Journal of Computational Physics 231 (2012) 795–821

Contents lists available at ScienceDirect

Journal of Computational Physics

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

Particle simulations of space weather

Giovanni Lapenta

Centrum voor Plasma-Astrofysica, Departement Wiskunde, Leuven Centre for Aero & Space Science, Technology, and Applications (LASA), Leuven Mathematical Modeling & Computational Science Research Centre (LMCC), Katholieke Universiteit Leuven, Celestijnenlaan 200B – bus 2400, B-3001 Heverlee, Belgium

ARTICLE INFO

Article history:
Available online 5 April 2011

Keywords:
Space weather
Particle-in-cell
Adaptive
Implicit

ABSTRACT

We review the application of particle simulation techniques to the full kinetic study of space weather events. We focus especially on the methods designed to overcome the difficulties created by the tremendous range of time and space scales present in the physical systems. We review the aspects of the derivation of the particle in cell (PIC) method relevant to the discussion. We consider first the explicit formulation highlighting its severe limitations due to the presence of stability constraints. Next we introduce implicit methods designed to remove such constraints. We describe both fully implicit methods based on the use of non-linear iteration solvers and semi-implicit methods based on the linearization of the coupling and on simpler linear solvers. We focus the discussion on the implicit moment method but remark its differences from the direct implicit method. The application of adaptive methods within PIC is discussed. Finally practical considerations about the implementation of the implicit PIC method on massively parallel computers to conduct studies of space weather events are given.

© 2011 Elsevier Inc. All rights reserved.

1. Introduction

According to the definition used by the US National Space Weather Strategic Plan [1], space weather refers to conditions on the sun and in the solar wind, magnetosphere, ionosphere, and thermosphere that can influence the performance and reliability of space-borne and ground-based technological systems and can endanger human life or health. Adverse conditions in the space environment can cause disruption of satellite operations, communications, navigation, and electric power distribution grids, leading to a variety of socioeconomic losses. The topic of space weather is becoming ever more important for its societal impact and great national and international efforts are being devoted to it [2].

Simulating space weather means literally simulating the magnetized plasma that permeates the whole solar system. A daunting task. Daunting because plasma physics includes a wide variety of space and temporal scales. A plasma is made of a collection of ionised atoms (nuclei), and electrons interacting via electric and magnetic fields. Near the Sun gravity plays a role too. At the core of the space weather simulation issue stands the vast gap between the electron and ion scales. An electron is 1836 times lighter than its nucleus in an hydrogen atom. This simple physical fact creates a great challenge in modeling the small electron scales embedded into solar system sized problems. For example, Fig. 1 shows the typical electron and ion scales observed in satellite crossing of the Earth magnetosphere (the magnetotail to be specific) [3]. The scales are presented in the form of a hourglass with the top part presenting the macroscopic scales of evolution of the Earth magnetotail that during quiescent times can be 10,000 km thick but at interesting times becomes as thick as the ion inertial length, d_i

E-mail address: giovanni.lapenta@wis.kuleuven.be
URLs: http://perswww.kuleuven.ac.be/~u0052182/

0021-9991/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jcp.2011.03.035

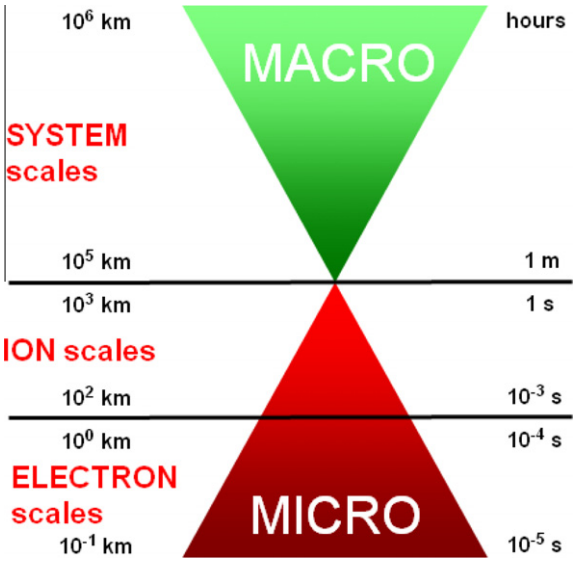

Fig. 1. Visual representation of the typical scales observed in the Earth environment (conditions common in the magnetotail).

or 1000 km. The ion scales are smaller and the electron scales much smaller, down to 100 m corresponding to typical electron Debye lengths. Similarly in time, the overall evolution of space weather events requires simulating hours, if not days, but the ion responses are on the scale of seconds to milliseconds and the electrons respond on scales of multiples of microseconds.

Simulating days of evolution of the plasma in the solar system with a time step of microseconds and with a resolution of 100 m is simply not possible and it will not be possible for a long time to come. The present paper reviews the efforts to overcome this problem and allow a simulation of space weather events at the particle level, retaining the full kinetic and electromagnetic treatment of the plasma.

Traditionally, the impossibility to resolve all scales has led to the use of fluid (MHD or more advanced models such as Hall-MHD and two-fluid) or hybrid methods where the electrons are described as a fluid while the ions are kinetic [4]. However, such reduced models loose the ability to capture the dissipation processes at the electron scales. Important examples for space weather are the regions of large energy releases such as shocks and reconnection regions where small scale processes at the ion and electron particle scale are key in determining the production of high energy particles (such as the solar energetic particles in the cosmic rays [2]) and in allowing macroscopic changes in the magnetic field configuration (e.g. magnetic reconnection during solar flares and coronal mass ejections [5] or in the Earth environment during magnetic storms and substorms [3]).

The present review starts in Section 2 with a brief overview of events involved in space weather and on the role that can be played by particle in cell kinetic simulation. We focus the attention then on full kinetic models based on the particle in cell (PIC) simulation technique. Section 3 presents an original (to the author's knowledge) derivation of the mathematical foundations of the PIC method, in a manner that helps identify the strengths and weaknesses for the application to space weather. Section 4 focuses on the explicit time discretisation of PIC methods and highlights the severe limitations of this approach when applied to space weather. Sections 5 and 6 provide additional information about the technical aspects of applying the explicit technique and developing codes based on it.

Section 7 introduces a different approach, that of implicit time differencing in PIC methods, an approach particularly suitable to space weather applications. The approach is not new but it is reviewed here especially in light of its more recent space weather applications. For this reason, we review in Section 6 the general idea of implicit PIC and provide in the subsequent sections its two possible implementations: fully implicit, requiring a solution of a set of non-linearly coupled equations and semi-implicit where the coupling is linearised and made more manageable. Section 8 gives an introduction of the recent developments in fully implicit PIC. This area is in its infancy with just a few initial works done but it is perhaps the most promising area of research in this field and it has been included despite the lack of a strong previous literature. Section 9 reviews the much more mature field of semi-implicit PIC methods.

Section 10 reviews effort to improve the spatial discretisation introducing methods that allow the use of adaptive grids.

Section 11 concludes the discussion with practical considerations about the implementation of particle methods in massively parallel computers and provides a discussion of the gains to be obtained by the use of implicit methods.

2. Goals of space weather simulation

A recent European initiative, defined as COST action 724, has considered the field of study of space weather and has recognised that "Space weather is the physical and phenomenological state of natural space environments" and that "the

associated discipline aims, through observation, monitoring, analysis and modelling, at understanding and predicting the state of the Sun, the interplanetary and planetary environments, and the solar and nonsolar driven perturbations that affect them". It is of primary importance to "forecast and nowcast the possible impacts on biological and technological systems". All quotations are from the concluding document available online [6].

The discipline of space weather is undergoing a great expansion. All countries active in space have a robust and growing space weather initiative. Within space weather, the modelling and forecasting efforts are increasingly becoming central. The present review focuses more on the research efforts than on the operational approached, even though the research efforts are continually transitioning to operational tools.

A brief overview of space weather is the best starting point to give the context to the modelling methods reviewed here. The main actors are the Sun with its magnetized solar wind continually emitted but highly variable and structured, the interplanetary medium where the solar ejecta interact and evolve and the Earth with its magnetic field and its plasma environment, originating in part from the Earth's atmosphere and in part from captured solar wind and cosmic rays particles. An excellent introductory textbook describing the solar systems plasmas and their evolution is Ref. [3], but many others present also the subject.

Modelling such enormous, complex and coupled system is a task requiring a whole community of researchers. For this reason, the Community Coordinated Modeling Center (CCMC) was created as a multi-agency partnership [7]. We draw here some paradigmatic illustrations from this community effort. The CCMC collects from the community of code developers models driven by actual observational data and capable of modelling realistic or idealised space weather events. The CCMC make the models available to everybody, not for download but to actually run the simulations and to visualise the results on the web site itself. The CCMC provides computing resources and makes their use completely transparent. Each user, even with no background in computing, can request simulations via a simple web interface. This approach makes CCMC a tool for research and for teaching available to all interested parties. As an example, all students of the space weather class at the Katholieke Universiteit Leuven use this tool to model one space weather event. The practical experience of three academic years shows that no student, even with the most rudimentary expertise on computers, finds insurmountable problems and all succeed in producing interesting and valuable space weather simulations.

The highly active space weather is characterized by singularly memorable events that occasionally acquire specific names such as the famous period of magnetic storms at the end of October 2003 referred to as the "Halloween storm" or the events that took place on the 211st anniversary of the French revolution (July 14, 2000) and referred to as the "Bastille Day Event". But besides the major events that acquire fame and name, many other types of smaller events must also be studied and have important impacts. Recently, a European effort led by the author of this review, SOTERIA [8], has investigated a list of a variety of events covering the different levels of severeness, from minor to major [9]. Below we illustrate the current status of space weather modelling using events from the SOTERIA list and using tools available to all interested readers at the CCMC. The interested reader can find much more information, more pictures and movies by following the links provided below.

We start our overview with a view of the solar system at the moment of arrival of the SOTERIA Event 3 (corresponding to the Halloween storm in October 2003), in Fig. 2. A large magnetised plasma cloud, called a coronal mass ejection (CME) is emitted from the Sun and travels outward, in this case reaching the Earth. Such processes and the related process of flaring are common events happening frequently, on a cycle of approximately 11 year ranging from more active times (solar maximum) to relatively quiet times (solar minimum). As this review is being written, we are starting a phase of increasing solar activity following a very unusually long and deep minimum that has puzzled the experts for its unusually long duration. In recorded history, there have been extended periods of quietness, the longest being the Maunder minimum, spanning from 1645 to 1715, a period of unusually cold winters in Europe, reminding us of the link between space weather and climate. But history records also unusually strong storms, such as the Carrington event of September 1859. At the time the event impacted the then latest modern marvel, the telegraph lines, creating serious consequences. There is no telling of the havoc it would create for our economy, very reliant on much more sensitive infrastructures such as power lines and satellite networks.

As magnetic disturbances reach the Earth environment, significant perturbations can severely disrupt space operations and even impact the ground. Solar energetic particles produced in conjunction with flares and CMEs can be a serious radiation hazard for humans and technology in space and even on regular airliners. Induced currents in the ionosphere can induce strong currents on the ground and especially on long conductors such as power grids, pipelines and communication networks. The Earth magnetosphere undergoes a restructuring where magnetic energy and plasma is transferred from the dayside facing the Sun to the nightside at its opposite end. Fig. 3 illustrates the conditions of the magnetosphere during the SOTERIA Event 3 (a major event during the latter part of July 2004) from a simulation conducted at the CCMC. The magnetised plasma of the solar wind interacts with the Earth magnetic field at the nose (right end) and the magnetic field lines of the two regions break and reconnect leading to the transfer of flux towards the magnetic tail (to the left). The process leads to an intensification of the tail current and its destabilisation with effects that propagate Earthward along the tail. These effects produce strong current near the Earth and particle precipitation leading to strong auroral activity.

The ultimate impact on the ionosphere is shown in Fig. 4 from the same simulation that self-consistently treats the coupled ionosphere-magnetosphere system driven by the arriving solar wind. The solar wind input is derived directly from the ACE satellite located at the L1 Lagrangian point at 1.5 million km towards the Sun.

In all these series of events there are key instances of localised processes strongly coupled with the overall evolution. Examples are the regions of reconnection, shocks and particle acceleration connected with the coronal mass ejections at

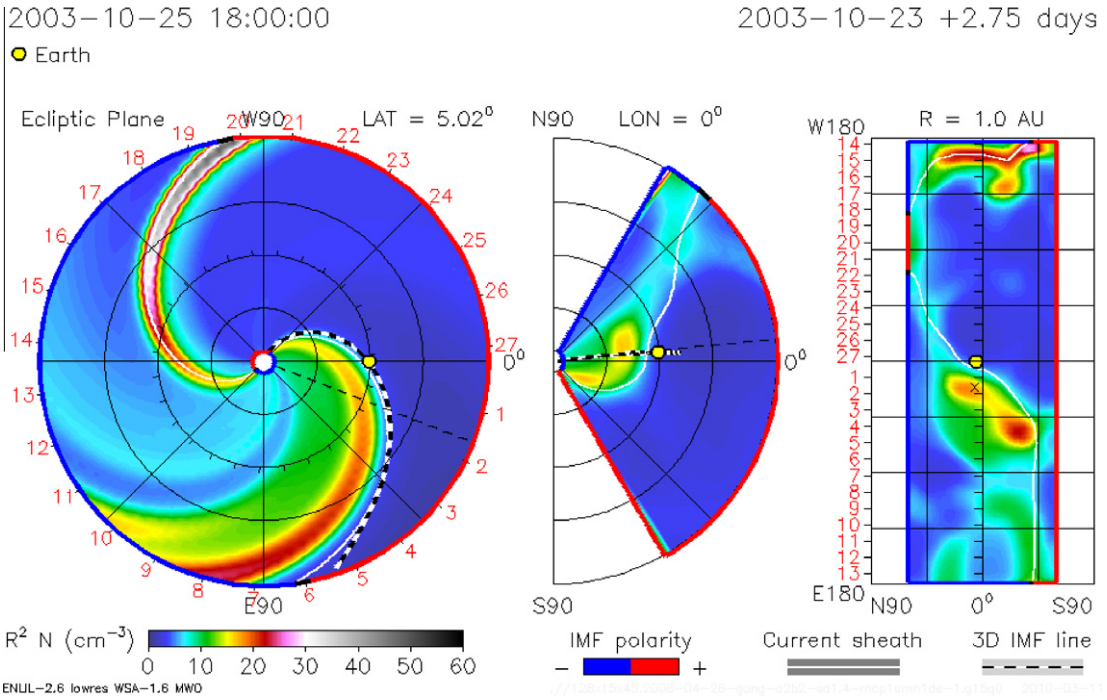

Fig. 2. SOTERIA Event 3. Results from the CCMC run Giovanni_Lapenta_031110_SH_1 done with ENLIL version 2.6 with cone model [10]. The left figure shows a top view of the plasma density in the solar system, with the Earth shown as a yellow circle and the sun in the center. The polarity of the magnetic field is shown in red/blue. The center picture shows the same information on a vertical plane at longitude 0 (defined as the longitude where the Earth is). The right figure shows a longitude–latitude plot at the radial distance of the Earth. The rotation of the Sun about its axis leads to a upward shift of the features shown in the figure. The movie for this run can be obtained directly on the CCMC web site using the name of the run. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

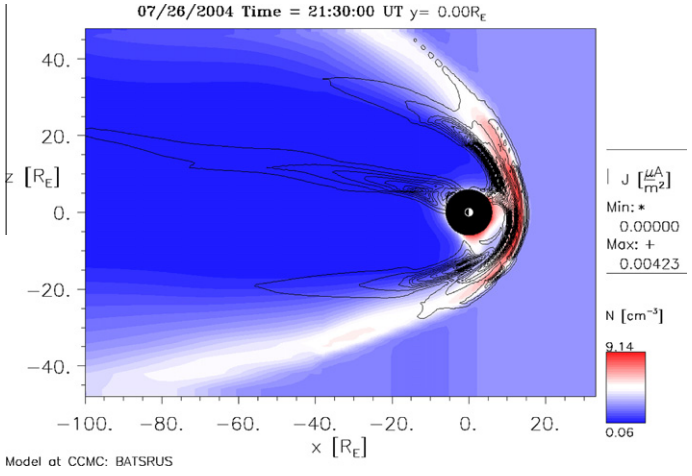

Fig. 3. SOTERIA Event 1. Conditions of the Earth magnetosphere. Results from the CCMC run Thea_Falkenberg_013009_1 done with the coupled BATS'R'US/RCM model version 8.01 [11]. The solar wind arrives from the Sun at the right. The density is shown in false colors and the current is represented with the contours of equal magnitude.

the Sun, the multiple strongly localised features of the Earth magnetosphere: the bow shock, the magnetopause and the magnetotail current sheet. The ionosphere and the auroral regions are also host to numerous localised small scales processes. For all these microscopic processes fluid magnetohydrodynamic (MHD) models are inadequate and utlimatly only the full kinetic treatment can provide truly predictive results. That does not mean that reduced limited models such as hybrid (fluid electrons and kinetic ions), two fluid and Hall MHD cannot play a very important role. But for the dissipation processes at shocks, in reconnection regions and in other wave-particle interaction processes, the kinetic model is needed.

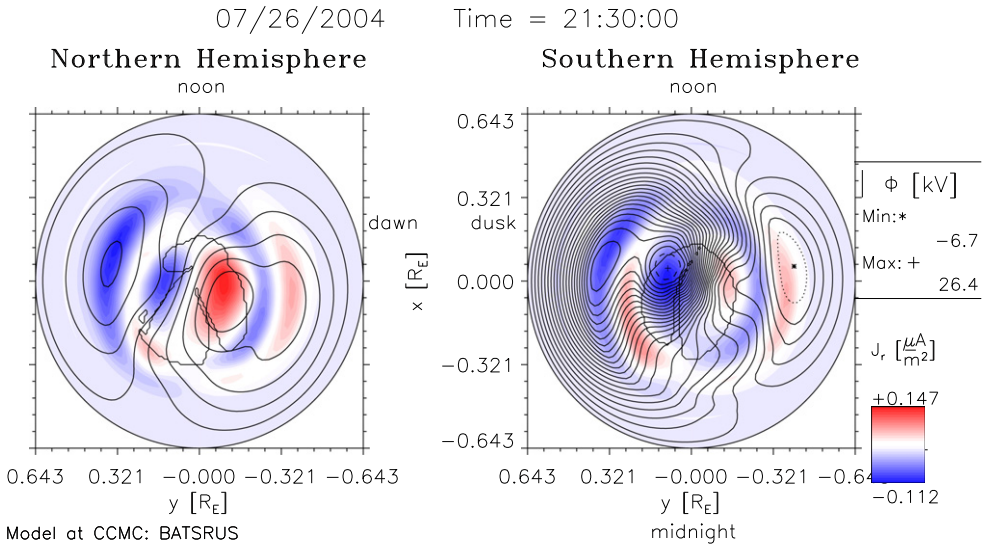

Fig. 4. SOTERIA Event 1. Conditions of the Earth ionosphere. Results from the CCMC run Thea_Falkenberg_013009_1 done with the coupled BATS'R'US/RCM model version 8.01 [11]. The northern emisphere is shown to the left and the southern emisphere to the right. In both figures the radial currents are shown in color with the red regions corresponding to positive and the blue regions to negative currents. The electric potential is represented with isopotential lines. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

An illustration of such an approach is provided in Fig. 5. A significant part of the magnetotail is simulated using the iPIC3D code [12]. A box with sides equalling about 6 Earth radii (R_E) in the Earth–Sun direction (x), $4R_E$ in the north–south direction (y) and $3R_E$ in the east–west direction (z) is modelled. As can be seen this is a relatively small region compared with the large scale MHD simulations reported in Fig. 3. But there the presence of kinetic effects is key: Fig. 5 shows to the expert the presence of many effects not present in the MHD simulations. A complete description of the evolution requires a deep understanding of the physics of reconnection [13,14] and is beyond the scope of the present work. So we limit the discussion to a generalistic overview that gives a flavour of the ideas, providing references for more in depth treatment.

Starting from the false color cut displaying the electric field amplitude perpendicular to the magnetic field, we observe the presence of the separatrices, two (top left and bottom right) being more prominent and associated with strong instabilities. These are the channels where the electrons flow in toward the center of the reconnection region [15]. The field lines undergoing reconnection are shown as flux tubes colored in yellow if going in the positive x direction and in violet for

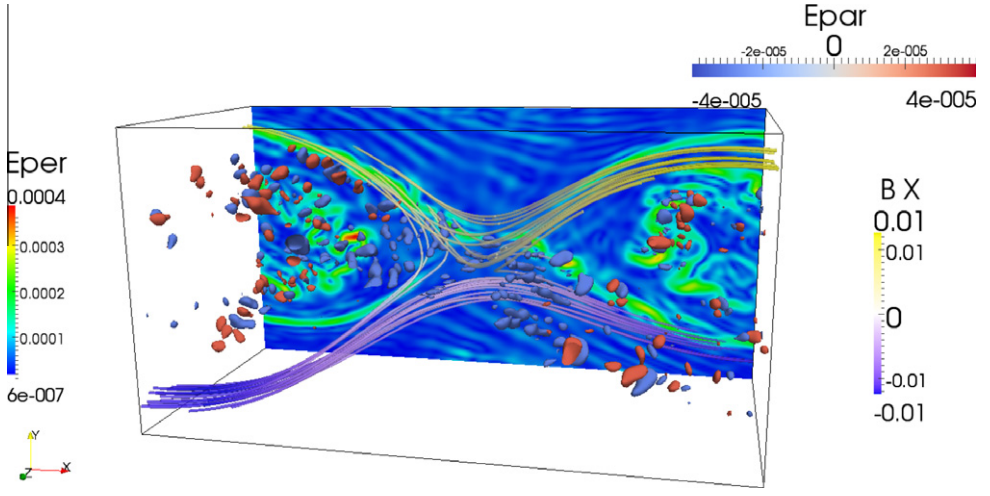

Fig. 5. Magnetotail reconnection simulation with iPIC3D [12]. Three quantities are shown at the same time with different color scales. The electric field parallel to the magnetic field is shown as isosurfaces of its absolute value colored by the signed value, positive (in red) or negative (in blue). The perpendicular component of the electric field is shown in the blue to green to red scale in the form of a cross section plot in false colors. The magnetic field is presented with selected field lines colored by the horizontal (along x) component with a the violet to yellow color scale. Note that only a reduced vertical section of the computational box of about $1.5R_E$ is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the opposite. In the center the field lines come together, break and reconnect. Along the two more active separatrices, also the parallel electric field is most active, presenting distinctive regions of opposite polarity in close proximity. These bipolar structures are well know and observed features of the reconnection process and are due to streaming instabilities forming so-called electron holes [16–18,15]. Notable bipolar structures are also present in the two regions inside the field lines and enclosed around them. These regions are also well observed and correspond to the dipolarisation fronts, fronts where the magnetic field acquires a more dipole-like structure with a important vertical component of the field [19]. Unlike MHD simulations, these regions of field dipolarisation become very unstable and both the parallel and perpendicular electric field are highly fragmented. All these effects are of great importance to obtain quantitative precision regarding the speed and consequences of the processes happening in the Earth magnetosphere [16].

We report below a kinetic approach that has the potential, as proved in Fig. 5, to be directly applicable to space weather: the particle in cell method.

3. Derivation of the particle method

As a key step in the study of the more advanced applications of PIC methods, we lay first the mathematical groundwork for the formulation of the particle in cell method starting from the mathematical continuous problem, the Vlasov–Maxwell system.

In most processes shown in the preceding section, the plasma temperatures and the particle speeds remain well below relativistic regimes. However, the currents are very strong and crucial to determine the evolution. A fully electromagnetic approach is preferable for the fields (although for selected applications reduced models such as the Darwin approximation are of limited use) but in most cases the particles do not need to be treated relativistically. A notable exception is that of the most energetic particles produced at shocks and in other space events. Those particles can travel in just 8 min the distance from the Sun to the Earth and are important to space weather application. For these reasons, the assumption is made in the derivations that classical physics can be used for the particle motion, but the extension to special relativity is in most cases straightforward and specific references are mentioned where appropriate.

The classic textbooks [20,21] report a heuristic derivation based on the physical properties of a plasma. Instead, we consider here a different approach aimed at making a clear mathematical link between the continuous Vlasov–Maxwell description of the plasma and its numerical solution.

The phase space distribution function $f_s(\mathbf{x},\mathbf{v},t)$ for a given species s (electrons or ions), defined as the number density per unit element of the phase space (or the probability of finding a particle in a $d\mathbf{x}$ and $d\mathbf{v}$ around a certain phase space point $(\mathbf{x},\mathbf{v})$), is governed by the Vlasov equation:

$$\frac{\partial f_s}{\partial t}+\mathbf{v}\cdot\frac{\partial f_s}{\partial \mathbf{x}}+\frac{q_s}{m_s}(\mathbf{E}+\mathbf{v}\times\mathbf{B})\cdot\frac{\partial f_s}{\partial \mathbf{v}}=0 \tag{1}$$

where q_s and m_s are the charge and mass of the species, respectively. The derivatives are written in the standard vector notation in the 3-dimensional configuration space $\mathbf{x}$ and 3-dimensional velocity space $\mathbf{v}$, forming together the classical phase space.

The electric and magnetic fields are given by the two curl Maxwell's equations,

$$\begin{aligned}\nabla\times\mathbf{E}&=-\frac{\partial\mathbf{B}}{\partial t}\\ \nabla\times\mathbf{B}&=\mu_0\epsilon_0\frac{\partial\mathbf{E}}{\partial t}+\mu_0\mathbf{j}\end{aligned} \tag{2}$$

supplemented by the two divergence equations:

$$\begin{aligned}\epsilon_0\nabla\cdot\mathbf{E}&=\rho\\ \nabla\cdot\mathbf{B}&=0\end{aligned} \tag{3}$$

where the net charge density is computed from the distribution functions as:

$$\rho(\mathbf{x},t)=\sum_s q_s\int_{\mathbb{V}} f_s(\mathbf{x},\mathbf{v},t)d\mathbf{v} \tag{4}$$

and the current density as:

$$\mathbf{j}(\mathbf{x},t)=\sum_s q_s\int_{\mathbb{V}}\mathbf{v}f_s(\mathbf{x},\mathbf{v},t)d\mathbf{v} \tag{5}$$

where $\mathbb{V}$ is the velocity space.

3.1. Discretisation of the distribution function

The PIC method can be regarded as a discretisation of the phase space as a superposition of moving elements, each representing a cloud of physical particles. The mathematical formulation of the PIC method is obtained by assuming that the distribution function of each species is given by the superposition of several elements (called computational particles or superparticles):

$$f_s(\mathbf{x},\mathbf{v},t) = \sum_p f_p(\mathbf{x},\mathbf{v},t) \tag{6}$$

Each element represents a large number of physical particles that are near each other in phase space. For this reason, the choice of the elements is made in order to be at the same time physically meaningful (i.e. to represent a bunch of particles near each other) and mathematically convenient (i.e. to allow the derivation of a manageable set of equations).

The PIC method is based upon assigning to each computational particle a specific functional form for its distribution, a functional form with a number of free parameters whose time evolution will determine the numerical solution of the Vlasov equation. The choice is usually made to have two free parameters in the functional shape for each spatial dimension. The free parameters will acquire the physical meaning of position and velocity of the computational particle. The functional dependence is further assumed to be the tensor product of the shape in each direction of the phase space:

$$f_p(\mathbf{x},\mathbf{v},t) = N_p S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p(t)) S_{\mathbf{v}}(\mathbf{v}-\mathbf{v}_p(t)) \tag{7}$$

where $S_{\mathbf{x}}$ and $S_{\mathbf{v}}$ are the *shape functions* for the computational particles and N_p is the number of physical particles that are present in the element of phase space represented by the computational particle.

A number of properties of the shape functions come from their definition:

1. The support of the shape functions is compact, to describe a small portion of phase space, (i.e. it is zero outside a small range).
2. Their integral over the domain is unitary:

$$\int_{V_\xi} S_\xi(\xi-\xi_p)d\xi = 1 \tag{8}$$

   where ξ stands for any coordinate or any velocity direction.
3. While not strictly necessary, Occam's razor suggests to choose symmetric shapes:

$$S_\xi(\xi-\xi_p) = S_\xi(\xi_p-\xi) \tag{9}$$

While these definitions still leave very broad freedom in choosing the shape functions, traditionally the choices actually used in practice are very few.

3.2. Selection of the particle shapes

The standard PIC method is essentially determined by the choice of S_v, the shape in the velocity direction, as a Dirac's delta in each direction:

$$S_{\mathbf{v}}(\mathbf{v}-\mathbf{v}_p) = \delta(v_x - v_{xp})\delta(v_y - v_{yp})\delta(v_z - v_{zp}) \tag{10}$$

This choice has the fundamental advantage that if all particles within the element of phase space described by one computational particle have the same speed, they remain closer in phase space during the subsequent evolution.

The original PIC methods developed in the 50s were based on using a Dirac's delta also as the shape function in space. But now for the spatial shape functions, all commonly used PIC methods are based on the use of the so-called b-splines. The b-spline functions are a series of consecutively higher order functions obtained from each other by integration. The first b-spline is the flat-top function $b_0(\xi)$ defined as:

$$b_0(\xi) = \begin{cases} 1 & \text{if } |\xi| < 1/2 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

The subsequent b-splines, b_l, are obtained by successive integration via the following generating formula:

$$b_l(\xi) = \int_{-\infty}^{\infty} d\xi' b_0(\xi-\xi') b_{l-1}(\xi') \tag{12}$$

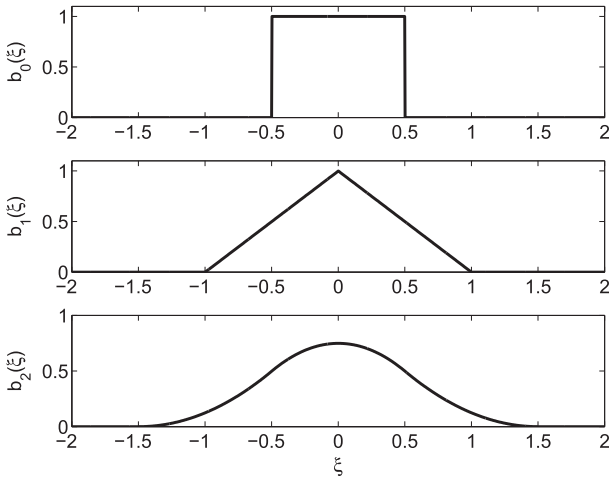

Fig. 6. First three b-spline functions.

Fig. 6 shows the first three b-splines.

Based on the b-splines, the spatial shape function of PIC methods is chosen as:

$$S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p) = \frac{1}{\Delta x_p \Delta y_p \Delta z_p} b_l\left(\frac{x-x_p}{\Delta x_p}\right) b_l\left(\frac{y-y_p}{\Delta y_p}\right) b_l\left(\frac{z-z_p}{\Delta z_p}\right) \tag{13}$$

where Δx_p, Δy_p and Δz_p are the lengths of the support of the computational particles (i.e. its size) in each spatial dimension. A few PIC codes use splines of order 1 but the vast majority uses b-splines of order 0, a choice referred to as cloud in cell because the particle is a uniform square cloud in space with infinitesimal span in the velocity directions.

3.3. Derivation of the equations of motion

To derive the evolution equations for the free parameters $\mathbf{x}_p$ and $\mathbf{v}_p$, we require the first moments of the Vlasov equation to be exactly satisfied by the functional forms chosen for the elements. This procedure requires some explanation. First, the Vlasov equation is formally linear in f_s and the equation satisfied by each element is still the same Vlasov equation. The linear superposition of the elements gives the total distribution function and if each element satisfies the Vlasov equation, the superposition does too. A caveat, the fields really depends on f_s making the Vlasov equation non-linear. As a consequence the fields used in each Vlasov equation for each element must be the total fields due to all elements, the same entering the complete Vlasov equation for f_s:

$$\frac{\partial f_p}{\partial t} + \mathbf{v}\cdot\frac{\partial f_p}{\partial \mathbf{x}} + \frac{q_s}{m_s}(\mathbf{E}+\mathbf{v}\times\mathbf{B})\cdot\frac{\partial f_p}{\partial \mathbf{v}} = 0 \tag{14}$$

Second, the arbitrary functional form chosen for the elements does not satisfy exactly the Vlasov equation. In analogy with the usual procedure for the finite element method, we require that the moments (rather than the local equation in each point of the domain) of the equation are satisfied. The moments of the Vlasov equation are obtained multiplying by powers of $\mathbf{x}$ and $\mathbf{v}$ and integrating. We indicate the integration over the spatial and velocity domain by the symbol $\langle \ldots \rangle \equiv \int d\mathbf{x} \int d\mathbf{v}$.

3.3.1. Moment 0

The zeroth order moment ($\langle Vlasov \rangle$) gives:

$$\frac{\partial \langle f_p \rangle}{\partial t} + \left\langle \mathbf{v}\cdot\frac{\partial f_p}{\partial \mathbf{x}} \right\rangle + \left\langle \frac{q_s}{m_s}(\mathbf{E}+\mathbf{v}\times\mathbf{B})\cdot\frac{\partial f_p}{\partial \mathbf{v}} \right\rangle = 0 \tag{15}$$

where we used the interchangeability of the integration in $d\mathbf{x}d\mathbf{v}$ and of the derivation over time. The second and third term are zero because f_p has finite support and vanishes at infinity in the coordinate space and in velocity. Note that the third term vanishes only because all terms of $\mathbf{v}\times\mathbf{B}$ do not include the same component of $\mathbf{v}$ as the partial derivative they multiply. Recalling that $\langle f_p \rangle = N_p$, it follows:

$$\frac{dN_p}{dt} = 0 \tag{16}$$

The application of the zeroth order moment leads to the establishment of the conservation of the number of physical particles per computational particle.

3.3.2. Three first order spatial moments

The application of the first order moment in $\mathbf{x}$, ($\langle \mathbf{x} \cdot Vlasov \rangle$), that implies the use of vector notation to treat simultaneously three equations for each spatial coordinate, gives:

$$\frac{\partial \langle f_p \mathbf{x} \rangle}{\partial t} + \left\langle \mathbf{x}\mathbf{v} \cdot \frac{\partial f_p}{\partial \mathbf{x}} \right\rangle + \left\langle \mathbf{x} \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \frac{\partial f_p}{\partial \mathbf{v}} \right\rangle = 0 \tag{17}$$

The last term is still zero by virtue of integration over $\mathbf{v}$, the other terms, instead, are new. The first term is:

$$\langle f_p \mathbf{x} \rangle = N_p \int S_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_p) d\mathbf{v} \int \mathbf{x} S(\mathbf{x} - \mathbf{x}_p) d\mathbf{x} \tag{18}$$

where the first integral is 1 by definition of $S_{\mathbf{v}}$ as a function of unitary integral and the second expresses the first order moment of $S_{\mathbf{x}}$. Recalling the assumption of symmetry of S_x, that moment equals $\mathbf{x}_p$:

$$\langle f_p \mathbf{x} \rangle = N_p \mathbf{x}_p \tag{19}$$

The middle term of Eq. (17) requires the integration of:

$$\iint \mathbf{x} \left(v_x \frac{\partial f_p}{\partial x} + v_y \frac{\partial f_p}{\partial y} + v_z \frac{\partial f_p}{\partial z} \right) d\mathbf{x}\, d\mathbf{v} \tag{20}$$

When integrated all terms vanish, except that with the derivative in the same direction as the moment. For example in the x-direction, only the first term remains,

$$\iint x v_x \frac{\partial f_p}{\partial x} d\mathbf{x}\, d\mathbf{v} = - \iint v_x f_p d\mathbf{x}\, d\mathbf{v} = -N_p \langle v_x \rangle \tag{21}$$

where the finite support of f_p, integration by part and the symmetry of $S_{\mathbf{v}}$ have been used. Using vector notation, the middle term of Eq. (17) leads to

$$\iint \mathbf{x} \left(v_x \frac{\partial f_p}{\partial x} + v_y \frac{\partial f_p}{\partial y} + v_z \frac{\partial f_p}{\partial z} \right) d\mathbf{x}\, d\mathbf{v} = -N_p \mathbf{v}_p \tag{22}$$

The end result of applying the first order moment in $\mathbf{x}$ is:

$$\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p \tag{23}$$

3.3.3. Three first order velocity moments

The application of the first order moment in $\mathbf{v}$, ($\langle \mathbf{v} \cdot Vlasov \rangle$) gives:

$$\frac{\partial \langle f_p \mathbf{v} \rangle}{\partial t} + \left\langle \mathbf{v}\mathbf{v} \frac{\partial f_p}{\partial \mathbf{x}} \right\rangle + \left\langle \mathbf{v} \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \frac{\partial f_p}{\partial \mathbf{v}} \right\rangle = 0 \tag{24}$$

The first term is computed in analogy to the first term in the spatial moments. The second term is still zero by virtue of integration over $\mathbf{x}$. The remaining term can be computed as above, recalling that all terms of $\mathbf{v} \times \mathbf{B}$ do not include the same component of $\mathbf{v}$ as the partial derivative they multiply:

$$\iint \mathbf{v} \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) \cdot \frac{\partial f_p}{\partial \mathbf{v}} d\mathbf{x}\, d\mathbf{v} = - \iint \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_p\, d\mathbf{x}\, d\mathbf{v} \tag{25}$$

using again integration by part and the finite support of the elements.

The resulting integral defines two new important quantities, the average electric field and magnetic field acting on a computational particle:

$$\iint \frac{q_s}{m_s} (\mathbf{E} + \mathbf{v} \times \mathbf{B}) f_p d\mathbf{x}\, d\mathbf{v} = -N_p \frac{q_s}{m_s} (\mathbf{E}_p + \mathbf{v}_p \times \mathbf{B}_p) \tag{26}$$

where the electric field on a computational particle is:

$$\mathbf{E}_p = \int S_{\mathbf{v}}(\mathbf{v} - \mathbf{v}_p) d\mathbf{v} \int S_{\mathbf{x}}(\mathbf{x} - \mathbf{x}_p) \mathbf{E}(\mathbf{x}) d\mathbf{x} \tag{27}$$

Recalling the property of $S_{\mathbf{v}}$, the formula for $\mathbf{E}_p$ simplifies to:

$$\mathbf{E}_p = \int S_{\mathbf{x}}(\mathbf{x} - \mathbf{x}_p) \mathbf{E}(\mathbf{x}) d\mathbf{x} \tag{28}$$

For the magnetic field,

$$\iint \frac{q_s}{m_s}\mathbf{v}\times\mathbf{B}f_p d\mathbf{x}\,d\mathbf{v} = \frac{q_s}{m_s}\int \mathbf{v}S_{\mathbf{v}}(\mathbf{v}-\mathbf{v}_p)d\mathbf{v}\times\int S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)\mathbf{B}(\mathbf{x})d\mathbf{x} \tag{29}$$

resulting in:

$$\mathbf{B}_p = \int S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)\mathbf{B}(\mathbf{x})d\mathbf{x} \tag{30}$$

The first order moment in $\mathbf{v}$ gives the final equation:

$$\frac{d\mathbf{v}_p}{dt} = \frac{q_s}{m_s}(\mathbf{E}_p + \mathbf{v}_p\times\mathbf{B}_p) \tag{31}$$

3.4. Equations of motion for the computational particles

The equations above give the following complete set of evolution equations for the parameters defining the functional dependence of the distribution within each element:

$$\begin{aligned}
&\frac{dN_p}{dt} = 0\\
&\frac{d\mathbf{x}_p}{dt} = \mathbf{v}_p\\
&\frac{d\mathbf{v}_p}{dt} = \frac{q_s}{m_s}(\mathbf{E}_p + \mathbf{v}_p\times\mathbf{B}_p)
\end{aligned} \tag{32}$$

It is a crucial advantage of the PIC method that its evolution equations resemble the same Newton equations as followed by the regular physical particles. The key difference is that the field is computed as average over the shape function based on the definition of $\mathbf{E}_p$ and $\mathbf{B}_p$.

Naturally, the electric field is itself given by Maxwell's equations which in turn need the charge density and the current density. The particle in cell approach provides them immediately as integrals of the distribution function over the velocity variable. Using the functional form for the distribution function of each computational element, the charge and current densities become:

$$\begin{aligned}
&\rho_s(\mathbf{x},t) = \sum_p q_s N_p S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)\\
&\mathbf{j}_s(\mathbf{x},t) = \sum_p q_s N_p \mathbf{v}_p S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)
\end{aligned} \tag{33}$$

The set of equations above provide a closed description for the Vlasov equation. Once accompanied by an algorithm to solve Maxwell's equations the full Vlasov–Maxwell system can be solved.

In summary, the particle method, as derived above, is a discretisation of phase space based on a finite set of computational particles. The continuum distribution function is replaced by a discrete mathematical representation provided by the superposition of moving fixed-shape elements. The pitfalls of this representation have been studied in depth in the past. A full discussion of this aspect is beyond the scope of the present review. We refer the reader to [21,20] for additional comments. In extreme summary, a element of phase space initially described correctly by the shape functions chosen for the discretisation would be distorted by non-uniform electric and magnetic fields. Instead, the computational particles have fixed shape and neglect this effect. This approximation introduces of course an error that can be interpreted as a modification of the phase space. It would be misleading to confuse this error with numerical diffusion, the distortion of phase space does not lead to any spatial propagation of information, as the Liouville theorem insures the conservation of the phase-space volume of each element, a process that is still correctly described by the discretised system. Alternatives to PIC have been developed to remove this error by introducing additional degrees of freedom for each computational element (superparticles) [22,23]. The additional degrees of freedom can describe the particle shape and represent its distortion in phase space. However, the additional complexity of such methods have so far prevented their widespread use.

4. Explicit temporal discretisation of the particle methods

The equations of motion for the computational particles are simple ordinary differential equations with the same form as the regular Newton equations. Of course, in the literature there are many algorithms to achieve the goal of solving the Newton equations. For the PIC algorithm a efficient choice is to use simple schemes: given the very large number of particles used (billions are now common in published works), the use of complex schemes may result in prohibitively long simulations. However, if more advanced schemes allow one to use large time steps, the additional cost per time step may be compensated by taking longer time steps [24].

The simplest algorithm and by far the most used is the so-called *leap-frog algorithm* based on staggering the time levels of the velocity and position by half time step: $\mathbf{x}_p(t = n\Delta t) \equiv \mathbf{x}_p^n$ and $\mathbf{v}_p(t = (n+1/2)\Delta t) \equiv \mathbf{v}_p^{n+1/2}$. The advancement of position

from time level n to time level $n+1$ uses the velocity at mid-point $\mathbf{v}_p^{n+1/2}$, and similarly the advancement of the velocity from time level $n-1/2$ to $n+1/2$ uses the mid point position x_p^n. This stepping of velocity over position and of position over velocity gives the method its name for its resemblance to the children's game bearing the same name (see Fig. 7).

The scheme is summarized by:

$$\begin{aligned}
&\frac{\mathbf{x}_p^{n+1}-\mathbf{x}_p^n}{\Delta t}=\mathbf{v}_p^{n+1/2}\\
&\frac{\mathbf{v}_p^{n+1/2}-\mathbf{v}_p^{n-1/2}}{\Delta t}=\frac{q_s}{m_s}\mathbf{E}_p(x_p^n)+\frac{q_s}{m_s}\left(\frac{\mathbf{v}_p^{n+1/2}+\mathbf{v}_p^{n-1/2}}{2}\right)\times\mathbf{B}_p\left(x_p^n\right)
\end{aligned}\tag{34}$$

where the use of $\mathbf{E}_p$ and $\mathbf{B}_p$ implies knowing the solution of Maxwell's equations given the particle positions at time level n. The first equation is clearly explicit, the second is trivially formulated explicitly as a roto-translation of the vector $\mathbf{v}_p^{n+1/2}$ [21].

Note that technically the leap-frog algorithm is second order accurate, when instead the regular explicit Euler-scheme is only first order. Nevertheless, the two differ in practice only for the fact that the velocity is staggered by half time step. This staggering is achieved by moving the initial velocity of the first time cycle by half a time step using an explicit method:

$$\frac{\mathbf{v}_p^{1/2}-\mathbf{v}_p^0}{\Delta t/2}=\frac{q_s}{m_s}\mathbf{E}_p(x_p^0)+\frac{q_s}{m_s}\left(\frac{\mathbf{v}_p^{1/2}+\mathbf{v}_p^0}{2}\right)\times\mathbf{B}_p\left(x_p^0\right)\tag{35}$$

4.1. Stability of explicit time differencing

The discretisation of the equations of motion using the leap-frog algorithm (or other similar explicit time differencing) introduces a stability constraint. As usual stability can be studied with the Von Neumann analysis [25,20,21]: the equations are linearized and their stability is studied by Fourier analysis. Note that the leap-frog algorithm is implicit in the Lorentz force term, but explicit on the electric force. As a consequence only the latter is of primary concern and stability can be studied in the limit of zero magnetic field. Assuming a linear (harmonic) dependence of the electric field,

$$\frac{q_s}{m_s}\mathbf{E}_p(\mathbf{x}_p)=-\Omega^2\mathbf{x}_p\tag{36}$$

the Von Neuman analysis searches for solutions of the form $\mathbf{x}_p^n=\widetilde{\mathbf{x}_p}e^{i\omega_N t}$ and $\mathbf{v}_p^n=\widetilde{\mathbf{v}_p}e^{i\omega_N t}$ with the numerical response of the system given by ω_N. Simple algebraic manipulations [20,21] lead to:

$$\left(\frac{\Omega\Delta t}{2}\right)^2=\sin^2\left(\frac{\omega_N\Delta t}{2}\right)\tag{37}$$

Notoriously the sin function of real arguments has values only in the $[-1,1]$ interval. As a consequence, for any values of $\Omega\Delta t>2$, the Von Neuman analysis leads to a complex numerical oscillation frequency, ω_N. Indeed being the problem real to begin with, two complex conjugate solutions arise. The numerical solution becomes unstable, with a numerical growth rate that has no physical equivalence. In practice, the particles heat unboundedly and quickly and the simulation fails within a few time steps.

The first cardinal rule of explicit plasma simulation is that the condition $\Omega\Delta t>2$ should never be violated, for any typical Ω in the system. Indeed, practice advises the use of a considerably smaller time step of order $\Omega\Delta t=0.1$, to avoid numerical heating [21].

The temporal stability constraint produces a first great challenge to the application of particle methods in space weather simulation (but also in most other applied plasma physics problems). As noted above, the range of time scales typical of space weather events is tremendous. What the stability constraint imposes is to choose a time step that is able to resolve with a cadence of about 1/10 the fastest frequency in the system. For space weather applications (see Fig. 1) this is commonly the electron plasma frequency, at least 5–7 orders of magnitude smaller than the typical scales of evolution of space weather phenomena.

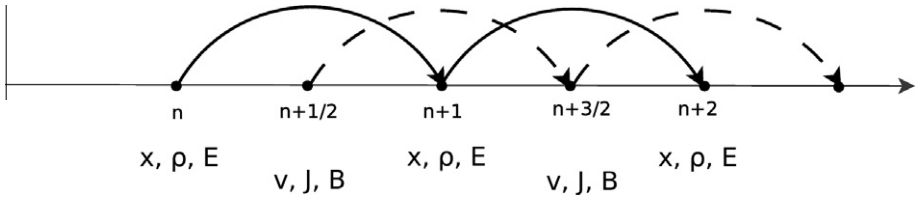

Fig. 7. Visual representation of the leap-frog algorithm. The time discretisation is staggered, with the electric field, charge density and particle positions at integer times and the magnetic field, current and particle velocity at half time steps.

5. Space–time discretisation of the field equations

So far the attention has focused on the particle discretisation of the Vlasov part of the Vlasov–Maxwell system. To complete the simulation method, the Maxwell's equations need to be discretised in space and time.

A wide variety of methods have been developed to do so. Essentially all methods for general electromagnetics can be coupled with particle methods [20,21,26]. Nearly every particle code uses (vastly or slightly) different methods from any other particle code. Below, some general considerations are made, applicable to most if not all of the methods in the literature.

The Maxwell's equations require two inputs, the sources ρ and $\mathbf{j}$ that come from the information carried by the particles. In turn, the outputs are the electric and magnetic fields (or, equivalently the vector and scalar potential). The main concern here is the coupling of Maxwell's equations with the particles and its consequences.

We assume that the fields are known at grid points, not necessarily the same for both fields and are defined as point values or averages over control volumes: $\mathbf{E}_g$, $\mathbf{B}_g$ where the index g labels the discrete points (possibly different for electric and magnetic field). Similarly, we assume that the chosen Maxwell scheme needs the sources at (possibly different) discrete grid points defined as averages over control volumes V_g defining each discrete point:

$$\begin{aligned}\rho_g &= \sum_s \frac{q_s}{V_g}\int_{\mathbb{V}}\int_{V_g} f_s\, d\mathbf{v}\, d\mathbf{x}\\ \mathbf{j}_g &= \sum_s \frac{q_s}{V_g}\int_{\mathbb{V}}\int_{V_g} \mathbf{v} f_s\, d\mathbf{v}\, d\mathbf{x}\end{aligned} \tag{38}$$

Recalling Eqs. (6) and (7), the computation of the sources for Maxwell's equations require computing a 3-dimensional integral of the particle shape over the control volume of each grid point:

$$\begin{aligned}\rho_g &= \sum_p \frac{1}{V_g}\int_{V_g} S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)d\mathbf{x}\\ \mathbf{j}_g &= \sum_p \mathbf{v}_p \frac{1}{V_g}\int_{V_g} S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)d\mathbf{x}\end{aligned} \tag{39}$$

where the integrations in velocity space were done easily recalling the definition of $S_{\mathbf{v}}$. The summation is over all particles of all species (for this reason the summation over the species is not indicated explicitly). The integral in space is straightforward but can involve complicated geometric operations if one allows general shapes for the control volumes. For this reason, the vast majority of PIC methods use an astute choice of the particle shapes and of the control volumes to simplify this effort. If the particle shapes are chosen as in Eq. (13) and the grid is chosen uniform, Cartesian and with cell sizes in each dimension equal to the particle sizes, a simple and elegant reformulation of the integration step can be obtained recalling property (12) of b-splines. Each dimension can be done separately thanks to the choice of Cartesian uniform grid and of the shape of particles as products of shapes in each direction. The integral needed in each direction can be reformulated as

$$\int_{\Delta x_g} S_x(x-x_p) = \int_{-\infty}^{\infty} S_x(x-x_p) b_0\left(\frac{x-x_g}{\Delta x_g}\right)dx \tag{40}$$

where Δx_g is the interval in x relative to control volume V_g. Choosing, $\Delta x_g = \Delta x_p$ (called simply Δx), recalling the definition of S_x, Eq. (13) and using Eq. (12):

$$\int_{-\infty}^{\infty} S_x(x-x_p) b_0\left(\frac{x-x_g}{\Delta x}\right)dx = b_{l+1}\left(\frac{x_g-x_p}{\Delta x}\right) \tag{41}$$

Collecting the integral in each direction, the so-called interpolation function can be defined:

$$W(\mathbf{x}_g-\mathbf{x}_p) = b_{l+1}\left(\frac{x_g-x_p}{\Delta x}\right) b_{l+1}\left(\frac{y_g-y_p}{\Delta y}\right) b_{l+1}\left(\frac{z_g-z_p}{\Delta z}\right) \tag{42}$$

The interpolation functions are a direct consequence of the choice made for the shape functions. Using the b-splines proves a powerful choice as it allows to write the interpolation functions just as b-splines one order higher. The calculation of the source for Maxwell's equations becomes then just the sum of a number of function evaluations without requiring geometrically complex integrals:

$$\begin{aligned}\rho_g &= \frac{1}{V_g}\sum_p q_p W(\mathbf{x}_g-\mathbf{x}_p)\\ \mathbf{j}_g &= \frac{1}{V_g}\sum_p q_p \mathbf{v}_p W(\mathbf{x}_g-\mathbf{x}_p)\end{aligned} \tag{43}$$

An analogous set of operations can be carried out for connecting the output of the Maxwell's equations to the particle equations of motion. The needed quantities are the electric and magnetic fields derived above in Eqs. (27) and (30). The definitions for the particle fields require continuum fields. These can be obtained once again using interpolation:

$$\begin{aligned}\mathbf{E}(\mathbf{x}) &= \sum_g \mathbf{E}_g S_{\mathbf{E}}(\mathbf{x}-\mathbf{x}_g)\\ \mathbf{B}(\mathbf{x}) &= \sum_g \mathbf{B}_g S_{\mathbf{B}}(\mathbf{x}-\mathbf{x}_g)\end{aligned} \tag{44}$$

where $S_{\mathbf{E}}(\mathbf{x}-\mathbf{x}_g)$ and $S_{\mathbf{B}}(\mathbf{x}-\mathbf{x}_g)$ are the fields interpolation functions. Upon substitution in the definition of the particle fields,

$$\begin{aligned}\mathbf{E}_p &= \sum_g \int S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)\mathbf{E}_g S_{\mathbf{E}}(\mathbf{x}-\mathbf{x}_g)d\mathbf{x}\\ \mathbf{B}_p &= \sum_g \int S_{\mathbf{x}}(\mathbf{x}-\mathbf{x}_p)\mathbf{B}_g S_{\mathbf{B}}(\mathbf{x}-\mathbf{x}_g)d\mathbf{x}\end{aligned} \tag{45}$$

The geometrical complexity of the integrals can be avoided choosing the interpolation functions for the electric and magnetic fields as b-splines of order 0:

$$S_{\mathbf{E},\mathbf{B}}(\mathbf{x}-\mathbf{x}_g) = b_0\left(\frac{x_g-x_p}{\Delta x}\right)b_0\left(\frac{y_g-y_p}{\Delta y}\right)b_0\left(\frac{z_g-z_p}{\Delta z}\right) \tag{46}$$

leading to

$$\begin{aligned}\mathbf{E}_p &= \sum_g \mathbf{E}_g W(\mathbf{x}_g-\mathbf{x}_p)d\mathbf{x}\\ \mathbf{B}_p &= \sum_g \mathbf{B}_g W(\mathbf{x}_g-\mathbf{x}_g)d\mathbf{x}\end{aligned} \tag{47}$$

with the same interpolation function used for the sources.

While the vast majority of schemes follows the choice of b-splines and uniform grids, more advanced schemes have also been considered. The cell positions can be chosen differently for different fields and different sources. Staggered schemes and Yee lattices are commonly used [27,28]. Also different order of interpolations can be used for different quantities [29]. A more radical solution is to consider non-uniform or even unstructured grids (see Section 10). In this case the interpolation cannot use the b-spline properties and the convenient definition of interpolation function. Nevertheless, the integral definitions above still hold and one can base the computation of sources and fields using directly the integral definition.

5.1. Space resolution and finite grid instability

The introduction of a grid to compute the field characterizes the particle methods for plasma physics as particle in cell (PIC) methods. The particles move in a continuum space, but their information is projected to grid points using integration over control volumes. This operation reduces the information, collapsing the continuum particle shapes to discrete contributions in a set of points (the grid values). The loss of information allows one numerical instability to creep in: the finite grid instability.

The mathematical study of the finite grid instability is complex, requiring a careful and complete analysis of all computational steps with the Laplace transformation of time and the Fourier transformation of space. The reader interested in the details is referred to the textbooks on the subject [21,20] and to the original work cited therein. The summary of the analysis is that to avoid the finite grid instability in explicit particles methods, the grid spacing must be chosen to satisfy the constraint

$$\Delta x/\lambda_{De} < \varsigma \tag{48}$$

where λ_{De} is the Debye length (an intrinsic plasma scale measuring the shilding length in a plasma) [20,21] and the parameter ς is of order one and depends on the details of the implementation. For the widely used choice of shape functions as b-spline of order 0 and consequently interpolation functions of order 1, the scheme referred to as cloud in cell (CIC), the literature reports, $\varsigma \approx \pi$.

The practical consequence of the finite grid instability is a tremendous numerical heating of the plasma characterized by an alternatively positive–negative variation of the electric field accompanied by a correlated zig-zag perturbation of the phase-space. Within a few cycles, the energy reaches the bounds of overflow in the machine representation of numbers. The finite grid instability must be avoided at all costs, requiring its stability constraint to be respected everywhere in the domain for all directions.

The consequences of this stability constraint for space weather simulations are devastating. The spatial scale to resolve is the Debye length in the highest density region of the domain where the Debye length is smallest. If using uniform grids, this constraint requires to use cell sizes several orders of magnitude smaller than the scales of interest. Note that this consideration has two sides.

First, often the processes of interest do not reach down to the Debye scale. For example most reconnection events in space are dominated by processes that are on the electron inertial length, $d_e = c/\omega_{pe}$. This scale is still two orders of magnitude larger than the Debye length. Having to resolve the Debye length purely for numerical stability reasons implies a waste of fully 2 orders of magnitude in each direction, or 10^6, a million times. If the time scale constraint encountered before hits the

explicit approach once, the spatial stability constraint hits it three times, once per spatial dimension. The combined effect of these two constraint makes it imperative to look beyond explicit particle methods.

Second, often space plasmas present regions of high density in contact with regions of lower density, resulting in a large range of Debye lengths. To avoid the finite grid instability, uniform grids need to be chosen to resolve the smallest Debye length in the system. Adaptive schemes can be devised to resolve in each region only the local Debye length, resulting in considerable savings [30]. However, even in this case, often the local scales of interest are much larger than the Debye length and the local grid resolution would be more desirably chosen according to other consideration than the need to resolve the Debye length.

5.2. Time discretisation of Maxwell's equations

The most common discretisation in space locates the electric and magnetic fields and the particle charges and currents at different locations. The Yee lattice in Fig. 8 finds by far the widest application but alternative approaches are widespread. Many approaches are based on finite differences [21,20] or finite volume [27,31]. The electric field and currents are vertex-centered while the magnetic field and the density are cell-centered. Finite element approaches have also been considered [24,32].

Even when the sources are assumed to be known from the explicit particle approach (i.e. are taken from the previous time step), still the Maxwell's equations are a linear coupled system for the fields. Whatever the scheme followed, the fields on the discrete grid points will be obtained by discretising numerically the Maxwell's equations that are linear in the fields:

$$\mathbb{M}\begin{pmatrix}\mathbf{E}_g\\ \mathbf{B}_g\end{pmatrix} = \begin{pmatrix}\rho_g\\ \mathbf{j}_g\end{pmatrix} \tag{49}$$

where the matrix $\mathbb{M}$ summarizes the linear coupling of the electric and magnetic fields. In the simplest approach the Maxwell's equations are discretised explicitly. The leap-frog algorithm is used again (see Fig. 7). The current and the magnetic field are computed at the half time steps: $\mathbf{B}_g^{n+1/2}$, $\mathbf{j}_g^{n+1/2}$ and the electric field and the density are computed at the integer time step: $\mathbf{E}_g^n$, ρ_g^n. We use here a generic notation where the spatially-discretised quantities are labeled with a index g that is generic and can mean cell centers, faces or nodes and can be different for different fields, depending on the scheme used. The operators are also indicated generically using the same symbol and name as for the corresponding continuum operators. The reader is referred to the cited material for the detailed implementation of any specific spatial discretisation.

A time centered marching algorithm follows simply. The magnetic field is computed from the $\nabla \times \mathbf{E}$ Maxwell's equation:

$$\mathbf{B}_g^{n+1/2} = \mathbf{B}_g^{n-1/2} - \Delta t \nabla \times \mathbf{E}_g^n \tag{50}$$

and the electric field is computed over a time interval half-shifted:

$$\mathbf{E}_g^{n+1} = \mathbf{E}_g^n + \frac{\Delta t}{\mu_0 \epsilon_0}\left(\nabla \times \mathbf{B}_g^{n+1/2} - \mu_0 \mathbf{j}_g^{n+1/2}\right) \tag{51}$$

The same simple marching order used for the particles can be used also for the fields resulting in a exceeding simple approach. However, the use of an explicit scheme within the field solver introduces an additional stability constraint: the Courant–Friedrichs–Lewy (CFL) [25] condition on the propagation of electromagnetic waves: $c\Delta t < \Delta x$.

This limitation is intrinsic to the Maxwell's equations and can be removed by discretising implicitly just the linear part of the equations. The coupling with the plasma can still be treated explicitly, i.e. the sources are kept frozen to their value at the

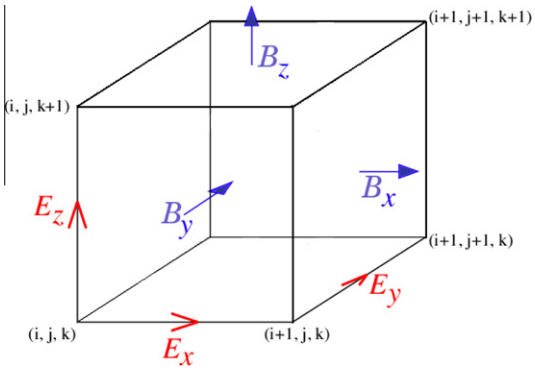

Fig. 8. Location of the electromagnetic fields in a typical PIC method based on the Yee lattice discretisation. The current density $\mathbf{j}_g$ is collocated with the electric field. The explicit PIC does not directly use the charge density, except for divergence cleaning (see text), and for that the charge density ρ_g and the correction potential $\delta\varphi_g$ are located on the vertices (i,j,k).

beginning of the time step, but the coupling between the two curl Maxwell's equations is kept implicitly. Again many methods can be applied. In practice (for historical more than practical reasons), the method of Ref. [33] based on the potential formulation is used in most codes where only the field solve is done implicitly but the particle part is still explicit. Instead, the second-order implicit formulation of the Maxwell's equations based on the electric field derived in Ref. [34] is used in codes where also the particles are treated implicitly.

It is worth insisting on the fact that removing the CFL condition on the light waves has no impact on the other two constraints that remain in place as long as the particle part of the algorithm and its coupling with Maxwell's equations is done explicitly.

So far the attention has focused on the two curl Maxwell's equations. In the continuum, the other two divergence conditions are automatically satisfied at all times provided that they are satisfied at the initial time and requiring that detailed charge conservation is satisfied:

$$\frac{\partial \rho}{\partial t} = \nabla \cdot \mathbf{j} \tag{52}$$

Once discretised, however, this convenient truth is no longer true. Most schemes enforce automatically the divergence condition on **B**. In fluid modeling, especially applied to space weather, great attention needs to go into enforcing the condition that $\nabla \cdot \mathbf{B} = 0$. In PIC methods, instead, this conditions is usually trivially satisfied by the judicious choice of a discretisation scheme where $\nabla \cdot \nabla \times \mathbf{E} = 0$ is satisfied. This is usually the case, except in some methods applied to non-uniform grids. In these rare cases, the vast literature developed for MHD can be of help [35,36]. Instead, a much more serious problem for PIC methods comes from the other divergence conditions, $\epsilon_0 \nabla \cdot \mathbf{E} = \rho$. This condition is normally not satisfied easily.

Two types of approaches are commonly followed.

First, the so-called charge conserving schemes can be used. The interpolation of charge and density is chosen in a way to conserve locally charge ensuring that the discretised version of Eq. (52) is valid [21]. The approach has two serious drawbacks: to enforce this property one is forced to sacrifice the conservation of momentum (see below) and reportedly the simulations display an increased level of noise requiring many more particles per cell (see Ref. [21] at p. 360 and Ref. [33]). The advantage is that no elliptic solver is needed as both divergence equations are automatically enforced. This feature leads to excellent parallel scaling when implemented on massively parallel computers [37].

Second, divergence cleaning approaches can be used by simply correcting the electric field for its electrostatic part to enforce Poisson's equation:

$$\nabla^2 \delta\varphi = \nabla \cdot \mathbf{E} - \rho/\epsilon_0 \tag{53}$$

where the corrected field is $\mathbf{E}' = \mathbf{E} - \nabla\delta\varphi$. Clearly this achieves the goal by adding a correction. But at the cost of solving an elliptic problem, a cpu-intensive operation. To circumvent this difficulty inexact versions have been proposed [38,39] where the elliptic equation is not solved to convergence but is in fact in a sense converged over multiple time steps, reducing the computational burden at the cost of allowing a controlled but non-zero error in the divergence equation.

6. Recapitulation of the application of explicit particle methods to space weather modeling

Collecting the steps gathered so far, the PIC algorithm is summarized by the series of operations depicted in Fig. 9.

The loop is followed step by step without iterations. Starting for example from the top, the particle equations of motion discretised with the leap-frog algorithm are advanced one step from $\mathbf{x}_p^n$, $\mathbf{v}_p^{n-1/2}$ to $\mathbf{x}_p^{n+1}$, $\mathbf{v}_p^{n+1/2}$, the new particle information is projected to the grid to obtain the density and current density, sources of Maxwell's equations. The latter are then solved for

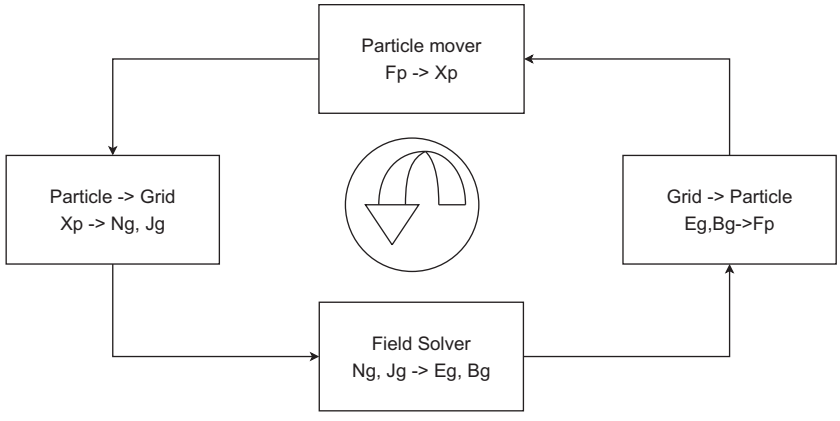

Fig. 9. Summary of a computational cycle of the explicit PIC method.

the new fields $\mathbf{E}^{n+1}$, $\mathbf{B}^{n+1/2}$, given the old fields $\mathbf{E}^{n}$, $\mathbf{B}^{n-1/2}$. This operation requires only the sources just computed. Once the new fields are obtained, the particle forces can be computed by interpolation from the grid and the cycle can restart.

The simplicity of the approach explains its wide success. However, as noted above the approach is plagued by two stability constraint coming from the explicit particle discretisation, $\omega_{pe}\Delta < 2$ and $\Delta x/\lambda_{De} < \varsigma$. Additionally, the condition $c\Delta t < \Delta x$ should also be considered when the field solver is also explicit.

The key complexity of a plasma comes from the non-linear coupling of particles and fields. In the real world, the four blocks displayed in Fig. 9 are constantly coupled. Instead, the explicit approach freezes temporarily one to apply the other: the particles move for one time step in frozen fields and the Maxwell's equations advance the fields in one time step based on frozen sources. Neither are true physically. This is the key limitation of the explicit approach, especially in multiple scale system such as is typical for space weather problems.

7. Implicit particle methods

To avoid the prohibitively large number of time steps and the exceedingly high resolution needed to run realistic problems with explicit particle methods, implicit methods have been considered for several decades [40–43].

The starting point of implicit particle methods replaces the explicit time differencing of the equations of motion with an implicit scheme. For example, the so-called θ scheme can be used [29,44]:

$$\begin{aligned}\mathbf{x}_p^{n+1} &= \mathbf{x}_p^n + \mathbf{v}_p^{n+1/2}\Delta t \\ \mathbf{v}_p^{n+1} &= \mathbf{v}_p^n + \frac{q_s\Delta t}{m_s}\left(\mathbf{E}_p^{n+\theta}\left(\mathbf{x}_p^{n+1/2}\right) + \mathbf{v}_p^{n+1/2}\times\mathbf{B}_p^{(n+*)}\left(\mathbf{x}_p^{n+1/2}\right)\right)\end{aligned} \tag{54}$$

where the flexibility of defining a decentering parameter θ is used to vary the properties of the scheme. The quantities at time level $n+\theta$ are computed as weighted averages between the old and new level, $\Psi^{n+\theta} = \Psi^n(1-\theta) + \Psi^{n+1}\theta$. A special word of caution is needed for the magnetic field time level. Using the old time level n is the simplest approach and does not lead to instability, but time centered discretisations or other more involved schemes have also been considered [44]. Other implicit schemes derived from the leap-frog algorithm have been proposed also [45,46].

The Eqs. (54) are implicit for two reasons.

First, to advance the position the new velocity is needed and to advance the velocity the new position is needed (because the fields need to be evaluated at the mid point between the new and old position). This implicitness is local to each particle and can be handled easily. In fact it can be eliminated with a clever decomposition of the velocity. The velocity equation is more conveniently rewritten as:

$$\mathbf{v}_p^{n+1/2} = \hat{\mathbf{v}}_p + \beta_s\widehat{\mathbf{E}}_p^{n+\theta}\left(\mathbf{x}_p^{n+1/2}\right) \tag{55}$$

where $\beta_s = q_p\Delta t/2m_p$ (independent of the particle weight and unique to a given species). For convenience, we have introduced hatted quantities obtained by explicit transformation of quantities known from the previous computational cycle:

$$\begin{aligned}\hat{\mathbf{v}}_p &= \boldsymbol{\alpha}_s^n\cdot\mathbf{v}_p^n \\ \widehat{\mathbf{E}}_s^{n+\theta} &= \boldsymbol{\alpha}_s^n\cdot\mathbf{E}_s^{n+\theta}\end{aligned} \tag{56}$$

The transformation tensor operators $\boldsymbol{\alpha}_s^n$ are defined as:

$$\boldsymbol{\alpha}_s^n = \frac{1}{1+(\beta_s B^n)^2}\left(\boldsymbol{I} - \beta_s\boldsymbol{I}\times\mathbf{B}^n + \beta_s^2\mathbf{B}^n\mathbf{B}^n\right) \tag{57}$$

and represent a scaling and rotation of the velocity vector. The tensor operator $\boldsymbol{\alpha}_s^n$ is the same for all computational particles of the same species s.

Second, and more importantly, Eqs. (54) are implicit because the new fields are needed also. If the electric and magnetic fields were given, the equations of motion could be solved simply. The challenge of the implicit method is that the electric and magnetic fields are not given. Unlike the explicit method, the particle equations of motion need the new fields that are not known yet. The implicit method reintroduces the coupling of the particle equations with the field equations. Implicit methods solve the Maxwell's equations also implicitly in time. Indicating with index g the spatial discretisation of the operators, the discretised Maxwell's equations can be written as:

$$\begin{aligned}\mathbf{B}_g^{n+1} - \mathbf{B}_g^n &= -\Delta t\nabla\times\mathbf{E}_g^{n+\theta} \\ \mathbf{E}_g^{n+1} - \mathbf{E}_g^n &= \frac{\Delta t}{\mu_0\epsilon_0}\left(\nabla\times\mathbf{B}_g^{n+\theta} - \mu_0\mathbf{j}_g^{n+\frac{1}{2}}\right) \\ \epsilon_0\nabla\cdot\mathbf{E}_g^{n+\theta} &= \rho_g^{n+\theta} \\ \nabla\cdot\mathbf{B}_g^{n+1} &= 0\end{aligned} \tag{58}$$

The same spatial discretisations discussed for the explicit method can be re-used in the implicit method. The interpolation formulas derived above provide the expression for the fields in the equations of motion and for the sources in Maxwell's equations. The coupling of the set of equations is evident.

7.1. Choice of θ and energy conservation

The parameter $\theta \in [1/2,1]$ is chosen in order to adjust the numerical dispersion relation for electromagnetic waves (for $\theta < 1/2$, the algorithm is shown to be unstable [42]).

Special attention requires the case of $\theta = 1/2$ where the scheme becomes second order accurate due to the central differencing in time of all quantities involved, $\overline{\Psi} = \Psi^{n+1/2} = (\Psi^n + \Psi^{n+1})/2$. This choice has an important consequence: exact energy conservation to round off precision [47–49].

The variation of the electric and magnetic field energy δW_E, δW_B during the time step Δt is:

$$\delta W_E + \delta W_B = \sum_g V_g \left(\frac{\left(\mathbf{E}_g^{n+1}\right)^2 - \left(\mathbf{E}_g^{n}\right)^2}{2\epsilon_0^{-1}} + \frac{\left(\mathbf{B}_g^{n+1}\right)^2 - \left(\mathbf{B}_g^{n}\right)^2}{2\mu_0} \right). \tag{59}$$

Expressing the new fields using Eq. (58):

$$\delta W_E + \delta W_B = -\frac{\Delta t}{\mu_0} \sum_g V_g \overline{\mathbf{E}}_g \cdot \mu_0 \bar{\mathbf{j}}_g + \frac{\Delta t}{\mu_0} \sum_g V_g \left(\overline{\mathbf{E}}_g \cdot \nabla \times \overline{\mathbf{B}}_g - \overline{\mathbf{B}}_g \cdot \nabla \times \overline{\mathbf{E}}_g \right). \tag{60}$$

The term in parenthesis in the right hand side of Eq. (60) is a discretisation of the divergence of the Poynting flux, $\overline{\mathbf{S}}_g = (\overline{\mathbf{E}}_g \times \overline{\mathbf{B}}_g)/\mu_0$ and the change in energy can be written as:

$$\delta W_E + \delta W_B = -\sum_g V_g (\overline{\mathbf{E}}_g \cdot \bar{\mathbf{j}}_g + \nabla \cdot \overline{\mathbf{S}}_g) \Delta t \tag{61}$$

Expressing the average current using its relation with the particle information,

$$\bar{\mathbf{j}}_g = \frac{1}{V_g} \sum_p q_p \bar{\mathbf{v}}_p W(\mathbf{x}_g - \bar{\mathbf{x}}_p) \tag{62}$$

the energy change can be written as:

$$\delta W_E + \delta W_B + \sum_g \nabla \cdot \overline{\mathbf{S}}_g \Delta t = -\sum_p q_p \Delta t \bar{\mathbf{v}}_p \cdot \sum_g \overline{\mathbf{E}}_g W(\mathbf{x}_g - \bar{\mathbf{x}}_p) \tag{63}$$

Recognizing the term $\overline{\mathbf{E}}_p$ from the summation over the cells in the right hand side and using the second of Eqs. (54), it follows:

$$\delta W_E + \delta W_B + \sum_g \nabla \cdot \overline{\mathbf{S}}_g \Delta t = -\sum_p m_p \bar{\mathbf{v}}_p \cdot \left(\mathbf{v}_p^{n+1} - \mathbf{v}_p^n - \bar{\mathbf{v}}_p \times \mathbf{B}_p^{n+*} \right), \tag{64}$$

where m_p is the rest mass of the computational particle. With simple vector operations, the final conservation statement follows as:

$$\delta W_E + \delta W_B + \sum_g \nabla \cdot \overline{\mathbf{S}}_g \Delta t = -\sum_p m_p \left(\left(v_p^{n+1}\right)^2 - \left(v_p^n\right)^2 \right), \tag{65}$$

where the right hand side is the change in kinetic energy δW_k.

In summary, it holds that

$$\delta W_E + \delta W_B + \delta W_k + \sum_g \nabla \cdot \overline{\mathbf{S}}_g \Delta t = 0, \tag{66}$$

and the total energy is conserved exactly.

Special attention requires the step where the assumption was made that the Poyting flux is defined as $\overline{\mathbf{S}}_g = (\overline{\mathbf{E}}_g \times \overline{\mathbf{B}}_g)/\mu_0$. In practice the property of energy conservation has so far only been demonstrated in simulation for 1D electromagnetic systems [47,49] (including the relativistic extension [48]) and for the electrostatic case [48].

7.2. Stability of implicit methods

The same Von Neuman stability analysis repeated for the implicit equations above leads to a very different solution,

$$\left(\frac{\Omega \Delta t}{2}\right)^2 = \tan^2\left(\frac{\omega_N \Delta t}{2}\right) \tag{67}$$

The tan function has values in the whole real continuum, allowing the use of any time step Δt. The stability constraint is eliminated and in principle the time step can be selected based only on the physics of interest.

All stability constraints are eliminated and replaced by the sole constraint of selecting the desired accuracy. For space weather simulations this is a great advantage. The scales of interest are often on the electron inertial length, nearly two

orders of magnitude larger than the Debye length that needs to be resolved by the explicit method. Similarly, time can be set to resolve the electron trajectories rather than the fastest waves. Several orders of magnitude can be gained in avoiding unnecessary spatial and temporal resolution. Alone this vastly overwhelms the significant additional cost per computational cycle of the implicit method over the explicit method. As shown below for each case considered, the additional cost of one computational cycle can be a factor of 3–10 higher for implicit method (depending wether fully implicit or semi-implicit). But the savings are instead of several orders of magnitude. In the example presented in Section 11.2 the gain of the implicit method is a staggering 10^6, overwhelming any additional cost per cycle.

But what happens then to the scales that are not resolved? The question was immediately felt as the implicit particle methods were developed and a considerable body of literature is available, reviewed in [21,50]. In summary, the scales that are not resolved remain present but in a distorted fashion. The frequency of all unresolved waves saturates towards the maximum resolved frequency, the Nyquist frequency of $1/\Delta t$. For a choice of a decentering parameter larger than $\theta > 1/2$, the unresolved waves are further damped, at a rate exceeding the natural Landau damping. Physically this allows the energy dissipation towards the unresolved scales, allowing to capture the gross contribution of such scales to the energy balance. A complete analysis of the fate of all plasma waves when the time step is not resolving them is described in Ref. [42].

The fundamental problem to address in developing an implicit PIC method is the coupling between the equations of motion and the field equations for the presence of the time advanced electric field (but not magnetic field, that is used from the previous cycle, as no instability is introduced) in the equations of motion and for the appearance [due to the presence of $\mathbf{x}_p$ and $\mathbf{v}_p$ in Eq. (33)] of the particle properties in the sources of the Maxwell equations. In both cases the coupling is implicit, so that the new particle properties need to be known before the fields can be computed and likewise the new fields need to be available before the new particle properties can be computed.

In recent years, the direct solution of the coupled system for fields and particles became within reach of modern computational methods and computer resources [51,52], but traditionally, alternative methods were developed to avoid dealing directly with the full system of coupled non-linear equations.

Below, we consider first the more modern fully implicit approach, and consider the more traditional semi-implicit approach next.

8. Fully implicit particle methods

The implicit method requires the coupled solution of the implicit equations of motion (54), where the new fields at the advanced time level are required, and of the implicit field Eqs. (58), where the new particle properties are needed to compute the sources. The coupling is non-linear, just as the original Vlasov–Maxwell system is also non-linear: the force on the particles is not given but rather depends non-linearly on the fields that in turn depend on the particles. Modern computational science has developed a powerful tool for solving such systems of non-linear equations: the preconditioned Jacobian-Free–Newton–Krylov (JFNK) approach.

In extreme summary and referring the reader to the relevant textbooks on the subject [53], given a non-linear set of equations, $\mathbf{f}(\mathbf{w}) = 0$for the set of unknowns $\mathbf{w}$, the JFNK method is based on solving iteratively the residual equation with the Newton method without having to form the Jacobian matrix. The new iteration $\mathbf{w}^{(k+1)} = \mathbf{w}^{(k)} + \delta\mathbf{w}$ is obtained by computing the displacement $\delta\mathbf{w}$ using the Newton method [53], i.e. using a Taylor series expansion truncated at first order (to ensure linearity of the resulting equation for $\delta\mathbf{w}$). To avoid forming and storing the Jacobian matrix needed for the Taylor series expansion, the equation for $\delta\mathbf{w}$ is approximated numerically by replacing the derivatives with differences:

$$\mathbf{f}(\mathbf{w}^{(k)}) + \frac{\mathbf{f}(\mathbf{w}^{(k)} + \epsilon\delta\mathbf{w}) - \mathbf{f}(\mathbf{w}^{(k)})}{\epsilon} = 0 \tag{68}$$

While formally non-linear, Eq. (68) is linearized by choosing ϵ small enough to ensure the accurate replacement of the first order derivative with the difference in the Taylor expansion. In practice, ϵ is chosen according to the machine precision [53]. The Jacobian Eq. (68) is expressed only in terms of residual function evaluations and there is no need to form or store the matrix. To solve Eq. (68), preconditioned Krylov-based schemes are used. Here, we use the GMRES algorithm since the Jacobian matrix can be non-symmetric [54].

In the present problem, the unknowns are in principle the whole set formed by the particle positions and velocities at the advanced time level: $\mathbf{x}_p^{n+1}, \mathbf{v}_p^{n+1}$ and the fields at the advanced level: $\mathbf{E}_g^{n+1}$ and $\mathbf{B}_g^{n+1}$. While viable, this choice leads in practice to the need to handle the application of the JFNK methods to an extremely large array. The typical particle simulation nowadays employs billions of particle and millions of grid points. To reduce the load on the computer memory, a new technique, the subsidiarity of the particle equations, has been introduced [51].

The starting point is the realization that the particle equations are not directly coupled with each other. Each set of 6 equations for a particle do not require the direct application of the information for the other particles except via the fields. So each set of 6 equations can be solved directly given the fields. Within the JFNK iteration scheme, the latest guess (Newton iteration k) is indeed available. Therefore the Newton method does not need to consider the particles as part of the unknown. Given a guess of the fields, the particle properties at the advanced time level are immediately computable. This is the core property of particle subsidiarity: the JFNK method considers as unknowns only the fields and leaves the particle computation as a user-supplied rule that is part of the residual evaluation for the fields.

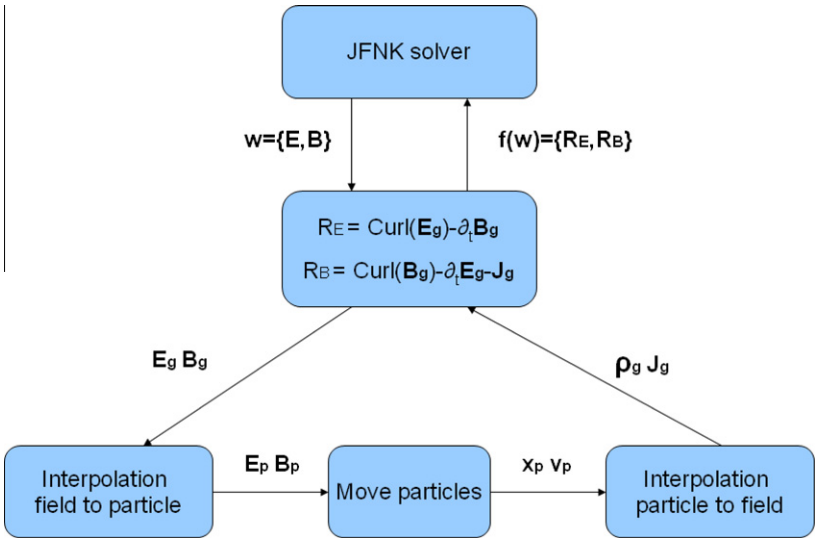

Fig. 10. Outline of the interaction of particle mover, field residual computation and JFNK solver in a fully implicit particle method using particle subsidiarity. The equations are indicated in a summarized incomplete notation just for reference.

In practice, the approach sets the following unknown vector

$$\mathbf{w} = \begin{pmatrix} \mathbf{E}_g^{n+1} \\ \mathbf{B}_g^{n+1} \end{pmatrix} \tag{69}$$

and the corresponding residual equation is formed by the first two of Eqs. (58). The particle equations become simply a rule on how to compute the sources $\rho_g^{n+\theta}$ and $\mathbf{j}_g^{n+\frac{1}{2}}$. The computational cycle of the explicit scheme is replaced by the interaction between the JFNK solver and the residual equation, as outlined in Fig. 10.

The JFNK method requires the user to provide the rule that produces the residual given a value of the unknown vector. The implicit particle method needs to procure such residual. This is done in steps:

1. The NK vector $\mathbf{w}$ is converted back into separate electric and magnetic fields: $\mathbf{E}_g^{n+1}$, $\mathbf{B}_g^{n+1}$.
2. These fields are interpolated to the particles: $\mathbf{E}_p^{n+1}$, $\mathbf{B}_p^{n+1}$.
3. The mover advances the particles in these fields provided by JFNK.
4. Once the particles are moved, the field sources can be computed: $\rho_g^{n+\theta}$ and $\mathbf{j}_g^{n+\frac{1}{2}}$.
5. With the sources, the field equations provide the residuals, $\mathbf{r}_{\mathbf{E}_g^{n+1}}$, $\mathbf{r}_{\mathbf{B}_g^{n+1}}$, needed by JFNK in response to its provided guess for the fields.
6. The NK residual vector is formed: $\mathbf{r}_{\mathbf{w}} = \left(\mathbf{r}_{\mathbf{E}_g^{n+1}}, \mathbf{r}_{\mathbf{B}_g^{n+1}}\right)$.

The operations to be conducted are still of the same order of complexity as if the JFNK was directly working also on the particle equations, however, the memory requirements for storing the intermediate Krylov vectors for the JFNK solver are greatly reduced.

The use of the JFNK makes the solution of the system of non-linear equations extremely simple for the user. Powerful libraries, ranging from the simple tool provided in the textbooks [53] to more sophisticated parallel libraries such as PETSc [55–57] remove the need for the user to code the JFNK method. All that is needed is to provide the simple procedures for computing the residuals. The approach has the advantage of minimizing the personal time of the programmers who develop the software. However, the computational cost of one compute cycle to bring the JFNK to completion is several times that of the simple explicit method described above. This additional cost is vastly overwhelmed by the gains in grid spacing and time step allowed by the removal of the stability constraint.

9. Semi implicit methods

The full implicit formulation described above has become tractable numerically only in recent years thanks to the progress in the JFNK approach. Previously, the task was considered beyond reach and methods were developed to simplify the non-linear coupling and reduce it to a more tractable linear problem. The semi-implicit approach still preserves many of the good properties of stability of the fully implicit method.

The research in the field has organized itself into two lines: the implicit moment method [42] and the direct implicit method [43]. While started from significantly different points of view, over the years the two approaches have converged towards a more common perspective and mathematical framework. The idea is to linearize the coupling between particle and fields and solve it as a coupled linear system instead of the true non-linear system it is. This approach is common in scientific computing and is referred to as semi-implicit. For space weather applications, so far, the implicit moment method has found the majority of applications, while the direct implicit methods is primarily used in inertial fusion research. Below we present in general the semi-implicit method, focusing on the moment implicit approach as the main example but pointing out the differences between implicit moment and direct implicit when appropriate.

The semi-implicit method reduces the number of equations that must be solved self-consistently to a set composed by the field equations and a simplified plasma response model that approximates the response of the particles to the field changes over the time step considered. The solution of this set of equations implicitly, and the subsequent explicit solution of the particle equations of motion in the resulting fields, is stable and accurate.

The idea is summarized in Fig. 11 where the logic of the sequence of operations involved in the semi-implicit method is shown. The coupling due to the implicit discretisation of both field and particle equations is broken by approximating the sources of the field equations using the moment equations instead of the particle equations directly. Once the field equations are solved within this approximation, the rest of the steps can be completed directly without iterations: with the new fields, the particle equations of motion can be solved and the new current and density can be computed for the next computational cycle.

The formulation used here is described in Refs. [42,29,58,43] and its relativistic extension in Ref. [59] The key step is to derive a suitable set of moment equations that can approximate the response of the plasma particles to the fields over a computational cycle. The approach is based on a series expansion of the interpolation function appearing in the expression for the field sources, Eq. (33). The expansion is done with respect to the particle position, choosing the center of the expansion as the particle positions at the start of the computational cycle:

$$W\left(\mathbf{x}-\mathbf{x}_p^{n+1}\right)=W\left(\mathbf{x}-\mathbf{x}_p^{n}\right)+\left(\mathbf{x}-\mathbf{x}_p^{n}\right)\cdot\nabla W\left(\mathbf{x}-\mathbf{x}_p^{n}\right)+\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_p^{n}\right)\left(\mathbf{x}-\mathbf{x}_p^{n}\right):\nabla\nabla W\left(\mathbf{x}-\mathbf{x}_p^{n}\right)+\cdots \tag{70}$$

where a tensor notation is used. The main difference between the implicit moment and direct implicit method is in the choice of the center of the series expansion. In the implicit moment method, the particle mover uses the θ-scheme and it is natural to center the expansion on the old particle position $\mathbf{x}_p^n$. The direct implicit method uses instead one in a number of different movers based on stabilizing the leap-frog algorithm with implicit variants [46]. In this case the expansion is done around a guess of the new advanced position [43]. The procedure outlined here can be extended (non trivially) to the case of special relativity both in the case of the implicit moment method [59] and of the direct implicit method [60].

The expansion (70) can be used to compute the field sources directly using particle information only from the previous computational cycle and removing the need to iterate the particle and field equations. The details of the simple but tedious algebraic manipulations are provided in Ref. [29], the final answer being:

$$\begin{aligned}\rho_s^{n+1}&=\rho_s^n-\Delta t\nabla\cdot J_s^{n+1/2}\\ \mathbf{J}_s^{n+1/2}&=\widehat{\mathbf{J}}_s-\frac{\Delta t}{2}\boldsymbol{\mu}_s\cdot E_\theta-\frac{\Delta t}{2}\nabla\cdot\widehat{\boldsymbol{\Pi}}_s\end{aligned} \tag{71}$$

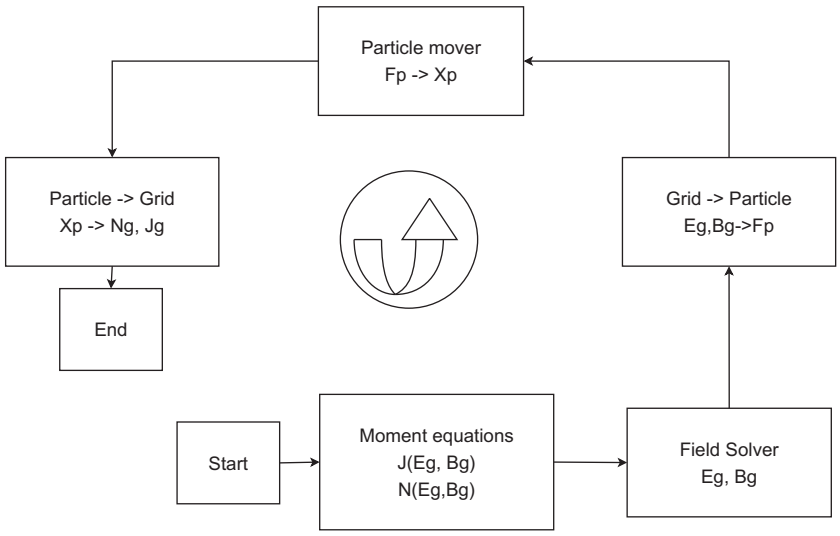

Fig. 11. Blocks needed in a cycle of the semi-implicit PIC scheme.

where the following expressions were defined:

$$\begin{aligned}\widehat{\mathbf{J}}_s &= \sum_p q_p \widehat{\mathbf{v}}_p W\left(\mathbf{x}-\mathbf{x}_p^n\right)\\ \widehat{\mathbf{\Pi}}_s &= \sum_p q_p \widehat{\mathbf{v}}_p \widehat{\mathbf{v}}_p W\left(\mathbf{x}-\mathbf{x}_p^n\right)\end{aligned} \tag{72}$$

with the obvious meaning, respectively, of current and pressure tensor based on the transformed hatted velocities. An effective dielectric tensor is defined to express the feedback of the electric field on the plasma current and density:

$$\boldsymbol{\mu}_s^n = -\frac{q_s \rho_s^n}{m_s}\boldsymbol{\alpha}_s^n \tag{73}$$

The expression (71) for the sources of the Maxwell's equations provide a direct and explicit closure of Maxwell's equations. When Eq. (71) is inserted in Eq. (58), the Maxwell's equations can be solved without further coupling with the particle equations. The procedure followed distilled from the equations of motion of the particles an average response that can be used to represent the plasma response within one time step. This is the key property of the semi-implicit method and allows it to retain the once-through approach typical of explicit methods and eliminates the need for expensive iteration procedures that would require to move the particles multiple times per each computational cycle.

In the moment implicit method, the plasma response takes the form presented above and represents a closure of the fluid equations using the particle information to express the pressure tensor. In the direct implicit method, a similar approach is used that relies on the mathematical linearization of the response rather than on the use of the moment method [50].

9.1. Field Solver

Different versions of the semi-implicit PIC method have relied on different solution procedures for the Maxwell equations. In CELESTE1D a potential formulation was applied [29], while early versions of CELESTE3D were based on a time-staggered solution of the two divergence Maxwell's equations and of a second order formulation of the two curl Maxwell's equations [34]. More recently [34] a consistent second order formulation has been presented and is summarized here.

In the continuum, the two curl Maxwell's equations give a solution with the property of satisfying also the two divergence equations at all times, provided that appropriate initial and boundary conditions are used. Based on this property [34], the following second order equation is obtained from the two time-discretised (but continuous in space) curl equations in Eq. (58):

$$(c\theta\Delta t)^2[-\nabla^2\mathbf{E}^{n+\theta} - \nabla\nabla\cdot(\boldsymbol{\mu}^n\cdot\mathbf{E}^{n+\theta})] + \boldsymbol{\epsilon}^n\cdot\mathbf{E}^{n+\theta} = \mathbf{E}^n + (c\theta\Delta t)\left(\nabla\times\mathbf{B}^n - \frac{4\pi}{c}\widehat{\mathbf{J}}^n\right) - (c\theta\Delta t)^2\nabla 4\pi\hat{\rho}^n \tag{74}$$

where $\boldsymbol{\mu}^n = \sum_s \boldsymbol{\mu}_s^n$ and $\boldsymbol{\epsilon}^n = \mathbf{I} + \boldsymbol{\mu}^n$, being $\mathbf{I}$ the identity tensor.

As shown in Ref. [34], the second order formulation for the electric field needs to be coupled with a divergence cleaning step to ensure that

$$\epsilon_0\nabla\cdot\mathbf{E}^n = \rho^n \tag{75}$$

holds for the initial field of each time step.

The boundary conditions for the second order formulation for $\mathbf{E}$ can be derived from the natural boundary conditions expressed in both $\mathbf{E}$ and $\mathbf{B}$ using theorems of classical electrodynamics [34]. The magnetic field is computed directly once the new electric field is computed, as:

$$\frac{\mathbf{B}^{n+1}-\mathbf{B}^n}{\Delta t} + \nabla\times\mathbf{E}^{n+\theta} = 0 \tag{76}$$

Using the temporal second-order temporal formulation above, the field equations are further discretised in space (see e.g. Ref. [34,58] for details). The discretised equations and their boundary conditions form a non-symmetric linear system that is solved using GMRES [54]. For the divergence cleaning equation, CG (or FFT on a uniform grids) [54] can be used since the discretised equation leads to a symmetric matrix.

9.2. Particle mover

Once the fields are computed from the semi-implicit moment method, the particles can be advanced. The set of equations of motions (54) have an inner coupling among themselves due to the appearance of the new velocity in the equation for the position and of the new position in the equation for the velocity. This coupling is purely local at the level of each particle and does not involve coupling with other particles. Instead of simply solving the 6 coupled equations of motions considering position and velocity as unknowns, the formal linearity of the explicit quadrature of the velocity provided in Eq. (55) can be used. Substituting the first of the Newton Eqs. (54) in the second, the nonlinear (vectorial) function expressing the discretised Newton equations is:

$$\mathbf{f}\left(\mathbf{x}_p^{n+1}\right) = \mathbf{x}_p^{n+1} - \mathbf{x}_p^n - \Delta t \hat{\mathbf{v}}_p - \Delta t \beta_s \hat{\mathbf{E}}_p^{n+\theta}\left(\mathbf{x}_p^{n+1/2}\right) \tag{77}$$

The three residual equations above have only three unknowns, the components of $\mathbf{x}_p^{n+1}$, all other quantities being known. The equations are non-linear due to the spatial dependence of the electric and magnetic fields.

Two methods have been used in the literature. First, a predictor–corrector (PC) approach can be used to solve the non-linear system of equations of motions [44]. Alternatively, less commonly, the Newton method [53] can be used to solve the set of equations of motion to convergence. While the JFNK method can be used for this task also, the simplicity of the 3 coupled equations allows the use of the full-fledged Newton method.

The use of the Newton-based mover ensure solution of the equations of motion to convergence of the non-linear iteration, when instead the PC mover uses typically a fixed number of iterations allowing for an uncontrolled error. This feature endows the Newton-based mover with an improved energy conservation in the overall scheme [58]. Both movers can be extended to the relativistic case [59].

9.3. Stability of the semi-implicit methods

The stability properties of the semi-implicit method described above have been studied extensively in the past [42]. The semi-implicit particle mover removes the need to resolve the electron plasma frequency, and the implicit formulation of the field equations removes the need to resolve the speed of light.

The time step constraints are replaced by an accuracy limit arising from the derivation of the fluid moment equations using the series expansion in Eq. (70). This limit restricts the mean particle motion to one grid cell per time step [42], i.e.

$$v_{th,e}\Delta t/\Delta x < 1, \tag{78}$$

The finite grid instability limit for the explicit method, $\Delta x < \varsigma\lambda_{De}$, is replaced by [42]

$$\Delta x < \varepsilon^{-1} v_{th,e}\Delta t, \tag{79}$$

that allows large cell sizes to be used when large time steps are taken.

This is a key feature of the semi-implicit method: time and space scales can be chosen freely according to the desired accuracy, but the ratio of spatial and length scales is not free. Rather it must stay within the bounds just outlined:

$$\varepsilon < v_{th,e}\Delta t/\Delta x < 1, \tag{80}$$

The upper limit is generally the main concern. The lower limit, ε, is in practical cases usually a very small number that can often be approximated by zero. The result is that the conditions can be satisfied by selecting large grid spacings and large time steps. In principle, the lower limit would prevent the choice of using large grid spacings with small time step, but in practice this limitation is weaker and often the time step can be chosen rather smaller than predicted by Eq. (80). However, the condition $v_{th,\,e}\Delta t/\Delta x < 1$ should never be violated.

The gain afforded by the relaxation of the stability limits is twofold.

First, the time step can far exceed the explicit limit. In typical space weather problems, the electron plasma frequency is far smaller than the time scales of interest and its accurate resolution is not needed. The processes developing at the sub-Δt scale are averaged and their energy is damped by a numerically-enhanced Landau damping. In other approaches, such as the gyrokinetic or hybrid approach [4], such processes are completely removed and the energy channel towards them is interrupted, removing for example the possibility to exchange energy between sub-Δt fluctuations and particles. In the semi-implicit approach, instead, the sub-Δt scales remain active and the energy channel remains open. This is a crucial feature to retain in a full kinetic approach. Furthermore, when additional resolution of the smallest scales is needed, the semi-implicit method can access the same accuracy of the explicit method simply using a smaller time step and grid spacing. This feature is not accessible to reduced model, e.g. gyroaveraged methods, that remove the small scales entirely.

Second, the cell sizes can far exceed the Debye length. Often the scales of interest are much larger than the Debye length. The ability to retain a full kinetic treatment without the need to resolve the Debye length results in a much reduced cost for the semi-implicit PIC method.

10. Adaptive gridding

As noted above the particle in cell method is essentially based on the assumption that the particles have a finite size, and that finite size is fixed and equal to the cell size. The approach is therefore naturally suited to uniform grids. The extension to non-uniform, adaptive grids is not straightforward. Nevertheless there are noteworthy attempts to expand the PIC method to non-uniform grids. We review below some of the most notable approaches of interest to space weather problems. As a matter of order, we organize the topic here into three groups.

First, we consider methods where the grid is non-uniform but still retain a regular structure for the connectivity that allow one to map the grid to a uniform logical grid. This is the case when all the nodes have equal connectivity, each node being in contact with 2 others in 1D, 4 in 2D and 6 in 3D. An example of such an arrangement is presented in 2D in Fig. 12. This approach can be defined as moving mesh adaptation (MMA) as it can be imagined as the end result of a process

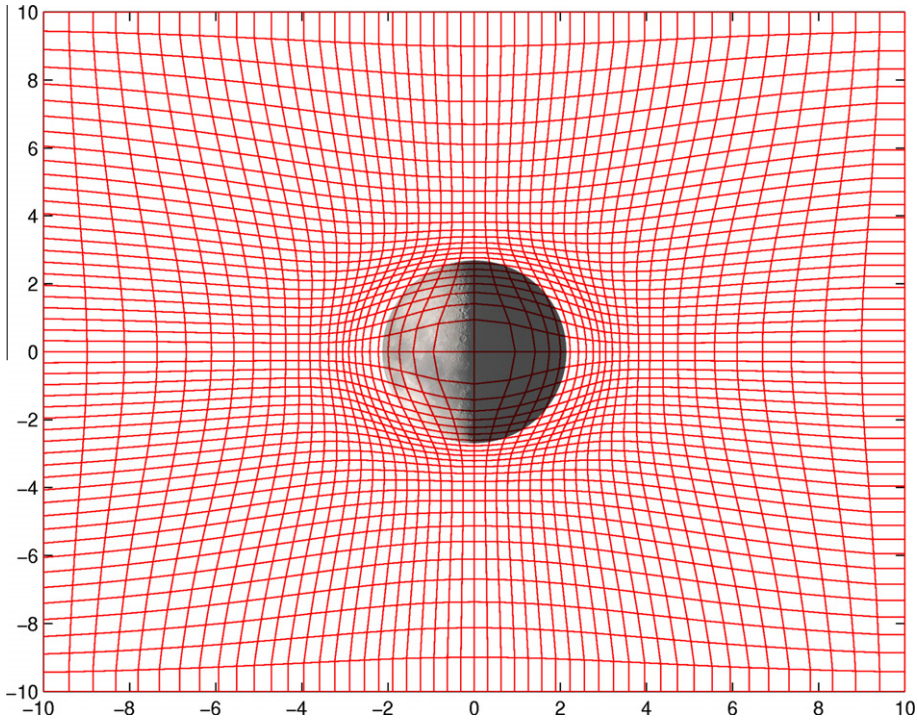

Fig. 12. Conceptual representation of moving grid adaptation. The grid in the physical domain **x** can be mapped to a uniform grid in the logical coordinates ξ. Example from a model of the Moon interaction with the solar wind.

that distorts the grid by attracting more points in the regions of interest and away from regions where the system is smooth an uneventful. This approach has been followed for example within the Celeste implicit moment method and is reviewed in [61–63].

Second, adaptive mesh refinement (AMR) [64] can be used for the field part of the PIC method. Main examples of this approach are [65,30]. The latter approach was developed especially for space weather application and successful investigations have been reported demonstrating the ability to resolve with the AMR approach the localized regions of reconnection embedded in larger systems [30].

Lastly, unstructured grids can be used [24] starting from finite element methods to discretise the fields.

All these approaches need to deal with the fact that as the grid spacing changes locally, the particles and the cells cannot always have the same size. As noted above the derivation presented is based on such an equality to provide a simple efficient interpolation between particles and cells. The first two approaches have relaxed the requirement that the particle size remain constant and have based the interpolation on the local grid spacing. The particles have the local size of the cells they are embedded in. For the first approach this is achieved by assuming that the particle shape function is constant in the logical space ξ rather than in the physical space **x**. If the shape functions are chosen in the logical space, the interpolation functions can still be computed with the b-spline chain rule and the same formulas of uniform grids can still be used. In the second approach, typically a patch-based AMR approach is used and the particles have the size of the cells in the patch where they reside and interpolation can proceed in each patch as if it was a uniform grid.

This simplification of the interpolation step on adaptive PIC has a serious consequence: the exact derivation breaks down, the changing particle shape introduces new terms proportional to the temporal derivative of the shape function. The derivation of the PIC method outlined in Section 3 assumed a fixed particle shape. All the integrations have been done using such fixed shapes. When this process is followed, a simple analysis proves that momentum is conserved exactly when the same interpolation function is used for all quantities [20,26] and provided that the solution of the Maxwell's equations also conserve momentum (i.e. the numerical Green's function reflects the properties of anti-symmetry of the exact continuum Green's function).

When the shape functions change in time the mathematical derivation of the PIC method is modified. In particular, the derivation of the equations of motion require additional terms related to the temporal variation of the shape function. However, in practice these additional terms are neglected and the resulting schemes do not conserve momentum. A side effect is that particles experience a self-force, due to their traversing regions of different sizes [66,67]. For the moving mesh approach, this error is controlled by using smooth grids where the spacing changes progressively. Being the neglected terms proportional to the changes in the shape function, when the changes are small the error is small. In the AMR approach, however, the change in particle shape is sudden at the interface between patches and leads to severe errors that need to be corrected with appropriately devised methods [67,66].

11. Practical use of particle simulation in space weather

We conclude the present review with a few practical considerations regarding the application of particle in cell methods to space weather problems. The focus will be on the use of implicit particle simulation as it has the better promise of being able to actually model space weather events at the kinetic level. Conversely, the explicit approach is more suitable to zero in on some detailed physics, a topic of interest to space weather but better described as fundamental space science and beyond the scope of the present review.

We report first on the latest developments on the implementation of the implicit methods in codes focusing especially on the latest results relative to iPIC3D, a massively parallel implicit moment PIC code [12]. Next, we consider the issue of computational costs in the particle simulation of space weather problems with multiple-scale coupling.

11.1. Scaling of massively parallel implicit PIC

Particle modeling has led to the development of many codes. For explicit modeling, the simplicity of the algorithm and the relative simplicity of its parallelization has led to the development of many codes based on significantly different implementations with different particle movers and field discretisation. Implicit codes are much fewer. Essentially there are only two families of codes. The direct implicit method led to the development of the code AVANTI [68], later developed into a commercial code, LSP aimed primarily at inertial fusion problems [69]. The implicit moment method has been applied primarily in space problems. Even the names are space-oriented: First came VENUS [42], followed by CELESTE [29,58] whose main improvement over VENUS is the use of modern preconditioned Krylov solvers for the field equations. CELESTE3D is available open source [70]. Parsek2D [71] and iPIC3D [12] have brought the implicit moment method to massively parallel computers.

Fig. 13 shows a scaling study conducted on NASA's Pleiades using up to 16,384 processors. The study uses one realistic problem, the study of reconnection in a 2D simulation box, using 32×32 cells per processor, each with 4 species having 196 particles per cell. The scaling is done by increasing the number of processors at a constant load per processor, thereby increasing the system size, all else (time step and number of cycles) remaining the same.

As can be noted, the measured scaling is only 20% lower than the ideal scaling when going from 32 to 16,384 processors. The details of the parallel implementation allowing such performances has been reported in [12].

11.2. Computational costs of particle modeling in space weather

The key difference between explicit and implicit methods for their application to space weather is the different selection criteria for grid spacing and time step. The explicit method must satisfy the stability constraints while the implicit method need only to resolve the scales of interest. An example is the best illustration of the concept.

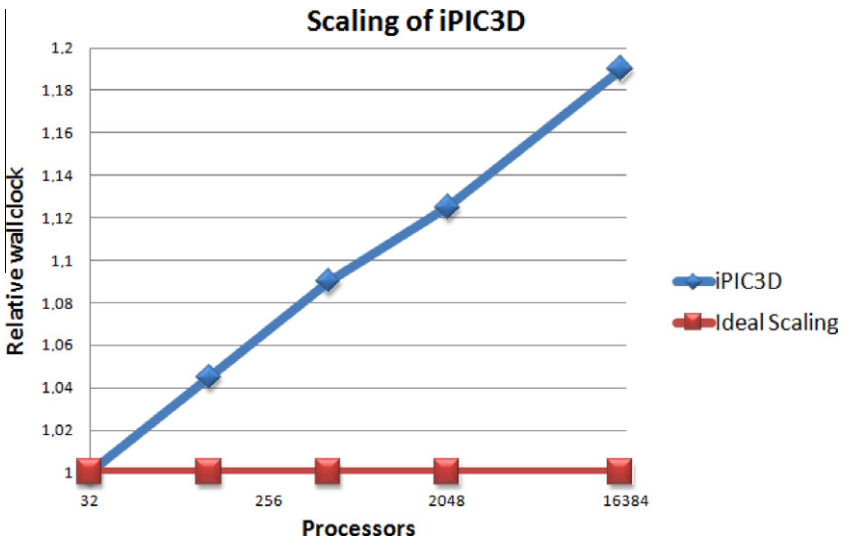

Fig. 13. Weak scaling study conducted on NASA's Pleiades supercomputer using the implicit moment code iPIC3D. In red, the ideal scaling is shown where the total CPU time remains constant as the number of processors and the system size are increased keeping the load per processor constant. In blue, the actual observed scaling is shown. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Paradigmatic choice of parameters for a magnetospheric simulation with explicit and implicit PIC. Spatial and temporal scales are shown in each row. The last column shows the gain in time step and grid spacing for each dimension of space–time. The last row summarizes the total gain.

	Explicit	Implicit	Gain
Δx	$\lambda_{De} = 250$ m	$d_e = 8$ km	32
Δy	$\lambda_{De} = 250$ m	$d_e = 8$ km	32
Δz	$\lambda_{De} = 250$ m	$d_e = 8$ km	32
Δt	$\Delta t = 0.1\omega_{pe}^{-1} = 2.5 \cdot 10^{-6}$ s	$\Delta t = 0.1\omega_{pi}^{-1} = 10^{-4}$ s	$\sqrt{1836}$
Total			$1.4 \cdot 10^6$

Considering again the case of the Earth magnetic tail whose scales are illustrated in Fig. 1, Table 1 shows the typical choices forced upon the explicit method by the stability constraints. The grid spacing must be chosen equal to the smallest Debye length in the system and the time step must resolve the electron plasma frequency.

The implicit method, instead, can select the accuracy desired and Table 1 shows some typical choices. The grid spacing is selected to resolve the scales of reconnection. The smallest scale of interest is the electron skin depth that corresponds to the dissipation layer where field lines decouple from the electrons, break and reconnect. The choice is then $\Delta x/d_e = 1$ (justified by the consideration that the reconnection region has a much larger electron skin depth then the reference value, see Ref. [15] for details), that corresponds to $\Delta x/\lambda_{De} = 32$. This gain increases by another order of magnitude in other colder regions, such as in the solar wind. The time step is constrained by $\Delta t < \Delta x/v_{the}$ or $\omega_{pe}\Delta t < \Delta x/\lambda_{De} = 100$. To resolve time well, a significantly smaller time step is typically used: $\omega_{pi}\Delta t = 0.1$. When those choices are made, a gain of order $1.4 \cdot 10^6$ is obtained in CPU time. For reference, if the same simulation is done on the same computer with the same number of processors, for an implicit run that took one day, the corresponding explicit run would require approximately 4000 years. The gain can be partially offset by the higher cost per computational time. Typically semi-implicit codes require about three times the cost per cycle. In that case, still a thousand years would be required for the corresponding explicit run. The gain achieved in lower temperature regions, common for example in the solar wind, are even more staggering, by several more orders of magnitude. Given these numbers, it is clear that the range of application of implicit PIC is vastly larger than explicit PIC.

The tremendous gain comes at the cost of loosing the information on the scales not resolved. Those scales are not eliminated but rather the waves developing there are distorted to the Nyquist frequency and damped. This leaves the channel open and allows the dissipations to develop. This is exactly what is needed for space weather applications. Not the detail of the most detailed scale but the ability to resolve only the sales of the most important processes.

For the demonstration of the application of the model above to space weather, we refer the reader to the published work done with Celeste3D [72–82,58,83,84,59,85–92] and iPIC3D [12,15] at the scales of interest outlined in Table 1. In particular, detailed comparisons of the application of extremely refined explicit simulations and coarser implicit simulations have proven the ability of the latter to resolve correctly the relevant dissipation scales in processes of reconnection [78] and in other space physics problems.

Acknowledgments

The author is grateful to Gianni Coppa for the fruitful discussions on the mathematical foundations of the PIC method. The author is also grateful to Jerry Brackbill for the many years of collaboration and stimulating exchanges of ideas on the implicit PIC method. Stefano Markidis is gratefully acknowledged for his key role as developer of the code iPIC3D. The present work is supported by the Onderzoekfonds KU Leuven (Research Fund KU Leuven) and by the European Commission's Seventh Framework Programme (FP7/2007-2013) under the Grant Agreement No. 218816 (SOTERIA project, www.soteria-space.eu) and No. 263340 (SWIFF project, www.swiff.eu). Simulations were conducted on the resources of the NASA Advanced Supercomputing Division (NAS), of the NASA Center for Computational Sciences Division (NCCS) and of the Vlaams Supercomputer Centrum (VSC) at the Katholieke Universiteit Leuven. Simulation results have been provided by the Community Coordinated Modeling Center at Goddard Space Flight Center through their public Runs on Request system (http://ccmc.gsfc.nasa.gov). The CCMC is a multi-agency partnership between NASA, AFMC, AFOSR, AFRL, AFWA, NOAA, NSF and ONR. The ENLIL Model was developed by D. Odstrcil at the University of Colorado at Boulder. The BATSRUS with Rice Convection Model was developed by Tamas Gombosi et al. [11] at the Center for Space Environment Modeling, University of Michigan.

References

[1] http://www.ofcm.gov/nswp-sp/text/a-cover.htm.
[2] V. Bothmer, I.A. Daglis, Space Weather – Physics and Effects, Praxis Publishing, 2007.
[3] M. Kallenrode, Space physics: an introduction to plasmas and particles in the heliosphere and magnetospheres, 2004.
[4] A.S.S. Lipatov, The Hybrid Multiscale Simulation Technology, Springer, Berlin, 2002.
[5] M.J. Aschwanden, Physics of the Solar Corona, An Introduction with Problems and Solutions, second ed., Praxis Publishing Ltd, Chichester, UK, 2005.
[6] Cost 724 final report. <http://www.costes0803.noa.gr/documents/cost-docs>.
[7] Community coordinated modeling center (CCMC). <http://www.ccmc.gsfc.nasa.gov/>.
[8] Solar terrestrial investigations and archives (SOTERIA). <http://www.soteria-space.eu>.

[9] T. Falkenberg et al., The catalogue of selected events, Soteria Report D4.1, 2008, pp. 1–115. <http://www.soteria-space.eu/doc/reports/SOTERIA_D4_1.pdf>.
[10] H. Xie, L. Ofman, G. Lawrence, Cone model for halo CMEs: application to space weather forecasting, Journal of Geophysical Research (Space Physics) 109 (2004) 3109, doi:10.1029/2003JA010226.
[11] G. Tóth, I.V. Sokolov, T.I. Gombosi, D.R. Chesney, C.R. Clauer, D.L. De Zeeuw, K.C. Hansen, K.J. Kane, W.B. Manchester, R.C. Oehmke, K.G. Powell, A.J. Ridley, I.I. Roussev, Q.F. Stout, O. Volberg, R.A. Wolf, S. Sazykin, A. Chan, B. Yu, J. Kóta, Space weather modeling framework: A new tool for the space science community, Journal of Geophysical Research (Space Physics) 110 (2005) 12226, doi:10.1029/2005JA011126.
[12] S. Markidis, G. Lapenta, Rzwan-udding, Multi-scale simulations of plasma with ipic3d, Mathematics and Computers in Simulation 80 (7) (2010) 1509–1519, doi:10.1016/j.matcom.2009.08.038.
[13] D. Biskamp, Magnetic Reconnection, Cambridge University Press, 2000.
[14] E. Priest, T. Forbes, Magnetic Reconnection: MHD Theory and Applications, Cambridge University Press, 1999.
[15] G. Lapenta, S. Markidis, A. Divin, M. Goldman, D. Newman, Scales of guide field reconnection at the hydrogen mass ratio, Physics of Plasmas 17 (8) (2010) 082106, doi:10.1063/1.3467503.
[16] J.F. Drake, M. Swisdak, C. Cattell, M.A. Shay, B.N. Rogers, A. Zeiler, Inherently three dimensional magnetic reconnection: a mechanism for bursty bulk flows?, Science 299 (2003) 873
[17] C. Cattell, J. Dombeck, J. Wygant, J.F. Drake, M. Swisdak, M.L. Goldstein, W. Keith, A. Fazakerley, M. André, E. Lucek, A. Balogh, Cluster observations of electron holes in association with magnetotail reconnection and comparison to simulations, Journal of Geophysical Research (Space Physics) 110 (A9) (2005) 1211, doi:10.1029/2004JA010519.
[18] M.V. Goldman, D.L. Newman, P. Pritchett, Vlasov simulations of electron holes driven by particle distributions from PIC reconnection simulations with a guide field, grl 35 (2008) 22109, doi:10.1029/2008GL035608.
[19] M.I. Sitnov, M. Swisdak, A.V. Divin, Dipolarization fronts as a signature of transient reconnection in the magnetotail, Journal of Geophysical Research (Space Physics) 114 (A13) (2009) 4202, doi:10.1029/2008JA013980.
[20] R. Hockney, J. Eastwood, Computer Simulation using Particles, Taylor & Francis, 1988.
[21] C. Birdsall, A. Langdon, Plasma Physics via Computer Simulation, Taylor & Francis, London, 2004.
[22] G.G.M. Coppa, G. Lapenta, G. Dellapiana, F. Donato, V. Riccardo, Blob method for kinetic plasma simulation with variable-size particles, Journal of Computational Physics 127 (2) (1996) 268–284, doi:10.1006/jcph.1996.0174.
[23] W.B. Bateson, D.W. Hewett, Grid and particle hydrodynamics: beyond hydrodynamics via fluid element particle-in-cell, Journal of Computational Physics 144 (2) (1998) 358–378, doi:10.1006/jcph.1997.5824.
[24] G.B. sJacobs, J.S. Hesthaven, High-order nodal discontinuous Galerkin particle-in-cell method on unstructured grids, Journal of Computational Physics 214 (2006) 96–121, doi:10.1016/j.jcp.2005.09.008.
[25] K. Morton, D. Mayers, Iterative methods for linear and nonlinear equations, SIAM, Philadelphia, 1995.
[26] Y.N. Grigoryev, V.A. Vshivkov, M.P. Fedoruk, Numerical particle-in-cell methods: theory and applications, VSP BV, AH Zeist, 2005.
[27] D. Sulsky, J.U. Brackbill, A numerical method for suspension flow, Journal of Computational Physics 96 (1991) 339.
[28] A. Taflove, S.C. Hagness, Computational Electrodynamics: The Finite-Difference Time-Domain Method, Artech House Publishers, Norwood, 2005.
[29] H.X. Vu, J.U. Brackbill, Celest1d: an implicit, fully-kinetic model for low-frequency, electromagnetic plasma simulation, Computer Physics Communications 69 (1992) 253.
[30] K. Fujimoto, R.D. Sydora, Electromagnetic particle-in-cell simulations on magnetic reconnection with adaptive mesh refinement, Computer Physics Communications 178 (2008) 915–923, doi:10.1016/j.cpc.2008.02.010.
[31] A.H. Mohammadian, V. Shankar, W.F. Hall, Computation of electromagnetic scattering and radiation using a time-domain finite-volume discretization procedure, Computer Physics Communications 68 (1991) 175–196, doi:10.1016/0010-4655(91)90199-U.
[32] F. Assous, P. Degond, E. Heintze, P.A. Raviart, J. Segre, On a finite-element method for solving the three-dimensional Maxwell equations, Journal of Computational Physics 109 (1993) 222–237, doi:10.1006/jcph.1993.1214.
[33] R.L. Morse, C.W. Nielson, Numerical simulation of the Weibel instability in one and two dimensions, Physics of Fluids 14 (1971) 830–840, doi:10.1063/1.1693518.
[34] P. Ricci, G. Lapenta, J. Brackbill, A simplified implicit Maxwell solver, Journal of Computational Physics 183 (2002) 117–141.
[35] G. Tóth, The $\nabla \cdot B = 0$ constraint in shock-capturing magnetohydrodynamics codes, Journal of Computational Physics 161 (2) (2000) 605–652. http://dx.doi.org/10.1006/jcph.2000.6519 .
[36] A. Dedner, F. Kemm, D. Kröner, C. Munz, T. Schnitzer, M. Wesenberg, Hyperbolic divergence cleaning for the MHD equations, Journal of Computational Physics 175 (2002) 645–673, doi:10.1006/jcph.2001.6961.
[37] K.J. Bowers, B.J. Albright, L. Yin, W. Daughton, V. Roytershteyn, B. Bergen, T.J.T. Kwan, Advances in petascale kinetic plasma simulation with VPIC and Roadrunner, Journal of Physics Conference Series 180 (1) (2009) 012055, doi:10.1088/1742-6596/180/1/012055.
[38] A.B. Langdon, On enforcing Gauss' law in electromagnetic particle-in-cell codes, Computer Physics Communications 70 (1992) 447–450, doi:10.1016/0010-4655(92)90105-8.
[39] C. Munz, P. Omnes, R. Schneider, E. Sonnendrücker, U. Voß, Divergence correction techniques for Maxwell solvers based on a hyperbolic model, Journal of Computational Physics 161 (2000) 484–511, doi:10.1006/jcph.2000.6507.
[40] R. Mason, Implicit moment particle simulation of plasmas, Journal of Computational Physics 41 (1981) 233.
[41] J. Denavit, Time-filtering particle stimulations with $\omega\_pe\Delta t \gg 1$, Journal of Computational Physics 42 (1981) 337.
[42] J. Brackbill, D. Forslund, Simulation of low frequency, electromagnetic phenomena in plasmas, Journal of Computational Physics 46 (1982) 271.
[43] A. Langdon, B. Cohen, A. Friedman, Direct implicit large time-step particle simulation of plasmas, Journal of Computational Physics 51 (1983) 107–138.
[44] H. Vu, J. Brackbill, Accurate numerical solution of charged particle motion in a magnetic field, Journal of Computational Physics 116 (1995) 384.
[45] B.I. Cohen, A.B. Langdon, D.W. Hewett, R.J. Procassini, Performance and optimization of direct implicit particle simulation, Journal of Computational Physics 81 (1989) 151, doi:10.1016/0021-9991(89)90068-5.
[46] A. Friedman, A second-order implicit particle mover with adjustable damping, Journal of Computational Physics 90 (1990) 292, doi:10.1016/0021-9991(90)90168-Z.
[47] S. Markidis, G. Lapenta, Coupling global and kinetic scales with the implicit particle-in-cell methods, in: Presented at 2010 Fall Meeting, AGU, San Francisco, Calif., 13–17 December.
[48] G. Lapenta, S. Markidis, Particle acceleration and energy conservation in particle in cell simulations. Physical Review Letters, submitted for publication.
[49] S. Markidis, G. Lapenta, The energy conserving particle-in-cell method, Journal of Computational Physics, submitted for publication.
[50] J.U. Brackbill, B.I. Cohen (Eds.), Multiple Time ScalesJ.U. Brackbill, D.W. Forslund (Eds.), Simulation of Low-Frequency, Electromagnetic Phenomena in Plasmas, Academic Press, Orlando, 1985, p. 271.
[51] H. Kim, L. Chacòn, G. Lapenta, Fully implicit particle-in-cell algorithm, Bulletin of the American Physical Society 50 (2005) 2913.
[52] S. Markidis, Development of Implicit Kinetic Simulation Methods, and Their Application to Ion Beam Propagation in Current and Future Neutralized Drift Compression Experiments, Ph.D. Thesis, University of Illinois at Urbana-Champaign, 2010.
[53] C.T. Kelley, Iterative Methods for Linear and Nonlinear Equations, SIAM, Philadelphia, 1995.
[54] Y. Saad, Iteractive Methods for Sparse Linear Systems, SIAM, Philadelphia, 2003.
[55] S. Balay, K. Buschelman, W.D. Gropp, D. Kaushik, M.G. Knepley, L.C. McInnes, B.F. Smith, H. Zhang, PETSc Web page, 2009. <http://www.mcs.anl.gov/petsc>.
[56] S. Balay, K. Buschelman, V. Eijkhout, W.D. Gropp, D. Kaushik, M.G. Knepley, L.C. McInnes, B.F. Smith, H. Zhang, PETSc users manual, Tech. Rep. ANL-95/11 – Revision 3.0.0, Argonne National Laboratory, 2008.

[57] S. Balay, W.D. Gropp, L.C. McInnes, B.F. Smith, Efficient management of parallelism in object oriented numerical software libraries, in: E. Arge, A.M. Bruaset, H.P. Langtangen (Eds.), Modern Software Tools in Scientific Computing, Birkhäuser Press, 1997, pp. 163–202.
[58] G. Lapenta, J.U. Brackbill, P. Ricci, Kinetic approach to microscopic–macroscopic coupling in space and laboratory plasmas, Physics of Plasmas 13 (2006) 055904.
[59] K. Noguchi, C. Tronci, G. Zuccaro, G. Lapenta, Formulation of the relativistic moment implicit particle-in-cell method, Physics of Plasmas 14 (4) (2007) 042308, doi:10.1063/1.2721083.
[60] M. Drouin, L. Gremillet, J.-C. Adam, A. Héron, Particle-in-cell modeling of relativistic laser-plasma interaction with the adjustable-damping, direct implicit method, Journal of Computational Physics 229 (12) (2010) 4781–4812, doi:10.1016/j.jcp.2010.03.015.
[61] J.U. Brackbill, An adaptive grid with directional control, Journal of Computational Physics 108 (1993) 38.
[62] J. Brackbill, An adaptive grid with directional control for toroidal plasma simulation, in: Proceedings of the 14th International Conference on the Numerical Simulation of Plasmas, American Physical Society, Annapolis MD, 1991.
[63] G. Lapenta, Adaptive Multi-Dimensional Particle In Cell. <arXiv:0806.0830>.
[64] M.J. Berger, J. Oliger, The AMR technique, Journal of Computational Physics 53 (1984) 484–512.
[65] J. Vay, P. Colella, A. Friedman, D.P. Grote, P. McCorquodale, D.B. Serafini, Implementations of mesh refinement schemes for particle-in-cell plasma simulations, Computer Physics Communications 164 (2004) 297–305, doi:10.1016/j.cpc.2004.06.075.
[66] J. Vay, P. Colella, P. McCorquodale, B. van Straalen, A. Friedman, D.P. Grote, Mesh refinement for particle-in-cell plasma simulations: applications to and benefits for heavy ion fusion, Laser and Particle Beams 20 (2002) 569–575, doi:10.1017/S0263034602204139.
[67] P. Colella, P.C. Norgaard, Controlling self-force errors at refinement boundaries for AMR-PIC, Journal of Computational Physics 229 (2010) 947–957, doi:10.1016/j.jcp.2009.07.004.
[68] D.W. Hewett, A.B. Langdon, Electromagnetic direct implicit plasma simulation, Journal of Computational Physics 72 (1987) 121–155, doi:10.1016/0021-9991(87)90075-1.
[69] D.R. Welch, D.V. Rose, B.V. Oliver, R.E. Clark, Simulation techniques for heavy ion fusion chamber transport, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 464 (1–3) (2001) 134–139, doi:10.1016/S0168-9002(01)00024-9.
[70] http://www.code.google.com/p/celeste/.
[71] S. Markidis, E. Camporeale, D. Burgess, Rizwan-Uddin, G. Lapenta, Parsek2D: an implicit parallel particle-in-cell code, in: N.V. Pogorelov, E. Audit, P. Colella, G.P. Zank (Eds.), Astronomical Society of the Pacific Conference Series, Astronomical Society of the Pacific Conference Series, vol. 406, 2009, p. 237.
[72] G. Lapenta, J. Brackbill, Contact discontinuities in collisionless plasmas: a comparison of hybrid and kinetic simulations, Geophysical Research Letters 23 (1996) 1713.
[73] G. Lapenta, J.U. Brackbill, A kinetic theory for the drift-kink instability, Journal of Geophysical Research (Space Physics) 102 (1997) 27099–27108, doi:10.1029/97JA02140.
[74] G. Lapenta, J. Brackbill, 3d reconnection due to oblique modes: a simulation of harris current sheets, Nonlinear Processes Geophysics 7 (2000) 151.
[75] G. Lapenta, J. Brackbill, Nonlinear evolution of the lower hybrid drift instability: current sheet thinning and kinking, Physics of Plasmas 9 (5) (2002) 1544–1554.
[76] G. Lapenta, J.U. Brackbill, W.S. Daughton, The unexpected role of the lower hybrid drift instability in magnetic reconnection in three dimensions, Physics of Plasmas 10 (2003) 1577–1587, doi:10.1063/1.1560615.
[77] P. Ricci, G. Lapenta, J. Brackbill, Gem challenge: Implicit kinetic simulations with the physical mass ratio, Geophysical Research Letters 29 (2002), doi:10.1029/2002GL015314.
[78] P. Ricci, J. Brackbill, W. Daughton, G. Lapenta, Collisionless magnetic reconnection in the presence of a guide field, Physics of Plasmas 11 (8) (2004) 4102–4114.
[79] P. Ricci, G. Lapenta, J. Brackbill, Structure of the magnetotail current: Kinetic simulation and comparison with satellite observations, Geophysical Research Letters 31 (2004) L06801, doi:10.1029/2003GL019207.
[80] P. Ricci, J.U. Brackbill, W. Daughton, G. Lapenta, Influence of the lower hybrid drift instability on the onset of magnetic reconnection, Physics of Plasmas 11 (2004) 4489–4500, doi:10.1063/1.1778744. <arXiv:physics/0403107> .
[81] G. Lapenta, A new paradigm for 3D collisionless magnetic reconnection, Space Science Reviews 107 (2003) 167–174, doi:10.1023/A:1025579923977.
[82] P. Ricci, G. Lapenta, J. Brackbill, Electron acceleration and heating in collisioneless magnetic reconnection, Physics of Plasmas 10 (9) (2003) 3554–3560.
[83] J. Birn, K. Galsgaard, M. Hesse, M. Hoshino, J. Huba, G. Lapenta, P.L. Pritchett, K. Schindler, L. Yin, J. Buchner, T. Neukirch, E.R. Priest, Forced magnetic reconnection, Geophysical Research Letters 32 (2005) L06105.
[84] E. Camporeale, G. Lapenta, Model of bifurcated current sheets in the Earth's magnetotail: Equilibrium and stability, Journal of Geophysical Research (Space Physics) 110 (A9) (2005) 7206, doi:10.1029/2004JA010779.
[85] M.E. Innocenti, G. Lapenta, Momentum creation by drift instabilities in space and laboratory plasmas, Plasma Physics and Controlled Fusion 49 (2007) 521, doi:10.1088/0741-3335/49/12B/S50.
[86] G. Lapenta, D. Krauss-Varban, H. Karimabadi, J.D. Huba, L.I. Rudakov, P. Ricci, Kinetic simulations of x-line expansion in 3D reconnection, Geophysical Research Letters 33 (2006) 10102, doi:10.1029/2005GL025124.
[87] G. Lapenta, J. King, Study of current intensification by compression in the Earth magnetotail, Journal of Geophysical Research (Space Physics) 112 (A11) (2007) 12204, doi:10.1029/2007JA012527.
[88] W. Wan, G. Lapenta, Micro–macro coupling in plasma self-organization processes during island coalescence, Physical Review Letters 100 (3) (2008) 035004, doi:10.1103/PhysRevLett.100.035004.
[89] W. Wan, G. Lapenta, Electron self-reinforcing process of magnetic reconnection, Physical Review Letters 101 (1) (2008) 015001, doi:10.1103/PhysRevLett.101.015001.
[90] W. Wan, G. Lapenta, Evolutions of non-steady-state magnetic reconnection, Physics of Plasmas 15 (10) (2008) 102302, doi:10.1063/1.2991406.
[91] W. Wan, G. Lapenta, G.L. Delzanno, J. Egedal, Electron acceleration during guide field magnetic reconnection, Physics of Plasmas 15 (3) (2008) 032903, doi:10.1063/1.2876465.
[92] G. Lapenta, Large-scale momentum exchange by microinstabilities: a process happening in laboratory and space plasmas, Physica Scripta 80 (3) (2009) 035507, doi:10.1088/0031-8949/80/03/035507.