput. Phys., submitted Computer Physics Communications.
[Decyk09] V. Decyk, T. Singh, and S. Friedman, Graphical Processing Unit-Based Simulations, in Proc. Intl. Computational Accelerator Physics Conf. (ICAP2009), 2009, [Decyk10] Viktor K. Decyk and Tajendra V. Singh, submitted to
[Esarey96] E. Esarey et al., IEEE Trans. Plasma Sci. 24, 252 (1996).
[Esarey97] E. Esarey et al., Phys. Rev. Lett. 79, 2682 (1997).
[Faure04] J. Faure et al., Nature 431, 541 (2004).
[Faure06] J. Faure et al., Nature 444, 737 (2006).
[Fonseca02] R.A. Fonscea et al., Lect. Notes Comp. Science 2329, III-342, Springer-Verlag, (2002).
[Geddes04] C. G. R. Geddes et al., Nature 431, 538 (2004).
[Geddes05] C.G.R. Geddes et al. “Production of high-quality electron bunches by dephasing and beam loading in channeled and unchanneled laser plasma accelerators,” Phys. Plasmas 12, 056709 (2005).
[Geddes07] C. G. R. Geddes et al., J. Phys. Conf. Ser. 78, 12021 (2007).
[Geddes08a] C.G.R. Geddes et al, Phys. Rev. Lett. 100, 215004 (2008).
[Geddes09] Geddes et al. “Laser Plasma Particle Accelerators: Large Fields for Smaller Facility Sources,” SciDAC Review 13 (Summer, 2009); http://www.scidacreview.org/0903/index.html.
[Gonsalves07] A.J. Gonsalves et al., in Proc. Particle Accel. Conf., IEEE, Piscataway, NJ (2007).
[Gordon00] Daniel F. Gordon et al., IEEE Trans. Plasma Sci, Vol. 28, 1135 (2000)
[Huang06] C.K. Huang et al. Journal of Computational Physics 217 (2006) 658–679
[Huang09] C. Huang, et al., “Recent results and future challenges for large scale Particle-In-Cell simulations of plasma-based accelerator concepts”, J. of Physics: Conference Series 180, 012005 (2009).
[Joshi06] C. Joshi, “Plasma accelerators,” *Scientific American*, Vol. 294, 41-47 (Feb. 2006).
[Kong10] X. Kong, et al., “Particle-in-cell Simulations with Charge- Conserving Current Deposition on Graphic Processing Units,” submitted to J. Computational Phys., 2010.
[Mangles04a] S. P. D. Mangles et al., Nature 431, 535 (2004).
[Leemans05a] W. P. Leemanset al., 2005 IEEE Trans. Plasma Sci. 33, 8 (2005).
[Leemans06a] W. P. Leemans et al., Nature Physics 2, 696 (2006).
[Leemans09] W.P. Leemans and E.H. Esarey, “Laser-driven plasma-wave electron accelerators,” Physics Today 2009.
[Lu07a] W. Lu et al in Phys. Rev. ST Accel. Beams 10, 061301 (2007)
[Martins10a] Samuel F. Martins et al., “Exploring laser-wakefield-accelerator regimes for near-term lasers using particle-in-cell simulation in Lorentz-boosted frames,” Nature Physics, **6**, 311–316 (2010)
[Martins10b] Samuel F. Martins et al., Computer Physics Communications, 181, 869-875 (2010).
[Michel08a] E. Cormier-Michel et al., accepted to PRE (2008).
[Mora97] P. Mora, T.M. Antonsen Jr., Kinetic modeling of intense, short laser pulses propagating in tenuous plasmas, Phys. Plasmas 4 (1997) 217.
[Nieter04] C. Nieter and J. Cary, J. Comp. Phys. 196, 448 (2004).
[Oz07] E. Oz, et al., “Ionization-induced electron trapping in ultra-relativistic plasma wakes,” *Physical Review Letters*. 98, 084801(2007).
[Paul08] K. Paul, et al., “Benchmarking the codes VORPAL, OSIRIS and QuickPIC with Laser Wakefield Acceleration Simulations,” Proc. Advanced. Accel. Concepts Workshop (2008).
[Pak10] Pak A., et al., “Injection and trapping of tunnel ionized electrons into laser-produced wakes, *Physical Review Letters*, Vol. 104, 025003 (January 2010).
[Pukhov02] A. Pukhov and J. Meyer-ter-Vehn , 2002 Appl. Phys. B 74, 355 .
[Rubel10] O. Rübel, et al., “Coupling Visualization and Data Analysis for Knowledge Discovery from Multi-dimensional Scientific Data,” Procedia Computer Science 1, 1751 (2010).
[Tajima79] Tajima T. and Dawson J. M. , Phys. Rev. Lett. 43, 267 (1979)

[Tsung04] F. S. Tsung et al., Phys. Rev. Lett. 93, 185002 (2004) .
[Tsung06] F. S. Tsung et al., Physics of Plamas **13**, 056708 (2006)
[Tzoufras08a] M. Tzoufras et al., Phys. Rev. Lett. vol. (2008).
[Tzoufras08b] M. Tzoufras, PhD Dissertation UCLA 2008.
[Vay07] Vay J.-L., Phys. Rev. Lett. 98, 130405 (2007).
[Vay10] J.-L. Vay, et al., “Modeling laser wakefield accelerators in a Lorentz boosted frame,” submitted to J. Comput. Phys.