WARWICK
THE UNIVERSITY OF WARWICK

Original citation:
Bird, Robert F., Gillies, P., Bareford, M. R., Herdman, J. A. and Jarvis, Stephen A.. (2016) Performance optimisation of inertial confinement fusion codes using mini-applications. International Journal of High Performance Computing Applications.
Permanent WRAP URL:
http://wrap.warwick.ac.uk/81789

Copyright and reuse:
The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Publisher's statement:
This had been posted ahead of publication.
Bird, Robert F., Gillies, P., Bareford, M. R., Herdman, J. A. and Jarvis, Stephen A.. (2016) Performance optimisation of inertial confinement fusion codes using mini-applications. International Journal of High Performance Computing Applications. Copyright © 2017 The Authors. Reprinted by permission of SAGE Publications.
https://doi.org/10.1177/1094342016670225

A note on versions:
The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

Performance Optimisation of Inertial Confinement Fusion Codes using Mini-applications

R. F. Bird[1], P. Gillies[2], M. R. Bareford[3], J. A. Herdman[2], and S. A. Jarvis[1]

[1]Department of Computer Science, University of Warwick, UK
[2]High Performance Computing, AWE plc Aldermaston, UK
[3]EPCC, University of Edinburgh, UK

August 26, 2016

Abstract

Despite the recent successes of nuclear energy researchers, the scientific community still remains some distance from being able to create controlled, self-sustaining fusion reactions. Inertial Confinement Fusion (ICF) techniques represent one possible option to surpass this barrier, with scientific simulation playing a leading role in guiding and supporting their development. The simulation of such techniques allows for safe and efficient investigation of laser design and pulse shaping, as well as providing insight into the reaction as a whole. The research presented here focuses on the simulation code *EPOCH*, a fully relativistic particle-in-cell plasma physics code concerned with faithfully recreating laser-plasma interactions at scale.

A significant challenge in developing large codes like EPOCH is maintaining effective scientific delivery on successive generations of high-performance computing architecture. To support this process, we adopt the use of *mini-applications* – small code proxies that encapsulate important computational properties of their larger parent counterparts. Through the development of a mini-application for EPOCH (called *miniEPOCH*), we investigate a variety of the performance features exhibited in EPOCH, expose opportunities for optimisation and increased scientific capability, and offer our conclusions to guide future changes to similar ICF codes.

1 Introduction

For decades, the UK has been a significant contributor to the research field of high-intensity laser-plasma interactions. The UK's Central Laser Facility is home to some of the world's most advanced high power lasers, which can deliver Petawatt focused beams, with approximately 10,000 times more power than the UK National Grid, during picosecond pulses. Developments in the deployment of relativistically intense 'long' laser pulses (to compress fuel) and fast 'short' pulses (for ignition) present significant challenges in computational plasma physics. Plasmas with intense electromagnetic fields require fully kinetic models of particle distribution in 7 dimensions (3 space, 3 momentum and time); and point design for targets requires the coupling of relativistic kinetic models with long time-scale radiation hydrodynamics codes. As future gyrokinetic codes continue to develop to support plasma turbulence studies, in order to exploit facilities such as ITER for example, the complexity of these simulations and the demands on the supporting supercomputers will also increase.

Particle-in-Cell (PIC) codes are amongst the most widely used computational tools in plasma physics research, and help develop further understanding of both inertial confinement fusion (ICF) and laser-plasma interactions in general. The research presented here focuses on the Extensible PIC Open Collaboration simulation codebase, named *EPOCH* [2], which is a nationally funded, fully relativistic EM particle-in-cell plasma physics code, developed by a leading network of over 30 UK researchers.

A significant challenge in developing large codes like EPOCH is maintaining effective scientific delivery on successive generations of high-performance computing architectures. In EPOCH, collections of physical particles are represented using a smaller number of pseudoparticles; the fields generated by the motion of these pseudoparticles are calculated using a finite difference time domain on an underlying grid of fixed spatial resolution. The forces on the pseudoparticles due to the calculated fields are used to update the velocities of the pseudoparticles, and these velocities are then used to update their positions. Using this approach it is possible to reproduce the full range of classical micro-scale behaviour of a collection of charged particles. Like many codes of this type, EPOCH is Fortran-based and MPI parallelised; dynamic load balancing options exist and MPI-IO allows checkpoint re-start on an arbitrary number of processors. Legacy simulation codes designed and implemented in this way exhibit poor utilisation of modern hardware features such as vector operations, and fail to fully exploit all levels of available parallelism – a problem which is exacerbated by the energy-efficient benefits available through heterogeneous computing.

The continued development, maintenance and future-proofing of EPOCH represents a significant software engineering challenge, EPOCH represents decades of development by skilled domain experts – the code is feature rich, but equally large and complex. Code porting to explore the potential benefits of new compute architectures represents a significant undertaking, and the resulting benefits of this effort may indeed be small. To help mitigate these problems, we adopt the use of *mini-applications* (commonly termed *mini-apps*) – small code proxies that encapsulate important computational properties of their larger parent counterparts [3, 12]. The existence of mini-apps is built on the premise that (i) although simulation codes may have millions of lines of source code, their performance is often dominated by a small subset of the code, and (ii) simulation codes may contain many physics models that are mathematically distinct, but in many cases exhibit similar performance characteristics. Mini-apps operate by encapsulating the most important computational operations and consolidating physics capabilities that have the same performance profiles; they will typically be orders of magnitude smaller than their parent code, and as a result be easier to port, easier to improve, easier to extend, and less likely to be subject to restrictive licensing governing their use or distribution.

This paper makes a number of contributions:

- We present the development of a new mini-app (*miniEPOCH*) for the parent code EPOCH. This mini-app is then used to explore known performance problems within EPOCH, and in particular (i) the increasing time-step duration during simulation runtime, and (ii) high levels of cache miss rates due to particle-store fragmentation;
- Using *miniEPOCH*, we explore opportunities for code optimisation, including the utilisation of shared memory, exploiting increasing vector width, increased vectorisation and improving memory locality.
- We demonstrate the use of a specialised kernel to exploit domain specific knowledge, leading to improved Particle-Per-Cell (PPC) scaling of the main PIC algorithm. To do this, a new kernel is introduced which is able to make additional assumptions about particle movements, falling back to the original kernel when these assumptions do not hold.
- Finally, we validate these EPOCH optimisations on ARCHER, a 1.6 PFlop/s Cray XC30, housed at the UK national supercomputing centre at EPCC. These improvements demonstrate a 2.02× speed-up in the core EPOCH algorithm and a 1.55× speed-up to the overall application runtime. Additionally we demonstrate the ability of an alternate algorithm to enable increased PPC.

The remainder of this paper is organised as follows: Section 3 provides additional background on EPOCH and its core algorithms; Section 2 documents an analysis of related work; Section 3 details performance and algorithmic characteristics of EPOCH and the resulting mini-app implementation; Section 4 presents a detailed study of code optimisation using the mini-app, and the translation of these optimisations to the parent code; Section 5 concludes the paper.

2 Related Work

Despite the benefits of mini-apps being identified as early as 1991 [3], the re-emergence and use of mini-apps has only gained greater traction in recent years [12]. There are now examples of mini-app use in co-design, code optimisation and porting, and in the exploration of new programming languages and paradigms. These works include our prior work with EPOCH, on which this work is directly built [5]. We refine our previous work by further expanding our investigation, most notably by including a discussion of techniques targeting increasing numbers of particles-per-cell. Below, we provide a discussion of mini-apps in the context of PIC research, and highlight some of the most relevant work to this study.

The use of mini-apps for code optimisation is exemplified in the work by Karlin et al. [16, 15], in which the authors use the mini-app LULESH (Livermore Unstructured Lagrangian Explicit Shock Hydrodynamics) to demonstrate the optimisation of multi-material hydrodynamics simulations, increasing both the performance of their mini-app (which solves a Sedov blast problem) and associated parent codes, such as Lawrence Livermore National Laboratory's (LLNL) arbitrary Lagrangian-Eulerian multi-physics code ALE3D. By using their mini-app to better focus optimisation efforts, they managed to increase LULESH's vector instruction utilisation by a factor of 8, reduce the number of memory reads by 62%, and reduce the overall application memory footprint by 19%. These improvements were then mapped back to ALE3D to achieve a 20% reduction in overall application runtime [23, 14, 20], performance gains that had remained undetected until the mini-app investigation.

Further studies demonstrate the use of mini-apps to rapidly investigate both hardware platforms and programming paradigms. Lavallée et al. [17] demonstrate how they were able to use the mini-app called HYDRO, developed by CEA Saclay, to investigate multiple hardware platforms available through the PRACE Tier-0 Research Infrastructure; similarly, the size of the mini-app greatly aided their development of several code variants of HYDRO, including those employing MPI, OpenMP, CUDA, OpenCL, HMPP and UPC – a task that would have been infeasible using the full production application. The availability of such a diverse range of code implementations allowed the authors to evaluate a variety of heterogeneous hardware and assess its viability as a future platform for the parent application RAMSES [26], an EU-funded computational astrophysics package used for the study of large-scale structure and galaxy formation. Such studies demonstrate the importance of mini-apps as a tool to enable studies in code portability, scaling, and performance.

While particle-in-cell codes are well understood [7, 8, 9, 24], the application of the mini-app software engineering methodology to the field of plasma physics and PIC remains largely unexplored. Previous work has been undertaken with the aim of providing flexible, concise environments for the development of PIC codes [21, 11]. GTC [18], at only 8,000 lines of code, is one such example of this; however, GTC is not associated with any parent code *per se*, and any findings associated with this code must still be translated to larger production codes in this code class through additional research. To the best of our knowledge, the research presented here is the first to develop and apply a PIC mini-app, which is associated with a large, production-code equivalent.

Mini-apps are increasingly being developed within open source frameworks, thus allowing the HPC community to benefit from their development. Projects such as Mantevo [12, 4] and the UK Mini-App Consortium (UKMAC) [1] exist to provide centralised repositories where collections of mini-applications can exist, selected to represent key scientific areas supported by high-performance computing simulation.

Finally, as point of clarification, the work presented here focuses on fully relativistic PIC codes, where Maxwell's equations are solved using FDTD methods; this is not the same as those solutions which utilise gyrokinetic equations, or those that employ Fourier transforms to operate in the time domain.

3 The Extensible PIC Open Collaboration simulation codebase (EPOCH)

EPOCH is a nationally funded, fully relativistic EM particle-in-cell plasma physics code, developed by a leading network of over 30 UK researchers. At the core of this simulation codebase are particle push and field update algorithms, developed by Hartmut Ruhl [25], extended to include advanced features such as collisions, ionisation and quantum electrodynamics- (QED) driven coherent radiation. EPOCH tracks the electric and magnetic fields generated by the motion of pseudoparticles, and is capable of reproducing the full range of classical microscale behaviour required to accurately simulate a collection of charged particles. Figure 1 depicts the core PIC algorithm used in EPOCH, which typically accounts for over 80% of the application runtime, and consists of the following three steps:

1. Move the particle across the physical domain, proportional to particle momenta;
2. Update the particle's momentum, based upon the local electric fields, magnetic fields, and particle shape;
3. Deposit the generated current onto the grid, to act as an intermediary for particle-particle interactions.

These steps represent a considerable computational workload and are currently expressed as a single code kernel in which the particle loop spans approximately 500 lines of source code.

To distribute work between processing elements, EPOCH uses a static n-dimensional MPI domain decomposition, assigning a rectangular region of the physical domain to each processing element. Each MPI task is then responsible for a distinct region of the domain, transferring control of particles as they leave the space, and accepting any particles which may enter. This method of decomposition is known to exhibit poor performance for problems which display strong load imbalance, so in the interests of presenting a fair study, we limit this investigation to problems that remain well balanced throughout their operation [19]. Techniques do, however, exist to combat such multi-node load imbalance, and include: i) advanced load balancing strategies [22]; ii) work stealing and migration [13]; and iii) advanced domain decomposition [10]. The

```
for all species do
    for all particles do

        ▷ Move particles.
        position ← position + momentum

        ▷ Update momentum based on field effects.
        e_cell ← ⌊position⌋
        for all neighbours of e_cell do
            calculate electric field effects
        end for

        b_cell ← ⌊position + 0.5⌋
        for all neighbours of b_cell do
            calculate magnetic field effects
        end for

        momentum ← momentum + electric and magnetic field effects

        ▷ Calculate and deposit currents.
        for all neighbour cells do
            calculate current
            deposit current
        end for

    end for
end for
```

Figure 1: Pseudocode depiction of EPOCH's core particle-in-cell (PIC) algorithm.

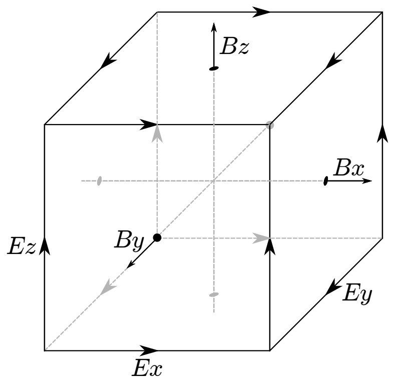

Figure 2: A so-called Yee Grid, applying centered finite difference operators on staggered grids in space and time for each electric and magnetic vector field component.

investigation of such methods resides outside of the scope of this study, and will be addressed in future work.

In the current implementation of EPOCH, particles are densely packed, with a single particle spanning multiple grid cells (typically 3×3). EPOCH employs a Finite-Difference Time-Domain method (FDTD), and represents electrical and magnetic fields on a staggered Yee grid [27], as shown in Figure 2. During the momentum update (step 2), a 25-point stencil in each of 3 dimensions is read per field, and used to update the particle momenta; the scale of these memory operations are a significant contributor to the overall application runtime. Once the current contributions have been calculated (step 3), they are then stored into a global array at indices determined by the grid vertices touched during particle movement. This write to multiple indices of a global array limits the possibility of particles simultaneously depositing current without the need for atomics or other concurrency control. It is this *current deposition* that most strongly differentiates PIC from alternative methods; unlike molecular dynamics, for example, PIC features no particle-particle interactions, instead approximating these using the Yee grid as an intermediary.

A known performance issue observed during the operation of EPOCH is that the duration of each time step increases as the simulation progresses. This problem is demonstrated in Figure 3, where we see a sharp increase in time-step duration until approximately 4,000 steps, after which time-step duration stabilises. This observation is counter-intuitive, as there is no change to the kernel during runtime. This issue has persisted through many generations of EPOCH; in Section 4 we build on our knowledge and understanding gained in previous research [6] to address this time-step scaling problem, as well as using the new EPOCH mini-app to explore further code optimisation opportunities.

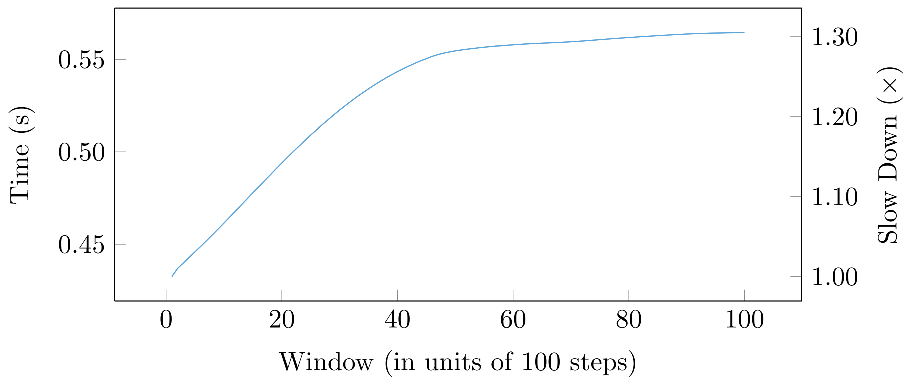

Figure 3: The duration of each time step in EPOCH as a simulation progresses. Each simulation window, of which there are 100 in total, contains 100 steps.

3.1 Optimisation Considerations

The original version of EPOCH uses a linked list to store its particles. While this itself does not present a problem, naïve implementations of linked lists offer no guarantees of contiguous memory access. This means that as data elements are inserted and deleted, memory allocations take place without consideration for locality of existing data, and considerable memory fragmentation can occur. This fragmentation can significantly impact performance, as modern hardware is optimised for contiguous memory loads, with each issued memory load fetching an entire cache line. By aligning data and promoting the grouping of data within cache lines, memory locality can be improved and one can reduce the effective bandwidth required to load the same amount of data from main memory. This effect can be seen when particles move across physical processor boundaries; as particles exit they are deleted and new particles added at arbitrary memory addresses. This means that although the initial particle allocation may be contiguous, the memory access pattern degrades over time, as a function of particle movement.

A second candidate for improving memory locality, and as a result the effective memory bandwidth, is to group data in memory such that it will cause subsequent loads to common addresses. In so doing, memory is more likely to be resident in cache when it is required, thus reducing cache eviction and thrashing. To achieve this in miniEPOCH, we implement a particle sort to group spatially local particles in memory. We then exploit domain specific knowledge to further improve this. During the second step of the algorithm, the momentum of particles is updated based on the surrounding magnetic and electric fields. Given the staggered nature of the Yee grid (Figure 2), it is possible to group particles which share a common vertex, and in so doing promote reuse within the 3-dimensional stencil update.

A further performance limiting factor in EPOCH is its inefficient use of

```
for all species do

    for all particles do
        ▷ Move particles.
        position ← position + momentum
    end for

    ▷ Optional Particle Sort.
    for all particles do
        ▷ Update momentum based on field effects.
        e_cell ← ⌊position⌋
        for all neighbours of e_cell do
            calculate electric field effects
        end for

        b_cell ← ⌊position + 0.5⌋
        for all neighbours of b_cell do
            calculate magnetic field effects
        end for

        momentum ← momentum + electric and magnetic field effects
    end for

    for all particles do
        ▷ Calculate and deposit currents.
        for all neighbour cells do
            calculate current
            deposit current
        end for
    end for

end for
```

Figure 4: Pseudocode depiction of EPOCH's modified particle-in-cell (PIC) algorithm.

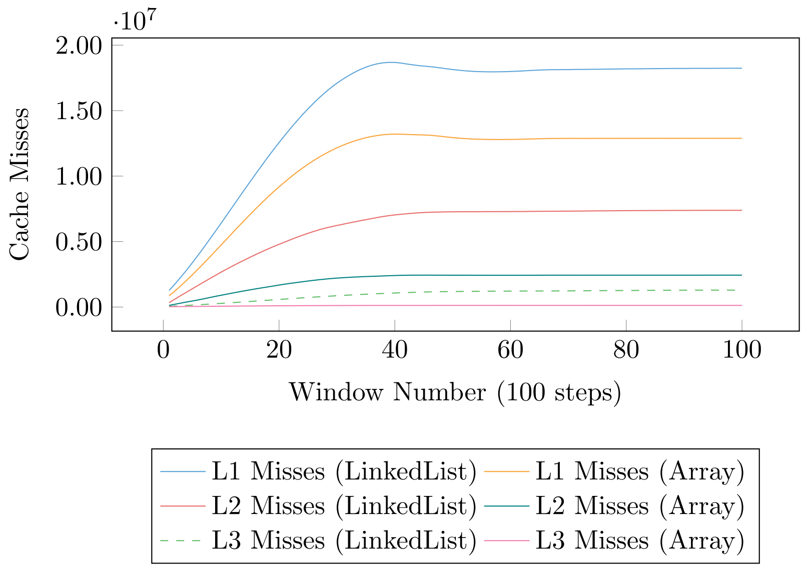

Figure 5: Cache misses during a simulation consisting of 100 simulation windows, each window containing 100 steps.

vector instructions (SIMD). Typically, it is desirable to vectorise over the most computationally intense code regions in order to fully exploit SIMD. In EPOCH this means vectorising over the *particles loop*. However, in the current expression of the algorithm, such auto-vectorisation of particles is not possible due to a classical update dependency within the current deposition. This update dependency is key to the PIC algorithm, so where the original code expresses the kernel as a single large loop, we adapt this in favour of expressing the code as three discrete steps. We explore this in our mini-app using loop fission, with pseudocode for this shown in Figure 4. As well as promoting vectorisation, splitting the code in this way also allows us to sort particles directly after they have been moved, giving us a stronger guarantee regarding particle reuse during the field-effect stencil. Hereafter we will refer to these three components as the *move-*, *stencil-* and *current- kernels*.

While the sort itself does increase the amount of computation required for the execution, this is mitigated by two factors: i) Much of the computation for the sort can be done while the particle is in cache from the particle move; and ii) For particles that are grouped together, we can avoid recomputing shared properties, which was not previously possible. Further to this, it is also possible to employ the sort after the *stencil* kernel, allowing us to place guarantees on the order in which the *current* kernel updates global memory, which can in turn be exploited to remove the classical update dependency and achieve vectorisation. As auto-vectorisation of all three kernels is possible, we can further increase vector efficiency by performing scalar replacement on arrays where possible,

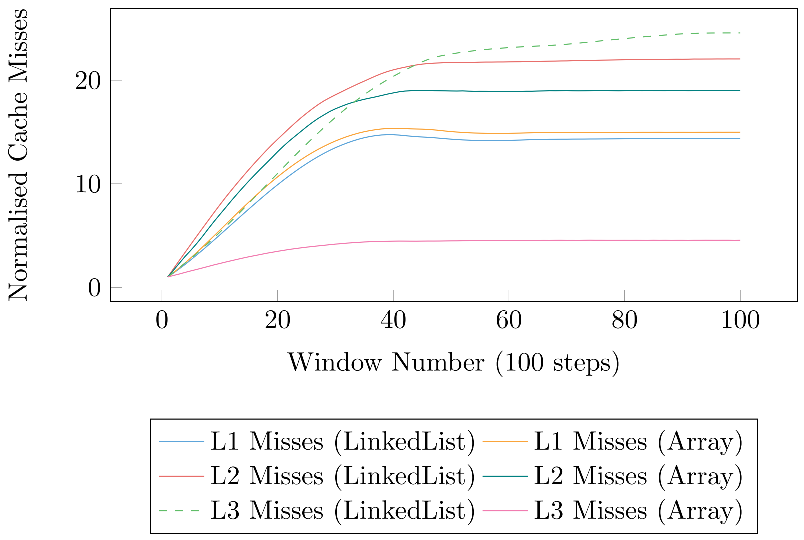

Figure 6: Normalised cache misses during a simulation consisting of 100 simulation windows, each window containing 100 steps.

and employing SIMD lane indexing to ensure all SIMD temporary arrays have coalesced data accesses.

3.2 Experimental Setup

Throughout this work we report numerical results for the periodic interaction of two densely packed electron streams. Each pseudoparticle is assumed to have an individual mass, and wide spanning third order *b*-spline stencil. Particle probes are disabled and, unless otherwise stated, a typical problem size of 128^2 grid-cells per core is used, with 32 particles per species, per cell, on a fully packed node. The results detailed in Section 4 were obtained from ARCHER, a 1.6 PFlop/s Cray XC30 housed at the UK national computing centre at EPCC. ARCHER features 4,920 dual socket Intel Xeon E5-2697v2 Ivy Bridge nodes, with 24 cores per node. Intel 15.0 was used to compile all code variants, with the highest level of code optimisation enabled (`-O3`), and with platform specific code generation (`-xHost`) enabled. PAPI 5.3.2.1 was used to gather the results of selected performance counters, including those used for recording cache misses and vector instruction counts.

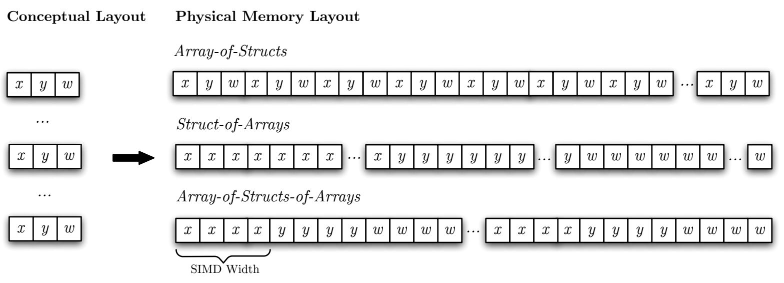

Figure 7: A comparison of Array-of-Structs, Structs-of-Arrays, and Array-of-Structs-of-Arrays data layouts.

4 Results

4.1 Known Performance Issues

During the initial investigation of the increasing duration of time-steps in an EPOCH execution, it was believed that the primary contributor to the poor time-step scaling was increasing fragmentation of the linked list. As the particles move between MPI ranks, it was expected that the particle store would become more fragmented. As previously discussed, this was due to new particles being added to the store as they entered the domain, whilst others were removed as they leave. Figure 5 shows the typical cache miss rates for an unsorted, linked-list implementation of EPOCH, while Figure 6 shows this data expressed as relative differences. We can clearly see that as the simulation progresses, the cache-miss-rate increase is strongly correlated with overall runtime, and peaks at approximately 4,000 steps. After 4,000 steps, the cache miss rates increase at a much reduced rate. While this data strongly suggests that the increase in runtime is caused by the change in cache miss rates, it provides no clue as to its origin.

To investigate this further, we used the mini-app to implement an alternative, array-based, particle store. This allows for contiguous access to particles, avoiding any fragmentation. This greatly improves the time-step scaling of the mini-app, with Figure 5 showing the representative changes in cache misses for a version of EPOCH with these changes applied. It is clear that this change in implementation only partially addresses the problem, with cache hit rates still scaling poorly as time-steps progress. It is interesting to note the marked decrease in L3 cache misses, which explains the substantial benefit to runtime scaling seen in Figure 8 for the array version of EPOCH. Whilst it is clear from Figure 8 that an array version of EPOCH benefits the overall runtime, the cache miss figures in Figure 5 identify that a secondary problem still exists. Further analysis of the cache miss rates directed our attention to the large stencil required when applying electromagnetic effects to the particle momenta. Increasing disorder in the particle list may explain the poor cache and memory

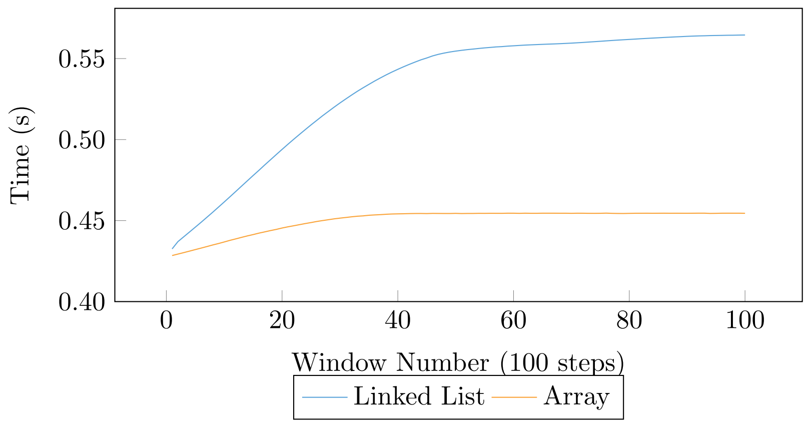

Figure 8: EPOCH Time-step duration, during a simulation consisting of 100 simulation windows, each window containing 100 steps.

behaviour, it would increase the probability of cache pressure, and the probability of spilling during this stencil-gather. Our strategy here was to periodically sort the particle store, and in so doing investigate the effect particle disorder had on both cache misses and runtime. Such a sort remedied the problem, and allowed for near perfect time-step scaling with a maximum deviation of 0.03% over ten thousand time-steps.

4.2 Vectorisation

Having arrived at a significantly improved array-based version of miniEPOCH, efforts could then be focused on optimising EPOCH to ensure that it was better able to utilise current and future hardware. At the outset of this study, EPOCH was unable to exploit vector operations in its main kernel. The reason for this was a classic update dependency when accumulating currents, with multiple particles possibly having to write to the same array location concurrently. Given the large size of the initial kernel (approximately 500 lines of code), loop fission could be used to great effect in order to separate out much of the non-dependant computation which was SIMD-parallel safe. The previously mentioned sort, combined with iterating over particles on a per vertex basis, allows us to hoist loop invariant calculations so that they could be performed once per vertex. This offered a decrease in the overall required computation, and allowed the compiler to better predict memory access patterns into global memory.

Through code modifications first tested in the mini-app, we were able to ascertain that the restructured fissioned kernel was able to successfully exploit vector instructions. Figure 9 shows the SIMD scaling of kernel runtime for 256-bit AVX (4 doubles), 128-bit AVX (2 doubles), and for the code operating without vectorisation. We see reasonable SIMD scaling for all kernels, except for *move*. This is because of its memory-stream-like structure, and the lack of float-

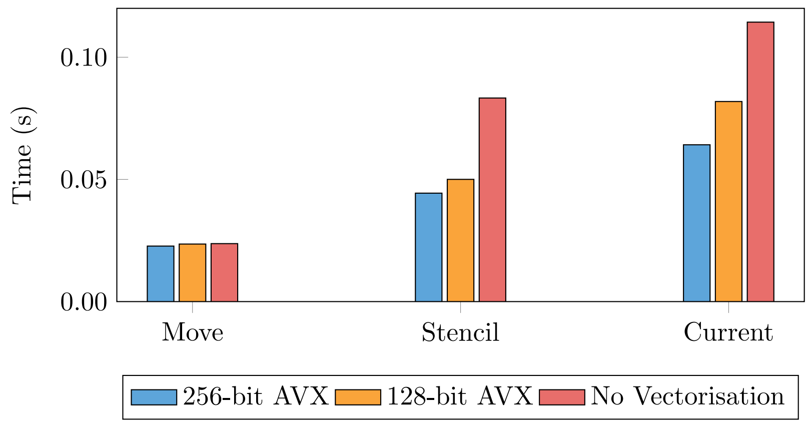

Figure 9: SIMD scaling of miniEPOCH AoS kernels for 256-bit AVX, 128-bit AVX, and for the code operating without vectorisation.

ing point operations to hide the memory accesses. The *stencil* and *current* kernels represent the majority of the time spent in miniEPOCH, and demonstrate significantly improved performance with SIMD width. Such improvements will yield further gains if the current trend of increasing SIMD width continues.

We can further improve performance by considering alternate memory layouts used for the array storage. Not only does this change how data is accessed, it also determines the number of concurrent memory streams the processor has to track during pre-fetching. Typically, particles are stored in an array of custom objects (or structs), a technique known as an *Array-of-Structs (AoS)*. By storing particles as an AoS, a single memory stream is required, with each particle loaded bringing with it all particle properties from main memory. This effect holds regardless of the number of properties used in the given kernel, and can represent a significant overhead for kernels which require few fields. When processing multiple array elements, as is typical in SIMD, loads must be gathered from memory, and any writes scattered, incurring a performance cost and increasing the latency of the memory operations. Alternative approaches include a *Struct-of-Arrays (SoA)* and a more complex hybrid, *Array-of-Structs-of-Arrays (AoSoA)* (which aims to combine the benefits of both SoA and AoS). A brief overview of these memory layouts is found in Figure 7, where in our experiments *SIMD Width* was typically 4, as we were operating on doubles with 256-bit AVX.

For the SoA data layout, single particle properties for multiple particles are stored together in an array. This means that under SIMD operation, single properties from multiple particles can be loaded in one contiguous and aligned load, at the expense of tracking a different memory stream per property required. This eliminates any potential for gather/scatters, and is often favourable when only a few particle properties are required. With AoSoA, groups of N elements of each property are stored together, in order, where N is typically a function of

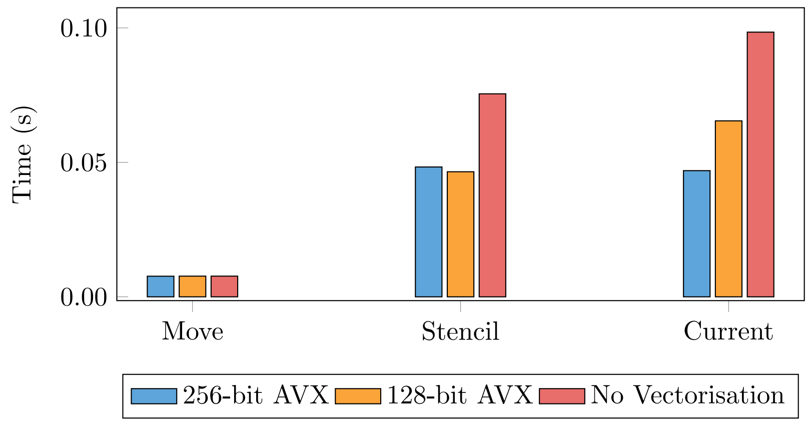

Figure 10: SIMD scaling of miniEPOCH SoA kernels for 256-bit AVX, 128-bit AVX, and for the code operating without vectorisation.

vector length. This approach attempts to combine the benefits of both SoA and AoS, but comes at the expense of vastly increased complexity and an indexing overhead. In the current kernel configuration, a particle has 6 properties and stores 1 intermediary value (all types are `doubles`). The *move* kernel requires 5 such properties, the *stencil* kernel requires 6 and the *current* kernel requires 4. To investigate enhanced SIMD scaling, we ported the mini-app to each of the alternative memory layouts – this provides an excellent example of where mini-apps allow rapid code exploration, which might not otherwise be possible on full production codes. Figure 10 shows the vector scaling of the SoA implementation, which compared to Figure 9 shows that the performance is favourable in the kernels requiring fewer particle field accesses, and generally favourable overall. This result is largely due to the more effective use of data loaded using SoA, as no bandwidth is wasted. Figure 11 shows the relative difference in instruction counts for varying SIMD widths, as recorded by PAPI. For good vector scaling one would expect to see the number of instructions executed decrease as a function of SIMD width. We see that both the `move` and `current` kernels scale as expected, but that the `stencil` kernel is only able to partially benefit from the increase in SIMD width.

Figure 12 shows the overall performance as a result of each memory layout. Again we observe that as fewer particle properties are required, SoA performs better and, as more particle properties are required, so AoS outperforms SoA. AoSoA pays the cost of masked hardware instructions due to the sparse particle grouping used, a cost which will be much reduced in future hardware generations and has already been much reduced on the Intel Xeon Phi product range. We see that each kernel benefits differently from the change in array layout, largely due to cache pressure and required memory bandwidth.

As with all mini-app optimisation studies, our goal was to map our improve-

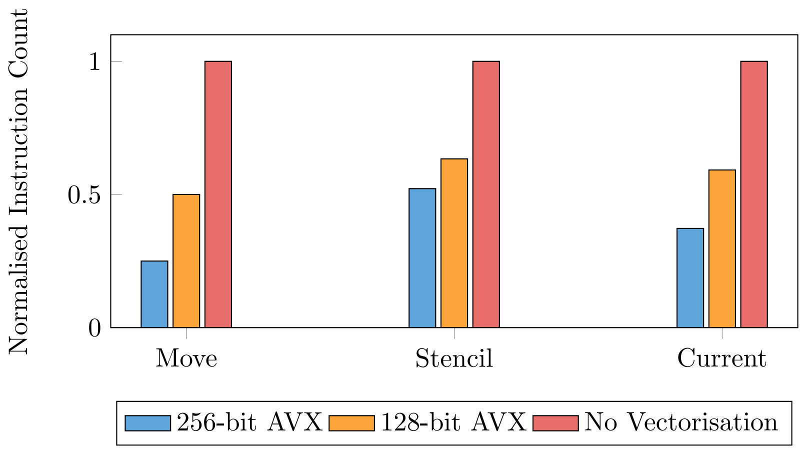

Figure 11: Normalised instruction counts per kernel for different SIMD widths for the *Struct-of-Arrays (SoA)* memory layout.

ments back to the parent code in order to facilitate improved scientific investigation. Figure 13 shows the overall runtime for an optimised version of EPOCH, as well as a comparison against the original EPOCH code base and an array-based implementation. The array versions show a small increase over the original implementation due to the lack of memory fragmentation and increased memory locality. The optimised version of EPOCH represents a considerable improvement in code performance. It includes all previously discussed improvements, including spatial sort, increased vectorisation, loop-invariant code hoisting, array scalarisation, and coalesced temporary SIMD arrays. These optimisations deliver a notable improvement to production code runtimes, and successfully demonstrate the value a mini-app based optimisation investigation. These improvements, recorded on ARCHER, a 1.6 PFlop/s Cray XC30, demonstrate a 2.02× speed-up in the core EPOCH algorithm and a 1.55× speed-up to the overall application runtime.

4.3 Particle-Per-Cell Scaling

The accuracy of PIC simulations is most strongly governed by the number of particles included in the simulation. We typically choose to describe this parameter as the number of Particles-Per-Cell (PPC). For a fixed problem, as the number of PPC increases, the number of real world particles each pseudo-particle in the simulation represents decreases. This increase in interacting bodies allows for more complex interactions to take place more frequently, allowing the simulation to be more true to real life experimentation. As the number of PPC increase however, so too does the amount of computational work the simulation is required to do. Whilst it is not guaranteed, this increase is typically linear with the number of PPC, as there are no direct particle-particle interactions,

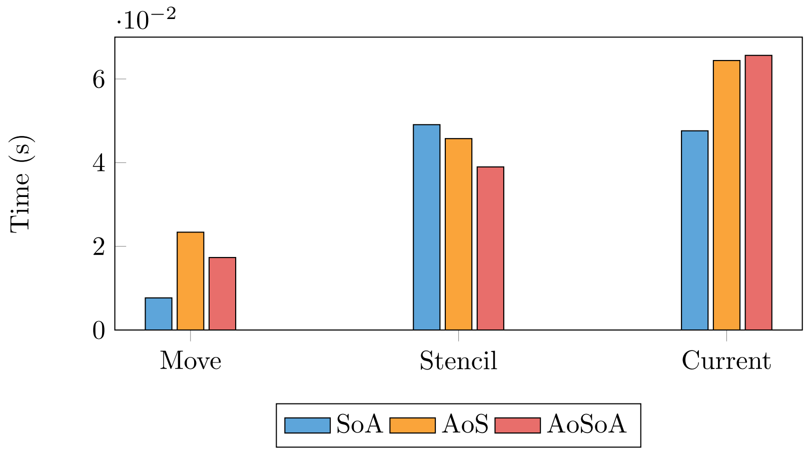

Figure 12: Kernel runtime for different data layouts.

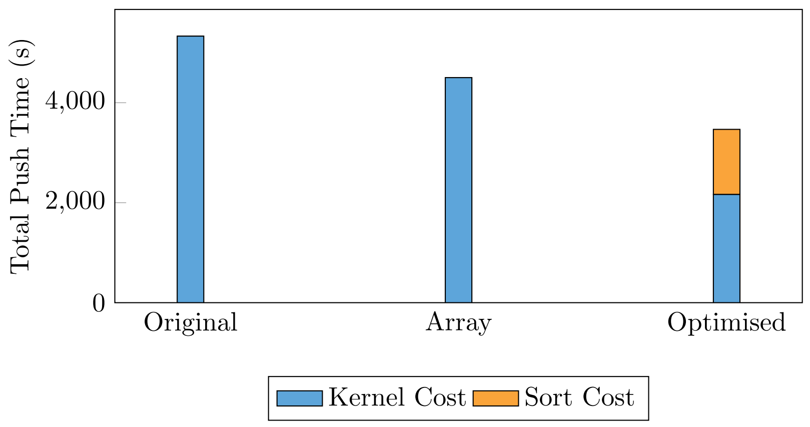

Figure 13: Overall runtime for the original and optimised versions of EPOCH for a 10,000 step run. The additional *sort* cost is highlighted for clarity.

each additional particle introduces an approximately constant amount of work to the system.

Here we present a technique to efficiently deal with increasing PPC counts, and the associated increase in computational work. We present a performance analysis in the context of miniEPOCH, and use it to demonstrate the viability of the approach to offer good performance on increasing scientific workloads. Such efficiency will become increasingly important as we continue towards the large runs required to facilitate exa-scale level science. The presented technique builds directly on top of the sorting work discussed, and aims to further exploit the spatial ordering of the particles in conjunction with domain specific knowledge. In order to achieve this, we implement an alternative kernel, with specialised logic, to operate on sorted regions of particles – specifically, those which do not cross grid boundaries during the particle push. In so doing, we can reduce the overall sort cost of the particles, whilst simultaneously seeking to increase algorithmic efficiency. By exploiting the knowledge that the particle groupings do not cross any grid boundaries, the specialised kernel can make a variety of simplifying assumptions.

In the default kernel, there is an overhead to tracking the motion of the particles. The largest observable over head of this, is maintaining the arrays of coefficients based on particle movement and shape – tracked as two arrays per dimension (`hx/hy`, `gx/gy`). As the particles move, these array of coefficients need to be matched based on the direction of particle movement, before any mathematical operations can be applied. In a kernel with the assumption of no particle movement, this matching is entirely deterministic and does not need to be explicitly calculated. This can be leveraged to decrease both the memory and compute needed to perform this, reducing the stencil from 7×7 to 5×5, whilst also guaranteeing static loop bounds.

Whilst this reduction may initially seem modest, it decreases the amount work by approximately 30% in each dimension. Furthermore this stencil is iterated across fully in a tight two dimensional loop, representing a reduction in work by nearly 50%. A range of additional benefits to this technique also exist, the most important of which is reduced memory bandwidth. Not only does the decrease in the per-particle array size mean a reduced memory footprint, it also reduces the traffic to main memory and cache. This in turn increases the ratio of floating point operations per byte, allowing for increased latency hiding. Finally, by establishing a tighter bound on the physical domain each particle can interact with (implemented as shared global memory), this can give more control and flexibility when multi-threading, as it increases the maximum amount of concurrency without overlap. This kernel can then be applied to all particles, with the result being masked out for those particles which do cross a cell boundary. A second pass of the full algorithm can then instead be applied to these particles.

By inspecting Figure 14, it can seen that the PPC scaling of the original EPOCH code is perfectly linear, with an increase in PPC work being mirrored exactly in runtime. This can be directly contrasted with the optimised version of the mini-app, which when comparing 32 PPC to 512 PPC takes only $9.91\times$

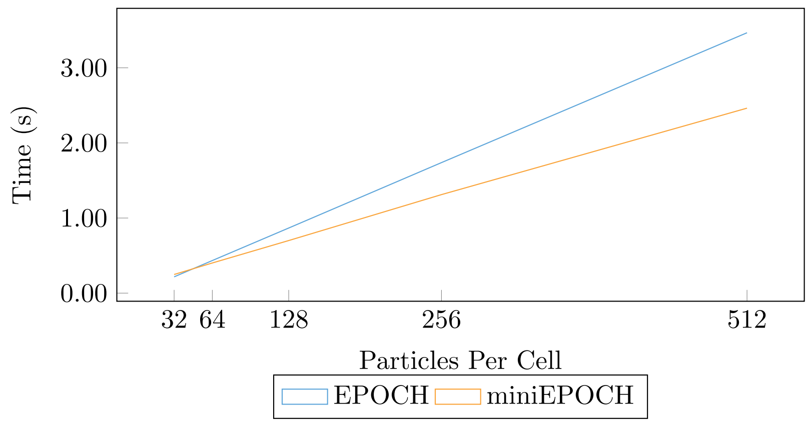

Figure 14: The PPC time scaling of the original EPOCH code base and the optimised mini-app.

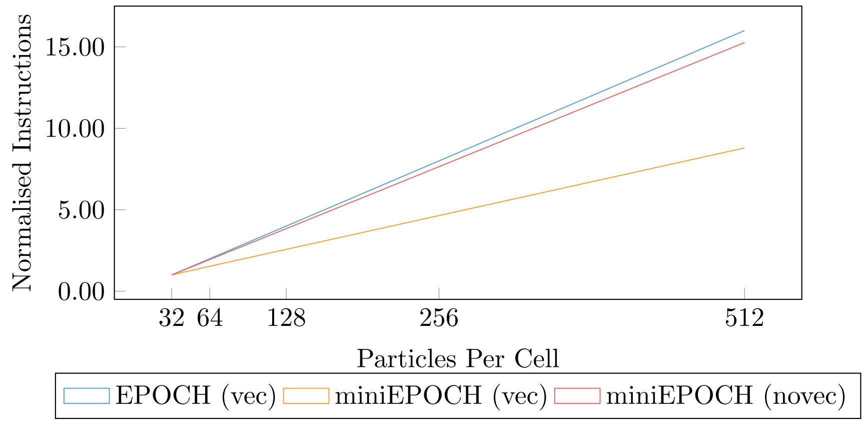

Figure 15: The PPC instruction count scaling of the original EPOCH code base, and the optimised mini-app.

as long to do a 16-fold increase in work. This scaling represents efficiency gains, with the runtime only increasing between 1.61× and 1.87× compared to the expected 2×, with an average of 1.78× over the sampled data. This super-linear scaling indicates that increased workloads are not only possible, but in fact favourable.

The root cause of these improvements is identified in Figure 15, which analyses the executed instruction counts for both EPOCH and miniEPOCH. We can see that with vectorisation disabled for miniEPOCH, the instruction throughput of both codes is very similar. However with vectorisation of miniEPOCH enabled, we can see that as the number of particles per cell increases, the vectorisation efficiency increases, leading to super-linear scaling in the number of instructions executed relative to the amount of work to be done.

5 Conclusions and Future Work

Despite recent successes at the large laser-based inertial confinement fusion device at the National Ignition Facility at LLNL, we remain some distance from being able to create controlled, self-sustaining fusion reactions. Inertial confinement fusion (ICF) represents one leading design for the generation of energy by nuclear fusion and computing simulations supporting ICF continue on some of the world's most powerful supercomputers. The research presented here focuses on *EPOCH*, a fully relativistic particle-in-cell plasma physics code. A significant challenge in developing large codes like EPOCH is maintaining effective scientific delivery on successive generations of high-performance computing architecture. To support this process, we adopt the use of *mini-applications* – small code proxies that encapsulate important computational properties of their larger parent counterparts.

Through the development of *miniEPOCH*, we investigate known time-step scaling issues within EPOCH and explore possible optimisations. In particular, this work:

- Presents the development of a new mini-app (*miniEPOCH*) for the EPOCH code. This mini-app is then used to explore known performance problems with EPOCH, and in particular (i) the increasing time-step duration during simulation runtime and, (ii) high levels of cache miss rates due to particle-store fragmentation;
- Explores opportunities for code optimisation in the context of the mini-app. These optimisations including the utilisation of shared memory, exploiting increasing vector width and improving memory locality.
- Investigates the use of a specialised kernel which is able to exploit domain specific knowledge and the spatial relationship between particles within a timestep in order to reduce the computational and memory requirements of the calculation. This enables the reuse of intermediate calculations, and ultimately allows for super-linear PPC scaling.

- Validates these findings on ARCHER, a 1.6 PFlop/s Cray XC30, housed at the UK national supercomputing centre at EPCC. These improvements demonstrate a 2.02× improvement to the core EPOCH algorithm and a 1.55× speed-up to the overall application runtime.

In future work we hope to optimise the OpenMP 4 variant of EPOCH, with the aim of delivering cross platform, portable performance, for future heterogeneous platforms. As OpenMP 4 compilers targeting accelerator platforms become available, we hope a finely tuned, single-source, version of EPOCH which uses the described optimisations will be able to successfully target both Intel Xeon Phi and NVIDIA GPU products. As part of this work an Intel Xeon Phi version of EPOCH has been developed, and this will form the basis of our efforts for such work.

Acknowledgments

This research is supported in part by the EPSRC grant “A Radiation Hydrodynamic ALE Code for Laser Fusion Energy” (EP/I029117/1), by The Royal Society, through their Industry Fellowship Scheme (IF090020/AM) and by AWE through the Warwick-hosted Centre for Computational Plasma Physics. Use of ARCHER was supported by several Resource Allocation Panel (RAP) awards, including “High-level Abstractions for Performance, Portability and Continuity of Scientific Software on Future Computing Systems”. Professor Stephen A. Jarvis is an AWE William Penney Fellow.

References

[1] UK Mini-App Consortium (UKMAC). `http://uk-mac.github.io`.

[2] T. D. Arber, K. Bennett, C. S. Brady, M. G. Ramsaym, N. J. Sircombe, P. Gillies, R. G. Evans, H. Schmitz, A. R. Benn, and C. P. Ridgers. Contemporary particle-in-cell approach to laser-plasma modelling. *Plasma Physics and Controlled Fusion*, 2015.

[3] D. H. Bailey, E. Barszcz, J. T. Barton, D. S. Browning, R. L. Carter, L. Dagum, R. A. Fatoohi, P. O. Frederickson, T. A. Lasinski, R. S. Schreiber, et al. The nas parallel benchmarks. *International Journal of High Performance Computing Applications*, 5(3):63–73, 1991.

[4] D. W. Barnette, S. D. Hammond, J. H. Laros, and J. Jayaraj. *Using Miniapplications in a Mantevo Framework for Optimizing Sandia's SPARC CFD Code on Multi-Core Many-Core and GPU-Accelerated Compute Platforms.* Dec 2012.

[5] R. Bird, P. Gillies, M. Bareford, J. Herdman, and S. Jarvis. Mini-app driven optimisation of inertial confinement fusion codes. In *Cluster Com-*

puting (CLUSTER), 2015 IEEE International Conference on, pages 768–776. IEEE, 2015.

[6] R. F. Bird, S. J. Pennycook, S. A. Wright, and S. A. Jarvis. Towards a Portable and Future-proof Particle-in-Cell Plasma Physics Code. *International Workshop on OpenCL*, 2014.

[7] C. K. Birdsall and A. B. Langdon. *Plasma physics via computer simulation.* CRC Press, 2014.

[8] K. J. Bowers, B. Albright, L. Yin, B. Bergen, and T. Kwan. Ultrahigh performance three-dimensional electromagnetic relativistic kinetic plasma simulation. *Physics of Plasmas (1994-present)*, 15(5):055703, 2008.

[9] H. Burau et al. PIConGPU: A Fully Relativistic Particle-in-Cell Code for a GPU Cluster. *IEEE Transactions on Plasma Science*, 38(10):2831–2839, 2010.

[10] P. M. Campbell, E. A. Carmona, and D. W. Walker. Hierarchical domain decomposition with unitary load balancing for electromagnetic particle-in-cell codes. In *Distributed Memory Computing Conference, 1990., Proceedings of the Fifth*, volume 2, pages 943–950. IEEE, 1990.

[11] G. Chen, L. Chacón, and D. C. Barnes. An efficient mixed-precision, hybrid CPU-GPU implementation of a nonlinearly implicit one-dimensional particle-in-cell algorithm. *Journal of Computational Physics*, 231(16):5374–5388, 2012.

[12] M. A. Heroux, D. W. Doerfler, P. S. Crozier, J. M. Willenbring, H. C. Edwards, A. Williams, M. Rajan, E. R. Keiter, H. K. Thornquist, and R. W. Numrich. Improving Performance via Mini-applications. *Sandia National Laboratories, Techical Report SAND2009-5574*, 2009.

[13] L. V. Kale and G. Zheng. Charm++ and ampi: Adaptive runtime strategies via migratable objects. *Advanced Computational Infrastructures for Parallel and Distributed Applications*, pages 265–282, 2009.

[14] I. Karlin, A. Bhatele, B. L. Chamberlain, J. Cohen, Z. Devito, M. Gokhale, R. Haque, R. Hornung, J. Keasler, D. Laney, et al. LULESH Programming Model and Performance Ports Overview. *Lawrence Livermore National Laboratory (LLNL), Livermore, CA, Tech. Rep*, 2012.

[15] I. Karlin, A. Bhatele, J. Keasler, B. L. Chamberlain, J. Cohen, Z. Devito, R. Haque, D. Laney, E. Luke, F. Wang, D. Richards, M. Schulz, and C. H. Still. Exploring traditional and emerging parallel programming models using a proxy application. In *Proceedings of the 2013 IEEE 27th International Symposium on Parallel and Distributed Processing*, IPDPS '13, pages 919–932, Washington, DC, USA, 2013. IEEE Computer Society.

[16] I. Karlin, J. McGraw, J. Keasler, and B. Still. Tuning the LULESH Mini-app for Current and Future Hardware. In *Nuclear Explosive Code Development Conference Proceedings (NECDC12)*, 2012.

[17] P.-F. Lavallée, G. C. de Verdière, P. Wautelet, D. Lecas, and J.-M. Dupays. Porting and optimizing HYDRO to new platforms and programming paradigms–lessons learnt.

[18] Z. Lin, T. S. Hahm, W. Lee, W. M. Tang, and R. B. White. Turbulent Transport Reduction by Zonal Flows: Massively Parallel Simulations. *Science*, 281(5384):1835–1837, 1998.

[19] H. Nakashima, Y. Miyake, H. Usui, and Y. Omura. OhHelp: A Scalable Domain-Decomposing Dynamic Load Balancing for Particle-in-Cell Simulations. In *Proceedings of the 23rd International Conference on Supercomputing*, pages 90–99. ACM, 2009.

[20] A. Nichols. Users manual for ALE3D: An arbitrary Lagrange/Eulerian 3D code system. *Lawrence Livermore National Laboratory*, 2007.

[21] J. Payne, D. Knoll, A. McPherson, W. Taitano, L. Chacon, G. Chen, and S. Pakin. Design and development of a multi-architecture, fully implicit, charge and energy conserving particle-in-cell framework. *Bulletin of the American Physical Society*, 58, 2013.

[22] S. J. Plimpton, D. B. Seidel, M. F. Pasik, R. S. Coats, and G. R. Montry. A load-balancing algorithm for a parallel electromagnetic particle-in-cell code. *Computer physics communications*, 152(3):227–241, 2003.

[23] H. C. Problem. Technical Report LLNL-TR-490254. *Lawrence Livermore National Laboratory.*

[24] F. Rossi, P. Londrillo, A. Sgattoni, S. Sinigardi, and G. Turchetti. Robust Algorithms for Current Deposition and Dynamic Load-balancing in a GPU Particle-in-Cell Code. In *ADVANCED ACCELERATOR CONCEPTS: 15th Advanced Accelerator Concepts Workshop*, volume 1507, pages 184–192. AIP Publishing, 2012.

[25] Ruhl, H. Classical Particle Simulations with the PSC Code. `http://www.physik.uni-muenchen.de/lehre/vorlesungen/wise_09_10/tvi_mas_compphys/vorlesung/Lecturescript.pdf`.

[26] R. Teyssier. Cosmological Hydrodynamics with Adaptive Mesh Refinement: a new high resolution code called RAMSES. *Astronomy & Astrophysics*, 385(1):337–364, 2002.

[27] K. S. Yee et al. Numerical solution of initial boundary value problems involving Maxwells equations in isotropic media. *IEEE Trans. Antennas Propag*, 14(3):302–307, 1966.