Monte Carlo particle-in-cell methods for the simulation of the Vlasov-Maxwell gyrokinetic equations

A. BOTTINO[1]† AND E. SONNENDRÜCKER[1]

[1]Max Planck Institut für Plasmaphysik, D-85748 Garching, Germany

The Particle-In-Cell (PIC) algorithm is the most popular method for the discretisation of the general 6-D Vlasov-Maxwell problem and it is widely used also for the simulation of the 5D gyrokinetic equations. The method consists of coupling a particle based algorithm for the Vlasov equation with a grid based method for the computation of the self-consistent electromagnetic fields. In this review we derive a Monte Carlo Particle In Cell Finite Element model starting from a gyrokinetic discrete Lagrangian. The variations of the Lagrangian are used to obtain the time continuous equations of motion for the particles and the Finite Element approximation of the field equations. The Noether theorem for the semi-discretised system, implies a certain number of conservation properties for the final set of equation. Moreover, the PIC method can be interpreted as a probabilistic Monte-Carlo like method, consisting of calculating integrals of the continuous distribution function using a finite set of discrete markers. The nonlinear interactions along with numerical errors introduce random effects after some time. Therefore, the same tools for error analysis and error reduction used in Monte-Carlo numerical methods can be applied to PIC simulations.

PACS codes:

---

1. Introduction

In this lecture note, we discuss a Monte Carlo Particle-In-Cell (PIC) Finite Element method based on the description of the equation of motions by a Lagrangian derived using the particle trajectories approach. Starting from the Lagrangian introduced by Low (1958), Sugama (2000) derived the gyrokinetic field theoretic Lagrangian that will be the base of our formulation. For the discretisation, we follow the procedure first proposed by Lewis (1970) to derived a Particle In Cell approximation based on a discrete Lagrangian, the variations of which are used to obtain the time continuous equations of motion for the particles and the Finite Element approximation of the fields. This feature implies a certain number of conservation properties thanks to the Noether theorem for the semi-discretised system. For these to be retained exactly after time discretisation, special care needs to be taken, implying implicit methods. This is inconvenient for large PIC simulations and is generally not done in practice.

Classical derivations of the PIC method are based on the approximate distribution function being a sum of δ functions. A detailed discussion of the method and a systematic derivation of the relevant equations were given by Hu & Krommes (1994). Unlike classical

† Email address for correspondence: bottino@ipp.mpg.de

derivation, we cast the PIC method from the beginning into the framework of Monte Carlo simulation, which is best adapted to the study of its approximation properties. This approach enables us to draw on the extensive literature on Monte Carlo methods in statistics to propose efficient noise reduction strategies. The aim of this lecture note, which follows a large literature on PIC method for gyrokinetics like Aydemir (1994); Hu & Krommes (1994); Tran *et al.* (1999); Allfrey & Hatzky (2003); Krommes (2007); Jolliet *et al.* (2007); Bottino *et al.* (2010); Kleiber *et al.* (2011) and many others, is to highlight on the one hand the semi-discrete variational principle on which these methods are based and on the other hand the Monte Carlo framework, in order to make a clear link with the Monte Carlo literature in statistics. The outline of the note is the following. First we will introduce the gyrokinetic approximation and in particular the phase space Lagrangian density on which it is based, derive the gyrokinetic Vlasov-Poisson equations and also present the special case of adiabatic electrons. Then we will describe our Finite Element Monte Carlo Particle in Cell discretisation the principle of which consists in obtaining a discrete Lagrangian by applying a Monte Carlo approximation in the particle part of the Lagrangian density and a Finite Element approximation for the fields. The PIC marker equations of motion and the discrete field equations then are obtained as the Euler Lagrange equations of the discrete Lagrangian. We will then explain how classical Monte Carlo noise reduction techniques can be applied. Finally in the last part we will present some typical gyrokinetic simulations and illustrate how the noise to signal ratio can be monitored in a PIC simulation.

2. Description of the gyrokinetic model

Gyrokinetic (GK) theory aims to describe plasma particle motion in terms of drifts of particle gyrocenters, rather than the usual combination of gyromotion and drifts of the particles, thus reducing the original 6D kinetic problem into a 5D problem. The development of GK theory was motivated by the need to describe complex plasma dynamics over time scales much longer than the short gyro-motion time scale. The typical example is the study of the low-frequency electromagnetic fluctuations (microturbulence) observed in inhomogeneous magnetized plasmas, characterized by $\Omega < \Omega_s$ where $\Omega_s = e_s B / m_s$ is the ion cyclotron frequency.

Several ways to construct gyrokinetic equations exist. Initially, GK equations were derived by gyro-averaging the Vlasov equation with recursive methods (see for example Frieman & Chen (1982) and Lee (1983)). Modern gyrokinetic theory is based on a Hamiltonian representation in which nonlinear gyrokinetic equations are derived from a systematic Hamiltonian theory as originally proposed by Dubin *et al.* (1983). The starting point of the derivation is the description of particle motion in a magnetic field in terms of a drift-kinetic Lagrangian. Gyrokinetic equations can be constructed when this Lagrangian is (Lie) transformed into a low-frequency form by means of an expansion based on a small parameter. The small parameter can be either the fluctuation amplitude of the drifts [Littlejohn (1981, 1983)] or the gyroradius compared to the dynamical scale lengths [Hahm (1988)]. In this formulation, a back transformation was used to obtain the self-consistent field equations for the potential [Dubin *et al.* (1983); Hahm (1988)]. In modern gyrokinetics, the so-called gyrokinetic field theory, the entire Lagrangian is constructed as the integral of a Lagrangian density, with the field equations obtained as the Euler-Lagrange equations by varying the potentials in the Lagrangian [Sugama (2000); Brizard (2000)]. These two approaches have been proved to be identical by Brizard & Hahm (2007) and both preserve the symmetry and conservation properties of the Vlasov equation.

Gyrokinetic field theory can be addressed in the context of a general Hamiltonian [Scott & Smirnov (2010)]. It consists in constructing a Lagrangian in which all the dependence on dynamical variables is in the Hamiltonian or in the free field terms. The symmetry with time through the Hamiltonian leads automatically to energy conservation. Indeed, self-consistent nonlinear gyrokinetic Vlasov-Maxwell equations can be derived starting from the Lagrangian, without any additional ordering assumption. This approach also naturally provides the exact energy conservation law these nonlinear GK equation satisfy and it is therefore particularly suited for numerical simulations. In the following, the simplest self-consistent and energy conserving gyrokinetic model, suited for gyrokinetic particle-in-cell simulations, is derived using general Hamiltonian based gyrokinetic field theory.

2.1. *Gyrokinetic Lagrangian*

As a starting point, the following Lie transformed low-frequency particle Lagrangian is assumed:

$$L_p \equiv \left(\frac{e}{c}\boldsymbol{A} + p_\parallel \boldsymbol{b} \right) \cdot \dot{\boldsymbol{R}} + \frac{mc}{e}\mu\dot{\theta} - H. \tag{2.1}$$

The velocity variables are the magnetic moment $\mu \equiv mv_\perp^2/(2B)$, the canonical parallel momentum, $p_\parallel$ and the gyroangle θ; $\boldsymbol{R}$ is the gyrocenter position. Upper dots denote total time derivatives. The perpendicular subscript denotes the component in the plane perpendicular to the background magnetic field $\boldsymbol{B} = \nabla \times \boldsymbol{A}$. Note that the symplectic part depends only on the background, while all the time varying fields are contained in the Hamiltonian H. In the sequel $\boldsymbol{A}$ will always denote the background vector potential, whereas $A_\parallel$ will denote the parallel component of the fluctuating field, the perpendicular part being assumed to be zero. Complete derivations of such a particle Lagrangian can be found, for example, in Hahm (1988); Brizard & Hahm (2007) and Miyato *et al.* (2009). Following Sugama (2000), the gyrokinetic total Lagrangian is given by:

$$L = \sum_{\text{sp}} \int \mathrm{d}W_0 \mathrm{d}V_0 f(\boldsymbol{Z}_0, t_0) L_p(\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t), \dot{\boldsymbol{Z}}(\boldsymbol{Z}_0, t_0; t), t) + \int \mathrm{d}V \frac{E^2 - B_\perp^2}{8\pi}, \tag{2.2}$$

where $\boldsymbol{Z} \equiv (\boldsymbol{R}, p_\parallel, \mu, \theta)$, $\mathrm{d}W \equiv (2\pi/m^2)B_\parallel^* \mathrm{d}p_\parallel \mathrm{d}\mu$ and $\mathrm{d}V$ denotes the volume element in physical space, with Jacobian $1/\sqrt{g}$. Here $f(\boldsymbol{Z}_0)$ is the distribution function for the species *sp* at an arbitrary initial time t_0. We will not use a specific notation to distinguish the distribution functions of the different species for simplicity. $\mathrm{d}W_0$ implies that the phase space Jacobian is function of $\boldsymbol{Z}_0$, i.e. $B_\parallel^*(\boldsymbol{Z}_0)$. L_p is the Lie transformed particle Lagrangian of Eq. (2.1), written in terms of the gyro-center coordinates of the particle, with the initial condition:

$$\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t_0) = \boldsymbol{Z}_0. \tag{2.3}$$

The first term in the total Lagrangian is the Lagrangian for charged particles. Note that here the integral is performed with respect to $\boldsymbol{Z}_0$: it is therefore assumed that the particles can be traced back in time to their position at t_0. The second term is the Lagrangian for the electromagnetic fields.

Writing that the distribution function is conserved along the particle trajectories

$$f(\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t), t) = f(\boldsymbol{Z}_0, t_0) \tag{2.4}$$

yields the GK Vlasov equation by taking the time derivative

$$\frac{d}{dt}f(\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t), t) = \frac{\partial}{\partial t}f(\boldsymbol{Z}, t) + \frac{\mathrm{d}\boldsymbol{Z}}{\mathrm{d}t} \cdot \frac{\partial}{\partial \boldsymbol{Z}}f(\boldsymbol{Z}, t) = 0. \tag{2.5}$$

The particle number conservation follows from Liouville's theorem for a time independent Jacobian:

$$\frac{\partial}{\partial \boldsymbol{Z}} \cdot (B_\parallel^* \frac{d\boldsymbol{Z}}{dt}) = 0,$$

which is a property inherited from the particle Lagrangian L_p that we shall verify later. Then, on the one hand the first integral defining L can be expressed by the change of variables $\boldsymbol{Z} = \boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t)$ as

$$\int \mathrm{d}W \mathrm{d}V f(\boldsymbol{Z}, t) L_p(\boldsymbol{Z}, \dot{\boldsymbol{Z}}, t) \tag{2.6}$$

and the Vlasov equation can be written in the conservative form

$$\frac{\partial}{\partial t} \left(\frac{2\pi}{m^2} B_\parallel^* f \right) + \frac{\partial}{\partial \boldsymbol{Z}} \cdot \left(\frac{2\pi}{m^2} B_\parallel^* \frac{d\boldsymbol{Z}}{dt} f \right) = 0. \tag{2.7}$$

The GK Hamiltonian in general depends on the electrostatic potentials Φ and on the parallel component of the fluctuation magnetic potential $A_\parallel$. In its simplest form, the GK Hamiltonian has the following form:

$$H = m\frac{U^2}{2} + \mu B + eJ_0\Phi - \frac{mc^2}{2B^2}|\nabla_\perp\Phi|^2. \tag{2.8}$$

The last term in the Hamiltonian is the negative of the kinetic energy associated with the $\boldsymbol{E} \times \boldsymbol{B}$ motion of the gyrocenters. The physical meaning and implications of this term are specifically addressed in a recent paper by Krommes (2013). A derivation of the gyrokinetic equations using higher order Hamiltonian can be find, for example, in Miyato & Scott (2013). Note that U, the parallel velocity, is not a coordinate. It is defined by $mU = p_\parallel - (e/c)J_0A_\parallel$ and is a function of the fluctuation magnetic potential $A_\parallel$. The gyroaveraging operator J_0 applied to an arbitrary function ψ in configuration space is defined by

$$(J_0\psi)(\boldsymbol{R}, \mu) = \frac{1}{2\pi} \int_0^{2\pi} \psi(\boldsymbol{R} + \boldsymbol{\rho}(\alpha)) \, d\alpha, \tag{2.9}$$

where $\boldsymbol{\rho}$ is the vector going from the guiding center position to the particle position. It depends on μ and spatial coordinates. In the context of GK field theory, this Lagrangian can be further approximated, without loosing self-consistency and energetic consistency of the final equations [Sugama (2000); Scott & Smirnov (2010)]. A possible approximation, widely used in GK simulations, is the so called *quasi-neutrality* approximation. It consists in neglecting the term proportional to E^2 in the free field term. This term is ordered small compared to the so called $E \times B$ energy, corresponding to the second order term in Φ in the Hamiltonian:

$$\int \mathrm{d}V \frac{E^2}{8\pi} + \int \mathrm{d}\Omega f \frac{m}{2} \frac{c^2}{B^2} |\nabla_\perp\Phi|^2 = \frac{1}{8\pi} \int \mathrm{d}V \left[E_\parallel^2 + \left(1 + \frac{\rho_S^2}{\lambda_d^2}\right) |\nabla_\perp\Phi|^2 \right], \tag{2.10}$$

having introduced the electron Debye length squared $\lambda_d^2 \equiv (k_B T_e)/(4\pi n e^2)$ and the ion sound Larmor radius $\rho_S^2 \equiv (k_B T_e m c^2)/(e^2 B^2)$; $\mathrm{d}\Omega \equiv \mathrm{d}V \mathrm{d}W$.
In fusion plasmas

$$\frac{\rho_S^2}{\lambda_d^2} = \frac{4\pi n m c^2}{B^2} = \frac{c^2}{v_a^2} \gg 1 \tag{2.11}$$

where v_a is the Alfvén velocity and $E_\parallel^2$ is even smaller. Therefore the E^2 term can be safely neglected. We will see in the following that this approximation will lead to a field equation for the electrostatic potential equivalent to a quasi-neutrality condition.

Recalling the definition of $p_\parallel$, the GK Hamiltonian can be rewritten as follows:

$$H = H_0 + H_1 + H_2, \tag{2.12}$$

$$H_0 \equiv \frac{p_\parallel^2}{2m} + \mu B, \tag{2.13}$$

$$H_1 \equiv e J_0 \Psi, \tag{2.14}$$

$$H_2 = \frac{e^2}{2mc^2}(J_0 A_\parallel)^2 - \frac{mc^2}{2B^2}|\nabla_\perp \Phi|^2, \tag{2.15}$$

having introduced the generalized potential $\Psi \equiv \Phi - (p_\parallel/mc)A_\parallel$. The GK total Lagrangian can be further approximated by assuming that only $(H_0 + H_1)$ multiplies f, while f is replaced by an equilibrium distribution function f_M independent of time in the term containing H_2:

$$L = \sum_{\text{sp}} \int d\Omega \left(\left(\frac{e}{c}\boldsymbol{A} + p_\parallel \boldsymbol{b} \right) \cdot \dot{\boldsymbol{R}} + \frac{mc}{e}\mu\dot{\theta} - H_0 - H_1 \right) f$$
$$- \sum_{\text{sp}} \int d\Omega H_2 f_M - \int dV \frac{B_\perp^2}{8\pi}. \tag{2.16}$$

Since this approximation leads to linearised field equations, it is traditionally known as *linearised polarisation* approximation. However, this choice of Lagrangian has a deeper meaning: already Sugama (2000) showed that in order to construct a gyrokinetic equation based on first order drift motion, only $H_0 + H_1$ must multiply f, thus contributing to the drifts via the variational derivatives. The term containing H_2 must act as a field term and should contribute to the field equations only, i.e. $H_2 f$ must be replaced by $H_2 f_0$. Conversely, if one desires to keep the dependent variable f in this term then H_2 must be kept in the drift motion. This is the basic statement of energetic consistency in a gyrokinetic global model as discussed in Scott & Smirnov (2010) . Note that this result was already present in both Sugama (2000) and Brizard (2000). In addition to this, Miyato *et al.* (2009) showed that the reduced MHD vorticity equation can be recovered by taking the time derivative, $\partial/\partial t$, of the linearised polarisation equation.

Although electromagnetic effects are important to correctly describe experimental plasmas, in the following we will neglect magnetic perturbations, setting $A_\parallel = 0$, which implies $B_\perp^2 = 0$ and $p_\parallel = mU$.

Finally, the electrostatic GK Lagrangian starting point for the derivation of our GK system is:

$$L = \sum_{\text{sp}} \int d\Omega \left(\left(\frac{e}{c}\boldsymbol{A} + p_\parallel \boldsymbol{b} \right) \cdot \dot{\boldsymbol{R}} + \frac{mc}{e}\mu\dot{\theta} - H_0 - H_1 \right) f + \sum_{\text{sp}} \int d\Omega \frac{mc^2}{2B^2}|\nabla_\perp \Phi|^2 f_M. \tag{2.17}$$

It is important to underline that from now on, any additional approximation or ordering will break the symmetry and conservation properties of the underlying dynamical system.

2.2. *Gyrokinetic electrostatic Vlasov-Poisson equations*

The gyrokinetic equations for the particle distribution function and the GK field equations can be derived from the GK Lagrangian using variational principles for the action functional I [Sugama (2000)]:

$$\delta I[\boldsymbol{Z}, \Phi] = \int_{t1}^{t2} \delta L[\boldsymbol{Z}, \Phi] dt = \int_{t1}^{t2} \left(\sum_{\alpha=1}^{6} \frac{\delta L}{\delta Z_\alpha} \cdot \delta Z_\alpha + \frac{\delta L}{\delta \Phi} \cdot \delta \Phi \right) dt, \tag{2.18}$$

where the functional derivative of L with respect to a function ψ is defined for an arbitrary variation $\delta\psi$ in the same function space as ψ by

$$\frac{\delta L}{\delta\psi}\cdot\delta\psi = \frac{d}{d\epsilon}_{|\epsilon=0}L(\psi+\epsilon\delta\psi) = \lim_{\epsilon\to 0}\frac{L(\psi+\epsilon\delta\psi)-L(\psi)}{\epsilon}. \tag{2.19}$$

The particle equations of motion can be obtained by taking the functional derivatives with respect to the particle phase space positions $\boldsymbol{Z}=(\boldsymbol{R},p_\parallel,\mu)$:

$$\frac{\delta I}{\delta\boldsymbol{Z}} = 0 \Rightarrow \frac{\delta L}{\delta\boldsymbol{Z}} = 0 \tag{2.20}$$

as t_1 and t_2 are arbitrary. This yields the GK Euler-Lagrange equations for the particle Lagrangian L_p, describing the drift motion of the gyrocenters, which can be expressed as

$$\frac{d}{dt}\frac{\partial L_p}{\partial\dot{y}} = \frac{\partial L_p}{\partial y}$$

for all the phase space components y. We follow here the idea introduced by Littlejohn (1983) for non canonical formulations. Let us first compute all the needed partial derivatives of L_p, denoting by $\nabla\boldsymbol{A}$ the Jacobian matrix associated to vector field $\boldsymbol{A}$:

$$\frac{\partial L_p}{\partial\dot{\boldsymbol{R}}} = \frac{e}{c}\boldsymbol{A}+p_\parallel\boldsymbol{b}, \quad \frac{\partial L_p}{\partial\boldsymbol{R}} = \frac{e}{c}(\nabla\boldsymbol{A})^T\dot{\boldsymbol{R}}+p_\parallel(\nabla\boldsymbol{b})^T\dot{\boldsymbol{R}}-\nabla(H_0+H_1), \tag{2.21}$$

$$\frac{\partial L_p}{\partial\dot{p}_\parallel} = 0, \quad \frac{\partial L_p}{\partial p_\parallel} = \boldsymbol{b}\cdot\dot{\boldsymbol{R}}-\frac{\partial(H_0+H_1)}{\partial p_\parallel}. \tag{2.22}$$

Taking the time derivative of the first term above, using that $\boldsymbol{A}$ and $\boldsymbol{b}$ do not depend on time

$$\frac{d}{dt}\frac{\partial L_p}{\partial\dot{\boldsymbol{R}}} = \frac{e}{c}(\nabla\boldsymbol{A})\dot{\boldsymbol{R}}+\dot{p}_\parallel\boldsymbol{b}+p_\parallel(\nabla\boldsymbol{b})\dot{\boldsymbol{R}} = \frac{e}{c}(\nabla\boldsymbol{A}^*)\dot{\boldsymbol{R}}+\dot{p}_\parallel\boldsymbol{b}, \tag{2.23}$$

where we introduce

$$\boldsymbol{A}^* = \boldsymbol{A}+p_\parallel\frac{c}{e}\boldsymbol{b}, \quad \boldsymbol{B}^* = \nabla\times\boldsymbol{A}^*. \tag{2.24}$$

We can now write the Euler-Lagrange equation for $\boldsymbol{R}$ and $p_\parallel$:

$$\frac{e}{c}(\nabla\boldsymbol{A}^*)\dot{\boldsymbol{R}}+\dot{p}_\parallel\boldsymbol{b} = \frac{e}{c}(\nabla\boldsymbol{A}^*)^T\dot{\boldsymbol{R}}-\nabla(H_0+H_1), \tag{2.25}$$

$$0 = \boldsymbol{b}\cdot\dot{\boldsymbol{R}}-\frac{\partial(H_0+H_1)}{\partial p_\parallel}. \tag{2.26}$$

Moreover $\partial L_p/\partial\dot{\theta}=\mu,\partial L_p/\partial\theta=0$, so that the Euler-Lagrange equation for θ yields $\mathrm{d}\mu/\mathrm{d}t=0$, which expresses that μ is an exact invariant. On the other hand, as the dependence on θ has been removed from the Lagrangian no evolution equation on θ is needed.

Let us introduce $\boldsymbol{F}=\nabla\boldsymbol{A}-(\nabla\boldsymbol{A})^T$ with the properties

$$\nabla\times\boldsymbol{b} = -\nabla\cdot\frac{\boldsymbol{F}}{B}, \quad \boldsymbol{C}\times(\nabla\times\boldsymbol{A}) = \boldsymbol{C}\times\boldsymbol{B} = \boldsymbol{F}\boldsymbol{C} \quad \text{for any vector } \boldsymbol{C}, \tag{2.27}$$

and similarly $\boldsymbol{F}^*=\nabla\boldsymbol{A}^*-(\nabla\boldsymbol{A}^*)^T$, with $\boldsymbol{C}\times\boldsymbol{B}^*=\boldsymbol{F}^*\boldsymbol{C}$ for any vector $\boldsymbol{C}$. Then in equation Eq. (2.25) we have

$$(\nabla\boldsymbol{A}^*)\dot{\boldsymbol{R}}-(\nabla\boldsymbol{A}^*)^T\dot{\boldsymbol{R}} = \boldsymbol{F}^*\dot{\boldsymbol{R}} = \dot{\boldsymbol{R}}\times\boldsymbol{B}^*, \tag{2.28}$$

and we can write Eq. (2.25) equivalently

$$\frac{e}{c}\dot{\boldsymbol{R}} \times \boldsymbol{B}^* + \dot{p}_\parallel \boldsymbol{b} = -\nabla(H_0 + H_1). \tag{2.29}$$

Finally taking the dot product of Eq. (2.29) with $\boldsymbol{B}^*$ yields an equation for $\dot{p}_\parallel$, and taking the cross product of $\boldsymbol{b}$ and Eq. (2.29), using that $\boldsymbol{b} \times (\dot{\boldsymbol{R}} \times \boldsymbol{B}^*) = (\boldsymbol{B}^* \cdot \boldsymbol{b})\dot{\boldsymbol{R}} - (\dot{\boldsymbol{R}} \cdot \boldsymbol{b})\boldsymbol{B}^*$ and equation Eq. (2.26), we get an equation for $\dot{\boldsymbol{R}}$. The equations of motion for the gyrocenters then read

$$\dot{\boldsymbol{R}} = \frac{\partial(H_0 + H_1)}{\partial p_\parallel}\frac{\boldsymbol{B}^*}{B_\parallel^*} - \frac{c}{eB_\parallel^*}\boldsymbol{b} \times \nabla(H_0 + H_1) \tag{2.30}$$

$$\dot{p}_\parallel = -\frac{\boldsymbol{B}^*}{B_\parallel^*} \cdot \nabla(H_0 + H_1) \tag{2.31}$$

where we denote by $B_\parallel^* = \boldsymbol{B}^* \cdot \boldsymbol{b}$. This is the formulation also obtained in Scott & Smirnov (2010) and Scott *et al.* (2010) who start from the same Lagrangian.

Those equations can be cast in a more familiar form by plugging in the values of H_0 and H_1:

$$\dot{\boldsymbol{R}} = \frac{p_\parallel}{m}\boldsymbol{b} - \left(\frac{p_\parallel}{m}\right)^2\frac{mc}{eB_\parallel^*}\boldsymbol{b} \times \frac{\nabla p}{B^2} + \left(\frac{\mu B}{m} + \left(\frac{p_\parallel}{m}\right)^2\right)\frac{mc}{eB_\parallel^*}\boldsymbol{b} \times \frac{\nabla B}{B} + \frac{c}{eB_\parallel^*}e\boldsymbol{b} \times \nabla J_0\Phi,$$

$$\dot{p}_\parallel = \mu B\nabla \cdot \boldsymbol{b} + \frac{\mu c}{eB_\parallel^*}p_\parallel\boldsymbol{b} \times \frac{\nabla p}{B^2} \cdot \nabla B + e\nabla J_0\Phi \cdot \left(-\boldsymbol{b} + \frac{c}{eB_\parallel^*}p_\parallel\left(\boldsymbol{b} \times \frac{\nabla p}{B^2} - \frac{\boldsymbol{b} \times \nabla B}{B}\right)\right),$$

$$\nabla p \equiv \frac{1}{4\pi}(\nabla \times \boldsymbol{B}) \times \boldsymbol{B}. \tag{2.32}$$

We can now check that the Liouville theorem, which is a fundamental property, is verified for the gyrocenter equations of motion Eqs. (2.30)-(2.31) with Jacobian $B_\parallel^*$. We denote by $H_e = H_0 + H_1$.

$$\nabla \cdot (B_\parallel^*\dot{\boldsymbol{R}}) + \frac{\partial(B_\parallel^*\dot{p}_\parallel)}{\partial p_\parallel} = \frac{\partial\nabla H_e}{\partial p_\parallel} \cdot \boldsymbol{B}^* - \frac{c}{e}\nabla H_e \cdot \nabla \times \boldsymbol{b} + \frac{c}{e}\nabla \times \boldsymbol{b} \cdot \nabla H_e - \boldsymbol{B}^* \cdot \frac{\partial\nabla H_e}{\partial p_\parallel} = 0, \tag{2.33}$$

where we made use of $\nabla \cdot \boldsymbol{B}^* = 0$ which is a necessary condition for Liouville's theorem to be satisfied. As $\boldsymbol{B}^*$ is defined as a curl this is obviously verified.

The equation for the electrostatic potential, the so called *polarisation equation*, or *GK Poisson equation*, is derived by computing the functional derivative of L (given by Eq. (2.17)) with respect to the electrostatic potential and setting it to zero. Let us compute this functional derivative using the definition Eq. (2.19). We first notice that the only dependence on Φ in the first integral defining L is in $H_1 = eJ_0\Phi$ in the electrostatic case and as the gyro average operator J_0 is a linear operator of Φ we simply have

$$\frac{\delta L}{\delta\Phi} \cdot \delta\Phi = -\sum_{\text{sp}}\int d\Omega \, eJ_0(\delta\Phi)f + \sum_{\text{sp}}\int d\Omega\frac{mc^2}{B^2}f_M\nabla_\perp\Phi \cdot \nabla_\perp\delta\Phi = 0 \quad \forall\delta\phi. \tag{2.34}$$

This is the variational formulation sometimes also called the weak form of the polarisation equation that defines the electrostatic potential Φ. This variational formulation can be used directly in a finite element discretisation as we will see later. Let us however single out the test function $\delta\Phi$ to express the polarisation equation in its usual form. For this we need on the one hand to shift the gyro average operator from $\delta\Phi$ to f, this is done

by defining its adjoint operator $J_0^\dagger$ by

$$\int \delta\Phi J_0^\dagger f \mathrm{d}W = \int f J_0(\delta\Phi)\mathrm{d}W. \tag{2.35}$$

Then using a Green's formula in the second integral of (2.34) assuming Φ vanishes on the boundary, and taking out the $B_\parallel^*$ that is hidden in the $\mathrm{d}W$, we get

$$-\sum_{\mathrm{sp}} \int \mathrm{d}V \delta\Phi \int \mathrm{d}W \left(eJ_0^\dagger f + \frac{1}{B_\parallel^*}\nabla \cdot (B_\parallel^* \frac{mc^2}{B^2} f_M \nabla_\perp \Phi) \right) = 0 \tag{2.36}$$

note that, if one wishes to apply directly J_0 to f instead of the test function, we need $J_0^\dagger = J_0$, which means that J_0 has to be Hermitian. Such a J_0 can be constructed as shown in Mcmillan *et al.* (2012) but is not necessary when we work directly on the variational formulation with Finite Elements, as then J_0 can be applied to the test function $\delta\Phi$.

The arbitrariness of $\delta\Phi$ implies that

$$\sum_{\mathrm{sp}} \int \mathrm{d}W \left(eJ_0^\dagger f + \frac{1}{B_\parallel^*}\nabla \cdot \left(B_\parallel^* \frac{mc^2}{B^2} f_M \nabla_\perp \Phi \right) \right) = 0 \tag{2.37}$$

since the integral with respect to $\mathrm{d}p_\parallel \mathrm{d}\mu$ commutes with ∇, the velocity integral on the second term can be performed, leading to the following linear polarisation equation:

$$\sum_{\mathrm{sp}} \left(\int \mathrm{d}W eJ_0^\dagger f + \nabla \cdot \frac{mn_0 c^2}{B^2}\nabla_\perp \Phi \right) = 0 \tag{2.38}$$

where n_0 is the density associated with the equilibrium Maxwellian f_M.

The expression of the polarisation equation clarifies the meaning of the approximations made: it is a linear equation and it has the form of $\sum_{\mathrm{sp}} en_{\mathrm{sp}} = 0$, where en_{sp} is the *particle* density, *i.e.* a quasi-neutrality condition.

In summary, the GK model used in the following is:

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{R}} \cdot \nabla f + \dot{p}_\parallel \frac{\partial f}{\partial p_\parallel} = 0, \tag{2.39}$$

$$\dot{\boldsymbol{R}} = \frac{p_\parallel}{m}\frac{\boldsymbol{B}^*}{B_\parallel^*} - \frac{c}{eB_\parallel^*}\boldsymbol{b} \times (\mu\nabla B + e\nabla J_0 \Phi), \tag{2.40}$$

$$\dot{p}_\parallel = -\frac{\boldsymbol{B}^*}{B_\parallel^*} \cdot (\mu\nabla B + e\nabla J_0 \Phi), \tag{2.41}$$

$$\sum_{\mathrm{sp}} \left(\int \mathrm{d}W eJ_0^\dagger f + \nabla \cdot \left(\frac{n_0 mc^2}{B^2}\nabla_\perp \Phi \right) \right) = 0. \tag{2.42}$$

Despite all the approximations made, this model is physically relevant and it can be used to describe a large class of micro-instabilities excited by the density and temperature gradients, like ion temperature gradient (ITG) driven modes or trapped electron modes (TEM).

Thanks to its derivation from a Lagrangian density which does not directly depend on time, there is a conserved energy. In our case the following energy is conserved [Dubin *et al.* (1983)]:

$$\mathcal{E}(t) = \sum_{\mathrm{sp}} \left(\int d\Omega (H_0 + H_1)f + \int d\Omega H_2 f_M \right). \tag{2.43}$$

Let us verify this by direct computation. As H_0 and f_M do not depend on time

$$\frac{\mathrm{d}\mathcal{E}}{\mathrm{d}t}(t) = \sum_{\mathrm{sp}} \left(\int \mathrm{d}\Omega (H_0 + H_1) \frac{\partial f}{\partial t} + \int \mathrm{d}\Omega \frac{\partial H_1}{\partial t} f + \int \mathrm{d}\Omega \frac{\partial H_2}{\partial t} f_M \right). \tag{2.44}$$

We first notice that

$$\frac{\partial H_i}{\partial t} = \frac{\delta H_i}{\delta \Phi} \cdot \frac{\partial \Phi}{\partial t}$$

so that

$$\sum_{\mathrm{sp}} \left(\int \mathrm{d}\Omega \frac{\partial H_1}{\partial t} f + \int \mathrm{d}\Omega \frac{\partial H_2}{\partial t} f_M \right) = \frac{\delta L}{\delta \Phi} \cdot \frac{\partial \Phi}{\partial t} = 0. \tag{2.45}$$

On the other hand, we have for each species independently, denoting by $H_e = H_0 + H_1$,

$$\int \mathrm{d}\Omega H_e \frac{\partial f}{\partial t} = 0.$$

This follows from the Hamiltonian structure of the Vlasov equation (Poisson brackets), but can also be verified as follows. As $\nabla \cdot \boldsymbol{B}^* = 0$ and $\partial \boldsymbol{B}^*/\partial p_\parallel = c/e \nabla \times \boldsymbol{b}$, we have

$$\nabla \cdot (\boldsymbol{B}^* \frac{\partial H_e^2}{\partial p_\parallel} f) + \frac{\partial}{\partial p_\parallel} (f \nabla \cdot (H_e^2 \boldsymbol{B}^*)) = \boldsymbol{B}^* \cdot \nabla f \frac{\partial H_e^2}{\partial p_\parallel} - \boldsymbol{B}^* \cdot \nabla H_e^2 \frac{\partial f}{\partial p_\parallel} - \frac{c}{e} f \nabla \times \boldsymbol{b} \cdot \nabla H_e^2. \tag{2.46}$$

Integrating over phase space, the terms on the left hand side vanish. On the other hand

$$f \nabla \times \boldsymbol{b} \cdot \nabla H_e^2 = \nabla H_e^2 \cdot (\boldsymbol{b} \times \nabla f + \nabla \times (f \boldsymbol{b})). \tag{2.47}$$

Then, as $\int \mathrm{d}V \mathrm{d}p_\parallel \mathrm{d}\mu \nabla H_e^2 \cdot \nabla \times (f\boldsymbol{b}) = 0$, using the GK Vlasov equation we get

$$2 \int \mathrm{d}V \mathrm{d}p_\parallel \mathrm{d}\mu B_\parallel^* H_e \frac{\partial f}{\partial t} = \int \mathrm{d}V \mathrm{d}p_\parallel \mathrm{d}\mu \nabla H_e^2 \cdot \frac{c}{e} \boldsymbol{b} \times \nabla f - \boldsymbol{B}^* \cdot \left(\nabla f \frac{\partial H_e^2}{\partial p_\parallel} - \nabla H_e^2 \frac{\partial f}{\partial p_\parallel} \right) = 0. \tag{2.48}$$

In the electrostatic, quasi-neutral limit that we consider, we have

$$H_0 = m \frac{U^2}{2} + \mu B, \quad H_1 = e J_0 \Phi, \quad H_2 = -\frac{mc^2}{2B^2} |\nabla_\perp \Phi|^2. \tag{2.49}$$

Then our conserved energy from Eq. (2.43) becomes:

$$\mathcal{E}(t) = \sum_{\mathrm{sp}} \int \mathrm{d}\Omega f \left(m \frac{U^2}{2} + \mu B + e J_0 \Phi \right) - \int \mathrm{d}\Omega f_M \frac{mc^2}{2B^2} |\nabla_\perp \Phi|^2$$

$$= \sum_{\mathrm{sp}} \int \mathrm{d}\Omega f \left(m \frac{U^2}{2} + \mu B \right) + \int \mathrm{d}\Omega f_M \frac{mc^2}{2B^2} |\nabla_\perp \Phi|^2 \tag{2.50}$$

using the variational form of the polarisation equation Eq. (2.34) with $\delta \Phi = \Phi$ which reads

$$\sum_{\mathrm{sp}} \int \mathrm{d}\Omega f_M \frac{mc^2}{B^2} |\nabla_\perp \Phi|^2 = \sum_{\mathrm{sp}} \int \mathrm{d}\Omega f e J_0 \Phi. \tag{2.51}$$

Note that the same relation can be obtained by multiplying the polarisation equation Eq. (2.38) by Φ and integrating (by part) over volume [Scott (2010)]. Because of this last

relation the energy can be written equivalently

$$\mathcal{E} = \sum_{\text{sp}} \int \mathrm{d}\Omega f (m\frac{U^2}{2} + \mu B + \frac{1}{2}eJ_0\Phi) \equiv \mathcal{E}_K + \mathcal{E}_F \tag{2.52}$$

with $\mathcal{E}_F \equiv 1/2 \sum_{\text{sp}} \int \mathrm{d}\Omega e f J_0 \Phi$.

The power balance equation, also called $E \times B$-thermal transfer equations is:

$$\frac{\mathrm{d}\mathcal{E}_k}{dt}(t) = -\frac{\mathrm{d}\mathcal{E}_F}{dt}(t) \tag{2.53}$$

It can be verified, using the Euler-Lagrange equations, that

$$\frac{\mathrm{d}\mathcal{E}_k}{dt}(t) = -\sum_{\text{sp}} \int \mathrm{d}\Omega f e \nabla (J_0\Phi) \cdot \dot{\boldsymbol{R}}_0, \tag{2.54}$$

where $\boldsymbol{R}_0$ represents the part of Eq.(2.40) which does not contain terms in Φ. This quantity can be compared to the time derivative of the field energy:

$$\frac{\mathrm{d}\mathcal{E}_F}{dt}(t) = \frac{\mathrm{d}}{dt}\left(\sum_{\text{sp}} \int \mathrm{d}\Omega f \frac{1}{2}eJ_0\Phi\right) \tag{2.55}$$

In numerical simulations it is particularly useful to consider the following form of the power balance equation:

$$\frac{1}{\mathcal{E}_F}\frac{\mathrm{d}\mathcal{E}_k}{dt}(t) = -\frac{1}{\mathcal{E}_F}\frac{\mathrm{d}\mathcal{E}_F}{dt}(t) \quad . \tag{2.56}$$

Figure 1 shows an example of power balance in a global nonlinear gyrokinetic simulation, for the CYCLONE base section described in section 4. The power balance equation not only gives an indication of the quality of the simulation, but also provides, in linear simulations, a measure of the instantaneous growth rate as

$$\gamma = \frac{1}{2}\frac{d}{dt}\log \mathcal{E}_F = \frac{1}{2}\frac{1}{\mathcal{E}_F}\dot{\mathcal{E}}_F.$$

Hence

$$\gamma = \frac{1}{2\mathcal{E}_F}\sum_{\text{sp}} \int \mathrm{d}\Omega f e \nabla (J_0\Phi) \cdot \dot{\boldsymbol{R}}_0. \tag{2.57}$$

The different contributions to the growth rate arising from the different terms in the gyrocenter velocity can be easily separated in the power balance equation and give a clear indication of the kind of instabilities present in the system in both linear and nonlinear simulations:

$$\gamma = \frac{1}{2\mathcal{E}_F}\sum_{\text{sp}} \int \mathrm{d}\Omega f e \nabla (J_0\Phi) \cdot (\boldsymbol{v}_\parallel + \boldsymbol{v}_{\nabla p} + \boldsymbol{v}_{\nabla B}),$$

$$\boldsymbol{v}_\parallel = \frac{p_\parallel}{m}\boldsymbol{b},$$

$$\boldsymbol{v}_{\nabla p} = -\left(\frac{p_\parallel}{m}\right)^2 \frac{mc}{eB_\parallel^*}\boldsymbol{b} \times \frac{\nabla p}{B^2},$$

$$\boldsymbol{v}_{\nabla B} = \left(\frac{\mu B}{m} + \left(\frac{p_\parallel}{m}\right)^2\right) \frac{mc}{eB_\parallel^*}\boldsymbol{b} \times \frac{\nabla B}{B} \quad .$$

The time evolution of the instantaneous growth rate for a typical toroidal ITG mode is
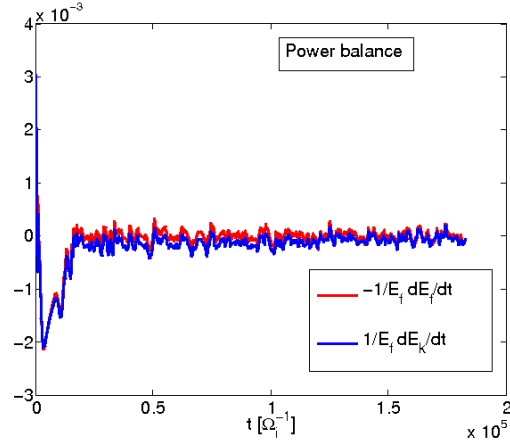
FIGURE 1. Time evolution of the right-hand-side and left-hand-side of the power balance equation (2.57) for the nonlinear CYCLONE base case described in Section 4, code ORB5.
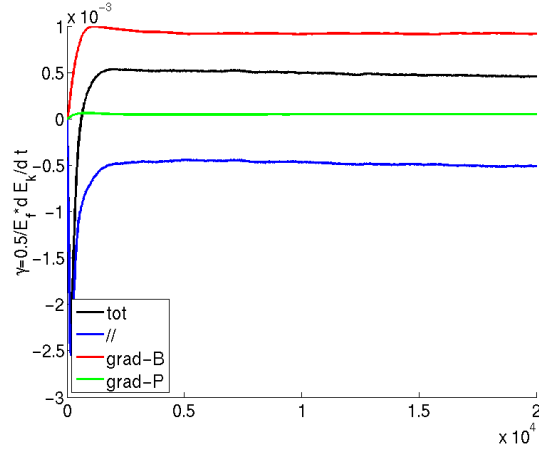

FIGURE 2. Time evolution of the different contributions to the instantaneous growth rate, equation 2.58, for the most unstable mode of linear CYCLONE base case described in Section 4, code ORB5.

illustrated in Figure 2. In general the ITG mode is driven unstable by particle magnetic drifts related to the inhomogeneity (gradient and curvature) of the tokamak magnetic field in the presence of a temperature gradient. Indeed, the power balance diagnostics clearly show that the magnetic drifts are the main destabilizing mechanism while the parallel motion has a stabilizing effect.

The gyrokinetic field theory approach to the derivation of the GK equations leads naturally to the known result on energetic consistency: the same Hamiltonian must be used to construct the polarisation equation and the gyrokinetic Vlasov equations. All the approximations are done in the Lagrangian L (hence H) and then the equations are derived without any further approximation or additional ordering. This also implies that the approximations made cannot be relaxed once the equations have been derived. For example, as already discussed, the linearised polarisation equation has been obtained by considering only $fH = f(H_0 + H_1)$ while the term fH_2 was replaced by $f_M H_2$, as a result, only $H_0 + H_1$ contributed to the Euler-Lagrange particle motion equations. Therefore,

if we want to construct a model with nonlinear polarisation, second order terms in the fields, related to $H_2 f$, must be included in the equation of motion. On the other hand, if we want to extend the model by including second order terms in the equations of motion, a nonlinear polarisation equation has to be used. A more detailed discussion of this topic can be find in Scott & Smirnov (2010) and Sugama (2000).

2.3. *Adiabatic electrons*

The electron contribution to the polarisation equation is often further approximated by replacing the electron density with a fluid approximation for the electron species. Using a fluid model for the electron motion, the equation for the parallel electron momentum, at the lowest order, is:

$$m_e \frac{\mathrm{d}v_{e\parallel}}{\mathrm{d}t} = e\nabla_\parallel \Phi - \frac{T_e}{n_e}\nabla_\parallel n_e. \tag{2.58}$$

We suppose that electrons react instantaneously to the electrostatic potential, therefore the inertial term in the previous equation can be neglected: this is equivalent to impose $m_e \to 0$. The general solution of Eq. (2.58), for vanishing m_e, is:

$$n_e(\boldsymbol{R}, t) = F \exp \frac{e\Phi(\boldsymbol{R}, t)}{T_e}. \tag{2.59}$$

In this model, known as adiabatic electrons approximation, the electron density is therefore proportional to the electrostatic potential. The value of the free function F must be fixed by additional constraints. The equilibrium configuration of a magnetically confined plasma consists in general of a sequence of nested magnetic surfaces, or flux surfaces. Almost each flux surface is covered ergodically by a single field line. In the case of electrostatic waves in a plasma with well defined magnetic flux surfaces, there is no net radial transport of particles if the electrons are exactly adiabatic, since the radial particle flux vanishes when averaged over a flux surface [Dorland (1983)]. As a result, the total number of electrons on each flux surface must be conserved. Following the derivation proposed by Dorland (1983), this constraint can be used to fix the value of the integration constant in Eq. (2.59) for magnetic configurations with well defined flux surfaces (without magnetic islands). This can be achieved by taking the flux-surface average of Eq. (2.59) :

$$\bar{n}_e(s, t) = \langle F \exp \frac{e\Phi(\boldsymbol{R}, t)}{T_e} \rangle, \tag{2.60}$$

having introduced the flux-surface average operator $\langle \rangle$ and the flux surface averaged density, $\bar{n}_e(s, t) = \langle n_e(\boldsymbol{R}, t) \rangle$, where the flux surface label s plays the role of a radial coordinate. A detailed description of the flux-surface average operator in tokamaks can be found in Hinton & Hazeltine (1976). The condition of conservation of the number of electrons on each flux surface is $\bar{n}_e(s, t) = n_{e0}(s)$, leading to

$$n_e(\vec{x}, t) = n_{e0}(s) \exp eT_e(s)(\Phi - \bar{\Phi}) \simeq n_{e0}(s) + \frac{en_{e0}}{T_e(s)}(\Phi - \bar{\Phi}), \tag{2.61}$$

where $\bar{\Phi}(s, t) = \langle \Phi(\boldsymbol{R}, t) \rangle$ is the electrostatic flux surfaced averaged potential. Finally, the linearised quasi-neutrality equation for adiabatic electrons is:

$$\frac{n_{e0}}{T_e}(\Phi - \bar{\Phi}) - \sum_{\mathrm{sp} \neq \mathrm{e}} \nabla \frac{n_0 m c^2}{B^2} \nabla_\perp \Phi = -n_{e0} + \sum_{\mathrm{sp} \neq \mathrm{e}} \int \mathrm{d}W e J_0 f \tag{2.62}$$

where $\bar{\Phi}$ is the flux surface average electrostatic potential. Although here the adiabatic electrons approximation was not done in the Lagrangian, it is possible to show that

the resulting set of equations can also be obtained via a Lagrangian formulation as described in Scott & Smirnov (2010). The adiabatic electron approximation largely reduces the difficulty and the numerical cost of gyrokinetic simulations. However non adiabatic electron effects, in particular the trapped electron dynamics, play an important role in experimentally relevant plasmas.

3. Particle-in-cell (PIC) discretisation

Different numerical methods can be used to solve the set of GK equations. One of the most common is the Lagrangian approach, based on a Monte-Carlo algorithm in which a finite collection of initial positions in phase-space is sampled by a set of particles, often called markers. This is the loading step in a PIC algorithm. Euler-Lagrange equations are solved to follow marker orbits in 5D, which is called pushing. On the other hand the fields are solved on a grid. To this aim the source terms for the field equations at every time step need to be computed on the fixed grid by a procedure called charge and current assignment. Field equations are then solved on the same (3D) grid by means of finite difference or finite element methods, thus providing the electric field on fixed grid. Interpolation techniques can then be used to get the value of the electric field at the marker position. The Lagrangian approach is often referred to as particle-in-cell (PIC). The PIC method for standard plasma physics applications is described in the textbooks Birdsall & Langdon (2004); Hockney & Eastwood (1988). It was introduced very early in the context of gyrokinetics by Lee (1983).

Let us now explain how the PIC method fits in the framework of Monte Carlo simulations and how the literature on this subject can be used to understand and improve PIC simulations. Good introductory textbooks on Monte Carlo simulations are Dunn & Shultis (2012) and Liu (2008). A discussion on the role of Monte Carlo sampling in gyrokinetic PIC methods can also be found in a review paper by Krommes (2007).

3.1. *Monte Carlo simulation*

The principle of a Monte Carlo simulation is to approximate the expected value of a random variable by an average over a large number of samples. For our purposes a random variable X is a function that can take values in $\mathbb{R}$ (or $\mathbb{R}^n$ for a vector valued random variable), which are distributed according to a probability density function. The random number generator available in numerical computation software typically generates a pseudo-random sequence uniformly distributed in $[0, 1]$, which corresponds to the uniform probability density $f = 1$. From this there are several techniques that enable to generate samples of random variables that are distributed according to any density f. In a particle-in-cell simulation, the probability density that will be used to generate samples will be the initial distribution of the particles f_0, normalized to one so that it becomes a probability density. This procedure is called marker loading in the PIC literature, and consist in generating an initial particle distribution in phase space. In this case the random variable Z is the phase space position of a marker, which is distributed according to the initial particle phase-space density f_0. Each marker can then be seen as one realization (*i.e.* one random draw) of the random variable Z. Another point of view which is more convenient for the Monte Carlo theory, is to consider only one realization for each of N random variables Z_i which are identically distributed, *i.e.* which are drawn according to the same probability density. With both interpretations, we get at the end a sample of N independent markers that can be used to represent the probability density f_0.

Once a random variable X of probability density function f is available. In our case

the phase space positions of the markers, arbitrary smooth functions ψ of these define new random variables, and we can compute expected values of these using the definition:

$$\mathbb{E}(\psi(X)) = \int \psi(x) f(x)\, \mathrm{d}x.$$

We refer to Dunn & Shultis (2012), for a more detailed background on the probability terminology for Monte Carlo simulations, explaining in particular also the important notion of independence of random variables.

Now, the Monte Carlo method simply consists in approximating the expected value of some random variables by a sample average. To approximate $\mathbb{E}(\psi(X))$ we consider a sequence of independent random variables $(X_i)_i$ distributed like X and approximate $\mathbb{E}(\psi(X))$ by the sample mean

$$M_N = \frac{1}{N} \sum_{i=1}^{N} \psi(X_i). \tag{3.1}$$

In order for this procedure to be useful, we need to check that the approximation we defined converges in some sense to the exact value and possibly estimate the speed of convergence.

Here the sample mean is an example of what is called an *estimator* in statistics, which is a rule for computing some statistical quantity, which is a function of the random variable, here the expected value, from sample data.

DEFINITION 1. *The difference between the expected value of the estimator and the statistical quantity it approximates is called bias. If this difference is zero, the estimator is said to be unbiased.*

Let us compute the bias of the sample mean given by Eq. (3.1). We easily get as the X_i are all distributed like X and thus have the same expected value that

$$\mathbb{E}(M_N) = \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}(\psi(X_i)) = \mathbb{E}(\psi(X))$$

so that the bias is zero and our sample mean is unbiased.

Useful statistical quantities in Monte Carlo simulations are variance and covariance, that are defined as follows:

DEFINITION 2. *Let X and Y be two square integrable real valued random variables. The* variance *of the random variable X is defined by*

$$\mathbb{V}(X) = \mathbb{E}(|X - \mathbb{E}(X)|^2) \geqslant 0$$

and

$$\sigma(X) = \sqrt{\mathbb{V}(X)}$$

is called standard deviation *of the random variable X. The* covariance *of X and Y is defined by*

$$\mathrm{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

We have obviously that $\mathbb{V}(X) = \mathrm{Cov}(X, X)$. The variance can be also expressed by $\mathbb{V}(X) = \mathbb{E}(|X|^2) - \mathbb{E}(X)^2$. Indeed, this follows from a simple computation, relying on the linearity of the expected value, which is an integral, and as $a = \mathbb{E}(X)$ is a number we

have, as f is a probability density which integrates to one

$$\mathbb{E}(a) = \int a f(x)\,\mathrm{d}x = a \int f(x)\,\mathrm{d}x = a.$$

In the same way

$$\mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y).$$

A useful result is Bienaymé's equality:

THEOREM 1 (BIENAYMÉ). *If $X_1, \ldots, X_m$ are independent real valued random variables with $\mathbb{V}(|X_i|) < +\infty$ then*

$$\mathbb{V}(X_1 + \cdots + X_m) = \mathbb{V}(X_1) + \cdots + \mathbb{V}(X_m).$$

Assuming the sample number $N \geqslant 2$ an unbiased estimator of the variance is given by the following sample variance

$$V_N = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - M_N)^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(X_i - \frac{1}{N} \sum_{i=1}^{N} X_i \right)^2. \tag{3.2}$$

Indeed, let us compute the expected value of V_N. Denoting by $a = \mathbb{E}(X_i)$ for $i = 1, \ldots, N$, we have

$$V_N = \frac{1}{N-1} \sum_{i=1}^{N} ((X_i - a) + (a - M_N))^2 = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - a)^2 - \frac{N}{N-1}(M_N - a)^2,$$

as $2 \sum_{i=1}^{N} (X_i - a)(a - M_N) = -2N(M_N - a)^2$. Hence

$$\mathbb{E}(V_N) = \frac{1}{N-1} \sum_{i=1}^{N} \mathbb{E}((X_i - a)^2) - \frac{N}{N-1} \mathbb{E}((M_N - a)^2) = \frac{1}{N-1} \sum_{i=1}^{N} \mathbb{V}(X_i) - \frac{N}{N-1} \mathbb{V}(M_N).$$

And because of Bienaymé's theorem

$$N^2 \mathbb{V}(M_N) = \mathbb{V}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathbb{V}(X_i) = N\mathbb{V}(X).$$

Hence

$$\mathbb{E}(V_N) = \frac{N}{N-1} \mathbb{V}(X) - \frac{1}{N-1} \mathbb{V}(X) = \mathbb{V}(X).$$

REMARK 1. *Note the $1/(N-1)$ factor in the variance estimator instead of the $1/N$ that one would expect at the first glance. Using $1/N$ instead would also yield an estimator of the variance, but this one would be biased, i.e. it would not have the right expected value.*

3.2. *Estimation of the error in a Monte Carlo simulation*

Let us first compute in a general way the mean square error (MSE) of an estimator. Assume $\hat{\theta}$ is an estimator for the statistical quantity θ which is a real number that can be computed as a function of a random variable X. The MSE is defined by

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2) = \int (\hat{\theta} - \theta)^2 dP.$$

Note that the root mean square error or RMS error, which is the square root of the MSE, is the classical L^2 error.

LEMMA 1. *Assume the random variable $\hat{\theta}$ is an estimator for θ and $\mathbb{E}(\hat{\theta}^2) < +\infty$. Then*

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2) = \mathbb{V}(\hat{\theta}) + \mathrm{Bias}(\hat{\theta})^2. \tag{3.3}$$

Proof. A straightforward calculation yields

$$\begin{aligned}
\mathrm{MSE}(\hat{\theta}) = \mathbb{E}((\hat{\theta} - \theta)^2) &= \mathbb{E}(\hat{\theta}^2) + \theta^2 - 2\theta\mathbb{E}(\hat{\theta}) \\
&= \mathbb{E}(\hat{\theta}^2) - \mathbb{E}(\hat{\theta})^2 + \mathbb{E}(\hat{\theta})^2 + \theta^2 - 2\theta\mathbb{E}(\hat{\theta}) \\
&= (\mathbb{E}(\hat{\theta}^2) - \mathbb{E}(\hat{\theta})^2) + (\mathbb{E}(\hat{\theta}) - \theta)^2 \\
&= \mathbb{V}(\hat{\theta}) + (Bias(\hat{\theta}))^2.
\end{aligned}$$

$\square$

Assume that the random variable X defining our Monte Carlo simulation verifies $\mathbb{E}(X^2) < +\infty$. Then we can apply the previous lemma to M_N as an estimator of $\mathbb{E}(X)$, which yields

$$\mathrm{MSE}(M_N) = \mathbb{V}(M_N) + (\mathbb{E}(M_N) - \mathbb{E}(X))^2.$$

So the RMS error is composed of two parts, the error coming from the variance of the sample and the possible bias on the sample occurring when the expected value of M_N is not exactly equal to the expected value of the random variable X being approximated.

In many cases the bias can be made to be zero, but in some cases it can be useful to introduce some bias in order to decrease the variance of the sample and the total error.

LEMMA 2. *Assume $\mathbb{E}(X^2) < +\infty$. Then the RMS error for an unbiased simulation based on the random variable X is*

$$e_{rms} = \sigma(M_N) = \frac{\sigma(X)}{\sqrt{N}}.$$

Proof. The formula Eq. (3.3) giving the mean squared error of an estimator shows that if the simulation is unbiased $\mathbb{E}(M_N) = \mathbb{E}(X)$ and

$$e_{rms} = \sqrt{\mathbb{V}(M_N)} = \sigma(M_N).$$

Now using Bienaymé's theorem we also have

$$N^2\mathbb{V}(M_N) = \mathbb{V}(\sum_{i=1}^{N} X_i) = \sum_{i=1}^{N} \mathbb{V}(X_i) = N\mathbb{V}(X).$$

And thus $\mathbb{V}(M_N) = \mathbb{V}(X)/N$, which gives the result. $\square$

The law of large numbers, strong or weak, implies that the sample mean converges towards the desired expected value, which justifies the Monte Carlo method.

Another major theorem of probability theory, the central limit theorem, gives a precise estimation of the error committed by an approximation.

THEOREM 2 (CENTRAL LIMIT THEOREM). *Assume $(X_1, X_2, \ldots, X_N)$ is a sequence of independent identically distributed random variables such that $\mathbb{V}(X) = \sigma^2(X) < \infty$. Then*

$$\lim_{N \to +\infty} P\left[\frac{|M_N - \mathbb{E}(X)|}{\sigma(X)/\sqrt{N}} \leqslant \lambda\right] = \frac{1}{\sqrt{2\pi}} \int_{-\lambda}^{\lambda} e^{-u^2/2} \, du. \tag{3.4}$$

This tells us that the asymptotic distribution of $(M_N - \mathbb{E}(X))/(\sigma(X)/\sqrt{N})$ is a unit normal distribution, or equivalently that M_N is a normal distribution with mean $\mathbb{E}(X)$ and standard deviation $\sigma(X)/\sqrt{N}$.

The right hand side of Eq. (3.4) is a number that can be computed explicitly, and that is called *confidence coefficient*. For $\lambda = 3$ the confidence coefficient is 0.9973 and for $\lambda = 4$ the confidence coefficient is 0.9999 (see e.g. Dunn & Shultis (2012) for other values). This is the probability that the true mean lies in the so-called *confidence interval* $[M_N - \lambda\sigma(X)/\sqrt{N}, M_N + \lambda\sigma(X)/\sqrt{N}]$. Note that deterministic error estimates are generally of the form h^p or $1/N^p$, where h is a cell size and N a number of discretisation points, and lie on a deterministic curve. As opposite to this, the error estimate in a Monte Carlo method is random, but it is always a normal distribution with variance which tends to 0 when the number of sample points tends to $+\infty$. In practice a good estimate of the error is given by $\sigma(X)/\sqrt{N}$, which is all the more interesting that the variance (or standard deviation) can be well estimated by the sample variance (or sample standard deviation), which is an a posteriori estimate that can be directly used in actual computations to measure the error.

3.3. *The Monte Carlo PIC algorithm*

We now derive the Monte Carlo PIC algorithm from a given Lagrangian as was first proposed by Lewis (1970) and also more recently by Evstatiev & Shadwick (2013). We consider a Lagrangian that is build using the single particle Lagrangian for each particle species L_{sp} and a field Lagrangian L_F:

$$L := \sum_{\mathrm{sp}} \int f_{\mathrm{sp}}(\boldsymbol{Z}_0, t_0) L_{\mathrm{sp}}(\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t), \dot{\boldsymbol{Z}}(\boldsymbol{Z}_0, t_0; t)) \, \mathrm{d}\boldsymbol{Z}_0 + L_F. \tag{3.5}$$

Here $f_{\mathrm{sp}}(\boldsymbol{Z}_0, t_0)$ denotes the initial phase space density of species sp, and $\boldsymbol{Z}(\boldsymbol{Z}_0, t_0; t)$ are the characteristics of the GK-Vlasov equations, *i.e.* the particle phase space trajectories, with initial condition $\boldsymbol{Z}_0$ at time t_0. Using the Liouville theorem the Lagrangian can also be written as

$$L = \sum_{\mathrm{sp}} \int f_{\mathrm{sp}}(\boldsymbol{Z}, t) L_{\mathrm{sp}}(\boldsymbol{Z}, \dot{\boldsymbol{Z}}) \mathrm{d}\boldsymbol{Z} + L_F. \tag{3.6}$$

We notice here that the particle contribution to the Lagrangian is written as a sum of integrals with densities f_{sp}. Using the definition of expected values, each of them can be replaced by an expected value of a random variable $\boldsymbol{Z}_{\mathrm{sp}}(t)$ of probability density $f_{\mathrm{sp}}(\cdot, t)$. This yields

$$L = \sum_{\mathrm{sp}} \mathbb{E}\left(L_{\mathrm{sp}}(\boldsymbol{Z}_{\mathrm{sp}}(t), \dot{\boldsymbol{Z}}_{\mathrm{sp}}(t)) \right) + L_F. \tag{3.7}$$

The particles contribution of the Lagrangian now being expressed as an expected value, the Monte Carlo method can be applied by replacing the expected value by a sample mean over a large number of samples being drawn independently according to the initial distributions $f_{\mathrm{sp}}(\cdot, t_0)$, which yields the Monte Carlo Lagrangian

$$L_{MC} = \sum_{\mathrm{sp}} \frac{1}{N_p} \sum_{k=1}^{N_p} L_{\mathrm{sp}}(\boldsymbol{z}_k(t), \dot{\boldsymbol{z}}_k(t)) + L_F. \tag{3.8}$$

Even though in does not appear explicitly in our notation, the number of particles N_p can be different for each species. We also remove the sp index from the marker positions to alleviate the notation.

There now remains to discretise the fields. In the variational framework, the most natural method to do this is given by the Galerkin approximation, which consists in doing the variations over functions constrained to remain in a finite dimensional function space. This then leads to a Finite Element approximation of the fields. Introducing the basis functions Λ_ν of the finite dimensional function space, all the functions in this space can be expressed as linear combinations of these basis functions:

$$\Phi_h(\boldsymbol{x}, t) = \sum_{\mu=1}^{N_g} \Phi_\mu(t) \Lambda_\mu(\boldsymbol{x}) \tag{3.9}$$

where $\Phi_\mu(t)$ are real numbers, and $\Lambda_\mu(\boldsymbol{x}) = \Lambda_j(x_1)\Lambda_k(x_2)\Lambda_l(x_3)$ is a product of polynomial basis functions, typically cubic B-splines. Replacing Φ by Φ_h in the Lagrangian Eq.(2.17), we get the following finite dimensional Lagrangian

$$L_{h,MC} = \sum_{sp} \frac{1}{N_p} \sum_{k=1}^{N_p} w_k \left(\left(\frac{e}{c} \boldsymbol{A}(\boldsymbol{R}_k) + p_{\parallel,k} \boldsymbol{b}(\boldsymbol{R}_k) \right) \cdot \dot{\boldsymbol{R}}_k + \frac{mc}{e} \mu_k \dot{\theta}_k - m\frac{U^2}{2} - \mu B(\boldsymbol{R}_k) \right.$$
$$\left. - eJ_0 \Phi_h(\boldsymbol{R}_k) \right) + \sum_{sp} \int \mathrm{d}\Omega \frac{mc^2}{2B^2} |\nabla_\perp \Phi_h|^2 f_M. \tag{3.10}$$

Note that we already here introduced an importance weight w_k that we shall explain in the next section. For the moment we have $w_k = 1$ for all k. We shall compute the Euler-Lagrange equations corresponding to this finite dimensional Lagrangian after having introduced variance reduction techniques which are essential for gyrokinetic PIC computations.

At this level no time discretisation has been performed. As the semi-discrete Lagrangian is still invariant with respect to a time translation, the corresponding Noether energy is exactly conserved. The equations of motions of the particles are discretised using a standard fourth order Runge-Kutta procedure, which breaks the exact energy conservation. However, as will be discussed later, numerical simulations still exhibit in practice very good energy conservation properties.

3.4. *Variance reduction techniques*

As we saw the Monte Carlo error for the approximation of the expected value of a random variable X is in $\sigma(X)/\sqrt{N}$. Apart from increasing the number of realizations N, the most efficient method to reduce the error is to use available information to replace X by another random variable with the same expected value but a lower variance. We shall describe two techniques that are used to this purpose in the context of PIC methods.

3.4.1. *Importance sampling*

We are interested in computing, for some given probability density f, quantities of the form

$$\int \psi(\boldsymbol{z}) f(\boldsymbol{z}) \, \mathrm{d}\boldsymbol{z}.$$

The standard Monte Carlo method for doing this is to define our integral as an expected value using a random variable $\boldsymbol{Z}$ of density f. Then

$$\int \psi(\boldsymbol{z})f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \mathbb{E}(\psi(\boldsymbol{Z})).$$

Depending on the function ψ it might not be the best approach to use directly the density f for drawing the random variable used in the simulation. Indeed if g is any other probability density that does not vanish in the support of f one can express our integral as an expectation using a random variable $\tilde{\boldsymbol{Z}}$ of density g:

$$\int \psi(\boldsymbol{z})f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \int \psi(\boldsymbol{z})\frac{f(\boldsymbol{z})}{g(\boldsymbol{z})}g(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \mathbb{E}(W(\tilde{\boldsymbol{Z}})\psi(\tilde{\boldsymbol{Z}})),$$

where the random variable $W(\tilde{\boldsymbol{Z}}) = f(\tilde{\boldsymbol{Z}})/g(\tilde{\boldsymbol{Z}})$ is called weight.

The Monte Carlo approximation using independent random variables distributed identically with density g can be expressed as

$$\tilde{M}_N = \frac{1}{N}\sum_{i=1}^{N} W(\tilde{\boldsymbol{Z}}_i)\psi(\tilde{\boldsymbol{Z}}_i),$$

from which we get

$$\mathbb{E}(\tilde{M}_N) = \mathbb{E}(W(\tilde{\boldsymbol{Z}})\psi(\tilde{\boldsymbol{Z}})) = \int \psi(\boldsymbol{z})f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}.$$

So $\tilde{M}_N$ is another unbiased estimator of the integral we wish to compute and the approximation error for a given number of samples N is determined by its variance.

Let us now investigate how g can be chosen to get a smaller variance. For this we need to compare the variance of $W(\tilde{\boldsymbol{Z}})\psi(\tilde{\boldsymbol{Z}})$ and the variance of $\psi(\boldsymbol{Z})$ knowing that both have the same expected value.

$$\mathbb{E}(W(\tilde{\boldsymbol{Z}})^2\psi(\tilde{\boldsymbol{Z}})^2) = \int \psi(\boldsymbol{z})^2 W(\boldsymbol{z})^2 g(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \int \psi(\boldsymbol{z})^2 W(\boldsymbol{z})f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}.$$

On the other hand

$$\mathbb{E}(\psi(\boldsymbol{Z})^2) = \int \psi(\boldsymbol{z})^2 f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z}.$$

So we see that there is a factor W difference between the two expressions and obviously if $W < 1$ in regions where ψ is larger, the procedure will lead to a smaller variance. Note that because f and g both have an integral one, we cannot have $W < 1$ everywhere.

We also remark that, assuming that $\psi(\boldsymbol{z})$ does not vanish, if we take $W(\boldsymbol{z}) = \mathbb{E}(\psi(\boldsymbol{Z}))/\psi(\boldsymbol{z})$ which corresponds to $g(\boldsymbol{z}) = f(\boldsymbol{z})\psi(\boldsymbol{z})/\mathbb{E}(\psi(\boldsymbol{Z}))$, we get

$$\mathbb{E}(W(\tilde{\boldsymbol{Z}})^2\psi(\tilde{\boldsymbol{Z}})^2) = \mathbb{E}(\psi(\boldsymbol{Z}))\int \psi(\boldsymbol{z})f(\boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \mathbb{E}(\psi(\boldsymbol{Z}))^2 = \mathbb{E}(W(\tilde{\boldsymbol{Z}})\psi(\tilde{\boldsymbol{Z}}))^2$$

so that $\mathbb{V}(W(\tilde{\boldsymbol{Z}})\psi(\tilde{\boldsymbol{Z}})) = 0$. This of course cannot be done in practice as $\mathbb{E}(\psi(\boldsymbol{Z}))$ is the unknown quantity we wish to approximate, but it can be used as a guideline to find a density g that reduces the variance as much as possible and tells us that the density g should be proportional to the integrand $f\psi$, *i.e.* that markers should be distributed according to the integrand.

3.4.2. *Control variates*

Consider the standard Monte Carlo problem of approximating $a = \mathbb{E}(X)$, for a given random variable X, by a sample mean.

Assume now that there exists a random variable Y the expected value of which is known, that is somehow correlated to X. For a given $\alpha \in \mathbb{R}$, let us define the new random variable

$$Z_\alpha = X - \alpha(Y - \mathbb{E}(Y)).$$

Obviously, we have for any α that $\mathbb{E}(Z_\alpha) = \mathbb{E}(X) = a$, which means that the sample mean of Z_α

$$M_{N,\alpha} = \frac{1}{N}\sum_{i=1}^{N}(X_i - \alpha(Y_i - \mathbb{E}(Y))) = \alpha\mathbb{E}(Y) + \frac{1}{N}\sum_{i=1}^{N}(X_i - \alpha Y_i)$$

could be used instead of the sample mean of X to approximate a. The random variable αY is called a control variate for X.

Let us now look under what conditions the variance of Z_α is lower than the variance of X. We assume that both $\mathbb{V}(X) > 0$ and $\mathbb{V}(Y) > 0$.

LEMMA 3. *If the random variables X and Y are not independent, there exists a value of α for which the variance of Z_α is smaller than the variance of X. More precisely*

$$\min_{\alpha \in \mathbb{R}} \mathbb{V}(Z_\alpha) = \mathbb{V}(X)(1 - \rho^2(X,Y)) = \mathbb{V}(Z_{\alpha^*}), \quad with \ \alpha^* = \frac{\mathrm{Cov}(X,Y)}{\mathbb{V}(Y)}.$$

Moreover

$$\mathbb{V}(Z_\alpha) < \mathbb{V}(X) \Leftrightarrow \left| \begin{array}{ll} \alpha < 2\alpha^* & if\ \alpha > 0, \\ \alpha > 2\alpha^* & if\ \alpha < 0. \end{array} \right.$$

Proof. As $Z_\alpha = X - \alpha Y + \alpha\mathbb{E}(Y)$, and $\mathbb{E}(Z_\alpha) = \mathbb{E}(X)$ we have

$$\mathbb{V}(Z_\alpha) = \mathbb{E}(Z_\alpha^2) - \mathbb{E}(X)^2,$$
$$= \mathbb{E}((X - \alpha Y)^2) + 2\alpha\mathbb{E}(Y)\mathbb{E}(X - \alpha Y) + \alpha^2\mathbb{E}(Y)^2 - \mathbb{E}(X)^2,$$
$$= \mathbb{E}(X^2) - 2\alpha\mathbb{E}(XY) + \alpha^2\mathbb{E}(Y^2) + 2\alpha\mathbb{E}(Y)\mathbb{E}(X) - 2\alpha^2\mathbb{E}(Y)^2 + \alpha^2\mathbb{E}(Y)^2 - \mathbb{E}(X)^2,$$
$$= \mathbb{V}(X) - 2\alpha\mathrm{Cov}(X,Y) + \alpha^2\mathbb{V}(Y),$$
$$= \sigma^2(X) - 2\alpha\sigma(X)\sigma(Y)\rho(X,Y) + \alpha^2\sigma^2(Y),$$

introducing the standard deviation of a random variable $\sigma^2(X) = \mathbb{V}(X)$ and the correlation coefficient of two random variables $\rho(X,Y) = \mathrm{Cov}(X,Y)/(\sigma(X)\sigma(Y))$.

So the variance of Z_α is a second order polynomial in α the minimum of which is reached for

$$\alpha^* = \frac{\sigma(X)}{\sigma(Y)}\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma^2(Y)},$$

and plugging this into the expression of $\mathbb{V}(Z_\alpha)$, we get

$$\mathbb{V}(Z_{\alpha^*}) = \sigma^2(X) - 2\sigma(X)^2\rho(X,Y)^2 + \sigma^2(X)\rho(X,Y)^2 = \mathbb{V}(X)(1 - \rho^2(X,Y)).$$

On the other hand

$$\mathbb{V}(Z_\alpha) - \mathbb{V}(X) = \alpha\sigma(Y)(\alpha\sigma(Y) - 2\sigma(X)\rho(X,Y)).$$

Hence for $\alpha > 0$,

$$\mathbb{V}(Z_\alpha) < \mathbb{V}(X) \Leftrightarrow \alpha < 2\frac{\sigma(X)}{\sigma(Y)}\rho(X,Y) = 2\alpha^*,$$

and for $\alpha < 0$, $\mathbb{V}(Z_\alpha) < \mathbb{V}(X) \Leftrightarrow \alpha > 2\alpha^*$. $\square$

REMARK 2. *This results means that provided* $\mathrm{Cov}(X, Y) \neq 0$, *i.e.* X *and* Y *are not independent, there is always an interval around the optimal value* α^* *for which* Z_α *has a lower variance than* X. *The more correlated* X *and* Y *are, the larger this interval is. So the most important is to find a random variable* Y *the expectation of which is known, that is as correlated with* X *as possible. Then if a good approximation of* $\mathrm{Cov}(X, Y)$ *can be computed, one can use this to get closer to* α^* *and minimize the variance as much as possible with the random variable* Y.

A typical example is when $X = Y + \epsilon \tilde{Y}$, where ϵ is small and $\mathbb{E}(Y)$ is known and for simplicity Y and $\tilde{Y}$ are independent. Plugging this in the expression of $\mathbb{V}(Z_\alpha)$ in the above proof yields

$$\mathbb{V}(Z_\alpha) = \mathbb{V}(Y) + \epsilon^2 \mathbb{V}(\tilde{Y}) - 2\alpha \mathbb{V}(Y) + \alpha^2 \mathbb{V}(Y) = (1 - \alpha)^2 \mathbb{V}(Y) + \epsilon^2 \mathbb{V}(\tilde{Y}).$$

So that taking $\alpha = 1$ yields that $\mathbb{V}(Z_\alpha)$ is of order ϵ^2 assuming $\mathbb{V}(\tilde{Y})$ of order 1. This is typically the form that is used in PIC simulations.

3.5. *Application of the variance reduction techniques to the PIC method.*

For the PIC method, we can combine the importance sampling method and the control variates method.

3.5.1. *Importance sampling*

The choice of a density for importance sampling depends on the expected value that we are interested in. There are many of those in a PIC code, but arguably the accurate computation of the electromagnetic field, which determines the self-consistent dynamics is the most important. Depending on the physical problem we want to deal with more particles will be needed in some specific phase space areas, like for example in some region of the tail for a bump-on-tail instability. For this reason, it is interesting in a PIC code to have the flexibility of drawing the particles according to any density, but one needs to be careful with the choice of this density as the results can become better or worse.

Initialization. Assume we know the density g_0 according to which we want to draw the markers. Then we initialize the marker's phase space positions $\boldsymbol{z}_i^0 = (\boldsymbol{x}_i^0, \boldsymbol{v}_i^0)$ as realizations of a random variable $\boldsymbol{Z}^0$ with density g_0.

Time stepping. The markers evolve along the characteristics of the Vlasov equation so that at time t the random variable $\boldsymbol{Z}(t) = (\boldsymbol{X}(t), \boldsymbol{V}(t))$ is distributed according to the density $g(t, \boldsymbol{z})$, that is the solution of the GK-Vlasov equation with initial condition g_0.

Then as we saw, the different quantities we need to compute using the Monte Carlo approximation are of the form

$$\int \psi(\boldsymbol{z}) f(t, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} = \int \psi(\boldsymbol{z}) \frac{f(t, \boldsymbol{z})}{g(t, \boldsymbol{z})} g(t, \boldsymbol{z}) \, \mathrm{d}\boldsymbol{z} = \mathbb{E}\left(\psi(\boldsymbol{Z}) \frac{f(t, \boldsymbol{Z})}{g(t, \boldsymbol{Z})} \right) \tag{3.11}$$

for some analytically known function $\psi(\boldsymbol{z})$. This means that we need to simulate the random variable $Y(t) = \psi(\boldsymbol{Z}(t)) f(t, \boldsymbol{Z}(t)) / g(t, \boldsymbol{Z}(t)) = \psi(\boldsymbol{Z}(t)) W$, where the random variable W is defined by $W = f(t, \boldsymbol{Z}(t)) / g(t, \boldsymbol{Z}(t))$. Because f and g are conserved along the same characteristics we have

$$W = \frac{f(t, \boldsymbol{Z}(t))}{g(t, \boldsymbol{Z}(t))} = \frac{f_0(\boldsymbol{Z}^0)}{g_0(\boldsymbol{Z}^0)},$$

so that the random variable W does not depend on time and is set once for all at the initialization.

Using importance sampling, we obtain the so-called weighted PIC method, in which the particles or markers are advanced like in the standard PIC method, but have in addition an importance weight which does not evolve in time. The drawback of this method is that the variance can increase when large importance weights and small importance weights are mixed close together in phase space which often happens in long nonlinear simulations.

3.5.2. *Control variates*

We combine here control variates with importance sampling for most generality, but it can also be used without importance sampling by taking $g_0 = f_0$.

In the PIC method expected values of the form Eq. (3.11) cannot be exactly computed because the particle density in phase space $f(t, \boldsymbol{z})$ is not analytically known except at the initial time. However in many problems, and in particular in problems we study in magnetic fusion applications, the distribution function stays close to an analytically known distribution function $\tilde{f}(t, \boldsymbol{z})$, which in our applications is typically the initial canonical Maxwellian. Next to the random variable $Y(t)$ associated to $f(t, \boldsymbol{z})$, this can be used to build the control variate $\tilde{Y}(t)$ associated to $\tilde{f}(t, \boldsymbol{z})$ such that

$$Y(t) = \psi(\boldsymbol{Z})\frac{f(t, \boldsymbol{Z})}{g(t, \boldsymbol{Z})}, \qquad \tilde{Y}(t) = \psi(\boldsymbol{Z})\frac{\tilde{f}(t, \boldsymbol{Z})}{g(t, \boldsymbol{Z})}.$$

Indeed we have

$$\mathbb{E}(\tilde{Y}(t)) = \int \psi(\boldsymbol{z})\frac{\tilde{f}(t, \boldsymbol{z})}{g(t, \boldsymbol{z})}g(t, \boldsymbol{z})\,\mathrm{d}\boldsymbol{z} = \int \psi(\boldsymbol{z})\tilde{f}(t, \boldsymbol{z})\,\mathrm{d}\boldsymbol{z}$$

which can be computed analytically for simple enough functions ψ and $\tilde{f}$. Moreover if $\tilde{f}$ is close enough to f then $\tilde{Y}(t)$ will be close to $Y(t)$ and from the previous discussion a variance reduction of the order of the squared distance between the two random variables can be expected.

Let us now explain how this can be implemented in a PIC simulation.

Initialization. As for importance sampling, the initial phase space positions of the markers are sampled as realizations $(\boldsymbol{z}_i^0)_{1 \leqslant i \leqslant N}$ of the random variable $\boldsymbol{Z}^0$ of density g_0. The importance weights are then defined by the corresponding realizations of the random variable $W = f_0(\boldsymbol{Z}^0)/g_0(\boldsymbol{Z}^0)$, *i.e.* $w_i = f_0(\boldsymbol{z}_i^0)/g_0(\boldsymbol{z}_i^0)$.

We also initialize the importance weights for $\delta f = f - \tilde{f}$, which are defined by the random variable

$$W_\alpha^0 = \frac{f_0(\boldsymbol{Z}^0) - \alpha\tilde{f}(t_n, \boldsymbol{Z}^n)}{g_0(\boldsymbol{Z}^0)} = W - \alpha\frac{\tilde{f}(0, \boldsymbol{Z}^0)}{g_0(\boldsymbol{Z}^0)}. \tag{3.12}$$

Time stepping. The markers $\boldsymbol{Z}$ are advanced by numerically solving the characteristics of the GK-Vlasov equation. This means that given their positions $\boldsymbol{Z}^n$ at time t_n, an ODE solver is used to compute an approximation of their position $\boldsymbol{Z}^{n+1}$ at time t^{n+1}. Because f and g satisfy the same GK-Vlasov equation, they are conserved along the same characteristics so that, as for importance sampling

$$W = \frac{f(t_n, \boldsymbol{Z}^n)}{g(t_n, \boldsymbol{Z}^n)} = \frac{f_0(\boldsymbol{Z}^0)}{g_0(\boldsymbol{Z}^0)}$$

is a random variable which does not depend on time. On the other hand, we know $\tilde{f}$ analytically and know that f and g are conserved along the characteristics, so that we can compute the importance weight for δf at time t_n from the phase space positions of

the markers at the same time:

$$W_\alpha^n = \frac{f(t_n, \boldsymbol{Z}^n) - \alpha \tilde{f}(t_n, \boldsymbol{Z}^n)}{g(t_n, \boldsymbol{Z}^n)} = \frac{f_0(\boldsymbol{Z}^0) - \alpha \tilde{f}(t_n, \boldsymbol{Z}^n)}{g_0(\boldsymbol{Z}^0)} = W - \alpha \frac{\tilde{f}(t_n, \boldsymbol{Z}^n)}{g_0(\boldsymbol{Z}^0)}. \qquad (3.13)$$

So W_α^n is a time dependent random variable which can be computed explicitly using the analytical functions $\tilde{f}$, f_0 and g_0. These values can be used to express the sample mean for the new simulated random variable $\tilde{Y}_\alpha = Y - \alpha(\tilde{Y} - \mathbb{E}(\tilde{Y}))$. This is defined by

$$M_{\alpha,N}^n = \frac{1}{N} \sum_{i=1}^{N} (Y_i^n - \alpha \tilde{Y}_i^n) + \alpha \mathbb{E}(\tilde{Y}).$$

Plugging in the values for Y_i^n and $\tilde{Y}_i^n$ we get

$$M_{\alpha,N}^n = \frac{1}{N} \sum_{i=1}^{N} \left(\psi(\boldsymbol{Z}_i^N) \frac{f(t_n, \boldsymbol{Z}_i^n) - \alpha \tilde{f}(t_n, \boldsymbol{Z}_i^n)}{g(t_n, \boldsymbol{Z}_i^n)} \right) + \alpha \mathbb{E}(\tilde{Y}) = \frac{1}{N} \sum_{i=1}^{N} W_{\alpha,i}^n \psi(\boldsymbol{Z}_i^N) + \alpha \mathbb{E}(\tilde{Y}).$$

This yields an estimator for $\psi(\boldsymbol{Z})$ based on the weights W_α^n and the expected value that can be computed analytically $\mathbb{E}(\tilde{Y})$. If no estimation of the optimal α^* is available this method is used with $\alpha = 1$.

This is classically known as the δf method in the PIC literature [Kotschenreuther (1998); Hu & Krommes (1994); Aydemir (1994); Allfrey & Hatzky (2003)], as its interest lies in the expression $f = \tilde{f} + \delta f$ with $\tilde{f}$ known. A large variance reduction for $\alpha = 1$ is obtained as long as $\delta f \ll \tilde{f}$, else one can also achieve some variance reduction by optimizing for α [Kleiber *et al.* (2011)].

3.6. *Monte Carlo PIC discretisation of the GK equations*

We are now ready to proceed with the Monte Carlo PIC discretization of the GK equation with the discrete Lagrangian Eq. (3.9) as a starting point.

Keeping the possibility to use importance sampling, we choose an initial marker distribution g_0 based on physics considerations rather than the actual initial particle distribution f_0. We then compute samples from a random variable $\boldsymbol{Z}$ distributed with probability density g_0. These will be the initial marker phase space positions $\boldsymbol{z}_k^0$. After that we compute the importance weights for each marker $w_k = f_0(\boldsymbol{z}_k)/g_0(\boldsymbol{z}_k)$ which do not evolve during the simulation.

In order to get the equations of motion of the markers, we compute the Euler-Lagrange equations associate to the Lagrangian Eq. (3.9) with respect to the components of the markers phase space positions z_k^α. This yields exactly the same equations as Eqs.(2.25)-(2.26) that can be inverted as in the continuous case, which yields the equations of motion for the markers

$$\dot{\boldsymbol{R}}_k = \frac{p_{\parallel,k}}{m} \frac{\boldsymbol{B}^*}{B_\parallel^*} - \frac{c}{eB_\parallel^*} \boldsymbol{b} \times [\mu_k \nabla B + e \nabla J_0 \Phi]$$

$$\dot{p_{\parallel k}} = -\frac{\boldsymbol{B}^*}{B_\parallel^*} \cdot [\mu_k \nabla B + e \nabla J_0 \Phi] \qquad (3.14)$$

given an initial condition $\boldsymbol{R}_k(0) = \boldsymbol{R}_k^0$, μ_k and $p_{\parallel,k}(0) = p_{\parallel,k}^0$. The value of the self-consistent gyroaveraged electrostatic potential at the marker position is needed to evolve the marker positions from the time t^n to the time t^{n+1}. We will see how to compute the electrostatic potential and the gyro average from the marker positions in the next two subsections.

This can be computed by solving the polarisation equation on a grid of physical space, after having constructed the charge density $\rho = \int dW e J_0 f$ on the grid.

If we consider the expression of the gyrocenter density:

$$n_{gc} = \int dW f_N \simeq \sum_{k=1}^{N} \frac{2\pi B_{\parallel k}^*(\boldsymbol{R})}{m^2} w_k \delta(\boldsymbol{R} - \boldsymbol{R}_k(t)) \tag{3.15}$$

3.7. *Polarisation equation using finite elements*

In the continuous case, the variational form of the polarisation equation is given by Eq. (2.34). Exactly the same equation, replacing Φ by its Finite Element approximation Φ_h, can be obtained by constraining the variations $\delta\Phi$ to be in the same Finite Element space as Φ_h. Another way to obtain the discrete equation, which might be more appealing for people not familiar with functional derivatives consists in taking the Euler-Lagrange equations with respect to the coefficients Φ_ν of

$$\Phi_h(\boldsymbol{x}, t) = \sum_{\mu=1}^{N_g} \Phi_\mu(t) \Lambda_\mu(\boldsymbol{x})$$

in the discrete Lagrangian Eq. (3.9). As $\dot{\Phi}_\mu$ does not appear in the Lagrangian this reduces to

$$0 = \frac{\partial L}{\partial \Phi_\nu} = \sum_{sp} \frac{1}{N_p} \sum_{k=1}^{N_p} w_k (-e J_0 \Lambda_\nu(\boldsymbol{R}_k)) + \sum_{sp} \int d\Omega \frac{mc^2}{B^2} \nabla_\perp \Lambda_\nu \cdot (\sum_{\mu=1}^{N_g} \Phi_\mu \nabla_\perp \Lambda_\mu) \tag{3.16}$$

which can be written, taking $\sum \Phi_m u$ out of the integral,

$$\sum_{\mu=1}^{N_g} \Phi_\mu \sum_{sp} \int d\Omega \frac{mc^2}{B^2} \nabla_\perp \Lambda_\nu \cdot \nabla_\perp \Lambda_\mu = \sum_{sp} \frac{1}{N_p} \sum_{k=1}^{N_p} w_k (e J_0 \Lambda_\nu(\boldsymbol{R}_k)). \tag{3.17}$$

The previous equation is actually a set of linear equations:

$$\sum_\mu A_{\mu\nu} \Phi_\mu = b_\nu \tag{3.18}$$

with:

$$A_{\mu\nu} = \sum_{sp} \int d\Omega \frac{mc^2}{B^2} \nabla_\perp \Lambda_\nu \cdot \nabla_\perp \Lambda_\mu \tag{3.19}$$

$$b_\nu = \sum_{sp} \frac{1}{N_p} \sum_{k=1}^{N_p} w_k (e J_0 \Lambda_\nu(\boldsymbol{R}_k)). \tag{3.20}$$

The main numerical advantage of this linearised model is that the matrix $A_{\mu\nu}$ does not change in time and it can be constructed (and factorized) at the beginning of the simulation.

When adiabatic electrons were used the discretised polarisation equation has the form:

$$\sum_\mu (A_{\mu\nu} + A_{\mu\nu}^{ZF}) \Phi_\mu = b_\nu \tag{3.21}$$

$$A_{\mu\nu} = \int \mathrm{d}V \left(\frac{en_0}{T} \Lambda_\nu(\boldsymbol{x}) \Lambda_\mu(\boldsymbol{x}) + \sum_{\mathrm{sp}} \frac{mc^2}{B^2} \nabla_\perp \Lambda_\nu(\boldsymbol{x}) \cdot \nabla_\perp \Lambda_\mu(\boldsymbol{x}) \right) \qquad (3.22)$$

$$A_{\mu\nu}^{ZF} = \int \mathrm{d}V \left(\frac{en_0}{T} \Lambda_\nu(\boldsymbol{x}) \bar{\Lambda}_\mu(s) \right) \qquad (3.23)$$

$$b_\nu = -\int \mathrm{d}V n_{0e} \Lambda_\nu + \sum_{\mathrm{sp}\neq\mathrm{e}} \sum_{\mathrm{sp}} \frac{1}{N_p} \sum_{k=1}^{N_p} w_k (e J_0 \Lambda_\nu(\boldsymbol{R}_k)). \qquad (3.24)$$

where $\bar{\Lambda}_\mu$ represents the flux surface average of the test function Λ_μ; $A_{\mu\nu}{}^{ZF}$ is usually called *zonal flow* matrix because in a tokamak gives nonzero contributions only for toroidally symmetric perturbations (zonal). Both $A_{\mu\nu}$ and $A_{\mu\nu}{}^{ZF}$ are sparse, symmetric and positive definite but $A_{\mu\nu}{}^{ZF}$ is computationally more intensive and contains many more non zero elements due to the integration along the poloidal and toroidal direction due to the flux surface average.

The computation of b_ν is called *charge assignment*: the charge density is obtained by assigning the weights to gyro-rings and projecting them on the B-spline basis. Let us now describe more precisely how this is done.

3.8. *Gyroaverage*

The calculation of ρ_N is complicated by the appearance of the gyroaverage operator J_0. In general, given a scalar field $\Phi(\boldsymbol{R})$, the gyroaverage operator $J_0\Phi$ is defined as:

$$J_0\Phi(\boldsymbol{R},\mu) \equiv \frac{1}{2\pi} \int_0^{2\pi} \Phi(\boldsymbol{R}+\boldsymbol{\rho}_i) \, \mathrm{d}\theta = \frac{1}{(2\pi)^3} \int \hat{\Phi}(\boldsymbol{k}) \, J_0(k_\perp \rho_i) \, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \, \mathrm{d}\boldsymbol{k}$$

where $\hat{\Phi}$ is the Fourier transformed potential and $\rho_i = (k_B T m c^2)(e^2 B^2)$. The previous equation shows that the operator J_0 has the form, in Fourier space, of a multiplication of Fourier coefficients by the zeroth Bessel functions $J_0(k_\perp \rho_i)$. This means that direct calculation of $J_0\Phi$ for each individual marker has to account for its interaction with all the waves in the system. Therefore, the exact calculation of J_0 is computationally prohibitive. Alternatively, the gyroaverage procedure can be approximated by an average over a number of points on the gyro-ring [Lee (1987)]. When four quadrature points are used, this procedure is rigorously equivalent to considering the Taylor expansion $J_0(k_\perp \rho_i) \simeq 1 - \frac{1}{4}(k_\perp \rho_i)^2$ and compute the transverse Laplacian using second order finite differences. A simple proof can be given when considering an uniform 2D grid, with grid spacing $h = \rho_i$ in both directions. Each point of the grid is defined by a pair of indexes (i,j):

$$J_0(\boldsymbol{R}) \simeq \Phi - \frac{1}{4}\rho_i^2 \nabla_\perp^2 \Phi(\boldsymbol{R})$$

$$\rightarrow J_0 \Phi_{i,j} \simeq \Phi_{ij} + \frac{\rho_i^2}{4h^2}(\Phi_{i+1,j} + \Phi_{i-1,j} + \Phi_{i,j+1} + \Phi_{i,j-1} - 4\Phi_{i,j})$$

$$= \frac{1}{4}(\Phi_{i+1,j} + \Phi_{i-1,j} + \Phi_{i,j+1} + \Phi_{i,j-1})$$

where the standard second order centers finite difference scheme $\nabla^2 \Phi_j = (-\Phi_{j+1} + 2\Phi_j - \Phi_j - 1)/h^2$ was used. In general, the number of points used in the quadrature is linearly increased with the gyroradius ρ_i to guarantee the same number of points per arc length

on the gyro-ring:

$$J_0\Phi(\boldsymbol{R},\mu) \equiv \frac{1}{2\pi}\int_0^{2\pi}\Phi(\boldsymbol{R}+\boldsymbol{\rho}_i)\,\mathrm{d}\theta \simeq \frac{1}{N_{gr}}\sum_{\beta=1}^{N_{gr}}\hat{\Phi}_\beta \qquad (3.25)$$

where $N_{gr}(\rho_i)$ is the number of points on the gyro-ring chosen for the average and $\hat{\Phi}_\beta$ represents the value of the electrostatic potential on the β point on the gyro-ring. In practice this gyro-average is applied to a finite element basis function Λ_ν to compute the right hand side b_ν of the discrete polarisation equation as defined in Eq. (3.20).

3.9. *Control variate*

In addition to importance sampling, considerable noise reduction can be achieved by applying the control variate technique discussed in subsection 3.5.2 in order to compute the right hand side b_ν of the discrete polarisation equation defined in Eq. (3.20).

The natural choice for the probability density used to define the control variate is $\tilde{f}$ a solution of the unperturbed equations of motion, typically a Maxwellian distribution function. We shall consider only the case $\alpha = 1$ which is used in practice most of the time. Then the initial importance weights for $\delta f = f - \tilde{f}$ of the markers that we denote by w_k^α become, using Eq. (3.12)

$$w_k^{\alpha,0} = \frac{f_0(\boldsymbol{z}_k^0) - \tilde{f}(t_n,\boldsymbol{z}_k^n)}{g_0(\boldsymbol{z}_k^0)} = w_k - \frac{\tilde{f}(0,\boldsymbol{z}_k^0)}{g_0(\boldsymbol{z}_k^0)}.$$

The right hand side of the discretised polarisation equation becomes:

$$b_\nu = \sum_{\mathrm{sp}}\int \mathrm{d}V\mathrm{d}W f_0\Lambda_\nu + \left(e\sum_{k=1}^{N}\left(w_k^\alpha \frac{1}{N_{gr,k}}\sum_{\beta=1}^{N_{gr,k}}\Lambda_\nu(\boldsymbol{x}_{k,\beta})\right)\right) \qquad (3.26)$$

The quasi-neutrality condition at $t = 0$ implies that the first term on the left hand side must vanish. When adiabatic electrons are used:

$$b_\nu = -\int \mathrm{d}V n_{0e}\Lambda_\nu + \sum_{\mathrm{sp}\neq\mathrm{e}}\int \mathrm{d}V\mathrm{d}W f_0\Lambda_\nu + \sum_{\mathrm{sp}\neq\mathrm{e}} e\sum_{k=1}^{N}\left(w_k^\alpha \frac{1}{N_{gr,k}}\sum_{\beta=1}^{N_{gr,k}}\Lambda_\nu(\boldsymbol{x}_{k,\beta})\right) \qquad (3.27)$$

Therefore, the fluid electron density n_{0e} must be carefully chosen to match:

$$-\int \mathrm{d}V n_{0e}\Lambda_\nu + \sum_{\mathrm{sp}\neq\mathrm{e}}\int \mathrm{d}V\mathrm{d}W f_0\Lambda_\nu = 0 \qquad (3.28)$$

The weights can now vary in time. They can be updated using Eq. (3.13)

$$w_k^{\alpha,n} = \frac{f(t_n,\boldsymbol{z}_k^n) - \tilde{f}(t_n,\boldsymbol{z}_k^n)}{g(t_n,\boldsymbol{z}_k^n)} = \frac{f_0(\boldsymbol{z}_k^0) - \tilde{f}(t_n,\boldsymbol{z}_k^n)}{g_0(\boldsymbol{z}_k^0)} = w_k - \frac{\tilde{f}(t_n,\boldsymbol{z}_k^n)}{g_0(\boldsymbol{z}_k^0)}. \qquad (3.29)$$

Another possibility to update the δf weights, especially useful for linear simulations, would be to derive an evolution equation for them using the GK Vlasov equation:

$$\frac{\mathrm{d}\delta f}{\mathrm{d}t} = -\dot{\boldsymbol{R}}\cdot\nabla\tilde{f} - \dot{p}_\parallel\frac{\partial\tilde{f}}{\partial p_\parallel}. \qquad (3.30)$$

4. Results

All the simulations presented in this work have been performed using the code NEMORB [Bottino *et al.* (2010)] a multi-species electromagnetic extension of the ORB5 [Jolliet

et al. (2007)]. The linearised field equations, the polarisation equation and the parallel Ampère's law, are discretised using B-splines. The code is based on straight-field-line coordinates and includes collision operators [Vernay *et al.* (2010)], shaped MHD equilibria and a hybrid model for the trapped electrons [Bottino *et al.* (2004), Bottino *et al.* (2006), Vernay *et al.* (2013)]. The NEMORB gyrokinetic model, in the electrostatic limit, is based on the gyrokinetic Vlasov-Poisson system of equations described in the previous section:

$$\frac{\partial f}{\partial t} + \dot{\boldsymbol{R}} \cdot \nabla f + \dot{p}_{\parallel} \frac{\partial f}{\partial p_{\parallel}} = C(f) + S, \tag{4.1}$$

$$\dot{\boldsymbol{R}} = \frac{p_{\parallel}}{m} \frac{\boldsymbol{B}^*}{B_{\parallel}^*} - \frac{c}{eB_{\parallel}^*} \boldsymbol{b} \times \left(\mu \nabla B + e \nabla J_0 \Phi \right), \tag{4.2}$$

$$\dot{p}_{\parallel} = -\frac{\boldsymbol{B}^*}{B_{\parallel}^*} \cdot \left(\mu \nabla B + e \nabla J_0 \Phi \right), \tag{4.3}$$

$$\sum_{\text{sp}} \left(\int \mathrm{d}W e J_0 f + \nabla \cdot \left(\frac{n_0 m c^2}{B^2} \nabla_{\perp} \Phi \right) \right) = 0. \tag{4.4}$$

where $C(f)$ is a collision operator and S is a heat source term. In the absence of heat sources, $S = 0$, transport processes tend to relax the temperature profile. In the simulations presented in this section, the heat source has the form of a Krook operator

$$S_H \propto -\gamma_H \tilde{\delta f}(\psi, v^2, t), \tag{4.5}$$

modified to enforce conservation of the number of particles. The symbol ˜ stands for the operator which projects a general function $A(\boldsymbol{R}, p_{\parallel}, \mu, t)$ to a reduced space $\tilde{A}(\psi, \epsilon, t)$, where ψ is the poloidal magnetic flux and ϵ is the kinetic energy. The complete definition of the projection operator and of the heat source S_H are given by Vernay *et al.* (2012). This source term tends to readjust the temperature profile toward the background profile. Note that small profile variations are still allowed during the simulation.

4.1. *Numerical analysis of the statistical noise*

As previously discussed, the Monte Carlo error for the approximation of the expected value of a random variable X is proportional to $\sigma(X)/\sqrt{N}$. In PIC simulations, this translates into the so-called "statistical noise" introduced when moments of the distribution function (for example, the charge assignment) are computed using a finite number of markers. The code ORB5 allows for a direct evaluation of the noise contribution to the charge assignment, ρ_{noise} [Bottino *et al.* (2007)]. This measure is based on the average amplitude of the contribution to the charge density, $|\rho_k|$, of the non-resonant (filtered) modes which are physically damped and whose amplitude arises merely from noise. Figure 3 shows an example of the time evolution of the volume averaged ρ_{noise} for a set of simulations with a different total number of marker. Through the comparison with the charge density of the "physical" modes, a signal to noise ratio can be constructed. This diagnostics provides a direct indicator of the quality of the numerical simulations during all the time evolution. The amplitude of ρ_{noise} can be analytically estimated in several ways (see, for example Nevins *et al.* (2005)). In the case of ORB5, a simple expression has been given by Bottino *et al.* (2007):

$$\rho_{\text{noise}}^2 \simeq \frac{N_G}{N_p} \langle w^2 \rangle G; \ \langle w^2 \rangle \equiv \frac{1}{N_p} \sum_{i=1}^{N_p} w_i^2 \tag{4.6}$$
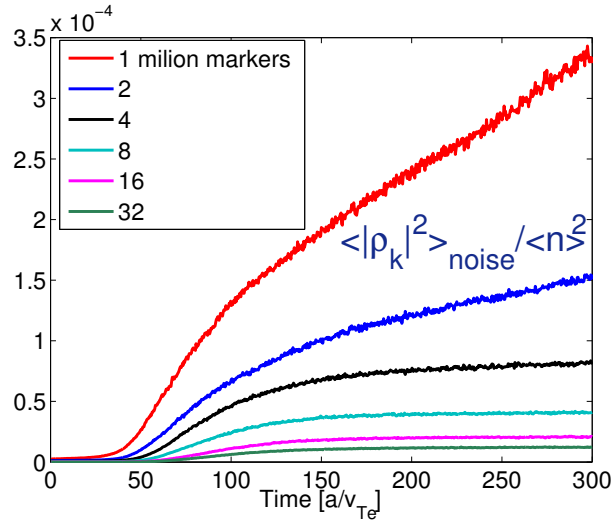
FIGURE 3. Time evolution of the spatial averaged ρ_{noise} for different number of markers for a circular cross-section plasma with $\rho^* = 1/80$. Numerical and physical parameters can be found in Bottino *et al.* (2007).

where N_p is the number of markers, N_G is the number of modes kept in the simulation and w_i is the weight of a single marker. The full derivation of the previous equation is reported in Appendix A. This equation indicates that the statistical noise can be reduced either by increasing the number of markers ($\sqrt{N_p}$ convergence) or by reducing the number of Fourier modes kept in the simulations (Fourier filtering of non-resonant modes). Several ORB5 simulations showed that the scaling of the noise in the number of particle per mode, (N_p/N_G), is in excellent agreement with the estimate, as it is illustrated in Fig.4, where $\rho_{\mathrm{noise}}^2/\langle w^2 \rangle$ is plotted as function of time for the simulations of Fig.4. Moreover, the scaling of the noise with the number of particles shows that the important parameter in PIC simulations is indeed the number of particle per Fourier mode and not the number of particles per grid cell (Fig. 5). The function G accounts for additional filtering coming through finite Larmor radius effects (FLR) and the grid projection algorithm. However, it is important to stress the role of the G function: although the number of numerical particles per mode is a universal scaling for the noise in PIC codes, the scaling factor, i.e. the G function, is strongly algorithm-dependent, and therefore code-dependent. For example, different projection algorithms in the charge assignment procedure can lead to very different level of noise: in ORB5 the level of noise is strongly reduced when moving from linear to cubic finite elements as illustrated in Bottino *et al.* (2007).

4.2. *Convergence in number of particles*

The NEMORB simulations presented in this section are based on parameters and profiles of the ITM Cyclone base case [Falchetto *et al.* (2008)]. This is a standard benchmark case based on a circular low-β equilibrium, deuterium plasma, $\rho* \simeq 1/185$ (mid radius), $T_e = T_D$ and flat initial R/L_T profiles between $0.2 < s < 0.8$, $s \propto \sqrt{\psi}$; $\rho* \equiv \rho_s/a$ is the ions sound Larmor radius normalized to the tokamak minor radius. The q profile is parabolic, matching the local value of the local Cyclone case ($q = 1.4$) at mid-radius. A detailed description of the physical parameters and profiles can be found in Falchetto *et al.* (2008) and Wersal *et al.* (2012). Figure 6 shows the results of a scan on the initial temperature gradient, with all the other parameters kept fixed. Those simulations are
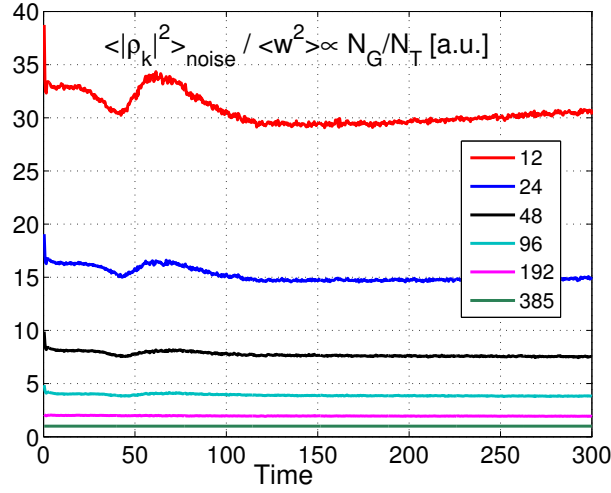
FIGURE 4. Scaling of the $\rho_{\mathrm{noise}}^2/\langle w^2 \rangle$ in N_g/N_p for a circular cross-section plasma with $\rho^* = 1/80$. The numbers in the caption indicate the number of markers per mode present in the different simulations. Numerical and physical parameters can be found in Bottino *et al.* (2007).
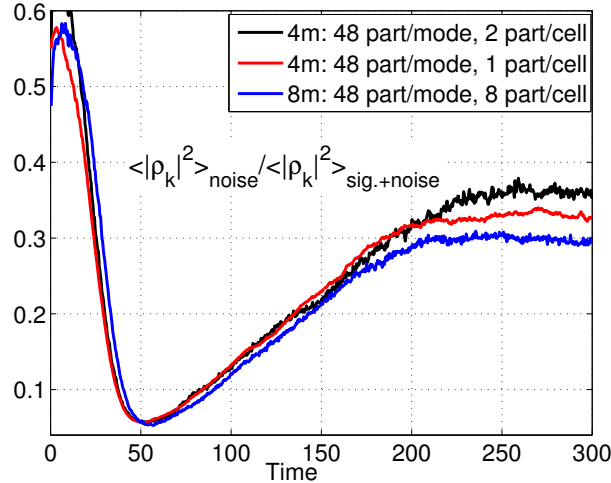

FIGURE 5. Noise to signal ratio for simulations with same number of markers per mode but different number of markers per grid points for a circular cross-section plasma with $\rho^* = 1/80$. Numerical and physical parameters can be found in Bottino *et al.* (2007).

equivalent in terms of physics and numerical parameters, to the standard ITM global case of Falchetto *et al.* (2008), except for the presence of the heat source preventing profile relaxation. The ion heat diffusivity in gyro-Bohm units ($\chi_{GB} \equiv \rho_s^2 c_s/a$, with $c_s^2 \equiv T_e/m_{\mathrm{D}}$ and $\rho_s^2 \equiv T_e m_{\mathrm{D}}/(eB_0)^2$) is plotted versus R/L_T, both the quantities correspond to radial averages between $0.4 < r/a < 0.6$. Here it is evident that the heat source prevents profile relaxation but still allows for local modifications of the gradient profile. The solid line represents the original Dimits fit for this curve, derived as a fit to the results of the LLNL gyrokinetic flux-tube PIC electrostatic turbulence code Dimits *et al.* (2000). The NEMORB results follow the Dimits curve, although with lower diffusion coefficients, as expected by global simulations. They also correctly reproduce the linear
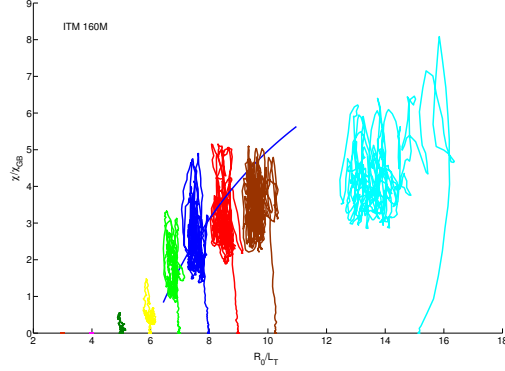
FIGURE 6. Radial averaged ion heat diffusivity in gyro-Bohm units versus R/L_T for the Cyclone base case with sources. During the saturation phase, χ/χ_{GB} lies close to the Dimits fit curve.*Published under license in Journal of Physics: Conference Series by IOP publishing Ltd.*
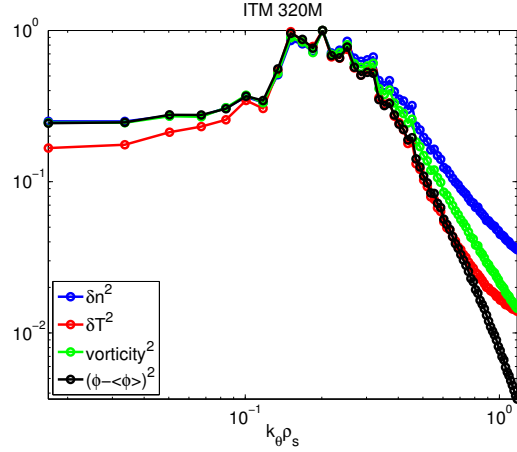

FIGURE 7. Different time averaged spectra for the initial $R/L_T \simeq 10.3$ simulations for the Cyclone base case with sources, calculated from 3D data. *Published under license in Journal of Physics: Conference Series by IOP publishing Ltd.*

$R/L_T|_{\mathrm{lin}} \simeq 4$ and nonlinear critical thresholds $R/L_T|_{\mathrm{nonlin}} \simeq 6$. This is particularly evident when considering the simulation with initial $R/L_T \simeq 5$ (dark green in Figure 6): after the initial linear phase, the heat flux drops rapidly to zero, showing that the mode was linearly unstable, but nonlinearly fully stabilized by zonal flow dynamics. The density, temperature, vorticity and potential spectra for a simulation with initial $R/L_T \simeq 10.3$ are presented in Figure 7. Time averaged spectra show evidence of the non-linear cascades to lower $k_\theta\rho_s$. The generalized vorticity, expressed as a frequency, is $\Omega \simeq eB/m_D(n_e - n_i)/n_0$; further details can be found in Scott *et al.* (2010) and in Wersal *et al.* (2012).

The simulations with initial $R/L_T \simeq 10.3$ has been chosen for a systematic study of the convergence in number of markers for real space time traces and for the corresponding spectra. The quantities considered are: density fluctuations, electrostatic potential, temperature fluctuations and vorticity. Spectra have been obtained by performing a time average during the saturation phase of the simulation, while real space data correspond to radial averages between $0.5 < r/a < 0.7$. The convergence results are illustrated in Fig.8.
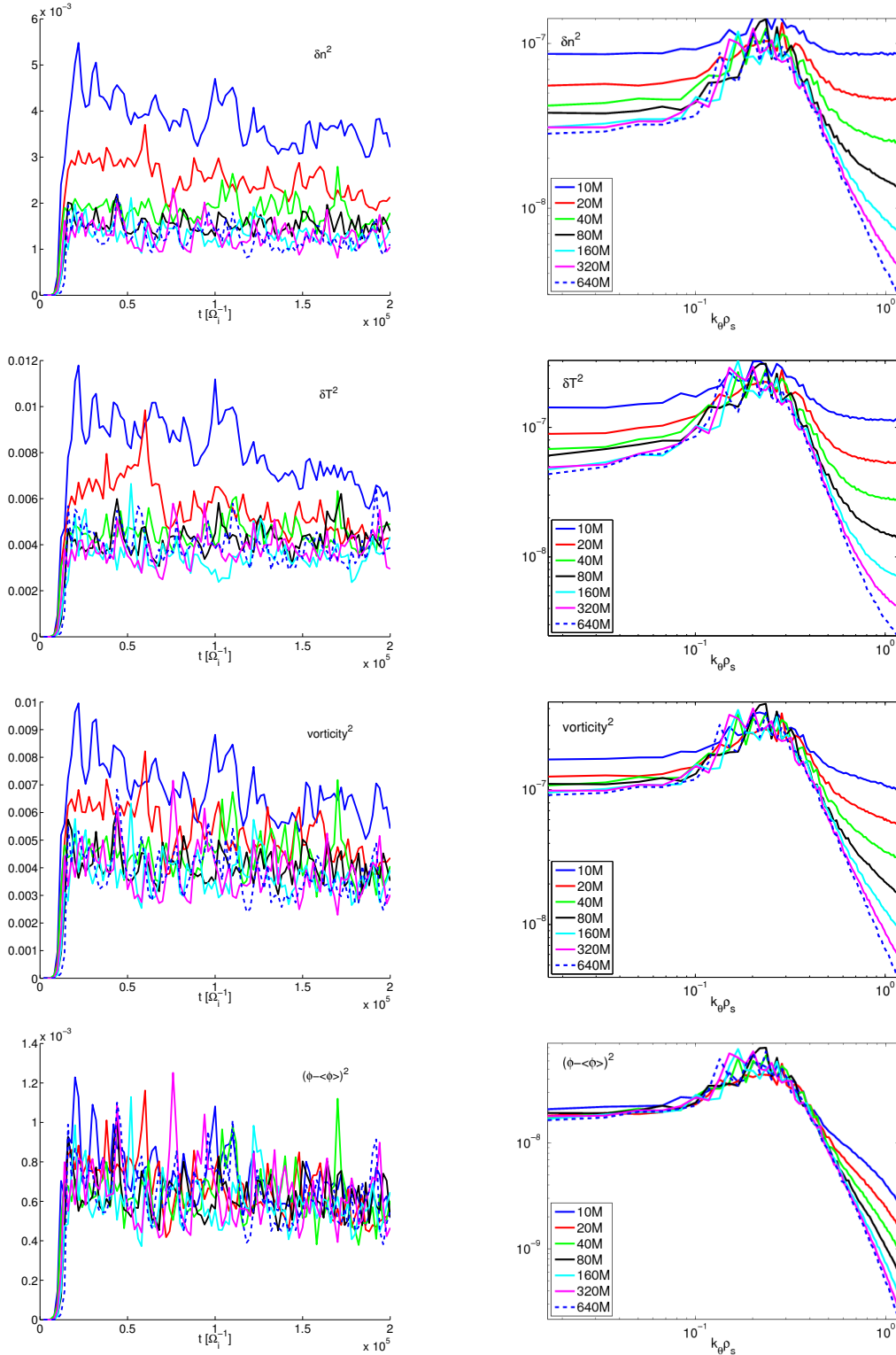
FIGURE 8. Time evolution of the radial averaged ($0.5 < r/a < 0.7$) signals (left) and time averaged spectra (right) for density, temperature, vorticity and non zonal electrostatic potential for the initial $R/L_T \simeq 10.3$ simulations. Non converged time traces correspond to flatter spectra. *Published under license in Journal of Physics: Conference Series by IOP publishing Ltd.*
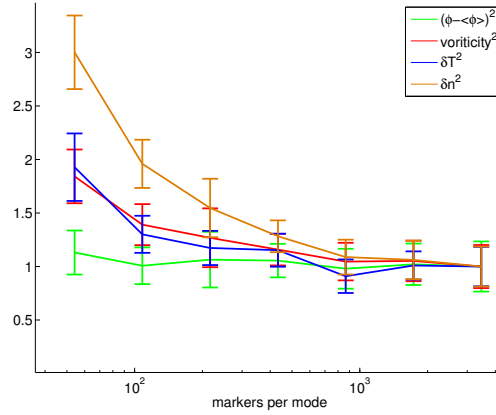
FIGURE 9. Radial and time averaged quantities of Fig.8, normalized to the 640M marker results, as a function of the number of markers per active mode.*Published under license in Journal of Physics: Conference Series by IOP publishing Ltd.*

For all the quantities considered, non converged time traces correspond to flat spectra at high $k_\theta \rho_s$. On the other hand, converged spectra always exhibit a clear power law down to the highest $k_\theta \rho_s$ values kept in the simulation. The most remarkable result is that not all the physically relevant quantities converge with the same rate. Figure 9 summarizes the results of Fig.8 by normalizing the time averaged quantities to the results of the simulations with the highest number of markers, plotted versus the number of markers per mode kept in the simulation. The electrostatic potential converges much faster than all the other quantities. The slowest converging field is the density fluctuation, δn. Remarkably, the ion heat flux diffusivity (and the heat fluxes) converge as fast as the potential. Although the heat flux, important for predicting turbulence induced transport in experiments, requires relatively few markers to converge (20M), comparisons with experimental measurements (for example, with reflectometry data) rely on accurate description of the density fluctuation spectra. According to our results, the latter requires at least a factor of 10 more markers then what convergence studies based on potentials or fluxes suggest.

5. Acknowledgments

Appendix A. Derivation of a simple analytical estimate for the noise

The amplitude of the statistical noise contribution to the charge density can be analytically derived. Before discussing the derivation in the case of the ORB5 code a simpler system is considered by neglecting FLR effects, equilibrium profile variation and zonal

flow effects and assuming purely adiabatic electrons. In this case, the quasi-neutrality equations becomes:

$$\frac{e\phi}{T_e} = \frac{n_0 V}{N_p} \sum_{i=1}^{N_p} w_i \delta(\mathbf{x} - \mathbf{x}_i), \tag{A 1}$$

where the ORB5 normalization for the weights (see Jolliet *et al.* (2007)) is assumed: n_0 is the averaged particle density V is the volume and N_p is the total number of markers. The continuous Fourier transform of the potential is:

$$\phi(\mathbf{p}) = \int \mathrm{d}^3 \mathbf{x}\, \phi(\mathbf{x}) \exp[-2\pi \mathrm{i}\mathbf{p} \cdot \mathbf{x}]. \tag{A 2}$$

Applying the Fourier transform to the Poisson equation directly yields:

$$\frac{e\phi_{\mathbf{p}}}{T_e} = \frac{V}{N_p} \sum_{i=1}^{N_p} w_i \exp[-2\pi \mathrm{i}\mathbf{p} \cdot \mathbf{x}_i]. \tag{A 3}$$

The absolute value squared then is:

$$\left| \frac{e\phi_{\mathbf{p}}}{T_e} \right|^2 = \frac{V^2}{N_p^2} \sum_{i=1}^{N_p} w_i^2 + \sum_{i=1}^{N_p} \sum_{j=1, j\neq i}^{N_p} w_i w_j \exp[\mathrm{i}2\pi \mathbf{p} \cdot (\mathbf{x}_j - \mathbf{x}_i)]. \tag{A 4}$$

The second term of the right-hand-side represents correlations between marker positions and weights and describes the contribution of waves and turbulence to the density. Following Nevins *et al.* (2005), we assume that the relevant physics is contained in that term while the noise is in the contained in the first term, which describes the random positioning of the markers. Therefore, the noise can be represented by

$$\left| \frac{e\phi_{\mathbf{p}}}{T_e} \right|^2 = \frac{V^2}{N_p^2} \sum_{i=1}^{N_p} w_i^2. \tag{A 5}$$

In order to compare with the simulation results, it is convenient to use discrete Fourier transforms (DFT). The relation between the continuous and discrete FT can easily be derived

$$F(\mathbf{p}) = \int \mathrm{d}^3 \mathbf{x}\, f(\mathbf{x}) \exp[-2\pi \mathrm{i}\mathbf{p} \cdot \mathbf{x}] \approx \sum_{i=1}^{N_T} f_i \exp[-2\pi \mathrm{i}\mathbf{p}_i \cdot \mathbf{x}] \Delta_x \Delta_y \Delta_z \tag{A 6}$$

where $\Delta_x \equiv L_x/N_x$, $\Delta_y \equiv L_y/N_y$ and $\Delta_z \equiv L_z/N_z$ describe the grid spacing; L_x, L_y, L_z are the size of the computational domain in the three spatial directions and $N_T \equiv N_x N_y N_z$ is the total number of grid points in configuration space.

The the sum $\sum_{i=1}^{N_T}$ symbol represents a sum over the different modes in x-, y- and z-direction. The discrete Fourier transform is defined as

$$F_q = \sum_{i=0}^{N-1} f_i \exp[2\pi \mathrm{i}q_i/N] \tag{A 7}$$

with q_i is an integer number satisfying

$$-\frac{N}{2} \leqslant q_i < \frac{N}{2}. \tag{A 8}$$

Using

$$N_x p_x \Delta_x = q_x \tag{A 9}$$

and similar relations for the other directions, as well as

$$\Delta_x \Delta_y \Delta_z = \frac{V}{N_T},$$ (A 10)

one arrives at

$$F_q \approx \frac{V}{N_T} F_{\mathbf{p}}\left(p_\alpha = \frac{q_\alpha}{N_\alpha \Delta_\alpha}\right).$$ (A 11)

In ORB5 the charge density is filtered in Fourier space in order to keep only $N_G < N_T$ modes, thus removing physically irrelevant modes. Typically, at every radial location only few poloidal modes are needed, due to the fact that the turbulence aligns with the magnetic field. If we identify the poloidal direction with y, only $N_G \equiv N_x N_y' N_z$, with $N_y' < N_y$, modes are kept in the charge density. More details about the Fourier filtering procedure can be found in Jolliet *et al.* (2007). Assuming that the statistical noise has a white noise character and it contributes to all the Fourier modes, the Fourier filtering procedure reduces the numerical noise contribution to the charge density.

$$F_q \approx \frac{V}{N_G} F_{\mathbf{p}}\left(p_\alpha = \frac{q_\alpha}{N_\alpha \Delta_\alpha}\right).$$ (A 12)

Using the previous relations, one can rewrite the noise formula as

$$\left|\frac{e\phi_{\mathbf{q}}}{T_e}\right|^2 = \frac{N_G^2}{N_p^2} \sum_{i=1}^{N_p} w_i^2.$$ (A 13)

We can also use the energy relation of the FFT

$$\sum_{i=1}^{N_G} f_i^2 = \frac{1}{N_G} \sum_{q=0}^{N_G-1} F_q^2.$$ (A 14)

To derive the relation for the fluctuations of ϕ in real space

$$\left\langle \left|\frac{e\phi}{T_e}\right|^2 \right\rangle = \frac{1}{N_G} \sum_i \left|\frac{e\phi}{T_e}\right|^2 = \frac{1}{N_G^2} \sum_q \left|\frac{e\phi_q}{T_e}\right|^2 = \frac{N_G}{N_p}\langle w^2 \rangle$$ (A 15)

where

$$\langle w^2 \rangle = \frac{1}{N_p} \sum_{i=1}^{N_p} w_i^2$$ (A 16)

is the averaged square weight.
This simple noise estimate can be extended to include FLR effects. In this case, the starting point is the right-hand-side of the gyrokinetic polarisation equation of ORB5:

$$b_\nu = \sum_{sp} \frac{n_0 V}{N_p} \sum_{k=1}^{N_p} w_k (eJ_0\Lambda_\nu(\boldsymbol{R}_k))$$ (A 17)

$$= \sum_{sp} \frac{n_0 V}{N_p} \sum_{k=1}^{N_p} w_k \int d\boldsymbol{x} \Lambda_\nu(\boldsymbol{x}) \int d\boldsymbol{p} \exp(2\pi i\boldsymbol{p}\cdot\boldsymbol{x})\exp(-2\pi i\boldsymbol{p}\cdot\boldsymbol{R}_k)J_0(\boldsymbol{p}\cdot\boldsymbol{\rho}_{ik}).$$

Multiplying the previous expression of b_ν by the complex conjugate and neglecting the correlation (physical) term, the noise contribution to the right-hand-side of the polarisa-

tion equation becomes:

$$|b_\nu|^2 = \sum_{\text{sp}} \frac{n_0 V}{N_p} \sum_{k=1}^{N_p} w_k^2 \int \mathrm{d}\boldsymbol{p}\mathrm{d}\boldsymbol{p}' \exp\left(-2\pi i(\boldsymbol{p}-\boldsymbol{p}')\cdot\boldsymbol{R}_k\right)$$

$$\exp\left(-\frac{k_\perp^2(\boldsymbol{p})\rho_{ik}^2 + k_\perp^{'2}(\boldsymbol{p}')\rho_{ik}^2}{2}\right) \tilde{\Lambda}_\nu(-\boldsymbol{p})\tilde{\Lambda}_\nu^\dagger(-\boldsymbol{p}') \tag{A 18}$$

where the following relation has been used:

$$\frac{1}{n_0}\int \mathrm{d}W f_0 J_0(k_\perp\rho_{ik}) = \exp\left(-\frac{k_\perp^2\rho_{ik}^2}{2}\right) \tag{A 19}$$

with f_0 Maxwellian distribution, and

$$\tilde{\Lambda}_\nu(-\boldsymbol{p}) \equiv \int \mathrm{d}\boldsymbol{x}\exp(2\pi i\boldsymbol{p}\cdot\boldsymbol{x}). \tag{A 20}$$

Due to the randomization of the marker positions, the previous expression can be further approximated assuming

$$\exp\left(-2\pi i(\boldsymbol{p}-\boldsymbol{p}')\cdot\boldsymbol{R}_k\right) \simeq \frac{1}{V}\int \mathrm{d}\boldsymbol{R}\exp\left(-2\pi i(\boldsymbol{p}-\boldsymbol{p}')\cdot\boldsymbol{R}_k\right) = \frac{1}{V}\delta(\boldsymbol{p}-\boldsymbol{p}'), \tag{A 21}$$

giving

$$|b_\nu|^2 = \sum_{\text{sp}} \frac{n_0 V}{N_p} \sum_{k=1}^{N_p} w_k^2 G \tag{A 22}$$

$$G \equiv \frac{1}{V}\int \mathrm{d}\boldsymbol{p}|\tilde{\Lambda}_\nu(-\boldsymbol{p})|^2 \exp\left(-\frac{k_\perp^2\rho_i^2}{2}\right). \tag{A 23}$$

This expression is similar to the simple estimate of Eq. (A 16) but contains an additional term G which includes filtering due to FLR effects and due to the spline representation of the potential. The full derivation can be find in Jolliet (2005).

REFERENCES

ALLFREY, S. J. & HATZKY, R. 2003 A revised δf algorithm for nonlinear PIC simulation. *Computer physics communications* **154** (2), 98–104.

AYDEMIR, A. Y. 1994 A unified Monte Carlo interpretation of particle simulations and applications to non-neutral plasmas. *Physics of Plasmas* **1** (4), 822–831.

BIRDSALL, C. K. & LANGDON, A. B. 2004 *Plasma physics via computer simulation*. CRC Press.

BOTTINO, A., PEETERS, A.G., HATZKY, R., JOLLIET, S., MCMILLAN, B.F., TRAN, T.M. & VILLARD, L. 2007 Nonlinear low noise particle-in-cell simulations of electron temperature gradient driven turbulence. *Physics of Plasmas* **14** (1).

BOTTINO, A., PEETERS, A.G., SAUTER, O., VACLAVIK, J. & VILLARD, L. 2004 Simulations of global electrostatic microinstabilities in ASDEX Upgrade discharges. *Physics of Plasmas* **11** (1), 198–206.

BOTTINO, A., SAUTER, O., CAMENEN, Y. & FABLE, E. 2006 Linear stability analysis of microinstabilities in electron internal transport barrier non-inductive discharges. *Plasma Physics and Controlled Fusion* **48** (2), 215–233.

BOTTINO, A., SCOTT, B., BRUNNER, S., MCMILLAN, B.F., TRAN, T.M., VERNAY, T., VILLARD, L., JOLLIET, S., HATZKY, R. & PEETERS, A.G. 2010 Global nonlinear electromagnetic simulations of tokamak turbulence. *IEEE Transactions on Plasma Science* **38** (9 PART 1), 2129–2135.

BRIZARD, A.J. 2000 Variational principle for nonlinear gyrokinetic Vlasov-Maxwell equations. *Physics of Plasmas* **7** (12), 4816–4822.

BRIZARD, A. & HAHM, T.S. 2007 Foundations of nonlinear gyrokinetic theory. *Rev. Mod. Phys.* **79**, 421.

DIMITS, A.M., BATEMAN, G., BEER, M.A., COHEN, B.I., DORLAND, W., HAMMETT, G.W., KIM, C., KINSEY, J.E., KOTSCHENREUTHER, M., KRITZ, A.H., LAO, L.L., MANDREKAS, J., NEVINS, W.M., PARKER, S.E., REDD, A.J., SHUMAKER, D.E., SYDORA, R. & WEILAND, J. 2000 Comparisons and physics basis of tokamak transport models and turbulence simulations. *Physics of Plasmas* **7** (3), 969–983.

DORLAND, W. 1983 Gyrofluid models of plasma turbulence. Ph.D. Thesis, Princeton University.

DUBIN, D.H.E., KROMMES, J.A., OBERMAN, C. & LEE, W.W. 1983 Nonlinear gyrokinetic equations. *Physics of Fluids* **26** (12), 3524–3535.

DUNN, W.L. & SHULTIS, J.K. 2012 *Exploring Monte Carlo Methods*. Academic Press.

EVSTATIEV, E.G. & SHADWICK, B.A. 2013 Variational formulation of particle algorithms for kinetic plasma simulations. *Journal of Computational Physics* **245**, 376–398.

FALCHETTO, G.L., SCOTT, B.D., ANGELINO, P., BOTTINO, A., DANNERT, T., GRANDGIRARD, V., JANHUNEN, S., JENKO, F., JOLLIET, S., KENDL, A., McMILLAN, B.F., NAULIN, V., NIELSEN, A.H., OTTAVIANI, M., PEETERS, A.G., PUESCHEL, M.J., REISER, D., RIBEIRO, T.T. & ROMANELLI, M. 2008 The European turbulence code benchmarking effort: Turbulence driven by thermal gradients in magnetically confined plasmas. *Plasma Physics and Controlled Fusion* **50** (12).

FRIEMAN, E.A. & CHEN, L. 1982 Nonlinear gyrokinetic equations for low frequency electromagnetic waves in general equilibria. *Physics of Fluids* **25**, 502.

HAHM, T.S. 1988 Nonlinear gyrokinetic equations for tokamak microturbulence. *Physics of Fluids* **31**, 2670.

HINTON, F.L. & HAZELTINE, R.D. 1976 Theory of plasma transport in toroidal confinement systems. *Reviews of Modern Physics* **48** (2), 239–308.

HOCKNEY, R.W. & EASTWOOD, J.W. 1988 *Computer simulation using particles*. CRC Press.

HU, G. & KROMMES, J.A. 1994 Generalized weighting scheme for δf particle-simulation method. *Physics of Plasmas* **1** (4), 863–874.

JOLLIET, S. 2005 Gyrokinetic PIC simulations of ITG and CTEM turbulence in tokamaks. Ph.D. Thesis No.**4326**, Ecole Polytechnique Fédérale de Lausanne.

JOLLIET, S., BOTTINO, A., ANGELINO, P., HATZKY, R., TRAN, T.M., McMILLAN, B.F., SAUTER, O., APPERT, K., IDOMURA, Y. & VILLARD, L. 2007 A global collisionless pic code in magnetic coordinates. *Computer Physics Communications* **177** (5), 409–425.

KLEIBER, R., HATZKY, R., KÖNIES, A., KAUFFMANN, K. & HELANDER, P. 2011 An improved control-variate scheme for particle-in-cell simulations with collisions. *Comput. Phy. Comm.* **182**, 1005–1012.

KOTSCHENREUTHER, M. 1998 Numerical simulation. *Bull. Am. Phys. Soc.* **34**.

KROMMES, J.A. 2007 Nonequilibrium gyrokinetic fluctuation theory and sampling noise in gyrokinetic particle-in-cell simulations. *Physics of Plasmas* **14** (9).

KROMMES, J.A. 2013 The physics of the second-order gyrokinetic magnetohydrodynamic hamiltonian: μ conservation, galilean invariance, and ponderomotive potential. *Physics of Plasmas* **20** (12).

LEE, W.W. 1983 Gyrokinetic approach in particle simulations. *Physics of Fluids* **26**, 556.

LEE, W.W. 1987 Gyrokinetic particle simulation model. *Journal of Computational Physics* **73**, 243.

LEWIS, H.R. 1970 Energy-conserving numerical approximations for Vlasov plasmas. *Journal of Computational Physics* **6** (1), 136–141.

LITTLEJOHN, R.G. 1981 Hamiltonian formulation of guiding center motion. *Physics of Fluids* **24**, 1730.

LITTLEJOHN, R.G. 1983 Variational principles of guiding centre motion. *J. Plasma Phys.* **29**, 111.

LIU, J.S. 2008 *Monte Carlo strategies in scientific computing*. Springer.

LOW, F.E. 1958 A Lagrangian formulation of the Boltzmann-Vlasov equation for plasmas. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences* **248** (1253), 282–287.

McMILLAN, B.F., HILL, P., JOLLIET, S., VERNAY, T., VILLARD, L. & BOTTINO, A. 2012

Gyrokinetic transport relations for gyroscale turbulence. *Journal of Physics: Conference Series* **401** (1).

MIYATO, N. & SCOTT, B. 2013 On the gyrokinetic model in long wavelength regime. *Plasma Physics and Controlled Fusion* **55**, 074011.

MIYATO, N., SCOTT, B. & STRINTZI, D. 2009 A modification of the guiding-centre fundamental 1-form with strong E×B flow. *J. Phys. Soc. Jpn.* **78**, 104501.

NEVINS, W.M., PARKER, S.E., CHEN, Y., CANDY, J., DIMITS, A., DORLAND, W., HAMMETT, G.W. & JENKO, F. 2005 Discrete particle noise in particle-in-cell simulations of plasma microturbulence. *Physics of Plasmas* **12**, Art. No. 122305.

SCOTT, B. 2010 Derivation via free energy conservation constraints of gyrofluid equations with finite-gyroradius electromagnetic nonlinearities. *Physics of Plasmas* **17** (10).

SCOTT, B., KENDL, A. & RIBEIRO, T. 2010 Nonlinear dynamics in the tokamak edge. *Contrib. Plasma Phys.* **50**, 228.

SCOTT, B. & SMIRNOV, J. 2010 Energetic consistency and momentum conservation in the gyrokinetic description of tokamak plasmas. *Physics of Plasmas* **17**, 112302.

SUGAMA, H. 2000 Gyrokinetic field theory. *Physics of Plasmas* **7**, 466.

TRAN, T.M., APPERT, K., FIVAZ, M., JOST, G., VACLAVIK, J. & VILLARD, L. 1999 Global gyrokinetic simulations of ion-temperature-gradient driven instabilities using particles. In *Proceedings of the Joint Varenna–Lausanne International Workshop 1998*, p. 45.

VERNAY, T., BRUNNER, S., VILLARD, L., MCMILLAN, B.F., JOLLIET, S., BOTTINO, A., GÖRLER, T. & JENKO, F. 2013 Global gyrokinetic simulations of tem microturbulence. *Plasma Physics and Controlled Fusion* **55** (7).

VERNAY, T., BRUNNER, S., VILLARD, L., MCMILLAN, B.F., JOLLIET, S., TRAN, T.M. & BOTTINO, A. 2012 Synergy between ion temperature gradient turbulence and neoclassical processes in global gyrokinetic particle-in-cell simulations. *Physics of Plasmas* **19** (4).

VERNAY, T., BRUNNER, S., VILLARD, L., MCMILLAN, B.F., JOLLIET, S., TRAN, T.M., BOTTINO, A. & GRAVES, J.P. 2010 Neoclassical equilibria as starting point for global gyrokinetic microturbulence simulations. *Physics of Plasmas* **17** (12).

WERSAL, C., BOTTINO, A., ANGELINO, P. & SCOTT, B.D. 2012 Fluid moments and spectral diagnostics in global particle-in-cell simulations. *Journal of Physics: Conference Series* **401** (1).