Contents lists available at ScienceDirect

Computer Physics Communications

journal homepage: www.elsevier.com/locate/cpc

An efficient and portable SIMD algorithm for charge/current deposition in Particle-In-Cell codes[☆]

H. Vincenti[a,b,*], M. Lobet[a], R. Lehe[a], R. Sasanka[c], J.-L. Vay[a]

[a] Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA
[b] Lasers Interactions and Dynamics Laboratory (LIDyL), Commissariat À l'Energie Atomique, Gif-Sur-Yvette, France
[c] Intel Corporation, OR, USA

ARTICLE INFO

Article history:
Received 9 January 2016
Received in revised form
10 August 2016
Accepted 19 August 2016
Available online 19 September 2016

Keywords:
Particle-In-Cell method
OpenMP
SIMD Vectorization
AVX2
AVX512
Tiling
Cache reuse
Many-core architectures

ABSTRACT

In current computer architectures, data movement (from die to network) is by far the most energy consuming part of an algorithm (≈20 pJ/word on-die to ≈10,000 pJ/word on the network). To increase memory locality at the hardware level and reduce energy consumption related to data movement, future exascale computers tend to use many-core processors on each compute nodes that will have a reduced clock speed to allow for efficient cooling. To compensate for frequency decrease, machine vendors are making use of long SIMD instruction registers that are able to process multiple data with one arithmetic operator in one clock cycle. SIMD register length is expected to double every four years. As a consequence, Particle-In-Cell (PIC) codes will have to achieve good vectorization to fully take advantage of these upcoming architectures. In this paper, we present a new algorithm that allows for efficient and portable SIMD vectorization of current/charge deposition routines that are, along with the field gathering routines, among the most time consuming parts of the PIC algorithm. Our new algorithm uses a particular data structure that takes into account memory alignment constraints and avoids gather/scatter instructions that can significantly affect vectorization performances on current CPUs. The new algorithm was successfully implemented in the 3D skeleton PIC code PICSAR and tested on Haswell Xeon processors (AVX2-256 bits wide data registers). Results show a factor of ×2 to ×2.5 speed-up in double precision for particle shape factor of orders 1–3. The new algorithm can be applied as is on future KNL (Knights Landing) architectures that will include AVX-512 instruction sets with 512 bits register lengths (8 doubles/16 singles).

Program summary

Program Title: vec_deposition

Program Files doi: http://dx.doi.org/10.17632/nh77fv9k8c.1

Licensing provisions: BSD 3-Clause

Programming language: Fortran 90

External routines/libraries: OpenMP > 4.0

Nature of problem: Exascale architectures will have many-core processors per node with long vector data registers capable of performing one single instruction on multiple data during one clock cycle. Data register lengths are expected to double every four years and this pushes for new portable solutions for efficiently vectorizing Particle-In-Cell codes on these future many-core architectures. One of the main hotspot routines of the PIC algorithm is the current/charge deposition for which there is no efficient and portable vector algorithm.

Solution method: Here we provide an efficient and portable vector algorithm of current/charge deposition routines that uses a new data structure, which significantly reduces gather/scatter operations. Vectorization is controlled using OpenMP 4.0 compiler directives for vectorization which ensures portability across different architectures.

☆ This paper and its associated computer program are available via the Computer Physics Communication homepage on ScienceDirect (http://www.sciencedirect.com/science/journal/00104655).

* Corresponding author at: Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, CA, USA.
E-mail addresses: hvincenti@lbl.gov (H. Vincenti), mlobet@lbl.gov (M. Lobet), rlehe@lbl.gov (R. Lehe), ruchira.sasanka@intel.com (R. Sasanka), jlvay@lbl.gov (J.-L. Vay).

http://dx.doi.org/10.1016/j.cpc.2016.08.023
0010-4655/© 2016 Published by Elsevier B.V.

Restrictions: Here we do not provide the full PIC algorithm with an executable but only vector routines for current/charge deposition. These scalar/vector routines can be used as library routines in your 3D Particle-In-Cell code. However, to get the best performances out of vector routines you have to satisfy the two following requirements: (1) Your code should implement particle tiling (as explained in the manuscript) to allow for maximized cache reuse and reduce memory accesses that can hinder vector performances. The routines can be used directly on each particle tile. (2) You should compile your code with a Fortran 90 compiler (e.g Intel, gnu or cray) and provide proper alignment flags and compiler alignment directives (more details in README file).

1. Introduction

1.1. Challenges for porting PIC codes on exascale architectures: importance of vectorization

Achieving exascale computing facilities in the next decade will be a great challenge in terms of energy consumption and will imply hardware and software developments that directly impact our way of implementing PIC codes [1].

Table 1 shows the energy required to perform different operations ranging from arithmetic operations (fused multiply add or FMADD) to on-die memory/DRAM/Socket/Network memory accesses. As 1 pJ/flop/s is equivalent to 1 MW for exascale machines delivering 1 exaflop (10^{18} flops/sec), this simple table shows that as we go off the die, the cost of memory accesses and data movement becomes prohibitive and much more important than simple arithmetic operations. In addition to this energy limitation, the draconian reduction in power/flop and per byte will make data movement less reliable and more sensitive to noise, which also push towards an increase in data locality in our applications.

At the hardware level, part of this problem of memory locality was progressively addressed in the past few years by limiting costly network communications and grouping more computing resources that share the same memory ("fat nodes"). However, partly due to cooling issues, grouping more and more of these computing units will imply a reduction of their clock speed. To compensate for the reduction of computing power due to clock speed, future CPUs will have much wider data registers that can process or "vectorize" multiple data in a single clock cycle (Single Instruction Multiple Data or SIMD).

At the software level, programmers will need to modify algorithms so that they achieve both memory locality and efficient vectorization to fully exploit the potential of future exascale computing architectures.

1.2. Need for portable vectorized routines

In a standard PIC code, the most time consuming routines are current/charge deposition from particles to the grid and field gathering from the grid to particles. These two operations usually account for more than 80% of the execution time. Several portable deposition algorithms were developed and successfully implemented on past generations' vector machines (e.g. CRAY, NEC) [2–6]. However, these algorithms do not give good performance on current SIMD architectures, that have new constraints in terms of memory alignment and data layout in memory.

To the authors' knowledge, most of the vector deposition routines proposed in contemporary PIC codes use compiler based directives or even C++ Intel intrinsics in the particular case of the Intel compiler, to increase vectorization efficiency (e.g. [8]). However, these solutions are not portable and require code re-writing for each new architecture.

1.3. Paper outline

In this paper, we propose a portable algorithm for the direct deposition of current or charge from macro particles onto a grid, which gives good performances on SIMD machines. The paper is divided into four parts:

(i) in Section 2, we quickly introduce the standalone 3D skeleton electromagnetic PIC code PICSAR-EM3D in which we implemented the different vector versions of the deposition routines presented in this paper,
(ii) in Section 3, we quickly remind the scalar deposition routine and show why it cannot be vectorized as is by the compiler. Then, we introduce a vector algorithm that performed well on former Cray vector machines but give poor performances on current SIMD machines. By carefully analyzing the bottlenecks of the old vector routine on current SIMD machines, we will derive a new vector routine that gives much better performances,
(iii) in Section 4 we present the new vector routines that were developed, based on the analysis in Section 3,
(iv) in Section 5, the new vector routines are benchmarked on the new Cori machine at the U.S. National Energy Research Supercomputer Center (NERSC) [9].

2. The PICSAR-EM3D PIC kernel

PICSAR-EM3D is a standalone "skeleton" PIC kernel written in Fortran 90 that was built using the main electromagnetic PIC routines (current deposition, particle pusher, field gathering, Yee field solver) of the framework WARP [10]. As WARP is a complex mix of Fortran 90, C and Python, PICSAR-EM3D provides an essential testbed for exploring PIC codes algorithms with multi-level parallelism for emerging and future exascale architectures. All the high performance carpentry and data structures in the code have been redesigned for best performance on emerging architectures, and tested on NERSC supercomputers (CRAY XC30 Edison and testbed with Intel Knight's Corner coprocessors Babbage).

2.1. PIC algorithm

PICSAR-EM3D contains the essential features of the standard explicit electromagnetic PIC main loop:

(i) Maxwell solver using arbitrary order finite-difference scheme (staggered or centered),
(ii) Field gathering routines including high-order particle shape factors (order 1—CIC, order 2—TSC and order 3—QSP),
(iii) Boris particle pusher,
(iv) Most common types of current depositions: Morse–Nielson deposition [1] (also known as direct $\rho\mathbf{v}$ current deposition)

Table 1
Energy consumption of different operations taken from [7]. The die hereby refers to the integrated circuit board made of semi-conductor materials that usually holds the functional units and fast memories (first levels of cache). This table shows the energy required to achieve different operations on current (Year 2015) and future (Year 2019) computer architectures. DP stands for Double Precision, FMADD for Fused Multiply ADD and DRAM for Dynamic Random Access Memory.

Operation	Energy cost (pJ)	Year
DP FMADD flop	11	2019
Cross-die per word access	24	2019
DP DRAM read to register	4800	2015
DP word transmit to neighbor	7500	2015
DP word transmit across system	9000	2015

and Esirkepov [11] (charge conserving) schemes. The current and charge deposition routines support high-order particle shape factors (1–3).

2.2. High performance features

Many high performance features have already been included in PICSAR-EM3D. In the following, we give a quick overview of the main improvements that brought significant speed-up of the code and that are of interest for the remainder of this paper. A more comprehensive description of the code and its performances will be presented in another paper.

2.2.1. Particle tiling for memory locality

Field gathering (interpolation of field values from the grid to particle positions) and current/charge deposition (deposition of particle quantities to adjacent grid nodes) account for more than 80% of the total execution time of the code. In the deposition routines for instance, the code loops over all particles and deposit their charges/currents on the grid.

One major bottleneck that might arise in these routines and can significantly affect overall performance is cache reuse.

Indeed, at the beginning of the simulations (cf. Fig. 1(a)) particles are typically ordered along the "fast" axis ("sorted case") that corresponds to parts of the grid that are contiguously located in memory. As the code loops over particles, it will thus access contiguous grid portions in memory from one particle to another and efficiently reuse cache.

However, as time evolves, the particle distribution often becomes increasingly random, leading to numerous cache misses in the deposition/gathering routines (cf. Fig. 1(b)). This results in a considerable decrease in performance. In 2D geometry, one MPI subdomain usually fits in L2 cache (256–512 kB per core) but for 3D problems with MPI subdomains handling $100 \times 100 \times 100$ grid points, one MPI subdomain does not fit in cache anymore and random particle distribution of particles can lead to performance bottlenecks (see Fig. 1(b)).

To solve this problem and achieve good memory locality, we implemented particle tiling in PICSAR-EM3D. Particles are placed in tiles that fit in cache (cf. Fig. 2). In the code, a tile is represented by a structure of array *Type(particle_tile)* that contains arrays of particle quantities (positions, velocity and weight). All the tiles are represented by a 3D Fortran array *array_of_tiles(:,:,:)* of type *particle_tile* in the code. Our data structure is thus very different from the one in [12], which uses one large Fortran $ppart(1 : ndims, 1 : nppmax, 1 : ntiles)$ array for all particles and tiles, where *ndims* is the number of particle attributes (e.g positions x, y, z), *nppmax* the maximum number of particles in a tile and *ntiles* the number of tiles. There are two reasons behind our choice:

(i) if one tile has much more particles than others, we considerably save memory by using our derived type compared to the array *ppart*. Indeed, in the latter case, if one tile has much more particles *np* than others, we would still need to choose $nppmax \geqslant np$ for all the tiles,
(ii) any tile can be resized as needed independently, without the need for reallocating the entire array of tiles.

Table 2
Speed-up of the whole PIC code brought by particle tiling. Tests were performed using a $100 \times 100 \times 100$ grid with 10 particle per cells. Particles are randomly distributed on the grid and have a thermal velocity of $v_{th} =\approx 0.1c$ with c being the speed of light in vacuum. For a time step imposed by the Courant condition of the Maxwell solver $dt = 0.57dx/c$ (where dx is the mesh size) and $n_p = 10$ particle per cells, the total number of particles leaving a cell after one time step is $n_p \times v_{th} \times dt/dx \approx 0.57$. After 20 time steps, nearly all particles have left their original cell. The reference time corresponds to the standard case of $1 \times 1 \times 1$ tile. The tests were performed on one MPI process and a single socket, on the Edison cluster at NERSC.

Tile size	Speed-up	L1 and L2 Cache reuse
$1 \times 1 \times 1$	$\times 1$	85%
$10 \times 10 \times 10$	$\times 3$	99%

Performance improvements of the whole code are reported in Table 2 for tests performed on Intel Ivy Bridge (Cray XC30 Edison machine at NERSC). These tests show a speed-up of x3 in case of a random particle distribution. Cache reuse using tiling reaches 99%. The optimal tile size ranges empirically between $8 \times 8 \times 8$ cells and $10 \times 10 \times 10$ cells. As will be shown later in the paper, having good cache reuse is crucial to increasing the flop/byte ratio of the proposed algorithm and obtaining improvements using vectorization.

Notice that at each time step, the particles of each tile are advanced and then exchanged between tiles. As particles move less than one cell at each time step, the amount of particles exchanged between tiles at each time step is low for typical tiles' sizes. (The surface/volume ratio decreases with tile size.) As a consequence, particle exchanges between tiles account in practice for a very small percentage of the total PIC loop (a few percents). Our particle exchange algorithm differs from the one used in [12] in that it avoids copying data into buffers. In addition, it can be efficiently parallelized using OpenMP (details are beyond the scope for this paper and will be presented in an upcoming publication).

2.2.2. Multi-level parallelization

PICSAR-EM3D also includes the following high performance implementations:

(i) vectorization of deposition and gathering routines,
(ii) OpenMP parallelization for intranode parallelisms. Each OpenMP thread handles one tile. As there are much more tiles than threads in 3D, load balancing can be easily done using the SCHEDULE clause in openMP with the guided attribute,
(iii) MPI parallelization for internode parallelism,
(iv) MPI communications are overlapped with computations. For particles, this is done by treating exchanges of particles with border tiles while performing computations on particles in inner tiles,
(v) MPI-IO for fast parallel outputs.

In the remainder of this paper, we will focus on the vectorization of direct charge/current deposition routines for their simplicity and widespread use in electromagnetic PIC codes. The Esirkepov-like current deposition is not treated in this paper but the techniques used here are very general and should apply in principle to any kind of current deposition.

3. Former CRAY vector algorithms and performance challenges on new architectures

In the following, we focus on the direct 3D charge deposition which can be presented in a more concise way than the full 3D

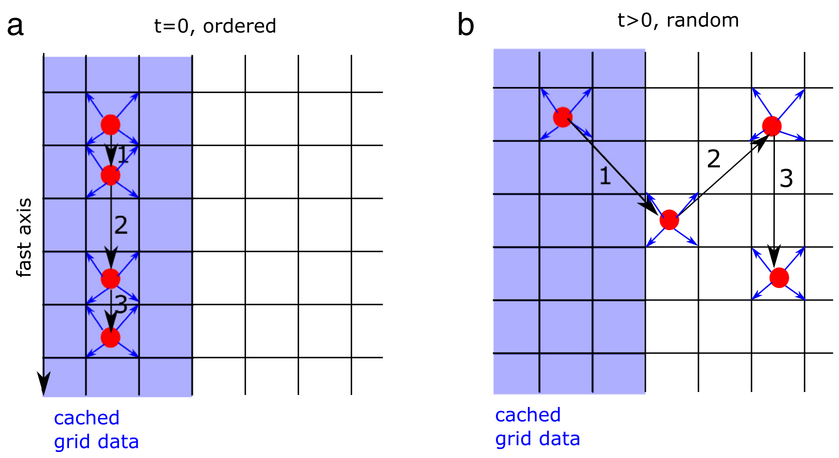

Fig. 1. Importance of cache reuse in deposition routines. Illustration is given in 2D geometry for clarity, with CIC (linear) particle shapes. Panel (a) shows a typical layout at initialization ($t = 0$) where particles are ordered along the "fast" axis of the grid, corresponding to grid cells (blue area) that are contiguous in memory. The loop on particles is illustrated with arrows and index of the loop with numbers 1–3. Using direct deposition, each particle (red point) deposits (blue arrows) its charge/current to the nearest vertices (4 in 2D and 8 in 3D for CIC particle shapes). Panel (b) illustrates the random case (at $t > 0$) where particles are randomly distributed on the grid. As the algorithms loop over particles, it often requires access to uncached grid data, which then results in a substantial number of cache misses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

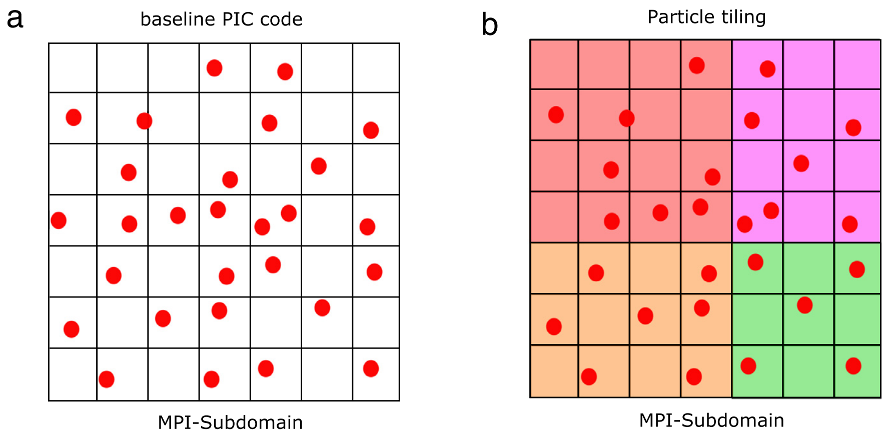

Fig. 2. Particle tiling for efficient cache reuse. Panel (a) shows the usual configuration used in standard codes. There is one big array for particles for each MPI subdomain. Panel (b) shows the data structure used in PICSAR. Particles are grouped in tiles that fit in cache, allowing for efficient cache reuse during deposition/gathering routines.

current deposition. The full 3D scalar/vector algorithm presented here for charge/current deposition can be found in the Fortran source file vec_deposition.F90 of the CPC Program Library archive associated to this paper. In this file, scalar charge deposition routines have the following prefix: "depose_rho_scalar_". Order is specified at the end. For instance order 1 charge scalar deposition routine is named depose_rho_scalar_1_1_1. For vector deposition routines, the prefix used is "depose_rho_vecHVv2_". The same notation is used for current deposition routines but with the prefixes "depose_jxjyjz_scalar_"/ "depose_jxjyjz_vecHVv2_" for scalar/vector routines.

3.1. Scalar algorithm

The scalar algorithm for order 1 charge deposition is detailed in listing 1. For each particle index *ip*, this algorithm (see line 5):

(i) finds the indices (j, k, l) of the cell containing the particle (lines 11–13),
(ii) computes the weights of the particle at the 8 nearest vertices $w1$ to $w8$ (line 15—not shown here),
(iii) adds charge contribution to the eight nearest vertices $\{(j, k, l), (j+1, k, l), (j, k+1, l), (j+1, k+1, l), (j, k, l+1), (j+1, k, l+1), (j, k+1, l+1), (j+1, k+1, l+1)\}$ of the current cell (j, k, l) (see lines 18–25).

As two different particles ip_1 and ip_2 can contribute to the charge at the same grid nodes, the loop over particles (line 5) presents a dependency and is thus not vectorizable as is.

3.2. Former vector algorithms and new architecture constraints

Several vector algorithms have already been derived and tuned on former Cray vector machines [2–6,13]. However, these techniques are not adapted anymore to current architectures and yield very poor results on SIMD machines that necessitate to comply with the three following constraints in order to enable vector performances:

(i) **Good cache reuse**. The flop/byte ratio (i.e. cache reuse) in the main loops of the PIC algorithm must be high in order to observe a speed-up with vectorization. Otherwise, if data has to be moved from memory to caches frequently, the performance gain with vectorization can become obscured by the cost of data movement. As we showed earlier, this is ensured by particle tiling in our code.
(ii) **Memory alignment**. Data structures in the code need to be aligned and accessed in a contiguous fashion in order to maximize performances. Modern computers read from or write to a memory address in word-sized chunks of 8 bytes (for 64 bit systems). Data alignment consists in putting

Listing 1: Scalar charge deposition routine for CIC particle shape factors

```fortran
SUBROUTINE depose_rho_scalar_1_1_1(...)
        ! Declaration and init
        ! ...........7
        ! Loop on particles
        DO ip=1,np
            ! --- computes current position in grid units
            x = (xp(ip)-xmin)*dxi
            y = (yp(ip)-ymin)*dyi
            z = (zp(ip)-zmin)*dzi
            ! --- finds node of cell containing  particle
            j=floor(x)
            k=floor(y)
            l=floor(z)
            ! --- computes weigths w1..w8
            ........
            ! --- add charge density contributions
            ! --- to the 8 vertices of current cell
            rho(j,k,l)          =rho(j,k,l)       + w1
            rho(j+1,k,l)        =rho(j+1,k,l)     + w2
            rho(j,k+1,l)        =rho(j,k+1,l)     + w3
            rho(j+1,k+1,l)      =rho(j+1,k+1,l)   + w4
            rho(j,k,l+1)        =rho(j,k,l+1)     + w5
            rho(j+1,k,l+1)      =rho(j+1,k,l+1)   + w6
            rho(j,k+1,l+1)      =rho(j,k+1,l+1)   + w7
            rho(j+1,k+1,l+1)    =rho(j+1,k+1,l+1)+ w8
        END DO
END SUBROUTINE depose_rho_scalar_1_1_1
```

the data at a memory address equal to some multiple of the word size, which increases the system's performance due to the way the CPU handles memory. SSE2, AVX and AVX-512 on x86 CPUs do require the data to be 128-bits, 256-bits and 512-bits aligned respectively, and there can be substantial performance advantages from using aligned data on these architectures. Moreover, compilers can generate more optimal vector code when data is known to be aligned in memory. In practice, the compiler can enforce data alignment at given memory boundaries (128, 256 or 512 bits) using compiler flags/directives.

(iii) **Unit-stride read/write**. If data are accessed contiguously in a do loop (unit-stride), the compiler will generate vector single load/store instructions for the data to be processed. Otherwise, if data are accessed randomly or via indirect indexing, the compiler might generate gather/scatter instructions that almost yield sequential performance or worse. Indeed, in case of a gather/scatter, the processor might have to make several different loads/stores from/to memory instead of one load/store, eventually leading to poor vector performances.

In the following, we investigate performance of one of the former vector algorithms for CRAY machines [4] and analyze its bottlenecks on SIMD architectures. This analysis will show a way to improve the vector algorithm and derive a new one that yields significant speed-up over the scalar version.

3.3. Example: the Schwarzmeier and Hewitt scheme (SH)

3.3.1. SH vector deposition routine

Listing 2 details the Schwarzmeier and Hewitt (SH) deposition scheme [4] that was implemented in PICSAR-EM3D and tested on Cori supercomputer at NERSC. In this scheme, the initial loop on particles is done by blocks of lengths *nblk* (cf. line 5) and split into two consecutive nested loops:

- A first nested loop (line 7) that computes, for each particle *nn* of the current block:
  (i) its cell position *ind*0 on the mesh (line 13),
  (ii) its contribution $ww(1, nn), \dots, ww(8, nn)$ to the charge at the 8 vertices of the cell and
  (iii) the indices $ll(1, nn), \dots, ll(8, nn)$ of the 8 nearest vertices in the 1D density array rho (cf. lines 14–19).

  Notice that 1D indexing is now used for *rho* to avoid storing three different indices for each one of the 8 vertices. The Fortran integer array *moff* (1:8) gives the indices of the 8 vertices with respect to the cell index *ind*0 in the 1D array *rho*. The loop at line 7 has no dependencies and is vectorized using the portable *$OMP SIMD* directive.
- A second nested loop (line 23) that adds the contribution of each one of the *nblk* particles to the 8 nearest vertices of their cell (line 26). As one particle adds its contribution to eight different vertices, the loop on the vertices at line 25 has no dependency and can also be vectorized using the *$OMP SIMD* directive.

Usually, *nblk* is chosen as a multiple of the vector length. Notice that using a moderate size *nblk*, for the blocks of particles, ensures that the temporary arrays ww and *ll* fit in cache.

The SH algorithm presented in listing 2 is fully vectorizable and gave very good performances on former Cray machines [4,6]. However as we show in the following section, it yields very poor performances on SIMD architectures.

3.3.2. Tests of the Schwarzmeier and Hewitt algorithm on Cori

The SH algorithm was tested on one socket of the Cori cluster at NERSC. This socket had one Haswell Xeon processor with the following characteristics:

(i) 16-core CPU at 2.3 GHz,
(ii) 256-bit wide vector unit registers (4 doubles, 8 singles) with AVX2 support,
(iii) 256 kB L2 cache/core, 40 MB shared L3 cache.

The Intel compiler was used to compile the code with option "-O3". The simulation was run using 1 MPI process and 1 OpenMP thread per MPI process, with the following numerical parameters:

(i) $100 \times 100 \times 100$ grid points with $10 \times 10 \times 10 = 1000$ tiles i.e 10 tiles in each direction,
(ii) Two particle species (proton and electron) with 10 particle per cells. The particles are randomly distributed across the simulation domain. Plasma electrons have an initial thermal velocity of $\approx 0.1c$.

Table 3
Performance comparisons of scalar and SH vector routines.

Routine	depose_rho_scalar_1_1_1		depose_rho_vecSH_1_1_1	
Compiler option	-no-vec	-xCORE-AVX2	-no-vec	-xCORE-AVX2
Time/it/part	14.6 ns	14.6 ns	21 ns	15.9 ns

Listing 2: Vector version of the charge deposition routine developed by SH for CIC particle shape factors

```
1  SUBROUTINE depose_rho_vecSH_1_1_1(...)
2   ! Declaration and init
3   .....
4   ! Loop on particles
5    DO ip=1,np,nblk
6       !$OMP SIMD
7        DO n=ip,MIN(ip+nblk-1,np) !!!! VECTOR
8              nn=n-ip+1
9               !- Computations relative to particle ip (cell position etc.)
10              ...
11              ! --- computes weight for each of the 8-vertices of the current cell
12              ! --- computes indices of 8-vertices in the array rho
13              ind0 = (j+nxguard+1) + (k+nyguard+1)*nnx + (l+nzguard+1)*nnxy
14              ww(1,nn) = sx0*sy0*sz0*wq
15              ll(1,nn) = ind0+moff(1)
16              ...
17              ...
18              ww(8,nn) = sx1*sy1*sz1*wq
19              ll(8,nn) = ind0+moff(8)
20        END DO
21        !$OMP END SIMD
22        ! --- add charge density contributions
23        DO m= 1,MIN(nblk,np-ip+1)
24            !$OMP SIMD
25             DO l=1,8  !!!! VECTOR
26                 rho(ll(l,m)) = rho(ll(l,m))+ww(l,m)
27             END DO
28             !$OMP END SIMD
29        END DO
30   END DO
31   ...
32 END SUBROUTINE depose_rho_vecSH_1_1_1
```

The results are displayed in Table 3 for order 1 scalar and SH routines, using two different compiler options in each case:

(i) -xCORE-AVX2 to enable vectorization,
(ii) -no-vec to disable auto-vectorization of the compiler. In this case, we also manually remove !$OMP SIMD directives to avoid SIMD vectorization of loops.

The scalar routine takes the same time for -xCORE-AVX2 and -no-vec options because the routine is not auto-vectorizable by the compiler.

For the vector routine, we see an improvement of 30% between -xCORE-AVX2 and -no-vec options, showing that vectorization is enabled and working in the -xCORE-AVX2 case. Nevertheless, the overall performance is poor, and the vector routine compiled with -xCORE-AVX2 is even 10% slower than the scalar routine.

By looking at the code on listing 2 and using compiler report/assembly code generated by the Intel compiler, we found two main reasons for this poor performance:

1. The first one comes from the strided access of the arrays ww and *ll* in the loop at line 7. Assuming cache line sizes of 64 bytes (8 doubles) and 256-bits wide registers, the four different elements $ww(1, nn_1)$ to $ww(1, nn_1 + 3)$ are thus on four different cache lines (ww is of size $(8, nblk)$) and this strided access necessitates 4 stores in memory at different cache lines ("scatter") instead of a single store if the accesses were aligned and contiguous. A solution would be to switch dimensions of ww but this might not bring any improvement at all because the loop on vertices (line 25) would then have strided access for ww ("gather"). Some PIC implementations choose contiguous access for ww/ll in the first loop and then use an efficient vector transpose of ww/ll before the second loop on vertices. However, this solution requires the use of "shuffle" Intel vector intrinsics to efficiently implement the transpose, which is not portable because this transpose will have to be re-written for a different processor. In addition, this transpose is done $8 \times np$ with np the number of particles and might thus add a non-negligible overhead if not done properly.
2. The second bottleneck comes from the indirect indexing for *rho* at line 26. The problem with the current data structure of *rho* is that the 8 vertices of one cell are not contiguous in memory, resulting in a rather inefficient gather/scatter instruction.

In the next section, we propose a portable solution for orders 1, 2 and 3 charge deposition that solves these two problems and yields a speed-up factor of up to ×2.5 in double precision over the scalar routine.

4. New and portable SIMD algorithms

In this section, we present vector algorithms that perform efficiently on SIMD architectures.

4.1. CIC (order 1) particle shape

4.1.1. Algorithm

The new vector algorithm is detailed in listing 3. Similarly to the SH routine, the main particle loop is done by blocks of *nblk*

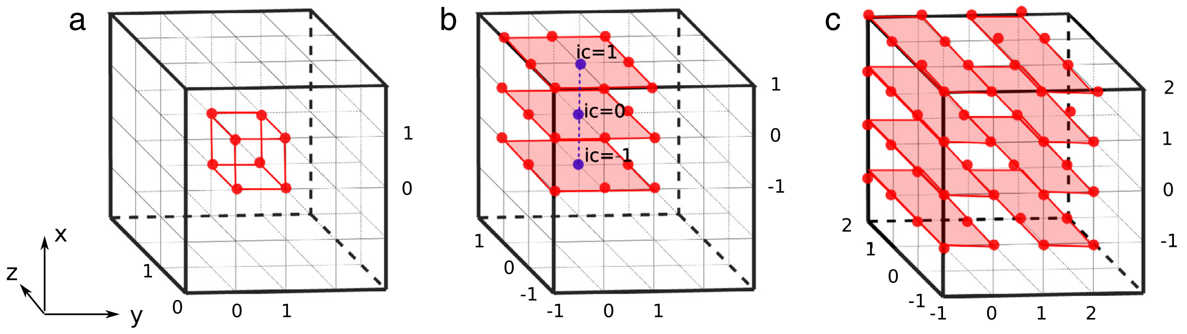

Fig. 3. Data structure used for the array *rhocells* **for different particle shape factors**. In each plot, the particle that deposits charge to its nearest vertices (red/blue points) is located in the cell at position (0, 0, 0). (a) **CIC (order 1) particle shape factor**. The particle deposits its charge to the eight nearest vertices (red points). For each cell $icell = (j, k, l)$, *rhocells* store the 8 nearest vertices (j, k, l), $(j+1, k, l)$, $(j, k+1, l)$, $(j+1, k+1, l)$, $(j, k, l+1)$, $(j+1, k, l+1)$, $(j, k+1, l+1)$ and $(j+1, k+1, l+1)$ contiguously. (b) **TSC** (**order** 2) **particle shape factor**. The particle deposits its charge to the 27 neighboring vertices (red and blue points). For a given cell $icell = (j, k, l)$*rhocells* store contiguously the 8 vertices (red points) $(j, k-1, l-1)$, $(j, k, l-1)$, $(j, k+1, l-1)$, $(j, k-1, l)$, $(j, k+1, l)$, $(j, k-1, l+1)$, $(j, k, l+1)$ and $(j, k+1, l+1)$. The blue points are not stored in rhocells and are treated scalarly in the algorithm. (c) **QSP** (**order** 3) **particle shape factor**. The particle deposits its charge to the 64 neighboring vertices (red points). For a given cell $icell = (j, k, l)$, *rhocells* store contiguously the 8 vertices (delimited by red areas) $(j, k-1, l-1)$, $(j, k, l-1)$, $(j, k+1, l-1)$, $(j, k+1, l-1)$, $(j, k-1, l)$, (j, k, l), $(j, k+1, l)$, $(j, k+1, l)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

particles and divided into two consecutive nested loops: (i) a first nested loop that computes particle weights and (ii) a second one that adds the particle weights to its 8 nearest vertices.

4.1.2. *Improvements brought by the new algorithm*

The new algorithm addresses the two main bottlenecks of the SH algorithm with the two following new features:

1. a new data structure for *rho* is introduced, named *rhocells*, which enables memory alignment and unit-stride access when depositing charge on the 8 vertices. In *rhocells*, the 8-nearest vertices are stored contiguously for each cell. The array *rhocells* are thus of size $(8, NCELLS)$ with *NCELLS* the total number of cells. The element $rhocells(1, icell)$ is therefore 64 bytes-memory aligned for a given cell *icell* and the elements $rhocells(1:8, icell)$ entirely fit in one cache line allowing for efficient vector load/stores. The array *rhocells* is reduced to *rho* once, after the deposition is done for all particles (cf. line 49). This step is easily vectorizable (see line 51) but might not lead to optimal performances due to the non-contiguous access in *rho* that leads to gather–scatter instructions. Notice however that this time, this operation is proportional to the number of cells *NCELLS* and not to the number of particles *np* as it was in the case of the SH algorithm. The overhead is thus proportionally lower when there are more particles than cells, which is the case in many PIC simulations of interest,
2. for each particle, the 8 different weights ww are now computed using a generic formula (see lines $41-42$) that suppresses gather instructions formerly needed in the SH algorithm. This also avoids implementing non-portable efficient transpose between the first and second loop, rendering this new algorithm fully portable.

4.2. *Higher particle shape factors*

Similar algorithms were derived for order 2 (TSC) and order 3 particle shape factors, and are detailed in the source file vec_deposition.F90 in the Program library. Corresponding current deposition algorithms can also be found in this file for orders 1, 2 and 3 depositions. In these algorithms, we use three structures *jxcells*, *jycells* and *jzcells* (analogous to *rhocells* for the deposition of *rho*) for the current components *jx*, *jy*, *jz* along directions *x*, *y* and *z*.

In the following, we detail the data structures used for *rhocells* for orders 2 and 3 particle shapes (cf. Fig. 3):

(i) **TSC** (**order** 2) **particle shape.** (cf. panel(b) of Fig. 3 and subroutine depose_rho_vecHVv2_2_2_2 in source file vec_deposition. F90 in Program Library). In this case, the particles deposit their charge to the 27 neighboring vertices. However, storing 27 contiguous vertices per cell in *rhocells* would not be efficient as the reduction of *rhocells* to *rho* would be much more expensive with potential cache-reuse inefficiency. Instead, while the same size for $rhocells(1:8, 1:NCELLS)$ is used, the vertices are now grouped in a different way. The new structure for $rhocells(1:8, 1:NCELLS)$ groups 8 points in a (y, z) plane for each cell *icell* (see red points in red areas). For each cell, each particle adds its charge contribution to 24 points in the three planes at $icell-1$, *icell* and $icell+1$. The three remaining central points (blue points) can be either treated scalarly for 512-bits wide vector registers or vectorized for 256-bits by artificially adding a virtual point that does not contribute to any charge. Notice that we did not find a generic formulation for the weights ww and we are therefore still performing a "gather" instruction for ww in the loop on the vertices (line 2089 in vec_deposition.F90). However, this gather is performed in the *y* and *z* directions for the first plane of 8 points (plane $ic = -1$ on panel (b)) and is subsequently reused on the two other planes $ic = 0$ and $ic = 1$ (see lines 2091–2095 in vec_deposition.F90). Gather is thus performed only 8 times out of 24 points and thus has a limited impact on performance, as shown below in the reported test results.

(ii) **QSP** (**order** 3) **particle shape.** (cf. panel (c) of Fig. 3 and subroutine depose_rho_vecHVv2_3_3_3 in source file vec_deposition.F90 in Program Library). In this case, particles deposit their charge to the 64 neighboring vertices. $rhocells(1:8, 1:NCELLS)$ also group 8 points in a (y, z) plane but differently from the TSC case (see red areas in panel (c)). For each cell, each particle adds its charge contribution to 64 points in the 8 different (y, z) planes at $icell-ncx-1$, $icell-ncx$, $icell-ncx+1$, $icell-ncx+2$, $icell+ncx-1$, $icell+ncx$, $icell+ncx+1$ and $icell+ncx+2$ where *ncx* is the number of cells in the *x* direction (see lines 2434–2450 in vec_deposition.F90). This might reduce the flop/byte ratio of the second loop when *nnx* is large enough so that elements $rhocells(1:8, icell)$ and $rhocells(1:8, icell+nnx-1)$ are not in *L*1 cache. The vertices could have been grouped in (y, z) planes of 16 points instead of 8 points but this would imply a bigger reduction loop of *rhocells* in *rho* and worst performances for a low number of particles. Notice that here again, we did not find an efficient

Listing 3: New vector version of charge deposition routine for CIC (order 1) particle shape factor

```
    SUBROUTINE depose_rho_vecHVv2_1_1_1(...)
        ! Declaration and init
        ...
        nnx = ngridx; nnxy = nnx*ngridy
        moff = (/0,1,nnx,nnx+1,nnxy,nnxy+1,nnxy+nnx,nnxy+nnx+1/)
        mx=(/1_num,0_num,1_num,0_num,1_num,0_num,1_num,0_num/)
        my=(/1_num,1_num,0_num,0_num,1_num,1_num,0_num,0_num/)
        mz=(/1_num,1_num,1_num,1_num,0_num,0_num,0_num,0_num/)
        sgn=(/-1_num,1_num,1_num,-1_num,1_num,-1_num,-1_num,1_num/)

        ! FIRST LOOP: computes cell index of particle and their weight on vertices
        DO ip=1,np,LVEC
            !$OMP SIMD
            DO n=1,MIN(LVEC,np-ip+1)
                nn=ip+n-1
                ! Calculation relative to particle n
                ! --- computes current position in grid units
                x = (xp(nn)-xmin)*dxi
                y = (yp(nn)-ymin)*dyi
                z = (zp(nn)-zmin)*dzi
                ! --- finds cell containing particles for current positions
                j=floor(x)
                k=floor(y)
                l=floor(z)
                ICELL(n)=1+j+nxguard+(k+nyguard+1)*(nx+2*nxguard) &
                +(l+nzguard+1)*(ny+2*nyguard)
                ! --- computes distance between particle and node for current positions
                sx(n) = x-j
                sy(n) = y-k
                sz(n) = z-l
                ! --- computes particles weights
                wq(n)=q*w(nn)*invvol
            END DO
            !$OMP END SIMD
            ! Charge deposition on vertices
            DO n=1,MIN(LVEC,np-ip+1)
                ! --- add charge density contributions to vertices of the current cell
                ic=ICELL(n)
                !$OMP SIMD
                DO nv=1,8 !!! - VECTOR
                    ww=(-mx(nv)+sx(n))*(-my(nv)+sy(n))* &
                        (-mz(nv)+sz(n))*wq(n)*sgn(nv)
                    rhocells(nv,ic)=rhocells(nv,ic)+ww
                END DO
                !$OMP END SIMD
            END DO
        END DO
        ! - reduction of rhocells in rho
        DO iz=1, ncz
            DO iy=1,ncy
                !$OMP SIMD
                DO ix=1,ncx !! VECTOR (take ncx multiple of vector length)
                    ic=ix+(iy-1)*ncx+(iz-1)*ncxy
                    igrid=ic+(iy-1)*ngx+(iz-1)*ngxy
                    rho(orig+igrid+moff(1))=rho(orig+igrid+moff(1))+rhocells(1,ic)
                    rho(orig+igrid+moff(2))=rho(orig+igrid+moff(2))+rhocells(2,ic)
                    rho(orig+igrid+moff(3))=rho(orig+igrid+moff(3))+rhocells(3,ic)
                    rho(orig+igrid+moff(4))=rho(orig+igrid+moff(4))+rhocells(4,ic)
                    rho(orig+igrid+moff(5))=rho(orig+igrid+moff(5))+rhocells(5,ic)
                    rho(orig+igrid+moff(6))=rho(orig+igrid+moff(6))+rhocells(6,ic)
                    rho(orig+igrid+moff(7))=rho(orig+igrid+moff(7))+rhocells(7,ic)
                    rho(orig+igrid+moff(8))=rho(orig+igrid+moff(8))+rhocells(8,ic)
                END DO
                !$OMP END SIMD
            END DO
        END DO

        ...
    END SUBROUTINE depose_rho_vecHVv2_1_1_1
```

generic formulation for the weights ww and we are therefore still performing a "gather" instruction (see lines 2432 and 2442 in vec_deposition.F90). However, this gather is performed in the y and z directions and gathered values are subsequently reused for computing the weights at different positions in x (see lines 2434–2440 and 2444–2450 in vec_deposition.F90). Gather is thus performed only 16 times out of 64 points and thus has a limited impact on performance, as shown below in the reported test results.

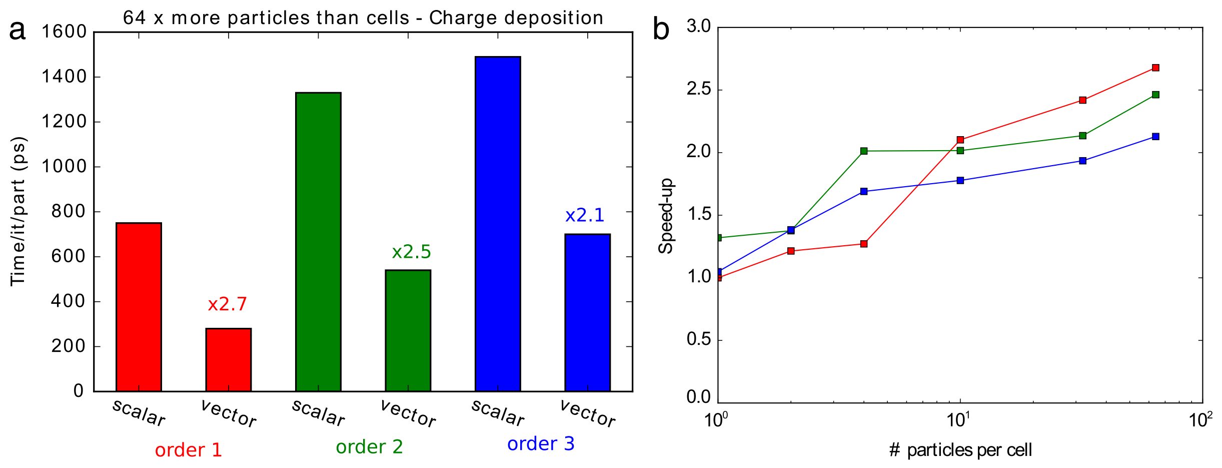

Fig. 4. **Speed-up brought by vectorization with the new charge deposition algorithm on Cori.** (a) Each bar plot shows the time/it/part in *ps* for different particle shape orders 1–3. Benchmarks were done with 64 times more particles than cells. (b) Evolution with the number of particles per cell of the speed-up gain brought by vector routines compared to scalar routines. Red/Green/Blue curves are evolution of the speed-up gain for orders 1/2/3 charge deposition respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

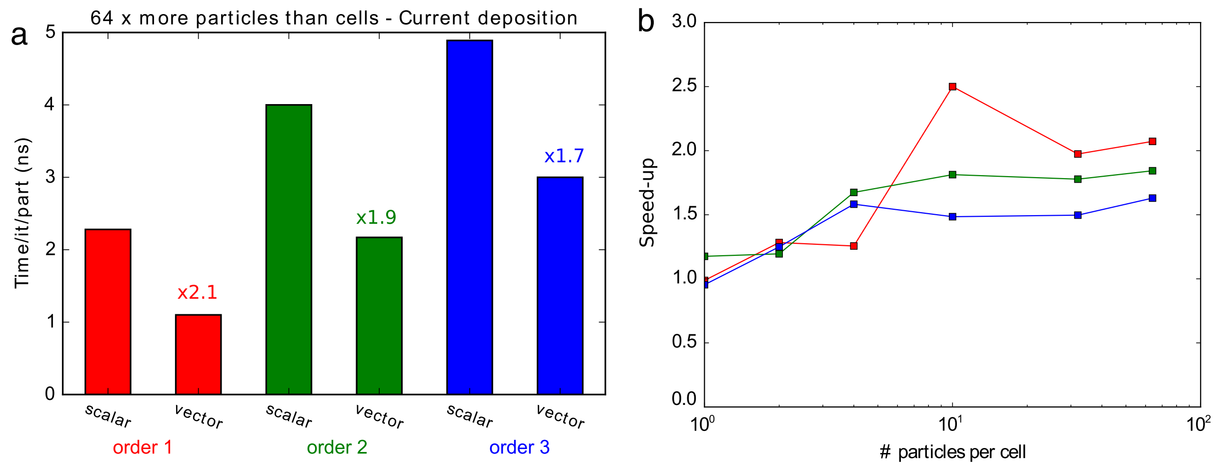

Fig. 5. **Speed-up brought by vectorization with the new current deposition algorithm on Cori.** (a) Each bar plot shows the time/it/part in *ps* for different particle shape orders 1–3. Benchmarks were done with 64 times more particles than cells. (b) Evolution with the number of particles per cell of the speed-up gain brought by vector routines compared to scalar routines. Red/Green/Blue curves are evolution of the speed-up gain for orders 1/2/3 current deposition respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Speed-up brought by the new vector algorithm on Cori

The new vector algorithms were benchmarked on one node (two sockets) of the Cori machine in the same numerical conditions than the ones used in Section 3.3.2 but with 2 MPI processes (one per socket) and 16 OpenMP threads per MPI process. For charge deposition, we use $10 \times 10 \times 10$ tiles in each direction. For current deposition, we use a larger number of tiles ($12 \times 12 \times 12$ tiles in each direction) so that the three structures *jxcells*, *jycells* and *jzcells* (equivalent of *rhocells* for current deposition) fit in cache. Results are shown in Fig. 4 for charge deposition and in Fig. 5 for current deposition. Panels (a) show the time/iteration/particle (in *ps* for Fig. 4 and ns for Fig. 5) for scalar and vector algorithms with different particle shape factors and when there are 64 times more particles than cells. Panels (b) show the evolution of the speed-up brought by the vector algorithm over the scalar one when the number of particles per cell is varied.

Even for a low number of particles per cell (see panels (b) of Figs. 4 and 5), the algorithm performs well, with speed-ups of up to $\times 1.8$. When the number of particles increases performances are even better because the reduction operation of *rhocells*/*jcells* structure in regular arrays *rho*/*j* becomes more and more negligible relatively to particle loops. For 64 times more particles than cells, performances for charge deposition now reach $\times 2.7$ with an order 1 particle shape factor. Order 3 charge/current deposition performs less efficiently than orders 1 and 2, because as we described in the previous section, the structure we chose for *rhocells*/*jcells* decreases the flop/byte ratio of the loop on vertices compared to orders 1 and 2. In the case of simulations using a lot of particles, for which the reduction of *rhocells* in *rho* is negligible, one might consider grouping vertices in *rhocells* by groups of 16 instead of 8 for order 3 deposition in order to increase the flop/byte ratio in loop on vertices. Notice finally that as we vectorize on vertices, there is no performance bottleneck related to a possibly inhomogeneous distribution of particles on the simulation domain.

If speed-ups from particle tiling ($3\times$) and from vectorization (up to $2.7\times$ for charge deposition at order 1) are combined, we can gain a cumulated speed-up of up to $8\times$ for deposition routines.

6. Conclusion and prospects

A new method is presented that allows for efficient vectorization of the standard charge/current deposition routines on current SIMD architectures, leading to efficient deposition algorithms for shape factors of orders 1, 2 and 3. This method uses a new data structure for grid arrays (charge/currents) ensuring data alignment and contiguity in memory which are essential to avoid many gather/scatter operations that can significantly hinder vector performances on modern architectures. The algorithms can be used on current multi-core architectures (with up to AVX2 support) as well as on future many-core Intel *KNL* processors that will support *AVX* −512. Further tests on KNL will be performed as the processor becomes available.

This work presents deposition routines that are fully portable and only use the *$OMP SIMD* directives that are provided by OpenMP 4.0. Efficient vectorization of the charge conserving current deposition from Esirkepov is being investigated, and will be detailed in future work.

Acknowledgments

We thank Karthik Raman from Intel inc for useful discussions on the optimization of the vector routines. This work was supported by the European Commission through the Marie Skłowdoska-Curie actions (Marie Curie IOF fellowship PICSSAR Grant Number 624543) as well as by the Director, Office of Science, Office of High Energy Physics, U.S. Dept. of Energy under Contract No. DE-AC02-05CH11231, the US-DOE SciDAC program ComPASS, and the US-DOE program CAMPA. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

References

[1] Birdsall, Langdon, Plasma Physics via computer simulation, 15-5.
[2] A. Nishiguchi, S. Orii, T. Yabe, J. Comput. Phys. 61 (1985) 519.
[3] E.J. Horowitz, J. Comput. Phys. 68 (1987) 56.
[4] J.L. Schwarzmeier, T.G. Hewitt, Proceedings, 12th Conf. on Numerical Simulation of Plasmas, San Francisco, 1987.
[5] A. Heron, J.C. Adam, J. Comput. Phys. 85 (1989) 284–301.
[6] G. Paruolo, J. Comput. Phys. 89 (1990) 462–482.
[7] P. Kogge, K. Bergman, S. Borkar, D. Campbell, W. Carlson, W. Dally, M. Denneau, P. Fran- zon, W. Harrod, K. Hill, J. Hiller, S. Karp, S. Keckler, D. Klein, R. Lucas, M. Richards, A. Scarpelli, S. Scott, A. Snavely, T. Sterling, R.S. Williams, K. Yelick, Exascale Computing Study: Technology Challenges in Achieving Exascale Systems. Technical Report, DARPA, 2008.
[8] R.A. Fonseca, J. Vieira, F. Fiuza, A. Davidson, F.S. Tsung, W.B. Mori, L.O. Siva, ArXiv, http://arxiv.org/pdf/1310.0930v1.pdf, 2013.
[9] http://www.nersc.gov.
[10] http://warp.lbl.gov.
[11] T. Esirkepov, Comput. Phys. Comm. 135 (2001) 144–153.
[12] Viktor K. Decyk, Tajendra V. Singh, Comput. Phys. Comm. 185 (2014) 708–719.
[13] David V. Anderson, Dan E. Shumaker, Comput. Phys. Comm. 87 (1995) 16–34.