Princeton Plasma Physics Laboratory

PPPL- 4941

PPPL- 4941

Optimization by Marker for delta *f* Particle Simulations

Wenjun Deng and Guo-Yong Fu

October 2013

PPPL PRINCETON PLASMA PHYSICS LABORATORY

Prepared for the U.S. Department of Energy under Contract DE-AC02-09CH11466.

Princeton Plasma Physics Laboratory
Report Disclaimers

Full Legal Disclaimer

Trademark Disclaimer

PPPL Report Availability

Princeton Plasma Physics Laboratory:

http://www.pppl.gov/techreports.cfm

Office of Scientific and Technical Information (OSTI):

http://www.osti.gov/bridge

Related Links:

U.S. Department of Energy

Office of Scientific and Technical Information

Fusion Links

Optimization by marker removal for δf particle simulations

Wenjun Deng*, Guo-Yong Fu

Princeton Plasma Physics Laboratory, Princeton, New Jersey 08543, USA

Abstract

A marker removal optimization technique is developed for δf particle simulations to optimize the marker distribution so as to save markers and computing time. The technique can be directly applied to single-mode linear simulations. For multi-mode or nonlinear simulations, the technique can still be directly applied if there is one most unstable mode that dominates the simulation and δf does not change too much in nonlinear stage, otherwise special care is needed, which is discussed in detail in this paper. The technique effectiveness, e.g., marker saving factor, depends on how localized δf is. In this paper, the technique is first demonstrated in a simple 2D bump-on-tail simulation, and then generalized to 5D gyrokinetic simulations. The technique saves markers by factors of 4 and 19 in our nonlinear 2D bump-on-tail and 5D toroidal Alfvén eigenmode (TAE) simulations, respectively. The technique can be used for phase space of arbitrary dimension, as long as the equilibrium motion constants can be found. The technique is not limited to particle-in-cell (PIC) simulations but could be applied to other approaches of

*Corresponding author
Email addresses: wdeng@wdeng.info (Wenjun Deng), fu@pppl.gov (Guo-Yong Fu)
URL: http://wdeng.info (Wenjun Deng)

marker particle simulations such as particle-in-wavelet (PIW) and treecode simulations.

Keywords: particle simulation, particle-in-cell, marker particle, optimization, arbitrary dimension

1. Introduction

Marker particle simulation (usually shortened as particle simulation) is a popular tool for plasma physics studies. It uses a small set (compared to the set of actual physical particles in plasma) of computational particles for Monte Carlo sampling of the distribution function of the physical particles [1]. These computational particles are called marker particles or shortened as markers. Sometimes they are also called super-particles or macroparticles. Some simulations use markers to sample the whole physical distribution function f, which are called full-f or total-f simulations. Some others decompose the distribution f into an equilibrium part f_0 that is usually time-independent and known analytically, and a perturbed part δf, and use markers to sample only the perturbed distribution δf. These are called δf simulations.

There are multiple simulation approaches that make use of marker particles. The oldest one is the particle-in-cell (PIC) approach that dates back to the 1950s [2]. The PIC approach is still very popular today and is used in a lot of currently active plasma simulation codes, such as GTC [3], GEM [4], XGC [5], AWECS [6], GYGLES [7], HMGC [8], MEGA [9], M3D-K [10], etc. There are also other approaches that are developed to improve the PIC approach, such as the particle-in-wavelet (PIW) approach [11] and the grid-free

treecode approach [12].

For better resolution in phase space, i.e., lower particle noise, in particle simulations, it is wise to load markers such that the marker distribution, which is denoted as g in this paper, is proportional to the physical particle distribution, which is called "importance sampling" in Ref. [1]. This can usually be done straightforwardly in full-f simulations. For example, if a simulation starts with a Maxwellian distribution $f|_{t=0} = f_{\mathrm{Maxw}}$, then markers are loaded with a Maxwellian distribution $g|_{t=0} = C \cdot f_{\mathrm{Maxw}}$, such that $f/g = 1/C$ is a constant, so as to achieve the "importance sampling." Note that the evolutions of both f and g are governed by the Vlasov equation (or the gyrokinetic equation in gyrokinetic simulations), i.e., $\mathrm{d}f/\mathrm{d}t = 0$ and $\mathrm{d}g/\mathrm{d}t = 0$, so f/g stays constant during the whole simulation when g is initially loaded with the "importance sampling," meaning that the "importance sampling" is effective during the whole simulation. Things become different for δf simulations, where $\mathrm{d}\delta f/\mathrm{d}t \neq 0$. Loading markers such that $\delta f/g$ is a constant initially does not guarantee $\delta f/g$ to stay constant during the simulation. To guarantee that δf is well-resolved during the whole simulation, typical δf simulations load markers in the phase space either uniformly or such that g is proportional to the total distribution f. Either loading method usually causes markers to spread all over the phase space.

Energetic particle physics is crucial to ITER and other fusion devices. Instabilities excited by energetic particles are being broadly studied in PIC simulations, such as toroidal Alfvén eigenmode (TAE) [13, 14, 15, 16, 17, 18, 19, 20, 21], reversed shear Alfvén eigenmode (RSAE) [22, 23, 24], β-induced Alfvén eigenmode (BAE) [25, 26, 27, 28, 29, 30], etc. These modes usually

exhibit localized δf structures, i.e., δf is nearly zero in a large portion of the phase space. The localized δf structures are usually but not always near the wave-particle resonance regions. Maintaining a large number of markers in the $\delta f \approx 0$ regions is a waste of computing time. In this work, we have developed an optimization technique by removing markers in the $\delta f \approx 0$ regions to achieve the "importance sampling" for δf simulations so as to save computing time.

This paper firstly demonstrates the marker removal optimization technique in a simple 2D PIC simulation in Sec. 2. Then the technique is generalized to 5D gyrokinetic PIC simulations and is tested in a nonlinear TAE simulation in Sec. 3. Further generalization to arbitrary phase space dimension and other possible extensions of the technique are discussed in Sec. 4. Finally, a summary is given in Sec. 5.

2. Optimization in 2D electrostatic bump-on-tail simulation

Before getting into a complicated high-dimension simulation, we start with a simple 2D electrostatic bump-on-tail simulation to ease our discussion of candidate optimization techniques and description of our marker removal technique.

2.1. Simulation model

The simulation model mainly consists of an electrostatic δf Vlasov equation and a Poisson equation:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\delta f_\alpha(x,v,t) &= -\frac{\mathrm{d}}{\mathrm{d}t}f_{0\alpha}(v) \\ &= -\frac{Z_\alpha}{m_\alpha}E(x,t)\partial_v f_{0\alpha}(v)\ , \qquad (1) \\ \partial_x E(x,t) &= \sum_\alpha Z_\alpha \int_{-\infty}^{\infty}\delta f_\alpha(x,v,t)\mathrm{d}v\ , \qquad (2) \end{aligned}$$

where the subscript α indicates the particle species; Z is the particle charge; m is the particle mass; δf is the perturbed distribution function and is a function of the position x, the velocity v and the time t; f_0 is the equilibrium distribution function and is assumed to be independent on x and t; and E is the electrostatic field. In the PIC simulation we use a two-weight scheme similar to that in Ref. [31]. The marker evolution equations therefore write:

$$\begin{aligned} \frac{\mathrm{d}x_{j_\alpha}}{\mathrm{d}t} &= v_{j_\alpha}\ , \qquad (3) \\ \frac{\mathrm{d}v_{j_\alpha}}{\mathrm{d}t} &= \frac{Z_\alpha}{m_\alpha}E(x_{j_\alpha})\ , \qquad (4) \\ \frac{\mathrm{d}p_{j_\alpha}}{\mathrm{d}t} &= 0\ , \qquad (5) \\ \frac{\mathrm{d}w_{j_\alpha}}{\mathrm{d}t} &= -(p_{j_\alpha}-w_{j_\alpha})\frac{Z_\alpha}{m_\alpha}E(x_{j_\alpha})\frac{\partial_v f_{0\alpha}}{f_{0\alpha}}\ , \qquad (6) \end{aligned}$$

where the subscript j_α indicates the marker index;

$$\begin{aligned} p &= f/g\ , \qquad (7) \\ w &= \delta f/g\ , \qquad (8) \end{aligned}$$

are the total weight and the perturbed weight, respectively; $f = f_0 + \delta f$ is the total distribution function and g is the marker distribution function.

All quantities in the above equations are already normalized as follows: time is normalized by $1/\omega_{\rm pe}$, where $\omega_{\rm pe}$ is electron oscillation frequency; length is normalized by Debye length $\lambda_{\rm De}$; mass is normalized by electron mass $m_{\rm e}$; charge is normalized by proton charge e; temperature is normalized by electron temperature $T_{\rm e}$; density is normalized by electron equilibrium density $n_{0\rm e}$; electric field is normalized by $e/(T_{\rm e}\lambda_{\rm De})$.

2.2. Electron bump-on-tail instability

Here we keep using the normalization in Sec. 2.1 and assume that ions do not move and stay as a fixed background and electrons have a bump-on-tail distribution:

$$f_{0\rm e}(v) = \frac{1}{\sqrt{2\pi}} \left\{ \underbrace{(1-n_{\rm b})\exp\left(-\frac{v^2}{2}\right)}_{\text{background Maxwellian}} + \underbrace{\frac{n_{\rm b}}{v_{\rm th,b}}\exp\left[-\frac{(v-v_{0\rm b})^2}{2v_{\rm th,b}^2}\right]}_{\text{beam shifted Maxwellian}} \right\} , \quad (9)$$

where $n_{\rm b}$ represents the electron density of the beam, $v_{0\rm b}$ represents the beam velocity, and $v_{\rm th,b}$ represents the thermal velocity of the beam. The linear dispersion relation for this scenario is:

$$0 = 1 + \frac{1}{k^2}\left\{[1+\zeta_{\rm e}\mathcal{Z}(\zeta_{\rm e})] + \frac{n_{\rm b}}{T_{\rm b}}[1+\zeta_{\rm b}\mathcal{Z}(\zeta_{\rm b})]\right\} , \quad (10)$$

$$\zeta_{\rm e} = \frac{\omega}{\sqrt{2}k} , \quad (11)$$

$$\zeta_{\rm b} = \frac{\frac{\omega}{k} - v_{0\rm b}}{\sqrt{2}v_{\rm th,b}} , \quad (12)$$

where $\mathcal{Z}()$ is the plasma dispersion function. This dispersion equation can be solved numerically using Muller's method [32]. Specifically, in this work we use these parameters: $n_{\rm b} = 0.1$, $v_{0\rm b} = 5$, $v_{\rm th,b}^2 = 0.1$, $k = 0.401$. The

Muller's method gives the real frequency and growth rate as: $\omega_{\rm r} = 1.4674$, $\gamma = 0.029120$. The expression for δf is:

$$\begin{aligned} & \delta f_{\rm e}(x,v,t) \\ = & -\left\{(1-n_{\rm b})\frac{v}{\sqrt{2\pi}}\exp\left(-\frac{v^2}{2}\right) + n_{\rm b}\frac{v-v_{0\rm b}}{\sqrt{2\pi}v_{\rm th,b}^3}\exp\left[-\frac{(v-v_{0\rm b})^2}{2v_{\rm th,b}^2}\right]\right\} \\ & \times\left[\frac{i}{\omega-kv}E_0 e^{ikx-i\omega t} + \text{complex conjugate}\right] , \end{aligned} \tag{13}$$

where E_0 is the initial electric field amplitude. The structure of δf in phase space for $t = 2\pi j/\omega$ with j being an arbitrary integer is shown in Fig. 1(a).

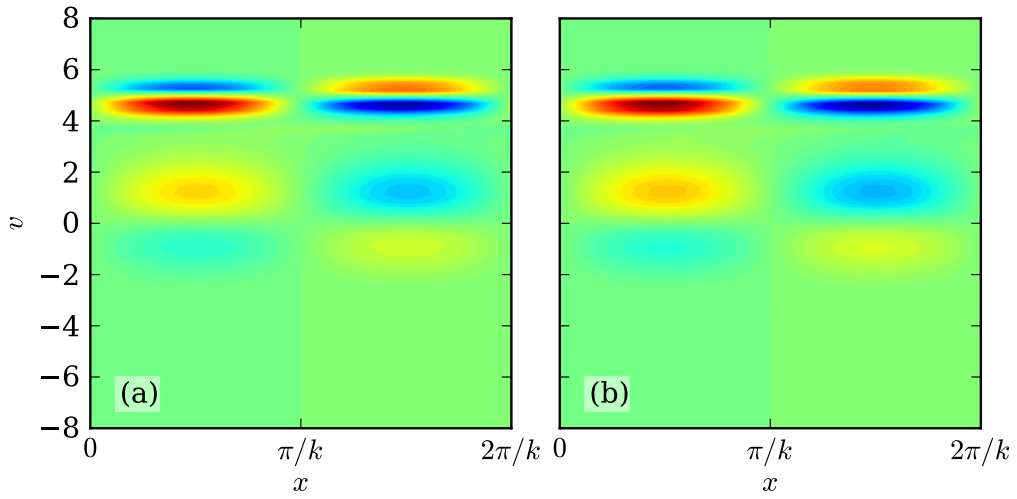

Figure 1: δf linear structure in (x, v) phase space. (a) Semi-analytic solution. (b) Simulation by PIC1D-PETSc.

The PIC simulations in this section are performed by the code PIC1D-PETSc [33]. The simulation result in the linear stage gives $\omega_{\rm r} \approx 1.47$, $\gamma \approx 0.0287$, and δf structure as shown in Fig. 1(b). The PIC simulation result agrees well with the semi-analytic solution.

2.3. Rigorous convergence test for marker particle simulations

Before introducing any optimization technique, we need a robust method for numerical error measurement to evaluate how good an optimization technique is. A marker particle simulation is essentially a Monte Carlo simulation [1], whose results, e.g., linear growth rate, nonlinear saturation level, etc., are random variables with certain expectations and standard deviations. The expectations should be equal to the physical results, and the standard deviations usually go as $\sim 1/\sqrt{N}$ due to the central limit theorem, where N is the number of sampling points, i.e., the number of markers in a particle simulation. In this case, the simulation results converge to the physical results as $N \to \infty$.

Because the statistical uncertainty is random for every run, it is not revealed by a traditional single-scan convergence test, which performs the simulation once for each marker number. For example, Fig. 2(a) gives a traditional single-scan convergence test, showing the effect of the number of markers on the linear growth rate of the bump-on-tail instability by PIC1D-PETSc simulation. Due to the random fluctuation in each run, this convergence test has three issues:

1. the convergence trend is not very clear;
2. the inverse-square-root relation between the statistical uncertainty and the number of markers is not shown;
3. it is not a consistent test as redoing the test with markers loaded with the same distribution but with a different random number sequence may give a very different image; Fig. 2(b) and (c) give another two possible results of such redoing.

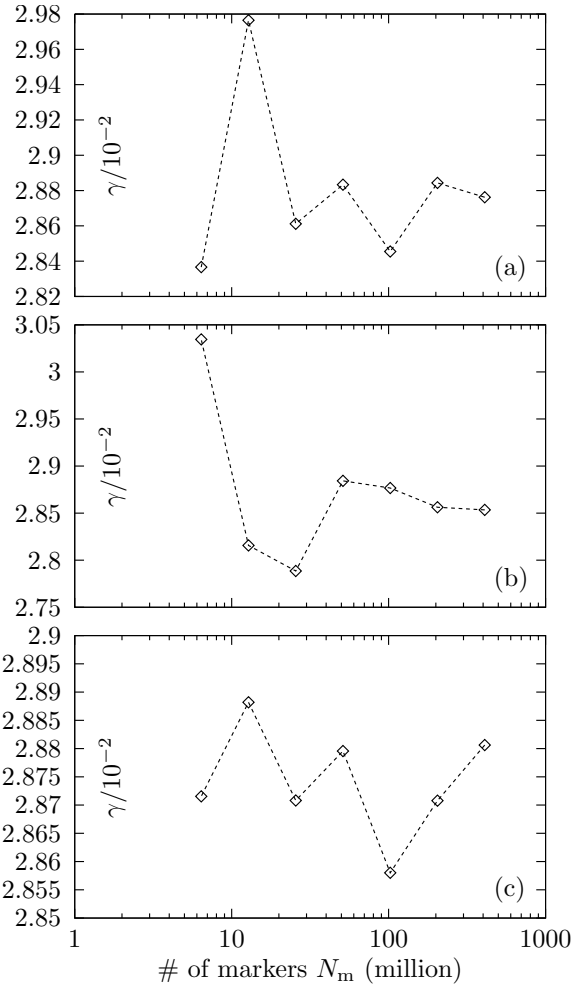

Figure 2: Traditional single-scan convergence tests of linear growth rate of bump-on-tail instability vs number of markers. (a), (b) and (c) are giving three possible test results when markers are loaded with the same distribution but with different random number sequences.

In contrast, a more rigorous convergence test is to perform the simulation multiple times as a sample, with markers loaded with different random number sequences, which can be achieved simply by changing the random seeds, then use the sample statistics to get a clearer error estimate as shown in Fig. 3. Here we assume that the distributions of the linear growth rate and other macroscopic simulation results are Gaussian without rigorous mathematical proof. From statistics we know that, if from a Gaussian random variable X we take a sample $\{X_j, j = 1..N\}$ which is of size N, the mean of the sample $\overline{X_j}$ is also a Gaussian distributed random variable with an expectation same as the original Gaussian distribution, i.e., $\mathfrak{E}(\overline{X_j}) = \mathfrak{E}(X)$, and a standard deviation $\sigma(\overline{X_j}) = \sigma(X)/\sqrt{N}$, where $\sigma(X)$ is the standard deviation of the original Gaussian distribution. According to the Cochran's theorem [34], the quantity $\sqrt{N-1}s(X_j)/\sigma(X)$ has a chi distribution with $N-1$ degrees of freedom, where $s(X_j)$ is the standard deviation of the sample. Based on these properties, we can estimate the expectation and the statistical uncertainty of the linear growth rate and other simulation results. Here we take the sample size to be 20, i.e., 20 simulations are performed for every choice of the marker numbers, with different random seeds. The sample means as estimations of the expectations of the linear growth rate are shown in Fig. 3(a). The error bars give the intervals of $\overline{\gamma_j} \pm 2s(\gamma_j)/\sqrt{20}$ and therefore are $\sim 95\%$ confidence intervals, where $\overline{\gamma_j}$ and $s(\gamma_j)$ are the mean and the standard deviation of the sample growth rate, respectively. It can be seen that these data points have approximately the same γ values within error bars. This indicates that finite marker number does not introduce apparent systematic error. The sample standard deviations as es-

timations of the standard deviations of the linear growth rate are shown in Fig. 3(b) (normalized by the estimated expectation). The error bars of the standard deviation give the intervals of $[s(\gamma_j)/\overline{\gamma_j}](100\% \pm 30\%)$ and therefore are $\sim 93\%$ confidence intervals. The fitting verifies the inverse-square-root relation between statistical uncertainty and marker number.

A correctly working optimization technique should:

1. make the data points in Fig. 3(a) to stay at the same level, so that no systematic error is introduced;
2. and meanwhile lower the data points in Fig. 3(b), so that the statistical uncertainty is reduced for the same marker number, or less markers are needed for the same statistical uncertainty level.

2.4. Candidate optimization techniques

There are at least two categories of techniques for optimizing the marker distribution during the simulation: marker increase and marker reduction. In the first category, the simulation initially loads small amount of markers. Then during the simulation when δf is known, in the regions where δf is localized, one (or a few) marker is split into multiple markers to increase resolution. Or alternatively, when δf is known, the initial set of markers is removed and a new set of markers in a larger amount is reloaded based on the calculated δf, to get better resolution. The reloading technique, also called phase space remapping, has been systematically developed for PIC simulations with 2D [35] and 4D [36] phase space. In the second category, the simulation initially loads large amount of markers, then merge or remove markers in the $\delta f \approx 0$ regions.

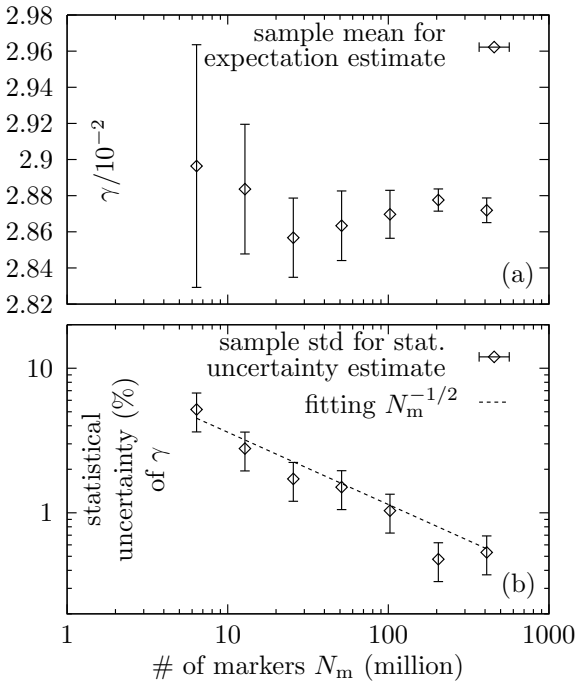

Figure 3: A more rigorous convergence test of linear growth rate of bump-on-tail instability vs number of markers. The growth rate is modeled as a Gaussian random variable. 20 simulations with different random seeds are performed for each choice of the marker numbers to sample this Gaussian random variable for estimation of (a) the expectation and (b) the standard deviation (normalized by the estimated expectation) and fit for the inverse-square-root relation. The error bars of the expectation give the intervals of $\overline{\gamma_j} \pm 2s(\gamma_j)/\sqrt{20}$ and therefore are $\sim 95\%$ confidence intervals. The error bars of the standard deviation give the intervals of $[s(\gamma_j)/\overline{\gamma_j}](100\% \pm 30\%)$ and therefore are $\sim 93\%$ confidence intervals.

The first category is a process of upsampling from a low-resolution sample and is difficult in general. The splitting technique is especially difficult because:

1. it can easily introduce non-conservation problem if not treated carefully [37], bringing in systematic error;
2. the children markers split from one parent marker are correlated; the statistical uncertainty of a Monte Carlo simulation goes $\sim 1/\sqrt{N}$ where N is the number of independent sampling elements; the splitting technique does not increase independent markers (unless after the decorrelation time, which could be very long) and therefore is hard to reduce statistical uncertainty.

The second category is a process of downsampling from a high-resolution sample and is relatively easier than the first category. For the merging technique, finding nearby markers to merge is not very easy, especially in high dimension phase space. In this work, a marker removal technique is developed and is presented below.

2.5. Marker removal technique

The marker removal is performed only once in the early linear stage, when linear eigenmode just forms. The removal procedures consist two marker loops.

The first marker loop is to calculate an importance function. Before normalization, the importance function is defined as:

$$\mathfrak{i}(v) = \frac{\int |\delta f(x,v)| \mathrm{d}x}{\int g(x,v) \mathrm{d}x} \ , \tag{14}$$

where $g(x, v)$ is the marker distribution. Making a grid on v, the importance function can be approximated by:

$$\mathfrak{i}(v) \approx \frac{\sum_j |w_j| S(v_j - v)}{\sum_j S(v_j - v)} \ , \tag{15}$$

where $S()$ is the shape function for marker-grid interpolation, and $\sum_j$ sums over all markers. Note that this approximation may overestimate $\mathfrak{i}(v)$ because:

$$\begin{aligned} \mathfrak{i}(v) &\simeq \frac{\int |\sum_j w_j S(v_j - v) S(x_j - x)| \mathrm{d}x}{\sum_j S(v_j - v)} \\ &\leq \frac{\int \sum_j |w_j| S(v_j - v) S(x_j - x) \mathrm{d}x}{\sum_j S(v_j - v)} \\ &\simeq \frac{\sum_j |w_j| S(v_j - v)}{\sum_j S(v_j - v)} \ . \end{aligned} \tag{16}$$

This overestimation should be small and is not important because the accuracy of the importance function is not important as will be discussed in Sec. 4.2. The importance function is normalized as:

$$I(v) = \frac{\mathfrak{i}(v)}{\max_v \{\mathfrak{i}(v)\}} \ . \tag{17}$$

This normalization makes the function in the range of $[0, 1]$.

The second marker loop uses $I(v)$ to identify the importance of each marker and determine whether each marker should be removed. For each marker indexed j,

1. its importance I_j is calculated directly by the importance function: $I_j = I(v_j)$;
2. then a uniform $[0, 1]$ random number called "dice" is generated;

3. if dice $> I_j$, remove this marker; otherwise, keep this marker and scale up its weights: $p_j \to \frac{1}{I_j} p_j$, $w_j \to \frac{1}{I_j} w_j$.

This way makes the marker distribution to have a same shape as the importance function, and only statistical error is introduced to the distribution function, as well as mass, momentum and energy. In the limit of infinite number of markers, the distribution function and its moments are conserved exactly.

The reason that the importance function integrates over x is that in this 2D phase space, v is an equilibrium motion constant while x is not. The importance function as well as the marker distribution after marker removal need to be a function of the equilibrium motion constants, so that in linear stage, the marker distribution is time independent to keep the optimization effective. In nonlinear stage, the optimization is still effective if the importance function does not change too much from the linear stage. Note that when there are multiple species in the system, each species has its own importance function and the above procedures need to be performed for each species.

2.6. Test on electron bump-on-tail simulation

The marker removal technique is tested on the electron bump-on-tail simulation described in Sec. 2.2. The time evolution of the electric field energy is shown in Fig. 4. As described in Sec. 2.5, the marker removal technique is applied in the early linear stage when the linear eigenmode just forms. The marker distributions before and after the marker removal are shown in Fig. 5.

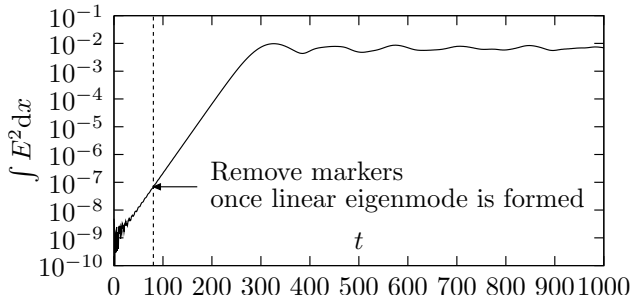

Figure 4: Time evolution of electric field energy in electron bump-on-tail simulation with marker removal applied in early linear stage once linear eigenmode is formed.

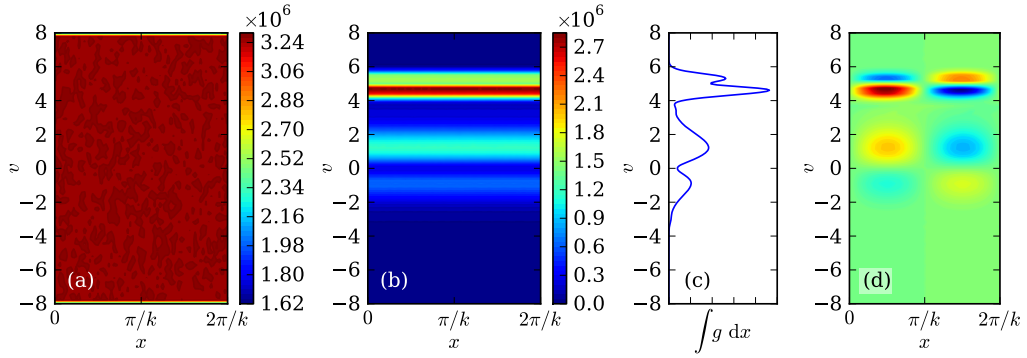

Figure 5: Marker distributions in bump-on-tail simulation and compare with δf: (a) initial uniform marker distribution; (b) marker distribution after removal; (c) marker distribution after removal in velocity space (x integrated); (d) δf for comparison.

The rigorous convergence test method described in Sec. 2.3 is used to test two common macroscopic quantities in simulation results, linear growth rate and nonlinear saturation level, so as to measure the optimization effect of the marker removal technique. The convergence results are shown in Figs. 6 and 7. It can be seen from the upper panels of both figures that "optimized" data points are vertically very close to the "regular" ones, indicating that the marker removal technique does not introduce apparent systematic error. The lower panels of these two figures show that the marker removal technique reduces the statistical uncertainty by a factor of ~ 2, indicating that retaining the same statistical uncertainty, the marker removal technique can save markers by a factor of ~ 4 due to the inverse-square-root relation. The optimization by marker removal is successful in this 2D simulation, and will be generalized to 5D gyrokinetic simulations in the next section.

3. Optimization in 5D gyrokinetic simulations

In this section, the marker removal optimization technique is generalized to 5D gyrokinetic simulations and is tested in a TAE simulation by the M3D-K code [10].

3.1. Marker removal technique in 5D gyrokinetic phase space

In the 5D gyrokinetic phase space, the phase space volume element is: $\mathcal{J}\,\mathrm{d}\mathcal{E}\,\mathrm{d}\mu\,\mathrm{d}P_\zeta\,\mathrm{d}\zeta\,\mathrm{d}\theta_{\mathrm{orb}}$, where $\mathcal{J}$ is the Jacobian and $\mathcal{E}$, μ, P_ζ, ζ, θ_{orb} are gyrocenter energy, magnetic moment, toroidal angular momentum, toroidal angle and orbit phase, respectively. The reason to choose this special coordinate system is that the importance function for marker optimization is a function of equilibrium motion constants, and $\mathcal{E}$, μ, P_ζ are the equilibrium motion

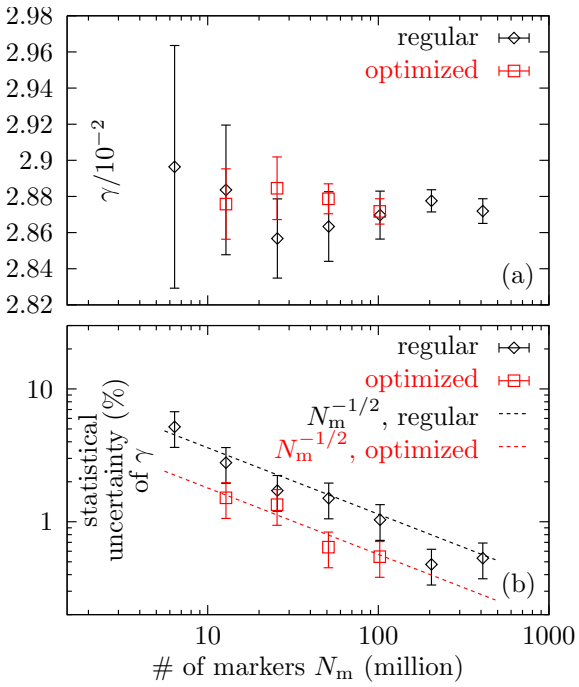

Figure 6: Convergence test of linear growth rate of bump-on-tail instability comparing the cases without and with marker removal optimization.

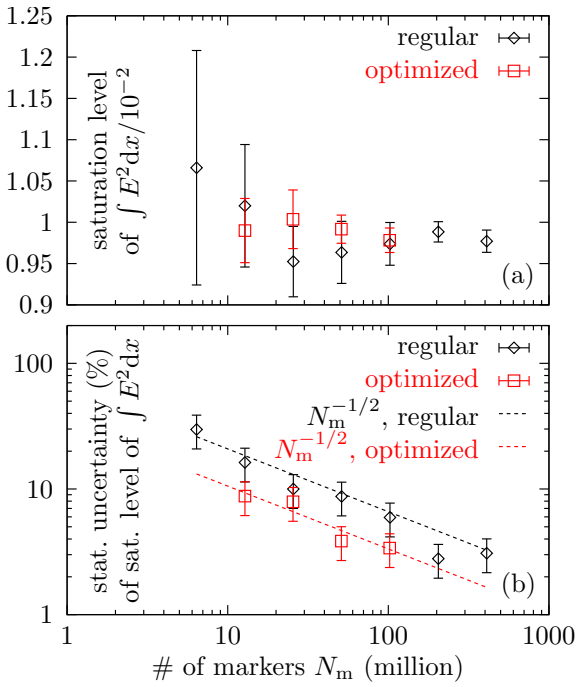

Figure 7: Convergence test of nonlinear saturation level of bump-on-tail instability comparing the cases without and with marker removal optimization.

constants in the 5D gyrokinetic phase space. The unnormalized importance function in the 5D phase space is defined as:

$$\mathfrak{i}(\mathcal{E},\mu,P_\zeta)=\frac{\int|\delta f\mathcal{J}|\,\mathrm{d}\zeta\,\mathrm{d}\theta_{\mathrm{orb}}}{\int g\mathcal{J}\,\mathrm{d}\zeta\,\mathrm{d}\theta_{\mathrm{orb}}}\ , \tag{18}$$

which can be approximated by:

$$\mathfrak{i}(\mathcal{E},\mu,P_\zeta)\approx\frac{\sum_j|w_j|S(\mathcal{E}_j-\mathcal{E})S(\mu_j-\mu)S(P_{\zeta,j}-P_\zeta)}{\sum_j S(\mathcal{E}_j-\mathcal{E})S(\mu_j-\mu)S(P_{\zeta,j}-P_\zeta)}\ . \tag{19}$$

Again this approximation may overestimate $\mathfrak{i}(\mathcal{E},\mu,P_\zeta)$, but this overestimation is not important. Similar to the case in the 2D phase space, the importance function is normalized as:

$$I(\mathcal{E},\mu,P_\zeta)=\frac{\mathfrak{i}(\mathcal{E},\mu,P_\zeta)}{\max_{\mathcal{E},\mu,P_\zeta}\{\mathfrak{i}(\mathcal{E},\mu,P_\zeta)\}}\ . \tag{20}$$

After the importance function is calculated, the procedures for marker removal in 5D phase space are the same as those in 2D phase space described in Sec. 2.5. Note that again, when there are multiple species, the procedures need to be performed for each species individually.

3.2. Test on a TAE simulation

M3D-K is a kinetic/MHD hybrid code that solves energetic particles in the 5D gyrokinetic phase space [10]. The marker removal technique has been implemented in the kinetic part of M3D-K and is tested on an $n = 1$ TAE simulation. The mode structures in linear and nonlinear stages are similar as can be seen from Fig. 8. The time history of the MHD perturbation energy is shown in Fig. 9. The third nonlinear peak is marked in this figure as this is used in the convergence tests for measuring the marker saving effect in the nonlinear stage. Similar to the 2D case, the marker removal is performed in

early linear stage when linear eigenmode just forms. Fig. 10 shows the marker distribution change from the removal in one $\mathcal{E}$–P_ζ slice ($\mu = 0$) of the motion constant space. In this simulation, initially the markers are loaded uniformly in $(R, Z, \zeta, v_\parallel, v_\perp)$ coordinates. As we look at the system in the motion constant coordinates, the coordinate transformation Jacobian $\mathcal{J}$ makes the marker distribution concentrated at the bottom of the phase space slice as shown in Fig. 10(a). Meanwhile, the δf structure locates mainly on the left side in the $\mathcal{E}$–P_ζ slice as shown in Fig. 10(b). After marker removal, the markers become concentrated in a narrow band on the left side in the $\mathcal{E}$–P_ζ slice as shown in Fig. 10(c). The rigorous convergence test method is applied to linear growth rate, nonlinear saturation level, and also the amplitude of the third nonlinear peak. The results are given in Figs. 11, 12 and 13, respectively. From these figures, it is again verified that numerical error introduced by finite marker number is mainly statistical. It can be also seen that the marker removal technique does not introduce apparent systematic error. The marker removal technique saves markers by a factor of 19, 34 and 24 with respect to linear growth rate, nonlinear saturation level, and third nonlinear peak, respectively. Overall, the saving factor is 19.

4. Possible generalizations and extensions of the marker removal optimization technique

4.1. Generalization to arbitrary dimension phase space

Our marker removal optimization technique can be straightforwardly generalized to arbitrary dimension phase space, as long as the equilibrium motion constants can be found in this phase space. The only thing that needs

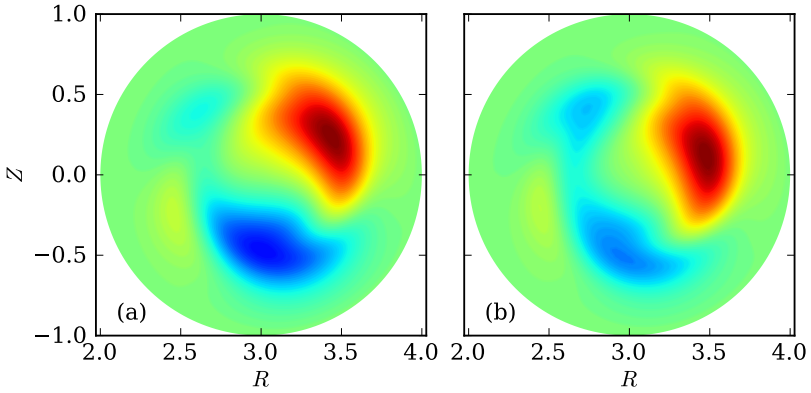

Figure 8: TAE mode structures in our simulation: (a) linear stage; (b) nonlinear stage

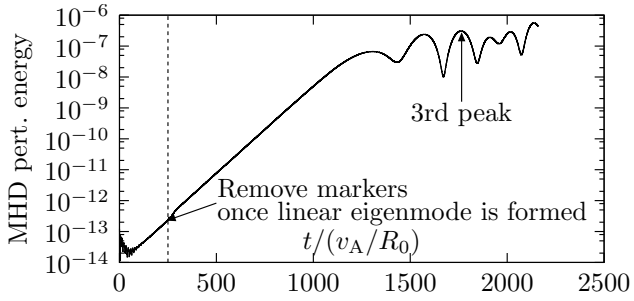

Figure 9: Time history of MHD perturbation energy in our TAE simulation. Marker removal is performed at early linear stage when linear eigenmode just forms.

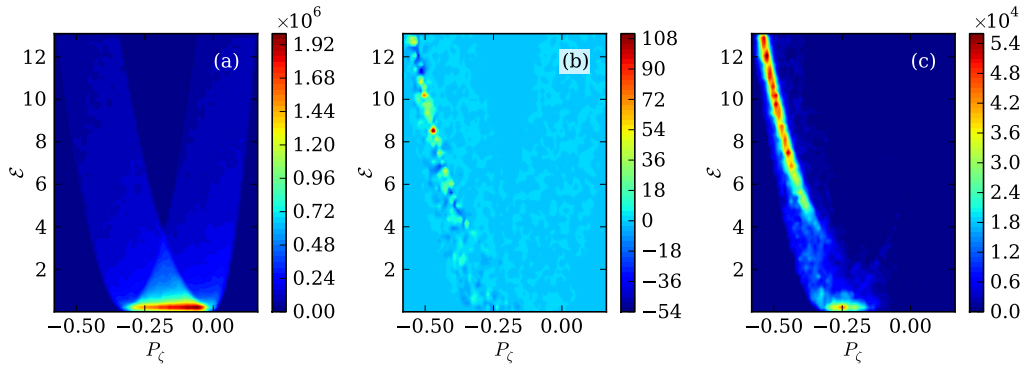

Figure 10: In the $\mu = 0$ slice of the motion constant space in our TAE simulation: (a) marker distribution ($\int g\mathcal{J}\mathrm{d}\zeta\mathrm{d}\theta_{\mathrm{orb}}|_{\mu=0}$) before removal; (b) perturbed distribution ($\int \delta f\mathcal{J}\mathrm{d}\zeta\mathrm{d}\theta_{\mathrm{orb}}|_{\mu=0}$) before removal; (c) marker distribution after removal.

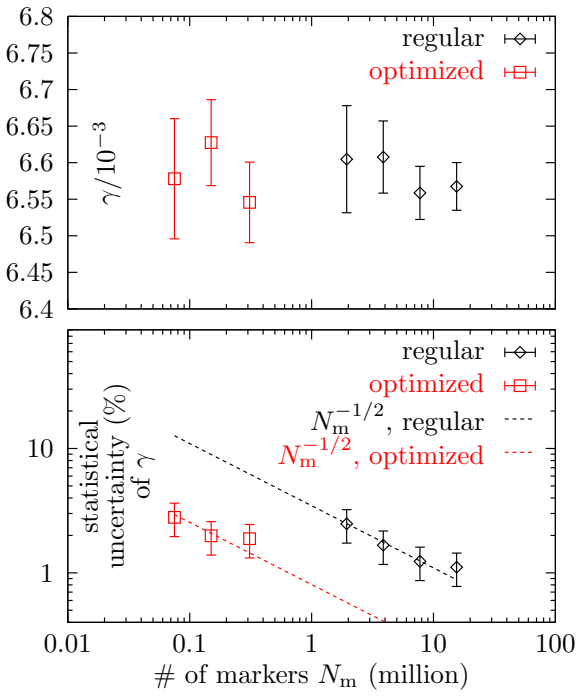

Figure 11: Convergence test of linear growth rate of TAE comparing the cases without and with marker removal optimization

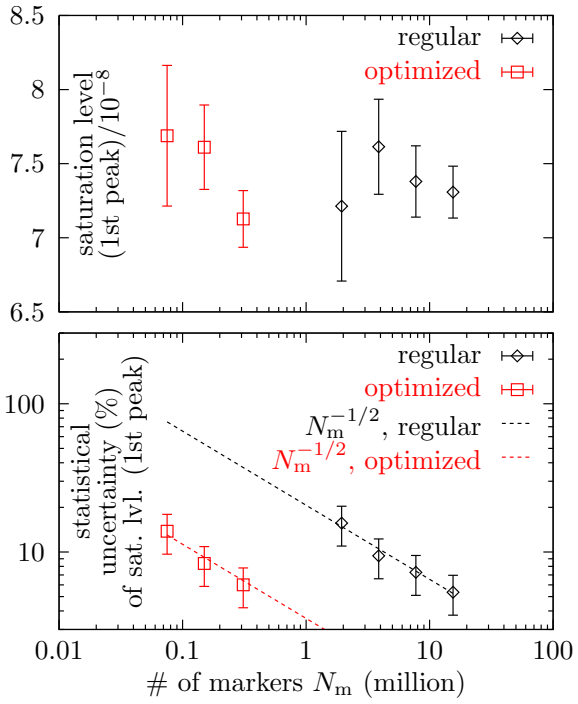

Figure 12: Convergence test of nonlinear saturation level of TAE comparing the cases without and with marker removal optimization

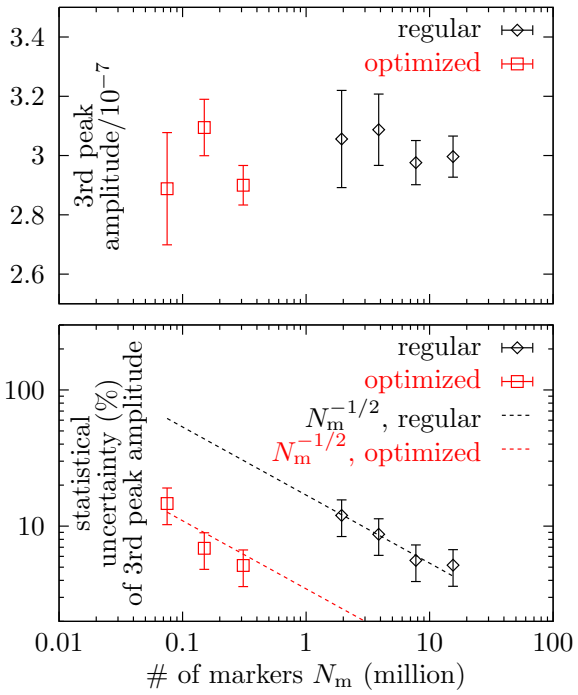

Figure 13: Convergence test of nonlinear third peak amplitude of TAE comparing the cases without and with marker removal optimization

to be changed in a different phase space is the importance function, which becomes:

$$\begin{aligned} \mathfrak{i}(\boldsymbol{z}_{\mathrm{c}}) &= \frac{\int |\delta f(\boldsymbol{z}_{\mathrm{c}}, \boldsymbol{z}_{\mathrm{n}}) \mathcal{J}| \mathrm{d}\boldsymbol{z}_{\mathrm{n}}}{\int g(\boldsymbol{z}_{\mathrm{c}}, \boldsymbol{z}_{\mathrm{n}}) \mathcal{J} \mathrm{d}\boldsymbol{z}_{\mathrm{n}}} \\ &\approx \frac{\sum_j |w_j| S(\boldsymbol{z}_{\mathrm{c},j} - \boldsymbol{z}_{\mathrm{c}})}{\sum_j S(\boldsymbol{z}_{\mathrm{c},j} - \boldsymbol{z}_{\mathrm{c}})} , \end{aligned} \tag{21}$$

where $\boldsymbol{z}_{\mathrm{c}}$ denotes the motion constants, and $\boldsymbol{z}_{\mathrm{n}}$ denotes the non-constant coordinates. The normalization of importance function does not change:

$$I(\boldsymbol{z}_{\mathrm{c}}) = \frac{\mathfrak{i}(\boldsymbol{z}_{\mathrm{c}})}{\max_{\boldsymbol{z}_{\mathrm{c}}}\{\mathfrak{i}(\boldsymbol{z}_{\mathrm{c}})\}} . \tag{22}$$

The removal procedures remain the same as those in the 2D simulation described in Sec. 2.5.

4.2. Flexibility of importance function

The importance function is only for controlling what portion of markers to be kept and it does not directly enter the simulation system. In the limit of infinite number of markers, no matter what importance function is used, the simulation result does not change. Therefore, the accuracy of the importance function is not crucial. As a result, the grid in motion constant space for calculation of importance function does not need to have such a small grid size that resolves all fine structures in δf since the grid size does not directly contribute to the numerical error of the simulation. Also, the shape function for calculation of importance function does not need to be of high order. A simple piecewise linear function is good enough. Furthermore, the choice of importance function is flexible and not unique. The importance function Eq. (21) leads to an aggressive marker removal, which brings in quite large

statistical uncertainty to a single run. Although the rigorous convergence test shows marker saving effect, it is not practical to do the rigorous convergence test for every simulation scenario as it needs to perform the simulation too many times. A less aggressive removal may be more practical. Such less aggressive removal can be achieved by adjusting the normalized importance function. Here are two examples of such less aggressive functions:

$$I_{\mathrm{la1}}(\boldsymbol{z}_{\mathrm{c}}) = [I(\boldsymbol{z}_{\mathrm{c}})]^{\xi} \qquad 0 < \xi < 1 \ , \tag{23}$$

$$I_{\mathrm{la2}}(\boldsymbol{z}_{\mathrm{c}}) = \frac{\tanh\left\{\eta[I(\boldsymbol{z}_{\mathrm{c}}) - 1/2]\right\}}{\tanh(\eta/2)} \qquad \eta > 0 \ , \tag{24}$$

where $I(\boldsymbol{z}_{\mathrm{c}})$ is the original normalized importance function given by Eq. (22). One can simply use one of these functions to calculate the importance of each marker instead of the original normalized importance function to achieve less aggressive removal effect. The aggressiveness can be adjusted by the parameters ξ and η. The optimal values of ξ and η need further investigation.

4.3. Applicability in multi-mode and nonlinear simulations

Our marker removal optimization technique is good for single-mode linear simulations (or linear stage of nonlinear simulations), because in this case the normalized importance function does not change over time in the linear stage, so the optimization effect, i.e., the "importance sampling", is sustained. In multi-mode linear simulations, usually there is one most unstable mode which dominates the simulation, in which case the technique is still good. In nonlinear stage, the technique is still good if the δf structures do not change or move too much; or the δf structures move, but markers also move along. This is why the rigorous convergence tests show that the tech-

nique is still good in nonlinear stage for both the 2D bump-on-tail simulation and the 5D TAE simulation.

If there are multiple modes that are comparably unstable and are all important in the simulation, or δf changes too much and the markers do not follow along in the nonlinear stage, then the marker removal technique is not recommended for the very first run; but if a second run is needed for whatever reason, the technique is useful by constructing an importance function from the previous run, which will be described in Sec. 4.4.

4.4. Construct importance function from known properties

When some properties of the simulation is known before the simulation is performed, the importance function may be constructed from those known properties instead of from the linear simulated δf. For example, if the linear δf can be calculated analytically, then the importance function can also be calculated before the simulation. In this case, the marker removal can be performed at the very beginning of the simulation instead of at the early linear stage. Or alternatively, the markers can be initially loaded to distribute as the importance function in the motion constant space and no removal needs to be performed. This alternative way needs to find the mapping of particle orbits between regular phase space coordinates, which are simply the position and the velocity, and motion constant space coordinates. This is solved in tokamak geometry [38], but may be unsolved and difficult in other geometries. In a geometry where the mapping is difficult to find, the marker removal is a good replacement.

Sometimes a simulation needs to be re-performed because a higher resolution is needed, or output of more data is needed, or some minor mistakes

were made in the previous simulation. If the previous simulation is still trustable, then it can be used to construct the importance function for the next simulation:

$$I_{\text{next}}(\boldsymbol{z}_{\text{c}}) = \max_{t}\{I_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t)\} \ , \tag{25}$$

where $I_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t)$ is the normalized importance function from previous run:

$$I_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t) = \frac{\mathfrak{i}_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t)}{\max_{\boldsymbol{z}_{\text{c}}}\{\mathfrak{i}_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t)\}} \ , \tag{26}$$

$$\begin{aligned} \mathfrak{i}_{\text{prev}}(\boldsymbol{z}_{\text{c}}, t) &= \frac{\int |\delta f(\boldsymbol{z}_{\text{c}}, \boldsymbol{z}_{\text{n}}, t)\mathcal{J}|\mathrm{d}\boldsymbol{z}_{\text{n}}}{\int g(\boldsymbol{z}_{\text{c}}, \boldsymbol{z}_{\text{n}}, t)\mathcal{J}\mathrm{d}\boldsymbol{z}_{\text{n}}} \\ &\approx \frac{\sum_j |w_j(t)| S[\boldsymbol{z}_{\text{c},j}(t) - \boldsymbol{z}_{\text{c}}]}{\sum_j S[\boldsymbol{z}_{\text{c},j}(t) - \boldsymbol{z}_{\text{c}}]} \ , \end{aligned} \tag{27}$$

where δf, g, w_j and $\boldsymbol{z}_{\text{c},j}$ are taken from the previous run. The importance function is time dependent because the previous simulation has been performed and the importance function can be obtained for any time.

4.5. Automatic selection of marker removal time

In the simulations described in Secs. 2 and 3, the marker removal times are specified manually and are determined based on regular simulations without marker removal. For a simulation of a different problem, the marker removal time can also be programmed to be automatically detected. Whether the linear eigenmode has formed can be detected by the stability of instantaneous growth rate and/or the stability of the normalized mode structure. Once the linear eigenmode has formed, the marker removal can and should be performed.

4.6. Applicability in simulations of microturbulence

PIC simulation is vastly used to study microturbulence. It is recommended to follow the discussions in Secs. 4.3 and 4.4 when applying the

marker removal optimization technique to microturbulence simulations. Note that microturbulence may occupy a large portion of the phase space, so the marker saving factor may be small for simulations of microturbulence.

4.7. Applicability in full-f simulations

It is frequently asked whether the marker removal optimization technique can be used in full-f simulations. In full-f simulations, the "importance sampling" can be simply achieved by loading markers such that f/g is a constant, as suggested in Ref. [1] and discussed in Sec. 1. When f is a simple function, e.g., Maxwellian, it is usually easy to directly load markers such that g is proportional to f. When f is a complicated function, direct loading may be difficult. In this case, once can first load markers arbitrarily (the simplest choice would be uniform loading), then use such a simple importance function $\mathfrak{i}(\boldsymbol{z}) = f(\boldsymbol{z})/g(\boldsymbol{z})$, normalize it, and then perform marker removal to achieve the "importance sampling."

4.8. Extension to other approaches of marker particle simulation

Although we applied the marker removal technique to particle-in-cell simulations in Sec. 2 and Sec. 3, the technique does not directly touch the "in-cell" part in the simulation. Therefore, the marker removal optimization technique can be directly applied to other approaches of marker particle simulation such as the PIW simulation [11] and the grid-free treecode simulation [12].

5. Summary

A marker removal optimization technique has been developed for δf particle simulations. The technique uses the linear eigenmode structure in the equilibrium motion constant space to construct an importance function by Eqs. (21) and (22), then remove some markers based on the importance function and adjust the weights of the leftover markers, to optimize the marker distribution. The technique is tested on a 2D bump-on-tail simulation and a 5D TAE simulation and can save markers by a factor of 4 and 19, respectively. The marker saving factor depends on how localized the δf structure is. The technique can be directly applied to single-mode linear simulations. For multi-mode simulations, if there is one most unstable mode that dominates the simulation, the technique can still be directly applied. For nonlinear simulations, if δf does not change too much, or particle orbits follow the change of δf, the technique can still be directly applied. If there are multiple modes that are comparably unstable and are all important in the simulation, or δf changes too much and particle orbits do not follow in nonlinear stage, the technique can be applied but the importance function needs to be constructed from other sources, such as a previous run or analytic theories. The technique is not limited to PIC simulations but could be applied to other approaches of marker particle simulations such as PIW and treecode simulations.

Acknowledgments

The authors acknowledge useful discussions with C. S. Chang, W. W. Lee, P. Porazik, and B. Wang. This work was supported by the US DOE SciDAC

Center for Nonlinear Simulation of Energetic Particles in Burning Plasmas (CSEP). Simulations were performed using the Hopper supercomputer at NERSC.

References

[1] A. Y. Aydemir, A unified monte carlo interpretation of particle simulations and applications to non-neutral plasmas, Physics of Plasmas 1 (4) (1994) 822–831. doi:10.1063/1.870740.

[2] F. H. Harlow, Fluid dynamics in Group T-3 Los Alamos National Laboratory: (LA-UR-03-3852), Journal of Computational Physics 195 (2) (2004) 414–433. doi:10.1016/j.jcp.2003.09.031.

[3] Z. Lin, T. S. Hahm, W. W. Lee, W. M. Tang, R. B. White, Turbulent transport reduction by zonal flows: Massively parallel simulations, Science 281 (5384) (1998) 1835–1837. doi:10.1126/science.281.5384.1835.

[4] Y. Chen, S. E. Parker, A δf particle method for gyrokinetic simulations with kinetic electrons and electromagnetic perturbations, Journal of Computational Physics 189 (2) (2003) 463–475. doi:10.1016/S0021-9991(03)00228-6.

[5] C. S. Chang, S. Ku, H. Weitzner, Numerical study of neoclassical plasma pedestal in a tokamak geometry, Physics of Plasmas 11 (5) (2004) 2649–2667. doi:10.1063/1.1707024.

[6] A. Bierwage, L. Chen, AWECS: A linear gyrokinetic delta-f particle-in-cell simulation code for the study of Alfvénic instabilities in high-beta tokamak plasmas, Commun. Comput. Phys. 4 (2008) 457–495.

[7] A. Mishchenko, R. Hatzky, A. Konies, Global particle-in-cell simulations of Alfvénic modes, Physics of Plasmas 15 (11) (2008) 112106. doi:10.1063/1.3021453.

[8] S. Briguglio, G. Vlad, F. Zonca, C. Kar, Hybrid magnetohydrodynamic-gyrokinetic simulation of toroidal Alfvén modes, Physics of Plasmas 2 (10) (1995) 3711–3723. doi:10.1063/1.871071.

[9] Y. Todo, T. Sato, Linear and nonlinear particle-magnetohydrodynamic simulations of the toroidal Alfvén eigenmode, Physics of Plasmas 5 (5) (1998) 1321–1327. doi:10.1063/1.872791.

[10] G. Y. Fu, W. Park, H. R. Strauss, J. Breslau, J. Chen, S. Jardin, L. E. Sugiyama, Global hybrid simulations of energetic particle effects on the n = 1 mode in tokamaks: Internal kink and fishbone instability, Physics of Plasmas 13 (5) (2006) 052517. doi:10.1063/1.2203604.

[11] R. Nguyen van yen, E. Sonnendrücker, K. Schneider, M. Farge, Particle-in-wavelets scheme for the 1D Vlasov-Poisson equations, ESAIM: Proc. 32 (2011) 134–148. doi:10.1051/proc/2011017.

[12] S. Pfalzner, P. Gibbon, A 3D hierarchical tree code for dense plasma simulation, Computer Physics Communications 79 (1) (1994) 24–38. doi:10.1016/0010-4655(94)90227-5.

[13] J. Lang, Y. Chen, S. E. Parker, G.-Y. Fu, Gyrokinetic δf particle simulations of toroidicity-induced Alfvén eigenmode, Physics of Plasmas 16 (10) (2009) 102101. doi:10.1063/1.3243493.

[14] A. Mishchenko, A. Konies, R. Hatzky, Global particle-in-cell simulations of fast-particle effects on shear Alfvén waves, Physics of Plasmas 16 (8) (2009) 082105. doi:10.1063/1.3207878.

[15] J. Lang, G. Y. Fu, Y. Chen, Nonlinear simulation of toroidal Alfvén eigenmode with source and sink, Physics of Plasmas 17 (4) (2010) 042309. doi:10.1063/1.3394702.

[16] J. Lang, G.-Y. Fu, Nonlinear simulation of toroidal Alfvén eigenmode with microturbulence-induced radial diffusion, Physics of Plasmas 18 (5) (2011) 055902. doi:10.1063/1.3574503.

[17] A. Mishchenko, A. Konies, R. Hatzky, Global particle-in-cell simulations of plasma pressure effects on Alfvénic modes, Physics of Plasmas 18 (1) (2011) 012504. doi:10.1063/1.3546021.

[18] Y. Todo, H. L. Berk, B. N. Breizman, Simulation of Alfvén eigenmode bursts using a hybrid code for nonlinear magnetohydrodynamics and energetic particles, Nuclear Fusion 52 (3) (2012) 033003. doi:10.1088/0029-5515/52/3/033003.

[19] Y. Todo, H. L. Berk, B. N. Breizman, Saturation of a toroidal Alfvén eigenmode due to enhanced damping of nonlinear sidebands, Nuclear Fusion 52 (9) (2012) 094018. doi:10.1088/0029-5515/52/9/094018.

[20] W. Zhang, I. Holod, Z. Lin, Y. Xiao, Global gyrokinetic particle simulation of toroidal Alfvén eigenmodes excited by antenna and fast ions, Physics of Plasmas 19 (2) (2012) 022507. doi:10.1063/1.3685703.

[21] C. Zhang, W. Zhang, Z. Lin, D. Li, Comparison of toroidicity-induced Alfvén eigenmodes and energetic particle modes by gyrokinetic particle simulations, Physics of Plasmas 20 (5) (2013) 052501. doi:10.1063/1.4803502.

[22] W. Deng, Z. Lin, I. Holod, X. Wang, Y. Xiao, W. Zhang, Gyrokinetic particle simulations of reversed shear Alfvén eigenmode excited by antenna and fast ions, Physics of Plasmas 17 (11) (2010) 112504. doi:10.1063/1.3496057.

[23] W. Deng, Z. Lin, I. Holod, Z. Wang, Y. Xiao, H. Zhang, Linear properties of reversed shear Alfvén eigenmodes in the DIII-D tokamak, Nuclear Fusion 52 (4) (2012) 043006. doi:10.1088/0029-5515/52/4/043006.

[24] Y. Chen, T. Munsat, S. E. Parker, W. W. Heidbrink, M. A. Van Zeeland, B. J. Tobias, C. W. Domier, Gyrokinetic simulations of reverse shear Alfvén eigenmodes in DIII-D plasmas, Physics of Plasmas 20 (1) (2013) 012109. doi:10.1063/1.4775776.

[25] X. Wang, F. Zonca, L. Chen, Theory and simulation of discrete kinetic beta induced Alfvén eigenmode in tokamak plasmas, Plasma Physics and Controlled Fusion 52 (11) (2010) 115005. doi:10.1088/0741-3335/52/11/115005.

[26] H. S. Zhang, Z. Lin, I. Holod, X. Wang, Y. Xiao, W. L. Zhang, Gyrokinetic particle simulation of beta-induced Alfvén eigenmode, Physics of Plasmas 17 (11) (2010) 112505. doi:10.1063/1.3498761.

[27] X. Wang, S. Briguglio, L. Chen, C. Di Troia, G. Fogaccia, G. Vlad, F. Zonca, Nonlinear dynamics of beta-induced Alfvén eigenmode driven by energetic particles, Phys. Rev. E 86 (2012) 045401. doi:10.1103/PhysRevE.86.045401.

[28] H. S. Zhang, Z. Lin, I. Holod, Nonlinear frequency oscillation of Alfvén eigenmodes in fusion plasmas, Phys. Rev. Lett. 109 (2012) 025001. doi:10.1103/PhysRevLett.109.025001.

[29] H. S. Zhang, Z. Lin, W. Deng, I. Holod, Z. X. Wang, Y. Xiao, W. L. Zhang, Nonlinear dynamics of beta-induced Alfvén eigenmode in tokamak, Physics of Plasmas 20 (1) (2013) 012510. doi:10.1063/1.4776698.

[30] L. Qi, J. Q. Dong, A. Bierwage, G. Lu, Z. M. Sheng, Thermal ion effects on kinetic beta-induced Alfvén eigenmodes excited by energetic ions, Physics of Plasmas 20 (3) (2013) 032505. doi:10.1063/1.4794287.

[31] W. W. Lee, T. G. Jenkins, S. Ethier, A generalized weight-based particle-in-cell simulation scheme, Computer Physics Communications 182 (3) (2011) 564–569. doi:10.1016/j.cpc.2010.10.013.

[32] D. E. Muller, A method for solving algebraic equations using an automatic computer, Mathematical Tables and Other Aids to Computation

10 (56) (1956) 208–215.
URL http://www.jstor.org/stable/2001916

[33] W. Deng, PIC1D-PETSc, http://wdeng.info/codes/pic1dp/ (Aug. 2012).
URL http://wdeng.info/codes/pic1dp/

[34] W. G. Cochran, The distribution of quadratic forms in a normal system, with applications to the analysis of covariance, Mathematical Proceedings of the Cambridge Philosophical Society 30 (1934) 178–191. doi:10.1017/S0305004100016595.

[35] B. Wang, G. Miller, P. Colella, A particle-in-cell method with adaptive phase-space remapping for kinetic plasmas, SIAM Journal on Scientific Computing 33 (6) (2011) 3509–3537. doi:10.1137/100811805.

[36] B. Wang, G. Miller, P. Colella, An adaptive, high-order phase-space remapping for the two dimensional Vlasov–Poisson equations, SIAM Journal on Scientific Computing 34 (6) (2012) B909–B924. doi:10.1137/120872954.

[37] G. Lapenta, Particle rezoning for multidimensional kinetic particle-in-cell simulations, Journal of Computational Physics 181 (1) (2002) 317–337. doi:10.1006/jcph.2002.7126.

[38] A. Bierwage, C. D. Troia, S. Briguglio, G. Vlad, Orbit-based representation of equilibrium distribution functions for low-noise initialization of kinetic simulations of toroidal plasmas, Computer Physics Communications 183 (5) (2012) 1107–1123. doi:10.1016/j.cpc.2012.01.013.

The Princeton Plasma Physics Laboratory is operated
by Princeton University under contract
with the U.S. Department of Energy.

Information Services
Princeton Plasma Physics Laboratory
P.O. Box 451
Princeton, NJ 08543

Phone: 609-243-2245
Fax: 609-243-2751
e-mail: pppl_info@pppl.gov
Internet Address: http://www.pppl.gov