Kinetic theory of particle-in-cell simulation plasma and the ensemble averaging technique

Michaël Touati\textsuperscript{1,*}, Romain Codur\textsuperscript{2}, Frank Tsung\textsuperscript{2}, Viktor K Decyk\textsuperscript{2}, Warren B Mori\textsuperscript{2} and Luis O Silva\textsuperscript{3}

\textsuperscript{1}Marvel Fusion Gmbh, Theresienhöhe 12, Munich, Germany
\textsuperscript{2}Department of Physics and Astronomy, University of California Los Angeles, Los Angeles, CA, United States of America
\textsuperscript{3}GOLP/Instituto de Plasma e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

E-mail: michael.touati@marvelfusion.com

Received 25 May 2022, revised 15 August 2022
Accepted for publication 7 September 2022
Published 13 October 2022

Abstract

We derive the kinetic theory of fluctuations in physically and numerically stable particle-in-cell (PIC) simulations of electrostatic plasmas. The starting point is the single-time correlation at the start of the simulation between the statistical fluctuations of the weighted densities of macroparticle centers in the plasma particle phase-space. The fluctuations are associated with different initial conditions, typically due to the random initial conditions (in velocity space) of the macroparticles/simulation plasma, assigned according to their initial distribution of probability. The single-time correlations at all time steps and in each spatial grid cell are then determined from the Laplace–Fourier transforms of the discretized Klimontovich-like equation for the macroparticles and Maxwell’s equations for the fields, as computed by modern PIC codes. We recover the expressions for the electrostatic field and the plasma particle density fluctuation autocorrelation spectra as well as the kinetic equations describing the average evolution of PIC-simulated plasma particles, first derived by Langdon (1970b \textit{Proc. 4th Conf. Numerical Simulation of Plasmas}) using a test macroparticle approach perturbing a discretized Vlasovian plasma and then averaging the obtained physical quantity over the initial macroparticle velocity distribution. We generalize and extend these results to the modern algorithms in PIC codes using arbitrary macroparticle weights. Analytical estimates of statistical fluctuation amplitudes are derived as a function of the plasma simulation parameters, using the central limit theorem in the limit of a large number of macroparticles per cell. The theory is then used to analyze the ensemble averaging technique of PIC simulations where statistical averages are performed over ensembles of PIC simulations, modeling the same plasma physics problem but using different statistical realizations of the initial distribution functions of the macroparticles. This method is illustrated by linear Landau damping uncovering (from noise, * Author to whom any correspondence should be addressed.

Original Content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence. Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.
which is usually considered numerical) the physical fluctuations driven by a single small amplitude electrostatic wave perturbing a PIC simulation plasma in equilibrium.

Keywords: kinetic theory, electrostatic plasma, particle-in-cell codes, numerical noise, fluctuations correlations, ensemble averaging technique

(Some figures may appear in colour only in the online journal)

1. Introduction

A plasma is a set of charged particles consisting of electrons and ionized atoms that behave collectively through their self-consistent long-distance electromagnetic fields. In this paper, we consider (the, apparently, less complex) nondegenerate, nonrelativistic, and noncoupled collisionless electrostatic plasmas without external forces. One possible approach to describe the plasma considers all particle equations of motion under the action of the plasma electrostatic fields, determined according to the self-consistent Maxwell equations. This fully deterministic approach allows us to deduce the particle positions and velocities at all times (and thus the temporal evolution of the plasma) if the full microscopic details of the system are known at a given time. Differentiating the phase-space densities associated with this detailed description with respect to time, the Klimontovich equation for each plasma species (Klimontovich 1958) is obtained. The Klimontovich equation describes the evolution in time of all plasma particles in the 6D phase-space under the action of self-consistent microscopic electrostatic fields. The detailed knowledge of the positions and velocities of the particles is needed to evaluate the microscopic electrostatic fields, which renders it considerably difficult due to the challenge of computing simultaneously the particle dynamics and the microscopic self-consistent fields, even resorting to today’s computer technology. The particle-in-cell (PIC) method relaxes this constraint by considering simulation particles with a finite spatial size, also called macroparticles (Burger et al 1965, Hockney 1965, 1966, Yu et al 1965, Buneman 1967, Birdsall and Fuss 1969, Dawson 1983). For dimensionless simulations mapped to a real physical system (or for simulations that use real physical units), each macroparticle usually represents several real particles—the relation between real particles and their macroparticle representation depends on the compromise between the plasma density of the system under study and the available computer technology (for instance, restrictions on computer time and memory/represent the plasma particles).

In the PIC method, macroparticle positions are interpolated onto a spatial grid used to compute the self-consistent discretized Maxwell equations. Therefore, for simulations of a plasma homogeneously distributed in the whole simulation box, the PIC algorithm is much more efficient than the direct Klimontovich model previously outlined, with a computational cost $C \sim O(N)$ compared to the computation of all binary Coulomb interactions between macroparticles that would need a computational cost $C' \sim O(N^2)$. In terms of the physics being captured, the PIC method smooths the fields and thus underestimates close-encounter Coulomb collisions (Okuda and Birksall 1970, Langdon 1970a).

Moreover, from classical thermodynamic theory, one can argue that the exact positions and velocities of all particles are never known/measured simultaneously. Instead, and for large $N$-body systems, knowledge of the system can be obtained via quantities averaged over time, infinitesimally small mesoscopic spatial volumes, or over different measurements performed under the same conditions. This naturally motivates a probabilistic approach, based on the temporal evolution of the probability distribution functions. Liouville’s theorem describes the temporal evolution of the full $N$-body probability distribution function of all particle positions and velocity random variables (Liouville 1838). As before, the information contained in the $N$-body distribution function is too large to allow for the computation of the Liouville equation coupled with the self-consistent microscopic Maxwell equations, even with today’s computer technology. Simplications are also possible in this approach: smaller order $s$-body distribution functions can be obtained by integrating the Liouville equation over the position and velocity random variables of all other particles $s + 1, \ldots, N$. In the resulting equation, describing the temporal evolution of the $s$-body distribution functions in the $s$-D phase-space, the presence of the ‘collision’ integral also poses challenges, and some additional assumptions are needed to close the problem. The first order equation ($s = 1$) in the BBGKY hierarchy of equations (Yvon 1935, Bogoliubov 1946, Born and Green 1946, Kirkwood 1946) can be truncated to obtain the Vlasov equation by assuming an infinite number of electrons in a Debye sphere, thus neglecting two-body correlations between one-body distribution functions (Vlasov 1938). However, the number of particles is necessarily finite in plasma, and particles are always correlated due to their long-range electrostatic interactions. The Vlasov equation remains valid only on time scales much smaller than this minimum interaction time. On larger time scales, the Bogoliubov hypothesis (de Boer and Uhlenbeck 1962) can be used to simplify the second order equation in the BBGKY hierarchy and estimate the two-body correlations between plasma particles. The Bogoliubov hypothesis assumes the two-body correlations relax on a time scale smaller than the time scale on which the one-body distribution function relaxes. Further assumptions are still needed to get an analytical expression

\footnote{We observe that this term is used, even if, sometimes, the ‘collision’ integral does not strictly describe close-encounter Coulomb collisions.}
of correlations e.g considering only binary electrostatic Coulomb interaction between plasma particles and neglecting their three-body correlations, it is possible to derive the Vlasov-Boltzmann equation or the Vlasov–Lenard–Balescu equation depending on the underlying assumptions (Balescu 1960, Guernsey 1960, Lenard 1960). Both the Vlasov–Boltzmann equation and the Vlasov–Lenard–Balescu equation reduce to the Vlasov–Fokker–Planck–Landau equation when considering only nonscreened and small angle Coulomb collisions between plasma particles (Landau 1937). The numerical computation of these kinetic equations onto the resulting 7D phase-space-time grid is still challenging with modern computer technology and further approximations are usually used to reduce the number of variables (Tzoufras et al 2011, Joglekar et al 2014, Touati et al 2014).

The theory of fluctuations in collisionless plasmas is related to (and connects) both the deterministic (Klimontovich-based) and probabilistic (Liouville-based) approaches. It is based on the study of single-time correlations between exact microscopic plasma quantity fluctuations, measuring the deviations of one statistical realization of the plasma consisting of a discrete number of self-interacting particles around the expected average, which is described by the continuous one-body distribution function. Rostoker (1961), Klimontovich and Silin (1962) and Dupree (1963) have shown that the study of single-time correlations between fluctuations also leads to the Vlasov–Lenard–Balescu equation describing the one-body distribution function in the case of homogeneous and stationary collisionless plasmas. This bears a striking resemblance to the fundamentals of PIC simulations: in PIC simulations, the initial macroparticle velocities are determined by random sampling of the (desired) probability distribution function. Therefore, this random sampling (via a random number generator) naturally produces statistical fluctuations around the initial (ensemble averaged) distribution function. The main motivation for this paper lies in this observation: we apply the theory of fluctuations to the discretized equations associated with the electrostatic PIC code (taking into account both the discretized equations of motion and the discretized field equations), and we explore this theory to determine the statistical properties (and possible advantages) of ensemble averages over sets of PIC simulations. The paper is organized as follows. In Section 2, we recapitulate the discretized equations of an electrostatic PIC code and we reformulate them within the mathematical framework of the kinetic theory of plasmas. In Section 3, we derive the theory of statistical fluctuations in electrostatic PIC simulations, for stationary, homogeneous, infinite, and fully ionized plasmas. We discuss the assumptions and the key results of the theory in section 4. In Section 5, we discuss the ‘ensemble averaging technique’ (Shanny et al 1967) within the theoretical framework developed in the previous sections, and we determine the conditions under which this technique relaxes the discrete particle effects inherent to the use of a finite and relatively small number of macroparticles resulting from the limits of computer technology. In Section 6, the main conclusions of the paper are presented.

2. Electrostatic PIC simulations

2.1. Discretized particle and field equations in an electrostatic PIC code

In a PIC simulation, the phase-space of each plasma species \( a \) is sampled according to \( N_a \) macroparticles of mass \( M_a \) and charge \( Q_a \). When the equations computed by a PIC code are renormalized to physical units, each macroparticle represents \( \delta N_a = M_a / m = Q_a / q_a \) real particles of mass \( m \) and electrical charge \( q_a \). \( \delta N_a \) is therefore a number, a lineal density or an areal density of real particles from species \( a \) in 3D, 2D or 1D PIC simulations, respectively. Macroparticle trajectories \( (r_{a,t}(t_n), v_{a,t}(t_n)) \) in phase-space are computed at each time step \( t_n = (n-1) \Delta t \) according to their discretized equation of motion

\[
\forall a, \forall \ell \in [1, N_a], \forall n, \quad \begin{cases}
\frac{dx_{a,t}}{dt} = v_{a,t}(t_n), \\
\frac{dv_{a,t}}{dt} = \frac{Q_a}{M_a} E_a(r_{a,t}(t_n), t_n),
\end{cases}
\]

\( \forall \ell \in [1, N_a] \)

(1)

where

\[
E_a(r_{a,t}(t_n), t_n) = \sum_{i,j,k} E_i(r_{i,j,k}, t_n) \delta f_{a,t}(r_{a,t}(t_n) - r_{i,j,k}) \Delta_i \Delta_j \Delta_k
\]

(2)

is the electric field, interpolated at macroparticle center locations. The electrostatic field is deduced according to the discretized Maxwell equations

\[
\begin{cases}
\frac{\partial}{\partial t} \cdot E_{i,j,k,n} = 4\pi \sum_a q_a \sum_{\ell=1}^{N_a} \delta f_{a,t}(r_{a,t}(t_n) - r_{i,j,k}(t_n)) \\
\frac{\partial}{\partial x} \times E_{i,j,k,n} = 0
\end{cases}
\]

(3)

where macroparticle centers \( r_{i,j,k}(t_n) \) are interpolated at the grid points \( \forall \{i,j,k\} \in [1,N_x] \times [1,N_y] \times [1,N_z] \), \( r_{i,j,k} = ((i-1)\Delta_x, (j-1)\Delta_y, (k-1)\Delta_z) \). Here, \( L_x = N_x\Delta_x, L_y = N_y\Delta_y \) and \( L_z = N_z\Delta_z \) are the simulation box dimensions and \( \Delta_x, \Delta_y, \Delta_z \) are the spatial grid spacings that are usually constant. The index \( k \) or indices \( j \) and \( k \) for 2D or 3D simulations, respectively, should be omitted in the equations but we will keep writing them in all the following to be as general as possible. A smoothed interpolating function

\[
S(r) = \begin{cases}
S_x(x) & \text{in 1D} \\
S_x(x)S_y(y) & \text{in 2D with } \forall \xi \in [x, y, z], \\
S_x(x)S_y(y)S_z(z) & \text{in 3D}
\end{cases}
\]

\[
S_{\xi}(\xi) = \left[ \Pi^{(a)}_{\Delta_x} * S_{0\xi}^{(a)} \right](\xi)
\]

(4)

is used where usually \( \Pi^{(a)}_{\Delta} = \Pi_{\Delta} * \cdots * \Pi_{\Delta} \) is the \( n \)-th order B-spline, i.e. the centered square distribution of support \( \Delta \), \( \Pi_{\Delta}(\xi) = 1/\Delta \) if \( |\xi| < \Delta/2 \) and 0 else, convoluted \( n \) times
with itself. $S_{0\ell}$ are optional filters or smoothing functions that mitigate spatial aliasing effects such as, for example, a Gaussian filter

$$S_{0\ell}(\xi) = \frac{1}{\sqrt{2\pi}\sigma_{t}^{2}} \exp\left[-\frac{\xi^2}{2\sigma_{t}^{2}}\right] \quad (5)$$

for spectral Maxwell solvers or such as a compensated $N$-th order binomial filter for finite-difference time-domain (FDTD) Maxwell solvers. Interpolating functions and filters vary a lot across codes. For example, in PIC codes using an FDTD scheme, the smoothing function is sometimes only applied to the electrostatic field after computing it, according to Maxwell equations (5). In all the following, we will consider the same interpolating function in both (2) and (3) so that the smoothed interpolating function (4) can be interpreted physically as the spatial shape of macroparticles. The generalization of our results to the case where different interpolating functions are used for (2) and (3) is straightforward. In general, the interpolating and smoothing functions have the following properties

$$\forall \mathbf{r}, \sum_{i,j,k} \Pi^{(n)}_{\Delta}(\mathbf{r}_{i,j,k} - \mathbf{r}) \Delta_{x} \Delta_{y} \Delta_{z} = 1,$$

$$\forall \mathbf{r}, \int_{\mathbb{R}^{3}} \Pi^{(n)}_{\Delta}(\mathbf{r}) d^{3}\mathbf{r} = 1$$

and

$$\int_{\mathbb{R}^{3}} S_{0}(\mathbf{r}) d^{3}\mathbf{r} = 1$$

$$\sum_{i,j,k} S_{0}(\mathbf{r}_{i,j,k}) \Delta_{x} \Delta_{y} \Delta_{z} = 1$$

for spectral Maxwell solver or

$$\forall \mathbf{r}, \sum_{i,j,k} \Pi^{(n)}_{\Delta}(\mathbf{r}_{i,j,k} - \mathbf{r}) \Delta_{x} \Delta_{y} \Delta_{z} = 1,$$

$$\forall \mathbf{r}, \int_{\mathbb{R}^{3}} \Pi^{(n)}_{\Delta}(\mathbf{r}) d^{3}\mathbf{r} = 1$$

and

$$\int_{\mathbb{R}^{3}} S_{0}(\mathbf{r}) d^{3}\mathbf{r} = 1$$

$$\sum_{i,j,k} S_{0}(\mathbf{r}_{i,j,k}) \Delta_{x} \Delta_{y} \Delta_{z} = 1$$

for FDTD Maxwell solvers

Here, we have noted $\Delta = (\Delta_{x}, \Delta_{y}, \Delta_{z})^{t}$ with the superscript $^t$ meaning the vector transposition, $\Pi^{(n)}_{\Delta}(\mathbf{r}) = \Pi^{(n)}_{\Delta}(\mathbf{x}) \Pi^{(n)}_{\Delta}(\mathbf{y}) \Pi^{(n)}_{\Delta}(\mathbf{z})$ and $S_{0}(\mathbf{r}) = S_{0}(x) S_{0}(y) S_{0}(z)$ to simplify the notations. Let us finally stress here that even in the case where the PIC codes compute Maxwell equations by using the second order FDTD scheme proposed by Yee (1966) coupled with the charge conserving scheme proposed by Villasenor and Buneman (1992) for first order B-spline macroparticle shapes or by Esirkepov (2001) for higher orders, the resulting numerical scheme can be written in the form (3) if the discretized Maxwell–Poisson equation is verified at the start of the simulation.

2.2. Initialization of a fully ionized, infinite and homogeneous plasma in PIC simulations

Let us consider the electrostatic PIC simulation of a fully ionized, infinite and homogeneous plasma of atomic number $Z$. We note $n_{i}$ the ion density and $n_{e} = Z n_{i}$ the electron density. We use $L_{x}, L_{y}$ and $L_{z}$-periodic boundary conditions so that

$$\forall a \in \{e, i\}, n_{a} = N_{a} \delta N_{a} / L_{a},$$

(7)

(In 1D simulations, $n_{e} = N_{e} \delta N_{e} / L_{x}$ and in 2D simulations, $n_{i} = N_{i} \delta N_{i} / L_{x} L_{y}$). We choose a number $N_{e}$ and a number $N_{i} = ZN_{e}$ of macro-ions and macro-electrons to sample the plasma particle phase-spaces such that $\delta N_{e} = \delta N_{i} = \delta N$. At the start of the simulation, we use a random number generator to initialize macroparticle velocities $V_{a,\ell}$ according to the desired probability distribution $F_{\delta}(V_{a})$. For example, we have for the non-drifting Maxwell–Boltzmann equilibrium probability distribution of velocities, that maximizes entropy:

$$F_{\delta}(V_{a}) = \frac{1}{(2\pi v_{T_{a}})^{3/2}} \exp\left(-\frac{V_{a}^{2}}{2v_{T_{a}}^{2}}\right) \quad (8)$$

where $v_{T}$ is the velocity random variable of macroparticles from species $a$, $v_{T_{a}} = \sqrt{k_{B} T_{a}/m_{a}}$ is the thermal velocity of particles, $k_{B}$ is the Boltzmann constant and $T_{a}$ are the species temperatures. However, in order to ensure homogeneous electrical charge neutrality in the whole PIC simulation plasma, macroparticle locations at the simulation start $R_{a,\ell}$ are not randomly chosen according to a uniform probability law. Instead, we deposit each macroparticle $\ell \in [1,N_{a}]$ homogeneously inside all spatial grid cells. If we denote the number of macroparticles distributed along the $x-$, $y-$ and $z$-axis respectively as $N_{x}, N_{y}, N_{z}$ such that $N_{a} = N_{x} N_{y} N_{z}$, the macroparticle locations read $\forall (\ell_{x}, \ell_{y}, \ell_{z}) \in [1,N_{a}] \times [1,N_{a}] \times [1,N_{a}]$,

$$R_{a,\ell} = \begin{pmatrix} X_{a,\ell} \\ Y_{a,\ell} \\ Z_{a,\ell} \end{pmatrix} = \begin{pmatrix} (\ell_{x} - \frac{1}{2}) \frac{L_{x}}{N_{x}} \\ (\ell_{y} - \frac{1}{2}) \frac{L_{y}}{N_{y}} \\ (\ell_{z} - \frac{1}{2}) \frac{L_{z}}{N_{z}} \end{pmatrix} \quad \quad (9)$$

where $\ell \in [\ell_{x} + (\ell_{x} - 1) N_{y} + (\ell_{y} - 1) N_{z}, N_{x}]$, and where $-1/2$ is used to avoid having half the box have more charge than the other half, which gives rise to nonphysical larger amplitude oscillations due to initial charge imbalances. The numbers of
macroparticles are related to the number of macroparticles per cell according to
\[ N_{a,\text{mpc}} = \frac{N_a}{N_i N_j N_z} = \frac{N_a}{N_i N_j N_z} \]

(In 1D simulations, \( N_{a,\text{mpc}} = N_a/N_i = N_a/N_i \) and in 2D simulations, \( N_{a,\text{mpc}} = N_a/N_i N_j = N_a/N_i N_j). \) The latter is a key parameter of PIC simulations, as we are going to see in the next sections.

### 2.3. Kinetic theory of PIC simulation electrostatic plasmas

One can reformulate equations (1) and (3) computed by an electrostatic PIC code by defining the phase-space densities of macroparticle centers multiplied by their weights
\[ f_n (\mathbf{r}, \mathbf{v}, t) = \delta N_a \sum_{i=1}^{N_a} \delta (\mathbf{r} - \mathbf{r}_{a,\ell} (t)) \delta (\mathbf{v} - \mathbf{v}_{a,\ell} (t)) \]

where we have noted \( \delta (\mathbf{r}) = \delta (x) \delta (y) \delta (z) \) (\( \delta (x) \delta (y) \delta (z) \) in 2D and \( \delta (\mathbf{r}) = \delta (x) \) in 1D) the Dirac distribution.

Firstly, by differentiating the latter with time and then discretizing it in space and time, we obtain
\[
\begin{align*}
\frac{\partial f_n}{\partial t} \bigg|_{i,j,k,n}^{i,j,k,n} + \frac{\partial}{\partial \mathbf{v}} \cdot \left( m_a \mathbf{E}_a (\mathbf{r}_{i,j,k}, t_n) f_n (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \right) &= 0 \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 0
\end{align*}
\]

at the grid points \( \mathbf{r}_{i,j,k} \) and time steps \( t_n \) according to macroparticle discretized equations of motion (1). Secondly, one may express Maxwell equations (3) as a function of weighted macroparticle center phase-space densities (11) according to
\[
\begin{align*}
\frac{\partial \mathbf{v}_n}{\partial \mathbf{r}} \cdot \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 4 \pi \sum_a q_a \int_{\mathbb{R}^3} f_n (\mathbf{r}', \mathbf{v}, t_n) S (\mathbf{r} - \mathbf{r}') d^3 \mathbf{r}' \\
\times (\mathbf{r}_{i,j,k} - \mathbf{r}') d^3 \mathbf{v} \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 0
\end{align*}
\]

Here, we recognize the phase-space densities of real particles
\[
\begin{align*}
f_n (\mathbf{r}, \mathbf{v}, t) &= \int_{\mathbb{R}^3} f_n (\mathbf{r}', \mathbf{v}, t) S (\mathbf{r} - \mathbf{r}') d^3 \mathbf{r}' \\
&= \delta N_a \sum_{i=1}^{N_a} S (\mathbf{r} - \mathbf{r}_{a,\ell} (t)) \delta (\mathbf{v} - \mathbf{v}_{a,\ell} (t))
\end{align*}
\]

as it is sampled according to the PIC method. By differentiating the latter with time in order to get the deterministic description of electrostatic PIC simulations, we obtain the discretized kinetic equation
\[
\begin{align*}
\frac{\partial f_n}{\partial t} \bigg|_{i,j,k,n}^{i,j,k,n} + \frac{\partial}{\partial \mathbf{v}} \cdot \left( m_a \mathbf{E}_a (\mathbf{r}_{i,j,k}, t_n) f_n (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \right) &= C_a (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 0
\end{align*}
\]

It is self-consistently coupled with the discretized Maxwell equations
\[
\begin{align*}
\frac{\partial \mathbf{v}_n}{\partial \mathbf{r}} \cdot \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 4 \pi \sum_a q_a \int_{\mathbb{R}^3} f_n (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) d^3 \mathbf{v} \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n}^{i,j,k,n} &= 0
\end{align*}
\]

The source term
\[
C_a (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) = \sum_a \delta \frac{\partial}{\partial \mathbf{v}} \cdot \mathbf{E} \left[ m_a N_a \sum_{i=1}^{N_a} \left( \mathbf{E}_a (\mathbf{r}_{a,\ell} (t_n), t_n) - \mathbf{E}_a (\mathbf{r}_{i,j,k}, t_n) \right) \right] \cdot \mathbf{v}_{a,\ell} (t_n)
\]
is due to the sum of forces \( (q_a/m_a) [\mathbf{E}_a (\mathbf{r}_{a,\ell} (t_n), t_n) - \mathbf{E}_a (\mathbf{r}_{i,j,k}, t_n)] \) acting on each macroparticle \( a, \ell \) because of the internal tension force which is constraining the macroparticle charge and inertia as a unit, when different parts of the macroparticle feel a different electrostatic field (Decyk 1982). According to its mathematical expression (17), these nonphysical forces decreases with decreasing spatial grid spacings and/or with increasing order of interpolation and smoothing functions. One must always check that
\[
\sum_{i,j,k} \int_{\mathbb{R}^3} \frac{m_a v^2}{2} C_a (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) d^3 \mathbf{v} \delta \Delta_1 \Delta_2
\]
remains sufficiently small in case of electrostatic oscillations on the macroparticle-size space scale by checking this energy flow is associated with this internal tension force at each time step \( t_n \) during the whole simulation.

The random number generator used to initialize macroparticle velocities \( \mathbf{v}_{a,\ell} \) introduces statistical fluctuations
\[
\delta f_n (\mathbf{r}, \mathbf{v}, t) = f_n (\mathbf{r}, \mathbf{v}, t) - \tilde{f}_n (\mathbf{r}, \mathbf{v}, t)
\]
between the weighted macroparticle center phase-space densities (11) from the simulation, seen here as a statistical realization of the plasma that one wants to simulate, and the corresponding weighted macroparticle center distribution functions

$$ f_{a_i}(\mathbf{r}, \mathbf{v}, t) = \mathbb{E} \{ f_{a_i}(\mathbf{r}, \mathbf{v}, t) \} \tag{20} $$

that ones would obtain on average over an infinite number of simulations and where we have thus noted $\mathbb{E}[X]$ the expectation of the random variable $X$ from the probability theory. According to the previous section 2.2, the latter reads

$$ \tilde{f}_{a_i}(\mathbf{r}, \mathbf{v}, t_i) = n_{a_i}(\mathbf{r}, t_i) F_{a_i}(\mathbf{v}) \tag{21} $$

at the simulation start $t_i$ where

$$ n_{a_i}(\mathbf{r}, t_i) = \delta N_{a_i} \sum_{\ell=1}^{N_{a_i}} \delta (\mathbf{r} - \mathbf{R}_{a_i,\ell}) \tag{22} $$

is the weighted macroparticle center spatial density. They lead to electrostatic field statistical fluctuations

$$ \delta \mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r}, t) - \bar{\mathbf{E}}(\mathbf{r}, t) \tag{23} $$

between the electrostatic field from the simulation and its expected value

$$ \bar{\mathbf{E}}(\mathbf{r}, t) = \mathbb{E} \{ \mathbf{E}(\mathbf{r}, t) \} \tag{24} $$

corresponding to the plasma macroscopic state that one wants to simulate. We remind the reader that we are dealing here with the fluctuations of statistical realizations of the plasma that one wants to simulate around its mean behaviour described by its distribution functions and not with spatial nor temporal fluctuations. In this sense, a PIC simulation is seen here as a numerical experiment. Taking the expected value of (12) and (13) at grid points $\mathbf{r}_{i,j,k}$ and time steps $t_n$, we obtain without any approximation

$$ \frac{\partial \delta f_{a_i}}{\partial t} \bigg|_{i,j,k,n} + \frac{\partial}{\partial \mathbf{r}} \cdot (\mathbf{v} \delta f_{a_i}) \bigg|_{i,j,k,n} + \frac{\partial}{\partial \mathbf{v}} \bigg( \frac{q_a}{m_a} \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) \tilde{f}_{a_i}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \bigg) = - \frac{q_a}{m_a} \frac{\partial}{\partial \mathbf{v}} \cdot \mathbb{E} \{ \delta \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) \delta f_{a_i}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \} \tag{25} $$

self-consistently coupled with the discretized macroscopic Maxwell equations

$$ \left\{ \begin{array}{l} \frac{\partial}{\partial t} \cdot \mathbf{E} \bigg|_{i,j,k,n} = 4\pi \sum_a q_a \int_{\mathbb{R}^3} \int_{\mathbb{R}^3} \tilde{f}_{a_i}(\mathbf{r}', \mathbf{v}, t_n) S \times (\mathbf{r}_{i,j,k} - \mathbf{r}') d^3\mathbf{r}' d^3\mathbf{v} \quad \tag{26} \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n} = 0 \end{array} \right. $$

Here,

$$ \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) = \sum_{i,j,k} \mathbb{E}(\mathbf{r}_{i,j,k}, t_n) S(\mathbf{r}_{a_i,\ell}(t_n) - \mathbf{r}_{i,j,k}) \Delta_x \Delta_y \Delta_z \tag{27} $$

is the expected electrostatic field interpolated at macroparticle center locations and

$$ \delta \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) = \sum_{i,j,k} \delta \mathbb{E}(\mathbf{r}_{i,j,k}, t_n) S(\mathbf{r}_{a_i,\ell}(t_n) - \mathbf{r}_{i,j,k}) \Delta_x \Delta_y \Delta_z \tag{28} $$

the interpolated electrostatic field fluctuating component from the simulation. Subtracting (25) from (12) and (13) from (26), we obtain the self-consistently coupled equations

$$ \left\{ \begin{array}{l} \frac{\partial \delta f_{a_i}}{\partial t} \bigg|_{i,j,k,n} + \frac{\partial}{\partial \mathbf{r}} \cdot (\mathbf{v} \delta f_{a_i}) \bigg|_{i,j,k,n} + \frac{\partial}{\partial \mathbf{v}} \bigg( \frac{q_a}{m_a} \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) \tilde{f}_{a_i}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \bigg) = - \frac{q_a}{m_a} \frac{\partial}{\partial \mathbf{v}} \cdot \mathbb{E} \{ \delta \mathbf{E}_{a_i}(\mathbf{r}_{i,j,k}, t_n) \delta f_{a_i}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \} \\
\frac{\partial}{\partial \mathbf{r}} \times \mathbf{E} \bigg|_{i,j,k,n} = 0 \end{array} \right. \tag{29} $$

They describe the evolution with time of statistical fluctuations introduced by different initial conditions implemented by random numbers. Knowing the statistical fluctuations $\delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_1)$ and $\delta f_{a_i}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_1)$ in all spatial grid cells $\mathbf{r}_{i,j,k}$ at the simulation start $t_1$, they allow us to deduce their values at any time steps $t_n$.

Let us emphasize here that the weighted macroparticle center phase-space density fluctuations (19) are not the statistical fluctuations

$$ \delta f_{a_i}(\mathbf{r}, \mathbf{v}, t) = f_{a_i}(\mathbf{r}, \mathbf{v}, t) - \tilde{f}_{a_i}(\mathbf{r}, \mathbf{v}, t) \tag{31} $$

between the expected plasma distribution functions

$$ \tilde{f}_{a_i}(\mathbf{r}, \mathbf{v}, t) = \mathbb{E} \{ f_{a_i}(\mathbf{r}, \mathbf{v}, t) \} \tag{32} $$

that one wants to simulate and the particle phase-space densities (14) from the PIC simulation. Therefore, one has to take the expected value of (15) and (16) at grid points $\mathbf{r}_{i,j,k}$ and time steps $t_n$ in order to get the probabilistic description of electrostatic PIC simulations. One thus obtains without any approximation the discretized kinetic equation
Liozoned plasma consisting of a stable PIC simulation plasma
We consider a stationary, homogeneous, infinite and fully
3.1 Assumptions and notations
3. Statistical fluctuations in electrostatic PIC simulations

(a) numerical consistency errors of numerical schemes used
to compute the equations are small compared to statistical fluctuations introduced by different initial conditions implemented by random numbers.

(b) the weighted macroparticle centers spatial densities are homogeneous at the simulation start
\[ \forall i, j, k, \bar{n}_{a_i} (r_{i,j,k}, t_1) = \bar{n}_a \] (35)
so that
\[ \forall i, j, k, f_{a_i} (r_{i,j,k}, v, t_1) = \frac{\bar{n}_a}{N_{a,mpc}} \sum_{\ell=1}^{N_{a,mpc}} \delta (v - V_{a_i,\ell}) \] (36)
where \( \ell_{i,j,k} = \ell + [(i - 1) + (j - 1) N_k + (k - 1) N_N] \) \( N_{a,mpc} \) and
\[ f_{a_i} (r_{i,j,k}, v, t_1) = \bar{n}_a F_{a_i} (v). \] (37)
Therefore, we have also in this case
\[ \forall i, j, k, \bar{E} (r_{i,j,k}, t_1) = 0 \] (38)
according to macroscopic Maxwell equations (26), the macroparticle shape properties (6) and assumption a.

(c) We neglect the quadratic terms
\[ \frac{q_a}{m_a} \frac{\partial}{\partial v} \cdot (\delta E_{r_i,j,k} (r_{i,j,k}, v, t_n) \delta f_{a_i} (r_{i,j,k}, v, t_n)) \] 
\[ - \delta E \{ \delta E_{r_i,j,k} (r_{i,j,k}, t_n) \} \delta f_{a_i} (r_{i,j,k}, v, t_n)) \] (39)
describing the influence of macroparticle close-encounter Coulomb collisions on the weighted macroparticle center discretized phase-space density statistical fluctuations (29).

(d) The weighted macroparticle center distribution functions and macroscopic electrostatic fields are stationary
\[ \forall t_n, f_{a_i} (r_{i,j,k}, v, t_n) = \bar{f}_{a_i} (r_{i,j,k}, v, t_1) \] and
\[ \bar{E} (r_{i,j,k}, t_n) = \bar{E} (r_{i,j,k}, t_1). \] (40)

In order to check assumption b, let us introduce the Fourier transform in the space of discrete functions
\[ \tilde{F} (k) = \Delta_x \sum_{i=-\infty}^{\infty} \Delta_y \sum_{j=-\infty}^{\infty} \Delta_z \sum_{k=-\infty}^{\infty} F (r_{i,j,k}) \exp (-i k \cdot r_{i,j,k}) \] (41)
such that
\[ F (r_{i,j,k}) = \int_{V_k} d^3 k \tilde{F} (k) \exp (i k \cdot r_{i,j,k}) \] (42)
where the integration domain is the first Brillouin zone \( V_k = [-k_x/2, k_x/2] \times [-k_y/2, k_y/2] \times [-k_z/2, k_z/2] \). The inverse Fourier transforms (42) can be seen as Fourier series components of (41). Indeed, according to its mathematical expression, the latter is necessarily \( k_{\ell}\) periodic where
\[ k_{\ell} = \left( \frac{2 \pi}{\Delta_x}, \frac{2 \pi}{\Delta_y}, \frac{2 \pi}{\Delta_z} \right) \] (43)
is the Nyquist frequency. It is defined as the highest spatial frequency that one can resolve in a PIC simulation due to discretization of space. According to (7), (9), (22), (41) and the Poisson summation formula applied first to the infinite sums over macroparticle indices \( \ell_x, \ell_y \) and \( \ell_z \) and then to the infinite sums over spatial indices \( i, j \) and \( k \), one finds
\[ \tilde{n}_{a_i} (k, t_1) = (2 \pi)^3 \bar{n}_a \sum_{p,q,r} (-1)^{p+q+r} \delta (k - \mathbf{I}_{p,q,r} k_x, k_y, k_z) \] (44)
where we have noted
\[ \delta (k, g_x) = \sum_{p,q,r} \delta (k - \mathbf{I}_{p,q,r} \cdot k_x) \] (45)
a cloud shape function (linear interpolating function has a Gaussian shape function (field generated by one macroparticle ily sees homogeneous weighted macroparticle center spatial of macroparticles per cell in each direction, the code necessar- way that, by depositing the same amount strictly greater than 1 in 2D and consequence of the condition (44) transforms of discrete functions while short range electrostatic interaction between macroparticles when $|r - r_{a,\ell}| < \Delta$ are underestimated (Langdon and Birdsall 1970, Okuda and Birdsall 1970). We can thus neglect Coulomb collisions between macroparticles and consider assumption c to be valid. We will check assumptions a and d later in sections 4.1 and 4.2, respectively. The following derivations are based on the theory of fluctuations (Rostoker 1961, Klimontovich and Silin 1962, Dupree 1963) well summarized in the Landau and Lifshitz textbook (Landau and Lifshitz 1981). Notations used are recapitulated in table 1.

3.2. Estimate of statistical fluctuations in a PIC simulation electrostatic plasma

According to their definitions (19), (20), (23) and (24), there is no electrostatic field statistical fluctuations

$$\mathbb{E}\{\delta E(r_{i,j,k}, t_n)\} = 0$$  \hspace{1cm} (46)

nor weighted macroparticle center phase-space density statistical fluctuations

$$\mathbb{E}\{\delta f_{mpc}(r_{i,j,k}, \mathbf{v}, t_n)\} = 0$$  \hspace{1cm} (47)

on average. Therefore, in order to get an estimate of the nonzero electrostatic field—usually called ‘numerical noise’—that one observes in PIC simulations due to the macroparticle initial conditions implemented by random numbers, we propose here to estimate it according to the square root of the single-time autocorrelation of its statistical fluctuations

$$|E_{\text{fluct}}| = \sqrt{\mathbb{E}\{\delta E(r_{i,j,k}, t_n)^2\}}.$$  \hspace{1cm} (48)

Indeed, as will be justified later in section 5.1, if the number of macroparticles per cell is sufficiently large, the electrostatic field statistical fluctuation $\delta E(r_{i,j,k}, t_n)$ at grid points $r_{i,j,k}$ and time steps $t_n$ follows a normal distribution law with mean (46) and variance (48) according to the central limit theorem. As a consequence,

$$|\delta E(r_{i,j,k}, t_n)| \leq \frac{\sqrt{2} \text{erf}^{-1}\left(\frac{p}{100}\right)}{N_{\text{elec}} > 1} |E_{\text{fluct}}|$$  \hspace{1cm} (49)

with a $p\%$-confidence and the square root of electrostatic field statistical fluctuations autocorrelation (48) thus provides a good estimate of statistical fluctuations.

To carry out such a program, we use the Laplace–Fourier transforms of discrete functions
Table 1. Recapitulation of main notations used in the text.

| Symbol | Definition |
|--------|------------|
| \( a \) | Index of plasma species |
| \( m_a \) | Particle mass |
| \( q_a \) | Particle charge |
| \( M_a \) | Macroparticle mass |
| \( Q_a \) | Macroparticle charge |
| \( \delta N_a \) | Macroparticle weight |
| \( S \) | Macroparticle shape function |
| \( N_a \) | Total number of macroparticles |
| \( \ell \) | Index of macroparticles |
| \( r_{a,\ell}(t) \) | Macroparticle center location |
| \( R_{a,\ell} \) | Macroparticle center location at the simulation start |
| \( v_{a,\ell}(t) \) | Macroparticle velocity |
| \( V_{a,\ell} \) | Macroparticle velocity at the simulation start |
| \( F_{a0}(V_a) \) | Probability distribution of macroparticle velocities at the simulation start |
| \( r \) | Space variable |
| \( v \) | Velocity variable |
| \( t \) | Time variable |
| \( E(r, t) \) | Self-consistent electrostatic field |
| \( E_s(r_{a,\ell}(t), t) \) | Self-consistent electrostatic field interpolated at macroparticle center locations |
| \( f_a(r, v_{a,\ell}, t) \) | Plasma species phase-space density sampled according to the PIC method |
| \( n_a(r, t) \) | Plasma species density |
| \( f_{a,\ell}(r_{a,\ell}, t) \) | Weighted phase-space density of macroparticle centers |
| \( n_{a,\ell}(r, t) \) | Weighted density of macroparticle centers |
| \( N_x, N_y, N_z \) | Number of spatial grid points along x-, y- and z-axis |
| \( i, j, k \) | Indices of spatial grid points along x-, y- and z-axis |
| \( \Delta x, \Delta y, \Delta z \) | Spatial cell size along x-, y- and z-axis |
| \( r_{i,j,k} \) | Spatial grid point location |
| \( N_t \) | Number of time iterations |
| \( n \) | Index of time iterations |
| \( \Delta_t \) | Time step |
| \( t_n \) | Iteration time |
| \( N_{a,mpc} \) | Number of macroparticles per cell |
| \( \mathcal{F}(k) \) | Fourier transform of discrete function \( F(r_{i,j,k}) \) |
| \( \mathcal{F}(\omega, k) \) | Fourier transform of continuous function \( F(r) \) |
| \( \hat{\mathcal{F}}(\omega, k) \) | Laplace–Fourier transform of discrete function \( F(r_{i,j,k}; t_n) \) |
| \( k_{g,x}, k_{g,y}, k_{g,z} \) | Nyquist spatial angular frequency along k\(_x\), k\(_y\), and k\(_z\)-axis |
| \( p, q, r \) | Indices of spatial angular frequency aliases along k\(_x\), k\(_y\), and k\(_z\)-axis |
| \( k_{p,q,r} \) | Spatial angular frequency vector alias |
| \( \{k\}_p \) | k-bracket depending on the Maxwell solver |
| \( \omega_g \) | Nyquist temporal angular frequency |
| \( m \) | Index of temporal angular frequency aliases |
| \( \omega_m \) | Temporal angular frequency aliases |
| \( \{\omega\}_t \) | \( \omega \)-bracket depending on the macroparticle pusher |
| \( \mathbb{E}\{F\} \) | Expectation of stochastic process \( F \) (sometimes called ensemble average) |
| \( \delta F = F - \mathbb{E}\{F\} \) | Fluctuations of stochastic process \( F \) |
| \( \mathbb{E}\{\delta F(t)^2\} \) | Single-time autocorrelation of fluctuations \( \delta F \) |
| \( \mathbb{E}\{\delta F(t)\delta G(t)\} \) | Single-time correlation between fluctuations of stochastic processes \( F \) and \( G \) |
| \( N_s \) | Total number of statistical realizations |
| \( s \) | Index of statistical realizations |
| \( F^{(s)} \) | Statistical realization of \( F \) |
| \( \langle F \rangle \) | Finite ensemble average over \( N_s \) statistical realizations |
that is commonly used to compute the discretized macroparticle equations of motions (1), and by associating the nonperturbed straight line orbits \((\mathbf{r}_{a,t}^{(0)}(t_n), \mathbf{v}_{a,t}^{(0)}(t_n)) = (\mathbf{R}_{a,t}, \mathbf{V}_{a,t}, \mathbf{V}_{a,t})\) with \(\hat{f}_{a,t}\) and the first order perturbed orbits \((\mathbf{r}_{a,t}^{(1)}(t_n), \delta \mathbf{v}_{a,t}(t_n))\) with \(\delta \hat{f}_{a,t}\) according to assumptions a, b, c and d (Lindeman 1970), the discrete Laplace–Fourier transform (50) of the resulting discretized equation (29) gives

\[
\begin{align*}
\frac{dF_{m}}{d\mathbf{v}} &= \delta \hat{f}_{a,t}(\mathbf{k}, \mathbf{v}, t) \\
\text{where}
\end{align*}
\]

(56)

\[
\omega_m = \omega - m\omega_g
\]

(58)

and

(59)

are the different temporal and spatial angular frequency aliases, following the same notation as in Langdon (1970a) and Birdsall and Langdon (1991). Maxwell equations (13) are usually computed by using a spectral solver or the second order FDTD method proposed by Yee (1966) coupled with the charge conserving scheme proposed by Villasenor and Buneman (1992) or Esirkepov (2001) rather than the Poisson equation solver considered in Langdon (1970b). In all cases, the Laplace–Fourier transform (50) of the resulting Maxwell equations (30) linearized according to assumptions a, b and d gives

\[
\begin{align*}
\{\mathbf{k}\}_{r} &= \left\{ \begin{array}{c}
(k_x, k_y, k_z)^	op \\
\sin (k_x \Delta_x / 2), \sin (k_y \Delta_y / 2), \sin (k_z \Delta_z / 2)
\end{array} \right\} \\
\text{for the spectral solver}
\end{align*}
\]

(61)

for the 2nd order FDTD scheme
and \(p, q, r \in \{k, k', q', r', v\}\). By substituting \(\hat{\delta}_{\text{d}}(\omega, k)\) from (66) into the expression of \(\hat{\delta}E(\omega, k)\) from (60), replacing \(k\) by \(k_{p,q,r}\), changing the indices \((p', q', r')\) into \((p' - p, q' - q, r' - r)\) in the resulting equation, one can impose then \((p, q, r) = (-p', -q', -r')\) and use the \(k_r\)-periodicity of \(\{k, r\}, \hat{\delta}E\) and \(\hat{\delta}_{\text{d}}(\omega)\) to obtain the discrete Laplace–Fourier transform (50) of electrostatic field statistical fluctuations (Langdon 1970a). It reads

\[
\hat{\delta}E(\omega, k) = -\frac{4\pi}{\varepsilon L(\omega, k)} \sum_{p,q,r} \hat{S}(k_{p,q,r}) \sum_a q_a \times \int_{\mathbb{R}^3} \delta_{\text{d}}(k, v, t_1) d^3v
\]

where the longitudinal permittivity of the PIC simulation plasma reads

\[
\varepsilon L(\omega, k) = 1 - \sum_a k_{p_a}^2 \sum_{p,q,r} \hat{S}(k_{p,q,r}) \quad \times \int_{\mathbb{R}^3} \frac{dF_{\alpha\delta}}{d\nu} d\nu
\]

with

\[
\omega_{p_a} = \sqrt{\frac{4\pi \bar{n}_{\alpha} q_a^2}{m_a}}
\]

the Langmuir angular frequencies of each plasma species. We can check that it tends to the longitudinal permittivity of collisionless plasmas in the limit \(\nu \xi \in \{x, y, z, t\}, \Delta_\xi \to 0\). The study of poles \(\omega = \omega(k)\) given by the dispersion relation \(\varepsilon L(\omega, k) = 0\) provides the stability criterion of the simulated plasma. We have assumed the simulated plasma remains stable, both physically and numerically. It means that all these poles have a negative imaginary part \(\text{Im}\{\omega(k)\} < 0\). In general, the stability criterion depends on the macroparticle shape and initial macroparticle velocity probability distributions \(F_{\alpha\delta}\). For example, it has been shown that electrostatic PIC simulations of plasmas at equilibrium using the first or greater orders B-spline interpolating function (and the same spatial spacings in all directions for simplicity) are stable if the condition

\[
\Delta_i \leq \Delta_i / \sqrt{\varepsilon_L} < 1 / \omega_p
\]

is respected (Birdsall and Langdon 1991) where

\[
\omega_p^2 = \omega_{p_a}^2 + \omega_{p_b}^2 \approx \omega_{p_a}^2
\]

is the Langmuir plasma angular frequency in the limit \(Zm_e / m_i \ll 1\).

One can now deduce the expression of electrostatic field statistical fluctuations autocorrelation from the discrete Laplace–Fourier transform of electrostatic field statistical fluctuations (62). It reads

\[
\mathbb{E}\left\{ \delta E(\mathbf{r}, \mathbf{k}, t_2) \right\} = \int \ldots \int \frac{d^3k}{v_{\mathbf{k}}} \int \ldots \int \frac{d^3k'}{v_{\mathbf{k}'}} \int \ldots \int \frac{d^3k''}{v_{\mathbf{k}''}} \times \mathbb{E}\left\{ \hat{\delta}E(\omega, \mathbf{k}) \hat{\delta}E(\omega', \mathbf{k}') \right\} \exp \left\{-i \left[ (\omega + \omega') t_2 - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}, t_2 \right]\right\}
\]

where

\[
\mathbb{E}\left\{ \hat{\delta}E(\omega, \mathbf{k}) \right\} = \frac{(4\pi)^3}{\varepsilon L(\omega, k)} \frac{\{k\} \cdot \{k\}}{\varepsilon L(\omega, k)}
\]

It depends on the weighted macroparticle center phase-space density fluctuation Fourier transform correlation at the simulation start \(t_1\). One can derive the exact analytical expression of the latter. One finds according to section 2.2

\[
\mathbb{E}\left\{ \delta_{\text{d}}(\mathbf{r}, \mathbf{v}, t_1) \delta_{\text{d}}(\mathbf{r}', \mathbf{v}', t_1) \right\} = \delta_{\alpha\beta} \delta N_{\alpha\beta} (\mathbf{r} - \mathbf{r}') n_{\alpha\beta} (\mathbf{r}, t_1) F_{\alpha\beta}(\mathbf{v}) [\delta (\mathbf{v} - \mathbf{v}') - F_{\alpha\beta}(\mathbf{v}')] \cdot \mathbf{v}'.
\]

The first term in the square bracket corresponds to the case where there is only one macroparticle in the infinitesimal phase-space volume located between \((\mathbf{r}, \mathbf{v})\) and \((\mathbf{r} + d\mathbf{r}, \mathbf{v} + d\mathbf{v})\). Replacing consequently \(\delta (\mathbf{r} - \mathbf{r}')\) by its equivalent expression \(\delta_{\alpha\beta} \delta_{\alpha\beta} \delta_{\alpha\beta} / \Delta_{\alpha\beta} \Delta_{\alpha\beta} \Delta_{\alpha\beta}\), in a discretized space, one can compute the Fourier transform (41) of the initial weighted macroparticle center phase-space densities correlation (69) expressed at the grid points \(\mathbf{r}_{i,j,k}\) and \(\mathbf{r}_{i',j',k'}\). It reads

\[
\mathbb{E}\left\{ \delta_{\text{d}}(\mathbf{r}, \mathbf{v}, t_1) \delta_{\text{d}}(\mathbf{r}', \mathbf{v}', t_1) \right\} = \delta_{\alpha\beta} \delta N_{\alpha\beta} (2\pi)^3 \hat{n}_{\alpha\beta} (\mathbf{k} + \mathbf{k}', \mathbf{k}) F_{\alpha\beta}(\mathbf{v}) [\delta (\mathbf{v} - \mathbf{v}') - F_{\alpha\beta}(\mathbf{v}')]
\]

according to assumption b. By performing successively the integration over \(\mathbf{k}', \mathbf{v}'\) and \(\omega'\) in (67) (see appendix A), one finds

\[
\mathbb{E}\left\{ \hat{\delta}E(\omega, \mathbf{k}) \hat{\delta}E(\omega', \mathbf{k}') \right\} = (2\pi)^4 \delta (\omega + \omega', \omega_{\mathbf{k}}) \times \delta (\mathbf{k} + \mathbf{k}', \mathbf{k}) \hat{\delta}E^2(\omega, \mathbf{k})
\]
with
\[ \widetilde{\delta E}^2(\omega, \mathbf{k}) = \int_{-\omega/2}^{\omega/2} \frac{d\omega'}{2\pi} \int_{V_k} \frac{d^3k'}{(2\pi)^3} \times \mathbb{E} \left\{ \delta \mathbf{E}(\omega, \mathbf{k}) \delta \mathbf{E}(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
\[ = \sum_{\omega, \mathbf{k}, \mathbf{k}'} \mathbb{E} \left\{ \delta \mathbf{E}(\omega, \mathbf{k}) \delta \mathbf{E}(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
in the expression of single-time electrostatic field statistical fluctuation autocorrelation (67). For example, we find for PIC simulation plasma at equilibrium the finite estimate (see appendix B)
\[ |\mathbf{E}_{\text{thcl}}| \sim \sqrt{\delta N_{k_B} T \rho_k (k_{\lambda D}) \frac{\delta N_{\lambda D}}{N_{e,\text{mpc}}}} \]
\[ = \left\{ \begin{array}{ll}
\frac{4}{\pi} \arctan \left( \frac{k_{\lambda D} \lambda}{\lambda_{\text{eff}}} \right) & \text{in 1D} \\
\frac{1}{\pi} \ln \left[ 1 + \left( k_{\lambda D} \lambda_{\text{eff}} \right)^2 \right] & \text{in 2D} \\
\frac{2}{\pi} k_{\lambda D} - 2 \arctan \left( k_{\lambda D} \lambda_{\text{eff}} \right) & \text{in 3D}
\end{array} \right. \] (73)
These estimates are compared with the amplitude of electrostatic field statistical fluctuations in electrostatic PIC plasma simulations in figure 2 considering \( \Delta_x = \Delta_y = \Delta_z \) for simplicity and using
\[ \delta N = \frac{\hat{n}_e \Delta^2}{N_{e,\text{mpc}}} \]
according to (7) and (10).

By performing the similar derivation for the single-time autocorrelation of statistical density fluctuations, one also finds
\[ \mathbb{E} \left\{ \delta n_a(\mathbf{r}_{i,j,k}, t_n)^2 \right\} \]
\[ = \int_{-\omega/2}^{\omega/2} \frac{d\omega'}{2\pi} \int_{V_k} \frac{d^3k'}{(2\pi)^3} \int_{-\omega/2}^{\omega/2} \frac{d\omega''}{2\pi} \int_{V_k} \frac{d^3k''}{(2\pi)^3} \times \mathbb{E} \left\{ \delta n_a(\omega, \mathbf{k}) \delta n_a(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
\[ = \sum_{\omega, \mathbf{k}, \mathbf{k}'} \mathbb{E} \left\{ \delta n_a(\omega, \mathbf{k}) \delta n_a(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
\[ = \sum_{\omega, \mathbf{k}, \mathbf{k}'} \mathbb{E} \left\{ \delta n_a(\omega, \mathbf{k}) \delta n_a(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
\[ = \sum_{\omega, \mathbf{k}, \mathbf{k}'} \mathbb{E} \left\{ \delta n_a(\omega, \mathbf{k}) \delta n_a(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
\[ = \sum_{\omega, \mathbf{k}, \mathbf{k}'} \mathbb{E} \left\{ \delta n_a(\omega, \mathbf{k}) \delta n_a(\omega', \mathbf{k}') \right\} \times \exp \left\{ -t (|\omega + \omega'| t_n - (\mathbf{k} + \mathbf{k}') \cdot \mathbf{r}_{i,j,k}) \right\} \]
and
\[ \delta n_a^2(\omega, \mathbf{k}) = \frac{2\pi \delta n_a \delta n_a}{|\omega_\rho t_n|^2} \sum_{\mathbf{k}_{p,q,r}} \mathbb{S} \left( \mathbf{k}_{p,q,r} \right)^2 \times \int_{\mathbb{R}^3} d^3\mathbf{v} F_{\mathbf{v},0} \delta (\mathbf{v} - \mathbf{k}_{p,q,r} \cdot \mathbf{v}, \omega_\rho) . \] (76)

Note that Eq. (76) is the single-time autocorrelation of plasma particle density fluctuations
\[ \delta n_a(\mathbf{r}_{i,j,k}, t_n) = \int_{\mathbb{R}^3} \delta f_a(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) d^3\mathbf{v} \] (77)
and not the one of weighted macroparticle center spatial densities
\[ \delta n_a(\mathbf{r}_{i,j,k}, t_n) = \int_{\mathbb{R}^3} \delta f_a(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) d^3\mathbf{v}. \] (78)

The single-time autocorrelation of plasma electrical charge fluctuations can be obtained starting from (72) or from (76). One can note that initial fluctuation correlation of differing plasma species quantities necessarily cancels according to (69). Fluctuation spectra (72) and (76) are the same as the ones obtained by Langdon (1970b, 1979), however generalizing them to simulations of multiple species plasmas using modern PIC algorithms and arbitrary macroparticle weight factors \( \delta N_a \). To derive these spectra, the authors used a test macroparticle formalism similar to the method used by Thompson and Hubbard (1960), Hubbard (1961) and later justified by Rostoker (1964) from the classical kinetic theory of plasmas. It consists of perturbing a discretized Vlasovian plasma by one macroparticle propagating along a straight-line orbit and averaging the resulting linear PIC simulated plasma quantity over all possible macroparticle initial conditions. The validity of their method is ensured by the first term, corresponding to the case where there is only one macroparticle in the infinitesimal volume of phase-space, in our results (69) of initial weighted macroparticle center phase-space density fluctuations correlations. While providing the physical interpretation of the results up to the first order, it may lead to confusion between Vlasov simulations and PIC simulations. Therefore, it seems to us that our method relates to fewer ambiguity equations computed by an electrostatic PIC code with the kinetic theory of plasmas. In addition, our approach adds the information that the statistical fluctuation spectra are only valid on a time scale \( \omega_\rho t_n \gg 1 \).

It means that one must let the simulation run during a time greater than \( \sim 10/\omega_\rho \) before obtaining comparable spectra by directly computing it from the PIC simulation results.

3.3. Kinetic equation of a PIC simulation electrostatic plasma

Performing the same reasoning as in the previous section, one can also find the single-time correlation between statistical fluctuations of interpolated electrostatic field and weighted macroparticle center phase-space density at the grid points \( \mathbf{r}_{i,j,k} \) and time steps \( t_n \) for each species. We find starting from (56), (62) and (70), assuming again assumptions a–d
Figure 2. Average over all spatial cells of electrostatic field fluctuation amplitudes at the discrete time $t_n = 10/\omega_p$ (Left) and their analytical estimates (73) (Right) in 1D (Top), 2D (Middle) and 3D (Bottom) periodic electrostatic PIC plasmas at Maxwell–Boltzmann equilibrium simulated with the spectral electrostatic PIC codes from the UPIC framework (Decyk 1995, 2007, Decyk and Singh 2014); each simulation out of the 30 000 uses a total of 4096 spatial cells and interpolation functions $\Pi_{\Delta_i}^{(1)}$ without filtering.
Plasma Phys. Control. Fusion 

and considering again only non-damped terms on time scales \(\omega_p t_n \gg 1\)

\[
\mathbb{E} \left\{ \delta \Theta_a \left( \omega', k' \right) \delta \Theta_a \left( \omega, k, v_a \right) \right\}
\]

\[
= \left( 2\pi \right)^4 \delta \left( \omega + \omega' \right) \delta \left( k + k' \right) \delta \left( k_0 \right) \delta \left( \omega_0 \right) \mathbb{E} \left\{ \delta \Theta_a \left( \omega, k, v_a \right) \right\}
\]

\[
(79)
\]

where

\[
\delta \Theta_a \left( \omega, k, v_a \right) = -i \sum_{p,q,r} \frac{8\pi^2}{\varepsilon_L} \left( k_{p,q,r} \right)^2 \left\{ k_{p,q,r} \right\}_r S \left( k_{p,q,r} \right) \left( k_{p,q,r} \right)_r \delta N_a \delta q_a
\]

\[
\times \int_{V_{k_0}} \delta \left( \omega - k_{p,q,r} \cdot v_a \right) \left\{ k_{p,q,r} \right\}_r
\]

\[
+ \frac{1}{m_a} \frac{\partial}{\partial \mathbf{v}_a} \cdot \left( q_a \frac{\delta \Theta_a}{\delta \mathbf{v}_a} \right) \left( v_{a,0} \right)
\]

\[
\times \left( k_{p,q,r} \right)_r \left\{ k_{p,q,r} \right\}_r \delta \left( k_{p,q,r} \cdot v_a - \omega_{0} \right)
\]

\[
\left( 2\pi \right)^4 \delta \left( \omega - k_{p,q,r} \cdot v_a \right) \mathbb{E} \left\{ \delta \Theta_a \left( \omega, k, v_a \right) \right\}
\]

\[
(80)
\]

It allows for estimating the discretized right-hand side of kinetic equations (25) depending on

\[
\mathbb{E} \left\{ \delta \Theta_a \left( r_{i,j,k}, t_n \right) \delta \Theta_a \left( r_{i,j,k}, t_n \right) \right\}
\]

\[
= \frac{1}{m_a} \frac{\partial}{\partial \mathbf{v}_a} \cdot \left( q_a \frac{\delta \Theta_a}{\delta \mathbf{v}_a} \right) \left( v_{a,0} \right)
\]

\[
\times \left( k_{p,q,r} \right)_r \left\{ k_{p,q,r} \right\}_r \delta \left( k_{p,q,r} \cdot v_a - \omega_{0} \right)
\]

\[
\left( 2\pi \right)^4 \delta \left( \omega - k_{p,q,r} \cdot v_a \right) \mathbb{E} \left\{ \delta \Theta_a \left( \omega, k, v_a \right) \right\}
\]

\[
(81)
\]

One finds after integration over \(\omega\) (see appendix C)

\[
- \frac{q_a}{m_a} \frac{\partial}{\partial \mathbf{v}_a} \cdot \mathbb{E} \left\{ \delta \Theta_a \left( r_{i,j,k}, t_n \right) \delta \Theta_a \left( r_{i,j,k}, t_n \right) \right\}
\]

\[
= - \frac{1}{m_a} \frac{\partial}{\partial \mathbf{v}_a} \cdot \sum_{p,q,r} \delta N_a \int_{V_{k_0}} \mathbf{Q} \left( \mathbf{v}_a, \mathbf{v}_b \right)
\]

\[
\cdot \left( \frac{\delta N_a}{m_a} \frac{d F_{a0} \left( \mathbf{v}_b \right)}{d \mathbf{v}_b} - \frac{\delta N_a}{m_b} \frac{d F_{a0} \left( \mathbf{v}_a \right)}{d \mathbf{v}_a} \right) d^3 \mathbf{v}_b
\]

\[
(82)
\]

depending on the tensor

\[
\mathbf{Q} \left( \mathbf{v}_a, \mathbf{v}_b \right) = -2q_a^2 q_b^2 \int_{R^3} \delta \left( k_{p,q,r} \right) \delta \left( \mathbf{k}_r \right) \delta \left( \mathbf{k}_r \right) \delta \left( k_{p,q,r} \right) \delta \left( k_{p,q,r} \right) \delta \left( \mathbf{k}_r \right) \delta \left( \mathbf{k}_r \right) \delta \left( \mathbf{k}_r \right)
\]

\[
\times \sum_{p,q,r} \delta \left( k_{p,q,r} \cdot \mathbf{v}_a - \omega_0 \mathbf{k}_r \cdot \mathbf{v}_a \right) \delta \left( k_{p,q,r} \cdot \mathbf{v}_b - \omega_0 \mathbf{k}_r \cdot \mathbf{v}_b \right)
\]

\[
\left( 2\pi \right)^4 \delta \left( \omega - k_{p,q,r} \cdot \mathbf{v}_a \right) \mathbb{E} \left\{ \delta \Theta_a \left( \omega, k, v_a \right) \right\}
\]

\[
(83)
\]

for 3D simulations (\(d^3 \mathbf{k}\) and the entire 3D space of integration \(R^3\) should be replaced by \((2\pi \omega) \cdot d\mathbf{k}\) and \(R\) or by \(2\pi \omega \cdot d\mathbf{k}\) and \(R^2\) for 1D or 2D simulations respectively). It tends to the LENARD–Balescu tensor for collisionless plasmas in the limit \(\forall \zeta \in \{x, y, z\}, \Delta \zeta \to 0\). According to assumptions b, d and the macroparticle shape function properties (6), we have necessarily

\[
\hat{n}_a \left( r_{i,j,k}, t_n \right) = \delta N_a \sum_{i,j,k} \mathbf{S} \left( r_{i,j,k} - r_{a,\ell} \right) \left( t_n \right) = \hat{n}_a
\]

\[
(84)
\]

and therefore

\[
\hat{f}_a \left( r_{i,j,k}, \mathbf{v}, t_n \right) = \hat{n}_a F_{a0} \left( \mathbf{v} \right) = \hat{f}_a \left( r_{i,j,k}, \mathbf{v}, t_n \right)
\]

\[
(85)
\]

Neglecting the expectation of the term depending on the internal tension forces of macroparticles (17), kinetic equations corresponding to an electrostatic PIC simulation follow immediately. They read

\[
\frac{\partial \hat{f}_a}{\partial t} + \frac{\partial}{\partial r} \cdot \left( \hat{f}_a \mathbf{v} \right) = \hat{n}_a F_{a0} \left( \mathbf{v} \right) = \hat{f}_a \left( r_{i,j,k}, \mathbf{v}, t_n \right)
\]

\[
(86)
\]

according to (25) and (82), in agreement with Langdon (1970b) and Birdsall and Langdon (1991). The component \((m, p, q, r) = (0, 0, 0, 0)\) in (82) is physical. As shown by Dawson (1983), it describes the average friction and diffusion of macroparticles in the fluctuating component of the electrostatic field which are direct consequences of the emission (Decyk 1987) and absorption (Landau 1946) of plasma waves by the macroparticles in the simulation. However, if each macroparticle represents more than one real particle such that \(\delta N_a = \delta N_a = \delta N \gg 1\), these physical processes are greatly overestimated by the weight factor \(\delta N_a\) compared to reality according to the expression of the ‘collision operator’ (82). While the component \((m, p, q, r) = (0, 0, 0, 0)\) of the latter cancels for single species 1D plasmas (Dawson 1964) and for plasmas at equilibrium for which \(\delta N_a = \delta N\), such that the diffusion of macroparticles compensates exactly their friction, this is never the case for its time aliases components \(m \neq 0\) and its spatial aliases \(p \neq 0, q \neq 0, r \neq 0\). Worst, they may even be at the origin of numerical instabilities (Langdon 1970b, Birdsall and Langdon 1991). Langdon (1970b), Birdsall and Langdon (1991) have shown that, in general, the ‘collision tensor’ (83) has the following properties

\[
\forall \mathbf{X}, \mathbf{Y} : \mathbf{Q} \left( \mathbf{v}_a, \mathbf{v}_b \right) \cdot \mathbf{X} < 0
\]

\[
\mathbf{Q} \left( \mathbf{v}_a, \mathbf{v}_b \right) = \mathbf{Q} \left( \mathbf{v}_b, \mathbf{v}_a \right)
\]

\[
(87)
\]

According to the kinetic equation (86), a consequence is that the average PIC simulation plasma entropy

\[
\mathbf{H} \left( r_{i,j,k}, t_n \right) = - \sum_a \int_{R^3} \hat{f}_a \left( r_{i,j,k}, \mathbf{v}_a, t_n \right) \text{ln} \hat{f}_a \left( r_{i,j,k}, \mathbf{v}_a, t_n \right) d^3 \mathbf{v}_a
\]

\[
(88)
\]
following average macroparticle straight line trajectories in space, necessarily increases

$$\left. \frac{dH}{dt} \right|_{\text{macroparticle}} > 0. \quad (89)$$

We conclude that whatever the velocity distributions $F_{\text{vel}}(v)$ used to initialize macroparticle velocities using a random number generator, if one lets the simulation run for a sufficiently long time duration, the PIC simulation plasma will necessarily tend on average to a quasi equilibrium state that maximizes entropy.

4. Discussion

4.1. Neglect of numerical consistency errors compared to statistical fluctuations

In this subsection, we discuss the simulation conditions for which assumption a is valid. First, (73) has to be compared with the numerical consistency errors associated with the leap-frog scheme (55) and with the Maxwell solver. For Maxwell spectral solvers, there is no numerical consistency error in space. The counterpart is the compulsory use of periodic boundary conditions. While it is not problematic here because we are considering an infinite and homogeneous plasma, it may become constraining under other circumstances. For the second order FDTD scheme, the numerical consistency error in space is given by

$$\varepsilon_s (r_{i,j,k}, t_n) = \varepsilon_{\text{MAX}} (r_{i,j,k}, t_n) = \frac{\Delta^3}{24} \left. \frac{\partial^3 E_\xi}{\partial \xi^3} \right|_{(r_{i,j,k}, t_n)} + O(\Delta^6), \quad (90)$$

The numerical consistency error on macroparticle locations due to the leap-frog scheme is

$$\Delta t \frac{d^3 r_{\text{num},t}}{dt^3} \bigg|_{t_n} + O(\Delta t^5). \quad (91)$$

Therefore, injecting it in Maxwell equations, we can estimate the consequent numerical consistency error on the electrostatic field as

$$\varepsilon_t (r_{i,j,k}, t_n) = 4\pi \frac{\Delta^3}{24} \sum_a Q_a n_a \sum_{\ell=1}^{N_t} \left. \frac{\partial E_\xi}{\partial t} \right|_{(r_{\text{num},t}, t_n)} S \times (r_{i,j,k} - r_{\text{num},t} (t_n)) + O(\Delta^5). \quad (92)$$

By normalizing time, space, electrical charge and mass according to $t = \omega_D t$, $r = r / \lambda_D$ where $\lambda_D = \lambda_D / \omega_D$ is the electron Debye screening length, $q_0 = q / e$ and $m_a = m_a / m_e$, respectively, and by assuming reasonably $Zm_e/m_i \ll 1$, we find the scalings

$$|\varepsilon_s| \approx \frac{\delta N}{4\pi n_e \Lambda D}, \quad |\varepsilon_t| \approx \frac{\delta N}{4\pi n_e \Lambda D} \quad (93)$$

where

$$\alpha = \max_{i,j,k} \left\{ \sum_{\ell=1}^{N_t} \left( \frac{\partial^3 E_\xi}{\partial \xi^3} \right) (r_{i,j,k}, t_n) \right\}$$

and

$$\beta = \max_{i,j,k} \left\{ \sum_{\ell=1}^{N_t} \left( \frac{\partial^3 E_\xi}{\partial \xi^3} \right) (r_{i,j,k}, t_n) \right\} \times \max_{i,j,k} \left\{ \frac{n_e (r_{i,j,k}, t_n)}{n_e} \right\}. \quad (94)$$

Assumption a is therefore valid if $|\varepsilon_s| \gg |\varepsilon_t|$ and $|\varepsilon_s| \gg |\varepsilon_s|$. Considering roughly $\alpha \sim \beta \sim 1$ and $\Delta t \sim \Delta n$, this is the case whenever

$$\frac{\Delta n^6}{\eta (2\pi / \Delta n) 24^4} \ll \frac{\delta N}{4\pi n_e \Lambda D} \quad (95)$$

which is not constraining in many cases. It means that statistical fluctuations play a crucial role in PIC simulations using a random number generator.

4.2. Stationarity of the plasma distribution function in PIC simulations

In this subsection, we discuss the simulation conditions for which assumption d is valid. It seems paradoxical that the right-hand side of kinetic equations (86) that we have obtained according to assumptions a–d never cancels while their left-hand side cancels according to assumptions a, b and d. The paradox is resolved if assumption d remains valid only on a time scale $\omega_D t_n$ during which the right-hand side of kinetic equations (86) is sufficiently small thus ensuring the distribution functions (32) remain homogeneous and stationary. Above a certain simulation time duration much greater than $\omega_D t_n$, the distribution functions begin to evolve with time. To estimate this simulation time threshold, let us consider the 3D electrostatic spectral PIC simulation of a plasma consisting of immobile ions and thermal electrons that uses equal weights $\delta N_e = \delta N_i = \delta N$ and a smoothing function filtering all spatial frequency alises. We also neglect time aliasing effects assuming a sufficiently small time step $\Delta t$ and we assume $\Delta_x = \Delta_y = \Delta_z = \Delta$ and $\alpha_x = \alpha_y = \alpha_z = a \ll \Delta$ such that $S (k) \approx \exp (-a^2 k^2 / 2)$. In this case, the PIC Coulomb logarithm, defined according to

$$\ln \Lambda (v_z) = \frac{\int_{k_{\text{min}}}^{k_{\text{max}}} \frac{\tilde{S} (k)}{k^2} \, dk}{k \varepsilon_L (k \cdot v_z, k)^4}, \quad (95)$$
only depends on macroelectron velocities \( \mathbf{v}_e \) if they have a velocity much greater than the thermal velocity \( |\mathbf{v}_e| \gg v_T^e \), implying the PIC simulation plasma permittivity is close to unity \( \varepsilon_L(k \cdot \mathbf{v}_e, k) \approx 1 \). Here, we have noted \( k = |k| \) and \( \kappa = k_0 = (\cos \theta, \sin \theta \cos \varphi, \sin \theta \sin \varphi) \). We have also introduced arbitrary cutoffs \( k_{\text{min}} \) and \( k_{\text{max}} \) that are not necessary for the convergence of the Coulomb logarithm contrary to the one obtained for point particles. In this particular case, the Lenard–Balescu-like tensor \((83)\) reduces to the Landau collision tensor

\[
Q(v_e, v_b) = -2e^2q_0^2 \frac{\ln \Lambda(v_e)}{|v_e - v_b|} \int_0^{2\pi} d\varphi \int_0^{\pi} \sin \theta d\theta \kappa \otimes \kappa \delta
\]

\[
= -2e^2q_0^2 \frac{\ln \Lambda(v_e)}{|v_e - v_b|} \left[ \mathbf{I} - \frac{(v_e - v_b) \otimes (v_e - v_b)}{|v_e - v_b|^2} \right]
\]

(96)

that describes small angle binary Coulomb collisions between point particles. It seems very strange to obtain such a collision integral since we have neglected close-encounter Coulomb collisions between macroparticles and consequently between real particles. In the textbook (Landau and Lifshitz 1981), they explain ‘it is sufficient to consider fluctuations in a collisionless plasma to calculate the collision integral because the important Fourier components of the electric field in collisions in plasmas are those with \(|k| \gg 1/\lambda_D \gg 1/\lambda_{\text{mfp}}\), so that collisions may be neglected’ where \( \lambda_{\text{mfp}} \) denotes real particle mean free paths due to close-encounter binary Coulomb collisions. As a consequence, ‘collisions’ are naturally taken into account in electrostatic PIC simulations of collisionless plasmas using a random number generator to initialize macroparticle velocities. Knowing from the kinetic theory of plasmas that the smallest relaxation rate is given by the electron–ion collision frequency, let us neglect in a first attempt the electron–electron fluctuations correlation term in the kinetic equation \((86)\). Considering then the limit \( Zn_e/m_e \ll 1 \), one can neglect the friction experienced by such electrons compared to their diffusion so that the right-hand side of the kinetic equation \((86)\) reduces to a Lorentz-like collision operator whose time scale is governed by the overestimated electron-ion collision frequency

\[
\nu_{ei}(v_e) = \delta N \frac{4\pi n_e Z_e^2 e^4}{m_e^2 |v_e|^3} \ln \Lambda(v_e) \lesssim \bar{\nu}_{ei} = \delta N \frac{4\pi n_e Z_e^2 e^4}{m_e^2 \bar{v}_{ei}^3} \ln \Lambda(\bar{v}_{ei}).
\]

(97)

We conclude that assumption \( d \) remains valid on a timescale \( t_n \) such that

\[
1 \ll \omega_p t_n \ll \frac{\omega_p}{\bar{\nu}_{ei}} \approx \frac{4\pi n_e \lambda_D^3}{Z \delta N \ln \Lambda(\bar{v}_{ei})} \approx \frac{\lambda_{\text{mfp}}}{\delta N \lambda_D}
\]

(98)

that is to say, for not too long electrostatic PIC simulations of ‘not too dense’ collisionless plasmas; ‘not too dense’ referring to the value of macroparticle weight factors \( \delta N \) constrained by today’s computer technology. Although overestimating them by the weight factor \( \delta N \), the term \((m, p, q, r) = (0, 0, 0, 0)\) of \((82)\) describes single-time correlations between the positions of real plasma particles due to their long-range Coulomb interactions. It is consequently not a numerical noise contrary to what is usually considered. Contrary to ‘not too dense’ collisionless plasmas, the distribution functions evolve faster than the statistical fluctuations correlation in electrostatic PIC simulations of ‘dense’ collisionless plasmas or collisional plasmas for which \( \omega_p/\bar{\nu}_{ei} \gtrsim 1 \). Assumption \( d \) is therefore not valid in this case. In addition, the physics of such fast evolution is not taken into account by electrostatic PIC codes because they underestimate close-encounter Coulomb collisions in this regime where \( 1/\lambda_D \approx 1/\delta N \lambda_{\text{mfp}} \). To model correctly this fast evolution of distribution functions, a Langevin force based on a Monte-Carlo model of multiple binary Coulomb collisions between real particles per time step \( \Delta t \), should be added into \((1)\) such as the one proposed by Nanbu (1997). Such a Monte-Carlo algorithm emulates statistical fluctuations due to binary Coulomb collisions between particles and ensures Bogoliubov hypothesis. However, these PIC simulations are out of the scope of this paper.

In order to give orders of magnitude, let us consider a laser-generated helium plasma \((Z = 2)\), heated to temperatures up to \( T_e \approx 1 \text{keV} \) and where the electron density is about \( n_e \approx 5 \times 10^{19} \text{cm}^{-3} \) \((\omega_p \approx 4 \times 10^{14} \text{rad}^{-1})\), the Debye length is about \( \lambda_D \approx 0.03 \mu \text{m} \). The resulting number of electrons in a Debye cube is about \( n_e \lambda_D^3 \approx 1776 \). If we choose reasonably \( \Delta t/\lambda_D = 0.01 \mu \text{m} \approx \lambda_D/3 \), we have necessarily \( N_{e,\text{mfp}}(\lambda_D/\Delta)^3 \approx 266 \) for assumption \( a \) to be valid. It leads to a maximum possible value of macroelectron weights \( \delta N_e \gtrsim 7 \) in 3D simulations for matching the correct plasma electron density. We deduce assumption \( d \) remains valid on a timescale \( \omega_p \bar{n}_e \ll 1670 \). Under such simulation conditions, assumption \( a \) is valid if one chooses \( \omega_p \Delta t/\lambda_D \lesssim \Delta/\lambda_D \) to ensure the numerical stability of the simulation. Indeed, \( \Delta^3/\eta(2\pi/\Delta) 2^{4/3} \approx 1 \times 10^{-15} \eta(2\pi/\Delta) \lesssim \Delta^3/4\pi n_e \lambda_D^3 \approx 3 \times 10^{-5} \) in this case. Considering a reasonable number of macroparticles simulated with today’s computer technology \( N_e \approx 1 \times 10^{11} \) using a highly parallelized PIC code and several \( 10^5 \) CPU hours with a computational cost typical of \( \approx 100 \text{ns/step/macroparticle} \) and a simulation time duration \( L_t = 1000/\omega_p \), we conclude that we are limited to plasma simulation boxes of \( L = L_t \approx 40 \mu \text{m} \) by using such a macroelectron weight \( \delta N_e \approx 7 \).

Performing the same reasoning on the main term considering the same assumptions for 2D electrostatic PIC simulations but defining a Coulomb logarithm

\[
\frac{\ln \Lambda^{3D}(v_e)}{\lambda_D} = \int_{k_{\text{min}}}^{k_{\text{max}}} \frac{S(k)}{\varepsilon_L(k \cdot \mathbf{v}_e, k)^7} dk
\]

(99)

instead, we find assumption \( d \) remains valid on a timescale \( t_n \) such that

\[
1 \ll \omega_p t_n \ll \frac{2\pi n_e \lambda_D^2}{Z \delta N \ln \Lambda^{3D}(\bar{v}_{ei})}
\]

(100)

in agreement with Montgomery and Nielsen (1970), Hockney (1971) and Hockney and Eastwood (1981). For 1D electrostatic PIC simulations, one finds the main term \((m, p) = (0, 0)\)
is identically zero owing to the Dirac delta function similarly to the 1D Lenard–Balescu or the 1D Landau collision terms in agreement with Dawson (1964). Indeed, according to the latter, the observed relaxation time
\[
\frac{\omega_p}{\nu_{1D}^{\delta N}} = \frac{1}{10} \left( \frac{\bar{n}_e \lambda_D}{\delta N} \right)^2
\]
(101)
of 1D electrostatic PIC simulation plasmas is due to the three-body collisions between macroparticles and we should have considered the fluctuation equations (29) and (30) without neglecting the quadratic terms (assumption c) to obtain such a result. However, the aliasing terms do not cancel in agreement with Birdsall and Langdon (1991) and Montgomery and Nielson (1970). One finds, for example, a relaxation time
\[
\frac{\omega_p}{\nu_{1D}^{(0,1)}} = \frac{2(k_b \lambda_D)^2 V \bar{n}_e \lambda_D}{\nu_T} \delta N
\]
(102)
where
\[
\frac{1}{\bar{V}_v} = \frac{1}{V_e} + \frac{Zm_e/m_i}{V_i}
\]
and
\[
\frac{1}{\bar{V}_b} = \int_0^\infty \frac{S(k_b \lambda_D)}{\bar{V}_v} \left( 1 - \frac{v_{e,x}}{v_{b,x}} \right)^2 \frac{F_{b,x}(v_{b,x})}{v_{b,x}} dv_{b,x}
\]
considering only the spatial aliasing term \((m, p) = (0, 1)\) when deriving the collision operator (82) that reads
\[
- \frac{q_e}{m_e} \frac{\partial}{\partial v_e} \mathbb{E} \{ \delta E_v (r_{i,j,k}, t_n) \delta f_{e,v} (r_{i,j,k}, v, t_n) \} \\
\approx \omega_{pee} \frac{\partial}{\partial v_{e,x}} \left( \nu_{1D}^{(0,1)} V_{T}^2 \frac{\partial f_{e,v}}{\partial v_{e,x}} \right)
\]
(103)
in agreement with Birdsall and Langdon (1991) and the empirical diffusion term proposed by Virtamo and Tuomisto (1979). As a conclusion for 1D electrostatic PIC simulations, we conclude assumption d remains valid on a timescale \(t_n\) such that
\[
1 \ll \omega_{p,T_n} \ll \omega_{p,T}^{1D}
\]
(104)
with \(\nu_{1D}^{1D}\) scaling as (102) for a large number of macroparticle centers per Debye length and, as (101) for a small number of macroparticle centers per Debye length, as already suggested by Birdsall and Langdon (1991).

4.3. Evolution of the empirical entropy in electrostatic PIC simulation plasmas

Landau damping of electrostatic waves is a time-reversible process (1946). Decyk (1987) has shown that the emission of electrostatic wakefield is also a time-reversible process. Therefore, it may seem paradoxical that we concluded that whatever velocity distributions \(F_v (v)\) that are used to initialize macroparticle velocities using a random number generator, if one lets the simulation run for a sufficiently long duration, the PIC-simulated plasma will necessarily tend on average to a quasi-equilibrium state that maximizes the entropy. The physical reason for this entropy increase is the average distribution function entropy increase: the correlations of deterministic and time-reversible statistical fluctuations around the average solution tends to make the average solution be the one that maximizes the entropy. Consequently, even if they are time-reversible, the statistical realizations also tend to the solution that maximizes entropy since they are supposed to be close to the average result in the limit \(N_{a,mpc} \gg 1\). Let us emphasize here that the macroscopic PIC-simulated plasma entropy (88) is not the empirical PIC-simulated plasma entropy
\[
\frac{H_{\pi} (r_{i,j,k}, t_n)}{\Pi_{\Delta} (r_{i,j,k}, v_{i,m,n}, t_n) \ln f_{a,T_n}} \times (r_{i,j,k}, v_{i,m,n}, t_n) \Delta v^3
\]
(105)
that one can compute directly from the PIC simulation by discretizing the velocity space according to bins of volume \(\Delta v^3\), located between velocity grid points \(v_{i,m,n} = \Delta v n\), \(\Delta v = (\Delta v_1, \Delta v_2, \Delta v_3)\) and by counting the number of macroparticles contained inside each bin
\[
f_{a,T_n} (r_{i,j,k}, v_{i,m,n}, t_n) = \int_{\mathbb{R}^3} d^3 \mathbf{v} f_a (r_{i,j,k}, v, t_n) \Pi_{\Delta}^0 (v - v_{i,m,n})
\]
(106)
Indeed, by differentiating the latter empirical distribution functions \(f_{a,T_n} (r, v, t)\) with time, one obtains the kinetic equation
\[
\frac{\partial f_{a,T_n}}{\partial t} + \frac{\partial}{\partial v} \cdot (\Pi_{\Delta}^0 (v - v_{i,m,n}) f_{a,T_n} ) = C_a' (r, v, t)
\]
(107)
where
\[
C_a' (r, v, t) = C_a (r, v, t) - \frac{\partial}{\partial v} \left[ \delta N_a \sum_{\ell=1}^{N_{\ell}} \left[ v_{a,\ell} (t) - v \right] S (r - r_{a,\ell} (t)) \Pi_{\Delta}^0 (v - v_{a,\ell} (t)) \right]
\]
(108)
according to the continuous macroparticle equations of motion. If \(\Delta v_1 = \Delta v_2 = \Delta v_3 = \Delta v\) is chosen sufficiently small such that there is a maximum of one macroparticle per bin volume, meaning \(\Pi_{\Delta}^0 (v - v_{i,m,n}) \rightarrow \delta (v - v_{i,m,n})\)
and \( f_{a,i}(\mathbf{r}_{i,j,k}, \mathbf{v}_{i,m,n}, t_n) \rightarrow f_a(\mathbf{r}_{i,j,k}, \mathbf{v}_{i,m,n}, t_n) \), the discretized empirical entropy (105) is only determined by the right-hand side (17) of Klimontovich-like equations (15) describing the effects of macroparticles internal tension forces. However, in practice, we do not have sufficient computer memory and there is always more than one macroparticle per velocity bin volume. As a consequence, we lose necessary information about the simulation results and the second term in the right-hand side (108) of equation (107) cannot be neglected. Finally, let us stress here that, if one does not use a random number generator to initialize macroparticle velocities \( \mathbf{V}_{a,t} \) at the simulation start \( t_1 \) by initializing, for example, a cold plasma where \( \forall a, t, \mathbf{V}_{a,t} = 0 \) instead, all equations are fully deterministic and there are no statistical fluctuations. For all PIC-simulated plasma quantities \( F, \mathbb{E}\{F\} = F \) and the kinetic equation (33) become meaningless. Only the deterministic description (15) applies in this case.

5. The ensemble averaging technique

5.1. Reduction of statistical fluctuations via ensemble averaging

Let us consider \( N_s \) PIC simulations of the same stationary, homogeneous, infinite and fully ionized plasma that we have studied in the previous section 3, the simulations differing only by the randomly generated numbers used to initialize macroparticle velocities. We note for all PIC simulation plasma quantity \( F \in \{ \delta (v - \mathbf{V}_{a,t} \cdot \mathbf{n}_a), f_a, f_{a,i}, \delta t_a, n_a, n_{a,i}, \delta n_a, f_a, f_{a,i}, \delta t_a, n_a, n_{a,i}, \delta n_a, \mathbf{E}, \mathbf{E}_s, \mathbf{E}_r, \delta \mathbf{E}_s, \delta \mathbf{E}_r \} \),

\[
\mathbb{E}\{F\} = \frac{1}{N_s} \sum_{j=1}^{N_s} F^{(j)} \rightarrow \mathbb{E}\{F\} \quad (109)
\]

its finite ensemble average is obtained from all PIC simulation results, each one denoted by the superscript \( (j) \), \( j \in [1, N_s] \). By linearity, the expected value of the ensemble average of PIC simulation plasma quantities is also their expected value

\[
\mathbb{E}\{\mathbb{E}\{F\}\} = \mathbb{E}\{F\} \quad (110)
\]
during one single PIC simulation. We thus have in particular

\[
\mathbb{E}\{f_a(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n)\} = f_a(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \quad (111)
\]

and

\[
\mathbb{E}\{\mathbf{E}(\mathbf{r}_{i,j,k}, t_n)\} = \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \quad (112)
\]

As a consequence, the deterministic description of the finite ensemble average of PIC simulations consists of the kinetic equation

\[
\frac{\partial \langle f_a \rangle}{\partial t} (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) + \frac{\partial}{\partial \mathbf{v}} \cdot \langle \langle \mathbf{F} \rangle \rangle (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) + \frac{\partial}{\partial \mathbf{v}} \cdot \langle \langle \mathbf{F} \rangle \rangle (\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) = - \frac{q_o}{m_a} \frac{\partial}{\partial \mathbf{v}} \left\{ \langle \delta \mathbf{E} \rangle (\mathbf{r}_{i,j,k}, t_n) \delta f_a(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \right\} + \langle C(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \rangle
\]

self-consistently coupled with the finite ensemble average of discretized Maxwell equations

\[
\left\{ \begin{array}{l}
\frac{\partial}{\partial \mathbf{v}} \cdot \langle \mathbf{E} \rangle \mid^{i,j,k,n} = 4\pi \sum_a \int_{\mathbf{E}_s, \mathbf{E}_r, \delta \mathbf{E}_s, \delta \mathbf{E}_r} d^3 \mathbf{v}
\frac{\partial}{\partial \mathbf{v}} \times \langle \mathbf{E} \rangle \mid^{i,j,k,n} = 0
\end{array} \right. \quad (114)
\]

according to (15), (16) and (109). Because all PIC simulations use different randomly generated numbers to initialize macroparticle velocities, they are all statistically independent. We thus deduce the relationship between the finite ensemble average of PIC simulation plasma quantity variances with their own variances from one single PIC simulation. For all PIC simulation plasma quantity \( G \in \{ n_{a,i}, n_{a,i}, \delta n_a, n_{a,i}, \delta n_a, f_a, f_{a,i}, \delta t_a, n_a, n_{a,i}, \delta n_a, \mathbf{E}, \mathbf{E}_s, \mathbf{E}_r, \delta \mathbf{E}_s, \delta \mathbf{E}_r \} \), we have

\[
\mathbb{E}\left\{ \left( G - \mathbb{E}\{G\} \right)^2 \right\} = \frac{1}{N_s} \mathbb{E}\left\{ \left( G - \mathbb{E}\{G\} \right)^2 \right\} \quad (115)
\]

We thus obtain in particular

\[
\mathbb{E}\left\{ \langle \delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle^2 \right\} = \frac{\mathbb{E}\left\{ \delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \right\}^2}{N_s} \quad (116)
\]

If \( N_s \) is sufficiently large, the finite ensemble average of electrostatic field statistical fluctuation \( \langle \delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle \) at grid points \( r_{i,j,k} \) and time steps \( t_n \) follows a normal distribution law with mean (112) and variance (116) according to the central limit theorem. One deduces

\[
\left| \langle \delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle \right| \leq \sqrt{2} \text{erf}^{-1} \left( \frac{P}{100} \right) \left| \mathbf{F}_{\text{fluct}} \right| \quad (117)
\]

with a \( p\% \)-confidence where we have noted

\[
\left| \mathbf{F}_{\text{fluct}} \right| = \left| \frac{\mathbf{F}_{\text{fluct}}}{\sqrt{N_s}} \right| \quad (118)
\]

assuming again assumptions a–d as in section 3.2. Considering again the particular case of PIC simulations of a collisionless plasma at equilibrium using equal weights \( \delta N_c = \delta N_i = \delta N \) and \( \Delta t / \Delta x / \sqrt{\nu T} = \Delta y / \sqrt{\nu T} = \Delta z / \sqrt{\nu T} < 1 / \omega_p \), it reads

\[
\left| \mathbf{F}_{\text{fluct}} \right| \sim \frac{\sqrt{\delta N \kappa_B T}}{\omega_p} \left( \frac{N_s}{N_d} \right) \left( k_B \lambda_D \right) \quad (119)
\]
A finite ensemble average of \( N_e \) PIC simulations thus reduces statistical fluctuation amplitudes of electrostatic fields and densities by a factor \( \sqrt{N_e} \) as compared to one single electrostatic PIC simulation. Also, since \( \delta N \propto 1/N_{e,mpc} \) according to its definition (7), we deduce there is an equivalence between running one PIC simulation using \( N_{e,mpc} \) macroparticles and a finite ensemble average of \( N_{e,mpc}/N'_{e,mpc} \) PIC simulations using each one \( N'_{e,mpc} \) macroparticles per cell. We can thus apply the central limit theorem to estimate the statistical fluctuations in one single PIC simulation using a sufficiently large number of macroparticles per cell \( N_{e,mpc} \gg N'_{e,mpc} \geq 2 \) in 1D, 4 in 2D and 8 in 3D as we did in section 3.2.

5.2. Kinetic equation for an ensemble average of electrostatic PIC simulations

Similarly, if we note \( F' \), another PIC simulation plasma quantity distinct from \( F \), the statistical independence of PIC simulations makes the finite ensemble average of PIC simulation plasma quantity covariance between \( F \) and \( F' \) decrease by a factor \( \sqrt{N_e} \) compared to their covariance from one single PIC simulation

\[
\mathbb{E}\left\{(F) - \mathbb{E}\{F\}\right\}\left\{(F') - \mathbb{E}\{F'\}\right\}
= \frac{1}{N_e}\mathbb{E}\left\{(F - \mathbb{E}\{F\})\left\{(F' - \mathbb{E}\{F'\}\right\}\right\}. \tag{120}
\]

We thus obtain in particular

\[
\mathbb{E}\left\{(\delta f_{a}(\mathbf{k}, \mathbf{v}, t_1))\left\{(\delta f_{b,c}(\mathbf{k}', \mathbf{v}', t_1))\right\}
= \frac{\delta N_{e}}{N_e} \left(2\pi\right)^3 \hat{n}_a \delta (\mathbf{k} + \mathbf{k}', \mathbf{k}_b) F_{ab}(\mathbf{v})
\times \left|\delta (\mathbf{v} - \mathbf{v}') - F_{b0}(\mathbf{v})'\right|. \tag{121}
\]

Therefore, by defining statistical fluctuations

\[
\langle \delta f_{a}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \rangle = \langle f_{a}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \rangle - \langle \bar{f}_{a}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \rangle \tag{122}
\]

and

\[
\langle \delta \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle = \langle \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle - \langle \bar{\mathbf{E}}(\mathbf{r}_{i,j,k}, t_n) \rangle \tag{123}
\]

between the finite ensemble average of PIC simulation results and the expected ones, by performing the same derivation as in section 3.3, using (121) instead of (70) and assuming again assumptions a–d, we obtain the probabilistic description of the finite ensemble average of PIC simulations. The kinetic equation reads

\[
\langle \frac{\partial \bar{f}_{a}}{\partial t} \rangle(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) + \frac{\partial}{\partial \mathbf{v}} \left( \mathbf{F}_{a} \right) = \frac{\partial}{\partial \mathbf{v}} \left( \bar{\mathbf{F}}_{a} \right)
+ \frac{\partial}{\partial \mathbf{v}} \cdot \left( \frac{q_{a}}{m_a} \left( \bar{f}_{a}(\mathbf{r}_{i,j,k}, t_n) - \bar{f}_{a}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \right) \right)
\times \left( \frac{\delta N_{a}}{N_a} \frac{\bar{f}_{a}(\mathbf{v})}{m_a} \frac{\delta N_{e}}{N_e} \bar{f}_{e}(\mathbf{v}) \right) \frac{\partial \bar{f}_{a}}{\partial \mathbf{v}}
\times (\mathbf{r}_{i,j,k}, t_n) d^3 \mathbf{v}. \tag{124}
\]

It is self-consistently coupled with the ensemble average of expected discretized Maxwell equations

\[
\begin{align*}
\frac{\partial}{\partial t} \langle \mathbf{E} \rangle | t_{i,j,k} &= 4 \pi q_{a} \int_{\mathbb{R}^3} \left\langle \mathbf{f}_{a}(\mathbf{r}_{i,j,k}, \mathbf{v}, t_n) \right\rangle d^3 \mathbf{v} \\
\frac{\partial}{\partial t} \times \langle \mathbf{E} \rangle | t_{i,j,k} &= 0 \tag{125}
\end{align*}
\]

via the ensemble average of expected electrostatic field interpolated at macroparticle center locations

\[
\langle \mathbf{E}_{e}(\mathbf{r}_{a,\ell}(t_n), t_n) \rangle = \sum_{i,j,k} \langle \mathbf{E}(\mathbf{r}_{i,j,k}, t_n) \rangle S
\times (\mathbf{r}_{a,\ell}(t_n) - \mathbf{r}_{i,j,k}) \Delta_{x} \Delta_{y} \Delta_{z}. \tag{126}
\]

Consequently, the kinetic equations (124) of the ensemble average of distribution functions tend to the Vlasov equation in the limit of an infinite number of simulations \( N_e/\delta N \to \infty \). More realistically, the ensemble average of \( N_e = \delta N \) independent electrostatic PIC simulations modeling the same macroscopic plasma with macroparticle weights \( \delta N = \delta N_{e} = \delta N_{i} \) is strictly equivalent to the numerical resolution of the Vlasov–Lenard–Balescu/Maxwell set of equations discretized in space and time only. Let us mitigate here these two important results (118) and (124) by reminding the reader that they are only valid assuming assumptions a–d, that is, only for simulation time durations \( L_t \) of collisionless plasmas using sufficiently small cell sizes and time steps such that

\[
\frac{(\omega_{pe} \Delta t)^6}{\eta(k_{p} \lambda_{D})24^2} \ll \frac{\delta N/N_{e}}{4\pi \hat{n}_{e} \lambda_{D}^2} \ll \frac{1}{Z_{\omega_{pe} L_t}} \tag{127}
\]

similarly to that obtained in the previous section 4.

Precautions have been taken in the previous paragraphs to distinguish between PIC simulation plasma quantities \( F, F' \) and \( G \). Indeed, the variance of plasma particle phase-space densities at the grid point \( \mathbf{r}_{i,j,k} \) and velocity \( \mathbf{v} \) naturally diverges due to the contribution of the term expressing the case where there is only one macroparticle in the phase-space volume located between \( \mathbf{r}_{i,j,k}(\mathbf{v}) \) and \( \mathbf{r}_{i,j,k} + \Delta_{\mathbf{v}} \mathbf{v} + d^3 \mathbf{v} \). To the best of our knowledge, there is no unbiased estimator of distribution function statistical realizations. However, to troubleshoot this issue, one may use the biased statistical estimator introduced in section 3.3 for the empirical distribution function (106) to express
where $\varDelta v = \varDelta v_x = \varDelta v_y = \varDelta v_z$ for simplicity, one obtains

$$
\tilde{f}_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t) = \tilde{E} \left\{ f_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t_1) \right\}
= \frac{n_a}{N_{a,\text{mpc}}} \varDelta_v \sum_{i,j,\xi} \Pi_{ij}^{(0)} (v_{l,m,n} - v_{a,i,j,\xi})
$$

(128)

according to assumption b. Assuming $\varDelta v = \varDelta v_x = \varDelta v_y = \varDelta v_z$, we deduce

$$
\tilde{f}_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t) = \tilde{E} \left\{ f_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t_1) \right\}
= \frac{n_a}{N_{a,\text{mpc}}} F_{a0} (v_{l,m,n}) + n_a \varDelta v \frac{2}{\varDelta_v} \sum_{i,j,\xi} \frac{d^2 F_{a0}}{dv_{i,j,\xi}}
$$

(129)

Therefore, in the limit of a large number of macroparticles per cell, the statistical estimator of particle phase-space densities (128) follows a normal distribution law with mean (129) and variance (130) according to the central limit theorem. Assuming the PIC simulation plasma is stable, we deduce in the limits $\omega_p t_0 \gg 1$, $\varDelta v \ll v_T$, $N_{a,\text{mpc}} \gg 1$ and $N_s \gg 1$

$$
|\delta f_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t_1)| \leq \frac{n_a}{N_{a,\text{mpc}}} \varDelta v \frac{2}{\varDelta_v} \sum_{i,j,\xi} \frac{d^2 F_{a0}}{dv_{i,j,\xi}} (v_{l,m,n})
$$

(131)

and

$$
|\delta f_{a,\Pi} (r_{ij,k}, v_{l,m,n}, t_1)| \leq \frac{n_a}{N_{a,\text{mpc}}} \varDelta v \frac{2}{\varDelta_v} \sum_{i,j,\xi} \frac{d^2 F_{a0}}{dv_{i,j,\xi}} (v_{l,m,n})
$$

(132)

with a $p\%$-confidence. Statistical fluctuations of phase-space densities depends consequently on a scaling different from $1/\sqrt{N_{a,\text{mpc}}}$ in single PIC simulations. Indeed, by using velocity bin sizes $\varDelta v \propto v_T / N_{a,\text{mpc}}$ that decrease with the increasing number of macroparticles per cell, the optimal choice minimizing the bias is $n = 1/(4 + \text{dim})$ where ‘dim’ denotes here the PIC simulation velocity space dimension. It provides instead a scaling $1/\sqrt{N_{a,\text{mpc}}^{2/(4 + \text{dim})}}$ for single PIC simulations and $1/\sqrt{(N_{a,\text{mpc}} \times N_s)^{2/(4 + \text{dim})}}$ for a finite ensemble average of $N_s$ PIC simulations.

### 5.3. Application to linear Landau damping

To illustrate how to decrease the amplitude of statistical fluctuations in PIC simulations by ensemble averaging, we consider here the linear Landau damping of a small amplitude electrostatic wave propagating in the 1D-3V PIC simulation plasma described in section 2. In order to generate the electrostatic wave, we perturb the PIC simulation plasma by driving an electrostatic field $E_0(x, t_1) \leq 0 = E_0 \cos k_0 (x_1 - v_0 (t_1 + \tau)) e_x$ from $\omega_p t_0 = -\omega_p \tau = -3.6$ to $\omega_p t_0 = 0$ at a spatial angular frequency $k_0 \lambda_D = 0.319$. Landau (1946) predicted that if particles move in a collisionless plasma at a velocity $v_x$ close to the wave phase velocity $v_p = \omega / k$ of an electrostatic wave, they see an almost constant electrostatic field and can thus interact with the wave. If $v_x \ll v_p$, trapped particles earn energy from the wave and if $v_x \gg v_p$, they lose energy to it. Consequently, if one considers a PIC simulation plasma consisting of immobile ions and thermal electrons with an initial Maxwellian distribution of velocities (8), more macroelectrons will earn energy from the wave than the electrostatic wave from macroelectrons since $\partial F_{e0} / \partial v_i (v_i > 0) < 0$ in this case. As a result, the driven electrostatic wave is linearly damped at the Landau damping rate $\gamma_L$ in a first-order approximation. If one uses a Gaussian smoothing function (5) and chooses a sufficiently small time step $\omega_p \varDelta t = 0.2$, aliasing effects are small in the expression of the PIC simulation plasma permittivity (63) and one may deduce from the resulting dispersion relation $\omega (\omega - i \gamma_L, k) = 0$

$$
\frac{\omega (k)}{\omega_p} \sim \frac{S (k)}{k \lambda_D} \left[ 1 + \frac{3}{2} (k \lambda_D)^2 \right]
$$

(133)

and

$$
\frac{\gamma_L (k)}{\omega_p} \sim \frac{S (k)}{8 (k \lambda_D)^2} \exp \left[ -\frac{1}{2 (k \lambda_D)^2} \right]
$$

(134)

in the limit $k \lambda_D \ll 1$. Actually, $\omega_L (k_0) \approx 1.18 \omega_p \bar{S} (k_0)$ and $\gamma_L (k_0) \approx 1.96 \times 10^{-2} \omega_p \bar{S} (k_0)$ at $k_0 \lambda_D = 0.319$ if one solves numerically the dispersion relation; cf the left panel of figure 3. By using spatial grid spacings $\varDelta x = \lambda_D$, a quartic interpolating function $\Pi_{\varDelta x} (\delta)$ and a smoothing parameter $\alpha_s = 0.866667 \lambda_D$ such that we lose the less physical information from spatial angular frequency spectra as possible, the macroparticle shape (4) is $\tilde{S} (k_0) \approx 0.903$ at the considered spatial angular frequency $k_0 \lambda_D = 0.319$ and it does not affect too much the theoretical results.

As macroelectrons damp the wave, they also earn energy. As a result, at the end of this linear stage more macroelectrons finish with a velocity $v_x \approx \bar{v}_p$ compared to those who finish with a velocity $v_x \approx \bar{v}_p$, so that the sign of the resulting distribution of electron velocities changes and macroelectrons give back the received energy to the wave.
By studying the energy conservation between the electrostatic wave and one trapped macroelectron, one deduces the nonlinear Landau damping occurs at the bounce angular frequency

$$\omega_b(k) = \sqrt{\frac{eE_0}{m_e}}$$  \hspace{1cm} (135)

where $E_0$ is the electrostatic wave amplitude. Consequently, the condition for observing the linear Landau damping of an electrostatic wave is given by the scaling $\omega \gg \gamma \gg \omega_b$, and the electrostatic wave amplitude is consequently constrained according to

$$\frac{eE_0}{m_0\omega_b v_T} \ll \frac{(\gamma/\omega_p)^2}{k\lambda_D} \ll \frac{(\omega/\omega_p)^2}{k\lambda_D}.$$  \hspace{1cm} (136)

Plots in figure 3 stop at the spatial angular frequency of $k\lambda_D = 1$. Indeed, above this value, the linear Landau damping rate has the same order of magnitude as the angular frequency of electrostatic wave oscillations. Therefore, the concept of electrostatic waves is meaningless in this regime. In the opposite regime where angular spatial frequencies $k\lambda_D \lesssim 0.2$, the linear Landau damping rate decreases drastically with decreasing spatial angular frequencies by many orders of magnitudes according to equation (134). In this regime, Landau damping is negligible and electrostatic waves may consequently propagate almost freely in the PIC simulation plasma. In the right panel of figure 3, we compare the Landau damping rate $\gamma_L(k)$ and bounce angular frequencies (135) for different values of initial perturbing electrostatic field amplitude $E_0$ in the interesting regime of Landau damping $0.2 \lesssim k\lambda_D \lesssim 1$. While the observation of nonlinear Landau damping of a perturbing electrostatic wave is relatively easy, the observation of linear Landau damping is much more challenging due to the smaller initial amplitude of the perturbing electrostatic field that is needed and that is usually below the PIC simulation plasma electrostatic field statistical fluctuation amplitudes (49). In our case, $k_0\lambda_D = 0.319$ leads to a value of $(\gamma_L/\omega_p)^2/k_0\lambda_D \approx 1.2 \times 10^{-3}$. In a first approximation and neglecting statistical fluctuations, the self-consistent PIC simulation-driven plasma electrostatic wave increases linearly with time $E_1(t, t_0 = 0) \approx E_0\alpha(t + \tau) \sin [k_0x - \omega_0(t + \tau))]$, under the action of the driver where $\alpha$ is the growing rate of the field. In order to give orders of magnitude, let us consider a laser-generated hydrogen plasma ($Z = 1$) at equilibrium with an electron density $n_e \approx 5.4 \times 10^{19} \text{ cm}^{-3}$ consisting of immobile ions ($m_e/m_i \ll 1 \Rightarrow v_i \ll v_T$) and thermal electrons at a temperature $T_e \approx 7 \text{ keV}$. For such a 1D collisionless plasma, the number of electrons in a Debye cube is about $n_e\lambda_D^3 \approx 32768$ and the Debye length about $\lambda_D \approx 0.085 \mu\text{m}$. We perform the simulations by using the open source spectral electrostatic PIC code developed from the UPIC framework (Decyk et al. 2007) available at https://github.com/UCLA-Plasma-Simulation-Group. We use a macro-electron weight $\delta N_e = 1$, a number of macroelectrons $N_e = 8388608$ and a simulation box $L_x = 256\lambda_D$ leading to a number $N_{e,mpc} = 32768$ of macro-electrons per cell and $k_0\lambda_D = 2\pi p_0/L_x$ with $p_0 = 13$. The choice of a macro-electron weight $\delta N_e = 1$ allows us to emphasize again that physical statistical fluctuations are also present in dedicated experiments and they are not a numerical noise. The input deck for single simulations can be found in appendix D. Given all these simulation parameters, we have

$$\frac{4\pi(\omega_p \Delta)x^6}{\arctan(k_0\lambda_D)^{242}} \approx 9.8 \times 10^{-7} \ll \frac{1}{4\pi n_e\lambda_D^3}$$

$$\approx 2.4 \times 10^{-6} \ll \frac{1}{\omega_p v_T} = 5 \times 10^{-3} \hspace{1cm} (137)$$

with $L_x = 200/\omega_p$ ($N_e = 1000$ time iterations). It means, according to the previous section 2, that all our estimates obtained in sections 3–5 are valid for these simulations.
particular, consistency errors due to the leap-frog scheme are smaller than statistical fluctuation amplitudes
\[
\left| \delta E\left(x_i, t_n\right) \right| \leq \frac{\sqrt{2} \text{erf}^{-1} \left( \frac{P}{100} \right)}{N_{\text{exp}}^{\gamma_p}} \left| E_{\text{fluct}} \right|
\]
(138)

with \( e \left| E_{\text{fluct}} \right| / m_e \omega_p v_T \approx 5 \times 10^{-4} \) and a \( p\%\)-confidence. We choose \( eE_0 / m_e \omega_p v_T = 10^{-3} \). According to simulations, we measure a growing rate of the field \( \alpha \approx 0.243 \omega_p \); cf figure 4. At \( t_n = 0 \), the driver is stopped and we expect the driven traveling electrostatic wave is Landau-damped according to
\[
E_t\left(x_i, t_n \geq 0 \right) \approx E_0 \sin \left[k_0 x_i - \omega_0 (t_n + \tau) \right] \exp \left(-\gamma_L t_n \right) e_x
\]
(139)

with \( eE_0 / m_e \omega_p v_T \approx eE_0 \alpha / m_e \omega_p v_T \approx 8.75 \times 10^{-4} \). It is just slightly above the statistical fluctuations level and leads to a bounce frequency \( \omega_{\text{b}} = \omega_0 (k_0) = 1.67 \times 10^{-2} \) slightly lower than the linear Landau damping rate \( \gamma_L \) in agreement with the simulations; cf figure 3. The observation of linear Landau damping is therefore difficult to observe in such electrostatic PIC simulations or dedicated experiments. A method proposed by Decyk (1987) has already been applied by Grismayer et al (2011) for resolving this problem. The subtraction technique consists in running two identical PIC simulations that use the same randomly generated numbers to initialize macroparticle velocities at the simulation start. Then, the nominal simulation without perturbation is subtracted from the one initialized with the small amplitude perturbing electrostatic field. Here, we use the ensemble average of \( N_e = 30 \) simulations and we compare the different results. In the left hand panel of figure 4, we have plotted the spatial mode \( p_0 \) of electrostatic field discrete Fourier transform
\[
\tilde{E}_t\left(k_0, t_n \right) = \frac{1}{N_t} \sum_{i=1}^{N_t} E_t\left(x_i, t_n \right) \exp \left(-i2\pi p_0 x_i / L_s \right)
\]
(140)

from one single simulation and from their ensemble average. We can see statistical fluctuation amplitudes (138) are effectively reduced by a factor \( \sqrt{30} \approx 5.7 \) in agreement with our estimate (118). After \( \omega_p t_n \approx 20 \), one single PIC simulation does not allow for the observation of the tiny electrostatic field amplitude (136) below statistical fluctuations (138) while the electrostatic wave amplitude continues to decrease until \( \omega_p t_n \approx 100 \) where it reaches the reduced statistical fluctuation amplitudes \( |E_{\text{fluct}}| \approx 9 \times 10^{-5} \). An advantage of the ensemble averaging technique is that simulations do not need to be parallelized compared to an equivalent but more-computationally-extensive simulation using \( N_e \) times more macroparticles and macroparticle weights \( \delta N_e / N_e \). In the right panel of figure 4, the ensemble average results is compared with the subtraction technique and the theoretical result
\[
\tilde{E}_t\left(k_0, t_n \right) \approx \frac{E_0}{2} \exp \left(-\gamma_L t_n \right).
\]
(141)

We find good agreements between the two different methods and the Vlasov/Maxwell theory.

6. Conclusions

PIC simulations are one of the most widely used and most powerful tools in plasma physics, being a core component and methodology of plasma physics as a scientific discipline. However, the statistical physics principles of a PIC code (and what is actually computed) are still a source of confusion among many users and developers. Conflicting views assume that a PIC code evolves the macroscopic Vlasov–Maxwell set of equations, others consider that PIC simulations describe the microscopic Klimontovich/Maxwell set of equations, while others claim that PIC simulations cannot be compared to (or used with) the kinetic theory of plasmas. In this paper, we clarify the kinetic properties of PIC simulations. We have derived the deterministic description of a PIC simulation electrostatic plasma, or PIC plasma. This description

---

**Figure 4.** (Left) Comparison of linearly Landau-damped electrostatic wave spatial Fourier mode \( p_0 \) from one single electrostatic 1D-3V PIC simulation (black curve) and an average of 30 simulations (red curve). (Right) Comparison of linearly Landau-damped electrostatic wave spatial Fourier mode \( p_0 \) from Landau theory (green curve), from an average of 30 electrostatic 1D-3V PIC simulations (red curve) and from the subtraction of one nominal simulation without perturbation from one PIC simulation with perturbation (blue curve).
consists of a Klimontovich-like equation discretized in space and time, but not in velocity space, with a source term (in the right hand side of the Klimontovich-like equation) due to the internal tension force of macroparticles. In the left hand side of the Klimontovich-like equation, the collective electrostatic force interpolated at macroparticle locations is self-consistently coupled with the discretized Maxwell equations describing the evolution of the (approximated) microscopic fields. The source terms of the discretized Maxwell’s equations are obtained by depositing the electrical charge and current of macroparticles onto the spatial grid such that the Coulomb electrostatic fields generated by macroparticles are reduced at distances smaller than the spatial grid spacing. If the minimum possible distance between particles due to their binary electrostatic interaction (the Landau length) is not resolved, then close-encounter binary Coulomb collisions between macroparticles are underestimated in PIC codes. To understand the statistical properties of PIC plasmas, we have applied the theory of fluctuations in collisionless plasmas derived by Rostoker (1961), Klimontovich and Silin (1962), and Dupree (1963) starting from the statistical fluctuations typically implemented by the random sampling used to initialize the velocities of macroparticles. By neglecting numerical consistency errors from the numerical schemes, we used the exact discrete Laplace–Fourier transform of the PIC kinetic equations coupled with the exact discrete Laplace–Fourier transform of self-consistent discretized Maxwell equations. We derived the single-time autocorrelation of the electrostatic field and plasma density fluctuations, as well as the single-time correlation between the fluctuations of weighted macroparticle center phase-space density and electrostatic field. This has allowed us to estimate the deviations from the expected results of the simulation by applying the central limit theorem to the statistical fluctuations in the limit of a large number of macroparticles per cell, as well as deriving the kinetic equations for each PIC simulation plasma species. The former is useful to determine the spatial spectrum seeding an eventual physical (or numerical) instability in a PIC simulation plasma or to study a physical process of small amplitude that can be hidden in the ‘numerical noise’ (fluctuations), and the latter allowed us to emphasize that ‘collisions’ are indeed taken into account in a PIC plasma, where ‘collisions’ here means the friction and diffusion of macroparticles in the fluctuating component of the electrostatic field due to Landau damping and plasma wave emission by macroparticles. The kinetic equations of the PIC plasma are indeed described by a Vlasov–Lenard–Balescu-like equation with nonphysical spatial aliases in the ‘collision’ integral (that can be mitigated by using a smoothing function with the cost of eventually losing a physical part of the plasma spatial spectrum) and with time aliases (that can be mitigated by using small time steps). Our theory recovers all results previously obtained (Langdon 1970b), confirming their approach of perturbing a discretized Vlasovian plasma with a macroparticle test and then averaging the obtained physical quantity over the initial macroparticle velocity distribution. Our theory also extends their results to multiple-species multiple-weights PIC plasmas. The underlying hypothesis and assumptions for the validity of the theory have been systematically justified and/or estimated, closely connecting this with the PIC plasma stability and the simulation time duration, with estimates for these bounds provided as a function of the dimensionality of the PIC plasma, the number of real electrons in the plasma Debye sphere and the macroparticle weight factors. It is important to observe that the physical statistical fluctuations are overestimated when weight factors $\delta N > 1$ are used i.e. when a much-smaller-than-in-reality number of macroparticles is chosen to represent the real plasma particles phase-space. We have used our theoretical framework to explore the ‘ensemble averaging technique’ as a possible technique to recover lower/physical statistical fluctuations. In this method, a series of PIC simulations, representing identical statistical realizations of the same macroscopic initial conditions, are averaged. The equivalent statistical realizations, corresponding to different initial microscopic conditions of the PIC plasma, are obtained from the random sampling from the same distribution function of initial macroparticle velocities. It is shown that the ensemble average of $N_e = \delta N$ electrostatic PIC simulations is equivalent to the Vlasov–Lenard–Balescu/Maxwell set of equations discretized in space and time. Even for macroparticle weights of unity, the statistical fluctuations may hide theoretical features predicted in the Vlasov limit of an infinite number of electrons in the Debye sphere. In this case, the ensemble averaging technique can also be used to study in detail such microphysics processes beyond the limits imposed by statistical fluctuations. In order to illustrate the ‘ensemble averaging technique’, we have applied it to the study of the linear Landau damping of one electrostatic wave perturbing a PIC plasma in equilibrium by averaging the results obtained from $N_s = 30$ simulations. The amplitude of the statistical fluctuations is reduced by a factor $\sqrt{N_s}$, thus allowing the ensemble averaging technique to identify short-lived very small amplitude perturbations that would be undetected in a single simulation. The method shows good agreement with the theory and the subtraction technique. This technique is also particularly suited for modern computer architectures, since all $N_s$ simulations that will be averaged can run in parallel, allowing for speed-ups by a factor $N_s$ the global simulation (including the individual simulations over which the ensemble averaging is going to be performed), as compared with an equivalent simulation that uses $N_s \times N_{e,mac}$ macroelectrons per cell and a lower macroelectron weight $\delta N_e/N_e$. All $N_s$ simulations can be run in a sequential way too. The ensemble averaging technique can thus also be used if the equivalent simulation needs more RAM memory than what is available. The flexibility of the technique, the theoretical understanding presented in this paper, and the outstanding computational resources now available, open the way to a broader and wider use of this technique in all areas of plasma physics using PIC simulations.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).
Appendix A. Derivation of the Laplace–Fourier transform of electrostatic field fluctuations. Single-time autocorrelation

For all functions \( g_\omega \),

\[
\int_{\omega - \omega_0/2}^{\omega + \omega_0/2} \frac{d\omega'}{2\pi} g(\omega) = \int_{\omega - \omega_0/2}^{\omega + \omega_0/2} \frac{d\omega'}{2\pi} g(\omega)\,.
\]

Then, since

\[
\frac{1}{(k_{p,q,r} \cdot \mathbf{v} - \omega_m)} \frac{1}{(k_{p,q,r} \cdot \mathbf{v} + \omega')} = \frac{1}{\omega_m + \omega'} \left[ \frac{1}{k_{p,q,r} \cdot \mathbf{v} - \omega_m} - \frac{1}{k_{p,q,r} \cdot \mathbf{v} + \omega'} \right],
\]

the only contributions that are not damped come from the residue at poles \( \omega' = -\omega_m \). In this sense, the factor \( 1/(\omega_m + \omega') \) is to be interpreted as \(-i\omega\delta(\omega_m + \omega')\) leading to

\[
W_\omega,\mathbf{k}(\omega, \mathbf{k}) = \sum_{\omega', \omega_0} \frac{2\pi}{\omega_L(\omega, \mathbf{k})} \frac{1}{(k_{p,q,r} \cdot \mathbf{v} - \omega_m)} \frac{1}{(k_{p,q,r} \cdot \mathbf{v} + \omega')} 
\]

\[
\times \sum_{p,q,r} \sum_{a} \frac{S(k_{p,q,r}) S(-k_{p,q,r}) \delta(n_{\omega_0} a_0 q_a^2)}{\omega_L(\omega, \mathbf{k})} \]

\[
\times \int_{a_0} d^3v \int_{a_0} d^3v' \frac{F_{\omega}(\mathbf{v}) [\delta(\mathbf{v} - \mathbf{v'}) - F_{\omega}(\mathbf{v})]}{(k_{p,q,r} \cdot \mathbf{v} - \omega_m)(k_{p,q,r} \cdot \mathbf{v'} + \omega')} \]

\[
\times \exp\left[-i(\omega + \omega') t_0\right].
\]

Since we have assumed that the PIC simulation plasma is stable, which means that all poles of the longitudinal PIC simulation plasma permittivity (63) have a negative imaginary part, the second term in (144) \( F_{\omega}(\mathbf{v}) F_{\omega}(\mathbf{v'}) \) has similarly poles \( \omega' \) in the lower half of the \( \omega' \)-planes only. Thus, the undamped part of the electrostatic field statistical fluctuations autocorrelation arises only from the contribution of the first term on large time scale \( \omega_p k_p \gg 1 \). We thus find after integration over \( \mathbf{v}' \) and by using again the integral-sum property (142)

\[
W_\omega,\mathbf{k}(\omega, \mathbf{k}) = \sum_{\omega', \omega_0} \frac{2\pi}{\omega_L(\omega, \mathbf{k})} \frac{1}{(k_{p,q,r} \cdot \mathbf{v} - \omega_m)} \frac{1}{(k_{p,q,r} \cdot \mathbf{v} + \omega')} 
\]

\[
\times \sum_{p,q,r} \sum_{a} \frac{S(k_{p,q,r}) S(-k_{p,q,r}) \delta(n_{\omega_0} a_0 q_a^2)}{\omega_L(\omega, \mathbf{k})} \]

\[
\times \int_{a_0} d^3v \frac{F_{\omega}(\mathbf{v})}{(k_{p,q,r} \cdot \mathbf{v} - \omega_m)(k_{p,q,r} \cdot \mathbf{v'} + \omega')} \]

\[
\times \exp\left[-i(\omega + \omega') t_0\right].
\]

Here, we have used the particle shape parity property \( \hat{S}(\mathbf{k}) = -\hat{S}(\mathbf{\bar{k}}) \) and the \( (\omega, \mathbf{k}) \)-periodicity property of the longitudinal PIC simulation plasma permittivity to simplify the equation. As a conclusion, one may express

\[
\mathbb{E}\left( \hat{\mathbb{E}}(\omega, \mathbf{k}) \hat{\mathbb{E}}(\omega', \mathbf{k}') \right) = (2\pi)^4 \delta(\omega + \omega', \omega_p) \delta
\]

\[
\times (\mathbf{k} + \mathbf{k'}, \mathbf{k}_p) \hat{\mathbb{E}}^2(\omega, \mathbf{k})
\]

in the expression of electrostatic field statistical fluctuations autocorrelation (67) where

\[
\hat{\mathbb{E}}^2(\omega, \mathbf{k}) = 8\pi W_\omega,\mathbf{k}(\omega, \mathbf{k})
\]

reads (72).

Appendix B. Derivation of the single-time autocorrelation of electrostatic field fluctuations in PIC simulation plasmas at equilibrium

According to (156), one obtains the fluctuation-dissipation theorem derived by Callen and Welton (1951)

\[
W_\omega,\mathbf{k}(\omega, \mathbf{k}) = \frac{\hat{\mathbb{E}}^2(\omega, \mathbf{k})}{8\pi} = \omega_p \frac{n_k T}{\left[\omega L(\omega, \mathbf{k})\right]^2}
\]

(150)
only in the particular case of PIC simulations of a plasma at equilibrium for which $\delta N_{c} = \delta N = \delta N_{\lambda}$ and using a spectral solver with a smoothing function filtering all spatial frequency aliases $p, q, r \neq 0, 0, 0$. Indeed, only in this particular case, $\delta N$ can be extracted from the sum over all species in (72) simplifying the electrostatic fluctuations energy spectrum into

$$W_{\omega, k}(\omega, k) = \frac{4\pi^{2}\delta N}{\varepsilon_{L}(\omega, k)^{2}} - \frac{S(k)}{k^{2}} \sum_{a} \tilde{n}_{a} q_{a}^{2}$$

while the imaginary part of PIC simulation electrostatic plasmas longitudinal permittivity (156) reduces to

$$\text{Im}\{\varepsilon_{L}(\omega, k)\} = \frac{4\pi^{2}j\omega}{k_{D}^{2}} \sum_{a} \tilde{n}_{a} q_{a}^{2}$$

thus providing the fluctuation-dissipation theorem (150) by recognizing (152) in the numerator of (151). Still, in this particular case, one can get a simple estimate of the electrostatic field fluctuations energy spatial spectrum

$$W_{k}(k) = \int_{\omega_{\lambda} > \omega / 2} d\omega \frac{d\omega}{2\pi} W_{\omega, k}(\omega, k).$$

It reads

$$W_{k}(k) = \frac{\delta N}{2\pi} \frac{k_{D}T}{k^{2}} \frac{S(k)}{k^{2}} (k\lambda_{D})^{2}$$

by performing the $\omega$-integration using the residue theorem, assuming $\Delta_{t} \ll \omega_{\lambda}/\omega_{c}$, where we have introduced the Debye screening length $\lambda_{D} = \omega_{c}/\omega_{\lambda}$. Assuming then $a_{1}$, $a_{2}$, and $a_{3} \ll \lambda_{D}$ and $\Delta = \Delta_{t} = \Delta_{3} \ll \lambda_{D}$ for simplicity, one may consider $S(k) \sim 1$ and estimate

$$\mathbb{E}\left\{\delta E(r_{i, j, k}, t_{n})\right\} = \frac{\int_{\nu_{D}} d^{3}k}{2\pi^{3}} W_{k}(k).$$

While diverging in 2D and 3D collisionless plasmas for which $\delta N = 1$, $N_{c} \rightarrow \infty$ and $\Delta = \Delta_{t} = \Delta_{3} = \Delta_{t} \rightarrow 0$, we find (73) for PIC simulation plasmas at equilibrium by replacing $d^{3}k/(2\pi)^{3}$ with $dk_{x}/2\pi$ for 1D simulations and with $dk_{x}dk_{y}/(2\pi)^{2}$ for 2D simulations as well as the domain of integration $V_{k}$ with the closed disk $\{(k_{x}, k_{y})\}$ such that $k_{x}^{2} + k_{y}^{2} \leq (k_{x}/2)^{2}$ for 2D simulations and with the closed sphere $\{(k_{x}, k_{y}, k_{z})\}$ such that $k_{x}^{2} + k_{y}^{2} + k_{z}^{2} \leq (k_{x}/2)^{3}$ for 3D simulations instead of $V_{k} = [-k_{x}/2, k_{x}/2]^{2}$, the closed square $[-k_{y}/2, k_{y}/2]^{2}$ and the closed cube $[-k_{y}/2, k_{y}/2]^{3}$ for 1D, 2D and 3D simulations respectively. An empirical factor 1/2 for 2D simulations and 2/3 for 3D simulations has been chosen to account for this latter approximation.

### Appendix C. Derivation of the ‘collision operator’

According to (63) and the Plemelj formula, one may express the imaginary part of PIC simulation electrostatic plasmas longitudinal permittivity according to

$$\text{Im}\{\varepsilon_{L}(\omega, k)\} = \frac{\pi}{k_{r}} \frac{\sum_{p, q, r} S(k_{p, q, r})^{2}}{\omega^{2}}$$

for 3D simulations has been chosen for the first term in (80) gives

$$\int_{\omega_{\lambda} > \omega / 2} d\omega \frac{d\omega}{2\pi} \frac{d^{3}k}{2\pi^{3}} \frac{32\pi^{4}n_{a}^{2}S(k_{p, q, r})^{2}}{x_{\omega}(\omega, k)^{2}} \frac{\omega^{2}}{m_{a}} \int_{R^{3}} d^{3}v F_{00}(v) \delta (\omega - k_{p, q, r} \cdot v, \omega_{B})$$

for the same reason, the contribution of the second term in (80) comes from its imaginary part that can be simplified into

$$- \int_{\nu_{D}} d^{3}k \frac{d\omega_{B}}{2\pi} \frac{d\omega_{B}}{2\pi} \text{Im}$$

$$\times \left\{\sum_{p, q, r} S(k_{p, q, r})^{2} \frac{\omega}{x_{\omega}(\omega, k)^{2}} \frac{\omega}{m_{a}} \int_{R^{3}} d^{3}v F_{00}(v) \delta (\omega - k_{p, q, r} \cdot v, \omega_{B})\right\}$$

$$\times \left\{\sum_{m} \frac{dF_{00}(v)}{dv_{B}} \frac{dF_{00}(v)}{dv_{B}} \delta (\omega - k_{p, q, r} \cdot v, \omega_{B})\right\}$$

$$\times \left\{\sum_{m} \frac{dF_{00}(v)}{dv_{B}} \frac{dF_{00}(v)}{dv_{B}} \delta (\omega - k_{p, q, r} \cdot v, \omega_{B})\right\}$$

according to the Plemelj formula. By performing the integration over $\omega$ using the integral sum properties (142), one finally deduces (82).
Appendix D. Landau damping BEPS input deck

pinput2
IDRUN = 1
INDX = 8, INDY = 1, NPX = 8388608, NPY = 2
NPXB = 0, NPYB = 0, INORDER = 2,
POPT = 2, DOPT = 2, DJOPT = 2
NTW = 1, NTP = 0, NTD = 10, NTA = 0, NTV = 0,
NTS = 0
NUSTRT = 1, NTR = 0, PSOLVE = 1, RELATIVITY = 0
TEND = 500.0, DT = 0.200, CI = 0.1
QME = -1.000, VTX = 1.000, VTY = 1.000, VTZ = 1.000
VX0 = 0.000, VYO = 0.000, VZO = 0.000
VDX = 0.000, VDY = 0.000, VDZ = 0.000
VTDX = 0.000, VTDDY = 0.000, VTDDZ = 0.000
AX = 0.866667, AT = 0.866667
NSRAND = 0, NDPROF = 0, AMPDX = 0.0,
SCALEDX = 0.0, SHIFTDX = 0.0
AMPDY = 0.0, SCALEDY = 0.0, SHIFTDY = 0.0
MOVION = 0, NPXI = 8192, NPYI = 2048
NPXB1 = 0, NPYB1 = 0
QMI = 1.000, RMASS = 100.0
REMPXI = 20.000, REMPYI = 20.000,
REMPZI = 20.000
VXI0 = 0.000, VYI0 = 0.000, VZI0 = 0.000
VDXI = 0.000, VDYI = 0.000, VDZI = 0.000
REMPDXI = 1.000, REMPDYI = 1.000,
REMPDZI = 1.000
NSRANDI = 0, NDPROFI = 0
AMPDXI = 0.0, SCALEDXI = 0.0, SHIFTDXI = 0.0
AMPDYI = 0.0, SCALEDYI = 0.0, SHIFTDYI = 0.0
SORTIME = 0, SORTIMI = 250
NPLOT = 0, IDPAL = 1, NDSTYLE = 1, NTASKS = 1,
ntw = 1, ntp = 50
pinput2 jf
driver select = 1
amp = 0.005, wavemode = 13, wavew = 1.18
rise = 0., flat = 0., fall = 0., yrise fall = 1000.0
timerise = 1.2, timeflat = 1.2, timefall = 1.2,
center1 = 200, center2 = 2, phase offset = 4
bow = 0., bow power = 0.
superGauss = 2
fvxmax = 12., fvyymax = 12.
nphbx = 150, nphby = 150
nphxx = 20, nphyx = 20
nphxy = 20, nphyy = 2

ORCID iDs
Michaël Touati https://orcid.org/0000-0001-7590-0941
Romain Codur https://orcid.org/0000-0001-7240-3843
Frank Tsung https://orcid.org/0000-0001-9257-7920
Viktor K Decyk https://orcid.org/0000-0003-1280-5001
Warren B Mori https://orcid.org/0000-0002-6394-5082

References
Abramowitz M and Stegun I 1965 Handbook of Mathematical Functions (New York: Dover)
Balescu R 1960 Phys. Fluids 3 52–63
Birdsall C K and Fus K D 1969 J. Comput. Phys. 3 494
Birdsall C K and Landgon 1991 Plasma Physics Via Computer Simulation (The Adam Hilger Series
on Plasma Physics) (Bristol: Adam Hilger)
Bogoliubov N N 1946 J. Exp. Theor. Phys. 16 691–702
Born M and Green H S 1946 Proc. R. Soc. A 188 10–18
Buneman O 1967 J. Comput. Phys. 1 517–35
Burger P, Dunn D A and Halsted A S 1965 Phys. Fluids 8 2263–72
Callen H B and Welton T A 1951 Irreversibility and generalized noise Phys. Rev. 83 34–40
Dawson J M 1964 Thermal relaxation in a one-species, one-dimensional plasma Phys. Fluids 7 419–25
Dawson J M 1983 Rev. Mod. Phys. 55 403–47
de Boer J and Uhlenseck G E 1962 Studies in Statistical Mechanics ed N N Bogoliubov (Amsterdam: North-Holland)
Decyk V K 1982 Energy conservation theorem for electrostatic systems Phys. Fluids 25 1205–6
Decyk V K 1995 Skeleton PIC codes for parallel computers Comput. Phys. Commun. 87 89–94
Decyk V K 2007 UPIC: a framework for massively parallel particle-in-cell codes Comput. Phys. Commun. 177 95–97
Decyk V K 1987 Simulation of microscopic processes in plasma Proc. 1987 Int. Conf. on Plasma Physics (Kiev, USSR, April 1987) vol II, ed A G Sitenko (Singapore: World Scientific) p 1075
Decyk V K and Singh T V 2014 Particle-in-cell algorithms for emerging computer architectures Comput. Phys. Commun. 185 708–19
Dupree T H 1963 Phys. Fluids 6 1714–29
Esirkepov T 2001 Comput. Phys. Commun. 135 144–53
Grismayer T, Fahlen J E, Decyk V K and Mori W B 2011 Plasma Phys. Control. Fusion 53 074011
Gurnsey R 1960 The kinetic theory of fully ionized gases PhD Thesis University of Michigan
Hockney R W 1965 J. ACM 12 95–113
Hockney R W 1966 Phys. Fluids 9 1826–35
Hockney R W 1971 Measurements of collision and heating times in a two-dimensional thermal computer plasma J. Comput. Phys. 8 19–44
Hockney R and Eastwood J 1981 Computer Simulation Using Particles (New York: McGraw-Hill)
Hubbard J 1961 Proc. R. Soc. A 260 114–26
Joglekar A S, Thomas A G R, Fox W and Bhattacharjee A 2014 Phys. Rev. Lett. 112 105004
Kirkwood J G 1946 J. Chem. Phys. 14 180
Klimontovich Y L 1958 J. Exp. Theor. Phys. 7 119
Klimontovich Y L and Silin V P 1962 J. Exp. Theor. Phys. 15 199
Landau L D 1937 J. Exp. Theor. Phys. 7 203
Landau L D 1948 Yad. Fiz. 10 25
Landau L D and Lifshitz E M 1981 Physical Kinetics (A Course of Theoretical Physics vol 10) (New York: Pergamon Press Inc.)
Landgon A B 1970b Proc. 4th Conf. Numerical Simulation of Plasmas (Washington, DC: Naval Research Laboratory) pp 467–95
Landgon A B 1979 Phys. Fluids 22 163
Landgon A B and Birdsall C K 1970 Theory of plasma simulation using finite-size particles Phys. Fluids 13 2115–22
Landgon A 1970a J. Comput. Phys. 6 247–67
Lenard A 1960 *Ann. Phys., NY* **10** 390–400
Lindman E 1970 *J. Comput. Phys.* **5** 13–22
Liouville J 1838 *J. Math. Pures Appl.* **3** 342–9
Montgomery D and Nielson C W 1970 Thermal relaxation in one- and two-dimensional plasma models *Phys. Fluids* **13** 1405–7
Nanbu K 1997 *Phys. Rev. E* **56** 7314
Okuda H and Birdsall C K 1970 Collisions in a plasma of finite-size particles *Phys. Fluids* **13** 2123–34
Rostoker N 1961 *Nucl. Fusion* **1** 101
Rostoker N 1964 Test particle method in kinetic theory of a plasma *Phys. Fluids* **7** 491–8
Shanny R, Dawson J M and Greene J M 1967 One-dimensional model of a Lorentz plasma *Phys. Fluids* **10** 1281–7
Thompson W B and Hubbard J 1960 *Rev. Mod. Phys.* **32** 714–8
Touati M, Feugeas J-L, Nicolai P, Santos J J, Gremillet V and Tikhonchuk V T 2014 *New J. Phys.* **16** 073014
Tzoufras M, Bell A, Norreys P and Tsung F 2011 *J. Comput. Phys.* **230** 6475–94
Villasenor J and Buneman O 1992 Rigorous charge conservation for local electromagnetic field solvers *Comput. Phys. Commun.* **69** 306–16
Virtamo J and Tuomisto H 1979 Verification of a simple collision operator for one-dimensional plasma by simulation experiments *Phys. Fluids* **22** 172–5
Vlasov A A 1938 *J. Exp. Theor. Phys.* **8** 291
Yee K 1966 *IEEE Trans. Antennas Propag.* **14** 302–7
Yu S P, Kooyers G P and Buneman O 1965 *J. Appl. Phys.* **36** 2550–9
Yvon J 1935 *La Théorie Statistique des Fluides et L’équation D’état* (Actualités Scientifiques et Industrielles) (Paris: Hermann)