

Scalable implicit solvers with dynamic mesh adaptation for a relativistic drift-kinetic Fokker–Planck–Boltzmann model [☆]

Johann Rudi ^{a,b,*}, Max Heldman ^{a,b,c}, Emil M. Constantinescu ^a, Qi Tang ^{d,**},
Xian-Zhu Tang ^d

^a Mathematics and Computer Science Division, Argonne National Laboratory, Lemont, IL 60439, United States of America

^b Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States of America

^c Department of Mathematics and Statistics, Boston University, Boston, MA 02215, United States of America

^d Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, United States of America

ARTICLE INFO

Keywords:

Relativistic Fokker–Planck–Boltzmann
Adaptive mesh refinement
Fully implicit time stepping
Runaway electrons

ABSTRACT

In this work we consider a relativistic drift-kinetic model for runaway electrons along with a Fokker–Planck operator for small-angle Coulomb collisions, a radiation damping operator, and a secondary knock-on (Boltzmann) collision source. We develop a new scalable fully implicit solver utilizing finite volume and conservative finite difference schemes and dynamic mesh adaptivity. A new data management framework in the PETSc library based on the p4est library is developed to enable simulations with dynamic adaptive mesh refinement (AMR), distributed memory parallelization, and dynamic load balancing of computational work. This framework and the runaway electron solver building on the framework are able to dynamically capture both bulk Maxwellian at the low-energy region and a runaway tail at the high-energy region. To effectively capture features via the AMR algorithm, a new AMR indicator prediction strategy is proposed that is performed alongside the implicit time evolution of the solution. This strategy is complemented by the introduction of computationally cheap feature-based AMR indicators that are analyzed theoretically. Numerical results quantify the advantages of the prediction strategy in better capturing features compared with nonpredictive strategies; and we demonstrate trade-offs regarding computational costs. The robustness with respect to model parameters, algorithmic scalability, and parallel scalability are demonstrated through several benchmark problems including manufactured solutions and solutions of different physics models. We focus on demonstrating the advantages of using implicit time stepping and AMR for runaway electron simulations.

[☆] This work was jointly supported by the U.S. Department of Energy's Fusion Theory Program of the Office of Fusion Energy Sciences and the SciDAC partnership on Tokamak Disruption Simulation between the Department of Energy's Office of Fusion Energy Sciences and the Department of Energy's Office of Advanced Scientific Computing; and through the FASTMath Institute. Los Alamos National Laboratory is operated by Triad National Security, LLC, for the National Nuclear Security Administration of U.S. Department of Energy (Contract No. 89233218CNA000001). Argonne National Laboratory is operated under contract DE-AC02-06CH11357.

* Corresponding author at: Department of Mathematics, Virginia Tech, Blacksburg, VA 24061, United States of America.

** Corresponding author.

E-mail addresses: jrudi@vt.edu (J. Rudi), maxh@vt.edu (M. Heldman), emconsta@mcs.anl.gov (E.M. Constantinescu), qtang@lanl.gov (Q. Tang), xtang@lanl.gov (X.-Z. Tang).

1. Introduction

Plasma disruptions and their mitigation strategies have been one of the most active research areas in tokamak fusion in recent years. Studying runaway electrons, which are a major cause for intolerable machine damages by tokamak disruptions, is of great interest. The runaway electrons are typically considered in the guiding center formulation, which has been averaged over the gyromotion. The resulting model is a relativistic Fokker–Planck–Boltzmann model when both small-angle and large-angle Coulomb collisions are accounted for. For more details on the physics, see the recent review papers [1,2] and references therein.

One of the main interests in studying the runaway electrons is to understand the impact of different collisions (large-angle, small-angle, etc.) and radiation damping under the dominant electric field along the magnetic field direction. The competition between parallel electric field acceleration and collisional/radiative drag sets the advective term, albeit in momentum space of the relativistic drift-kinetic equation, while the collisional energy diffusion and pitch-angle scattering appear in the form of diffusion operators in energy and pitch of the momentum space. There is a natural time-scale separation between the advective and diffusive time scales, with the latter much longer than the former at suprathermal energy. Therefore it is critical to step over the advective scale due to parallel electric field acceleration of the runaway electrons and study the collision or diffusion scale. As a result, implicit time stepping is typically used.

The existing work based on a continuous representation of distribution functions includes conservative finite difference and finite volume schemes [3–6]. They typically solve the distribution function in the 2D momentum space. Several numerical challenges are associated with this model. First, the advection due to the parallel electric field is dominant in most of the domain, but the diffusion coefficient is nonuniform and becomes larger when getting close to the thermal velocity. Since we are generally interested in the impact of collisions, a desired algorithm should step over the fast advection time scale. Second, the distribution function becomes very anisotropic in pitch and has a nontrivial fat (in energy) tail as time evolves, which is commonly referred to a “runaway tail.” A corresponding treatment in meshing is necessary to resolve the highly anisotropic (in pitch) tail. Third, the distribution function can span over 10 or more orders of magnitude. An accurate numerical algorithm is needed to resolve such a large range of values. This work addresses all these issues. For the first issue, we will rely on implicit time stepping, nonlinear solvers, and algebraic multigrid preconditioners. The second issue will be resolved through developing an adaptive mesh refinement (AMR) capability, which is a major focus of the current work. The third issue will be resolved through the accurate numerical schemes in both space and time as well as a modeling choice to chop off a large portion of a Maxwellian close to the thermal velocity.

In both fluid and kinetic models, AMR plays a key role in fusion device modeling because of localized interesting structures such as thin current sheets, plasmoids, and very anisotropic distributions. Many of the previous related studies focus on fluid plasma models such as extended magnetohydrodynamics (MHD) or its equilibrium (see, e.g., [7–11] and the references therein), while there are only a limited number of studies for continuous kinetic simulations [12–14]. In the broader context of statistical physics for a variety of applications, AMR has found usage in improving the computational efficiency since it “allows resolving important regions of phase space where the particles are present and reduce the number of cells in the regions with no particles” [15]. In Ref. [16], a 1D1V electron kinetic solver with AMR was explicitly demonstrated for the thermal electron runaway acceleration, also known as Dreicer acceleration, in what amounts to be an electrostatic glow discharge for low-temperature plasma generation. The prospective advantage of AMR in mesh-based kinetic solvers, as pointed out in these previous works, holds even greater promise and importance in the current problem of Ohmic-to-runaway current conversion in a magnetized plasma. This is because the strong energy dependence of plasma collisionality introduces a distinct set of numerical challenges that AMR must be tailored to overcome in this paper. First, the energy range of the electrons can easily span 6–7 orders of magnitude between thermal electrons (eVs) and the runaway tail (> MeV), which produces huge variations in plasma collisionality. Second, extreme anisotropy in the momentum space, in the electron pitch ξ , is increasingly aggravated for higher electron energies. Third, the runaway tail distribution, which is the quantity of interest for this type of problems, can be 12 orders of magnitude, or more, lower than the bulk electron density, but the resolution of this tail is critical to evaluate the primary runaway seeds for the follow-up avalanche growth. Much of our considerations in AMR adaptation strategy, spatial and temporal discretization, and the necessary implicit solvers for vastly varying plasma collisionality as a function of electron energy, is geared toward overcoming these extreme numerical challenges.

One of the most important pieces in any of those AMR algorithms is an indicator to determine where to refine and coarsen the mesh. There are two types of AMR indicators in the literature: error-estimator-based and feature-based. For instance, an indicator based on the features in the current and vorticity is proposed in the adaptive MHD solver [8], while a weighted sum of several finite element error estimators is used as the indicator for the same resistive MHD system in [11]. Other common error-estimator-based approaches for finite difference/finite volume include Richardson extrapolation [17], solution smoothness via derivatives [18], and goal-oriented strategies [19]. These probe certain aspects of the solution error but can become expensive if one considers frequent regridding. Another common issue is that the AMR indicators are based on the present solution, which results in adaptive meshes being lagged compared with features of interest because regridding often happens every ten or more time steps. For another class of problems in wave propagation, explicit time stepping is computationally attractive; and in this explicit case one time step with explicit Euler can be taken [20] to approximate mesh refinement one time step in the future. In this work, where we focus on implicit time stepping, we explore several different types of AMR indicators and propose predicting these AMR indicators for multiple time steps.

The computational challenges of solving for the runaway electron distribution from relativistic Fokker–Planck and Fokker–Planck–Boltzmann equations—despite being linear partial differential equations (PDEs)—are: (i) the ability to resolve the dynamically evolving runaway distribution both near the runaway vortex regime and at a parallel electric field that is significantly beyond the onset of the runaway vortex; (ii) the solution being a small but positive function that varies 10–20 orders of magnitude in mo-

mentum space; (iii) the competing effects of advection- and diffusion-dominated regimes that are shifting throughout the momentum space.

Challenges (i) and (ii) are addressed with an AMR discretization to resolve the dynamical structure of the electron distribution in momentum space; while (iii) is tackled with an algorithmically scalable solver for implicit time stepping on the AMR discretization. Specifically, we develop adaptive, scalable, fully implicit solvers for a time-dependent relativistic drift-kinetic Fokker–Planck–Boltzmann model in the momentum space. The octree-based AMR package `p4est` [21] is used as the basic framework to provide a parallel dynamic AMR algorithm. The numerical schemes are based on conservative finite difference or finite volume algorithms, such as, the QUICK and MUSCL schemes [22–24]. The solvers are developed under the PETSc framework [25], and our focus is fully implicit time stepping, nonlinear solvers, and preconditioning strategies. Diagonally implicit Runge–Kutta (DIRK) [26] or explicit first-stage singly diagonally implicit Runge–Kutta (ESDIRK) integrators [27] are chosen. The nonlinear solver is based on a Jacobian-free Newton–Krylov (JFNK) method [28], which is preconditioned by a finite difference coloring Jacobian and hypre’s algebraic multigrid preconditioner [29,30]. While developing the adaptive solvers, we present and analyze several feature-based AMR indicators for the distribution function. We further propose predicting these AMR indicators for multiple time steps, which can be done at a fraction of the computational cost when reducing the original advection–diffusion PDE to a pure-advection PDE. We quantify the benefits of predictive AMR on the resolution of distribution functions and we describe the limitations when trading off computational costs for accuracy; this will be done with extensive numerical experiments. We demonstrate the robustness of our numerical solvers and preconditioners for the computationally challenging runaway electron models within a simulation with dynamic AMR. We achieve problem-robust and algorithmically scalable methods. Additionally, we document the possibilities and limitations for parallel scalability, where the existence of limitations is inherent to any implicit scheme.

The rest of the paper is organized as follows. Section 2 gives an overview of the relativistic Fokker–Planck–Boltzmann model. The discussions focus on the forms of the collision operator, synchrotron radiation, and the full equation in the momentum space. Section 3 describes the key contributions related to AMR algorithms. In particular, a prediction approach for AMR indicators is proposed. Section 4 presents the details of further key contributions: the robust and algorithmically scalable numerical schemes and their implementation in our code framework. Section 5 focuses on the experimental setup and reproducibility. We describe several manufactured solutions for accuracy tests and some details on computational environments. We also discuss the impact of the field and knock-on collision operators. Section 6 presents several numerical examples, which include algorithmic scalability, parallel scalability, and a physics study. The conclusions and discussion of future work are given in Section 7.

2. Governing equations

We study the phase-space dynamics of a runaway electron distribution and consider a simplified version of the relativistic drift-kinetic Fokker–Planck–Boltzmann equation [31]. The governing equations we are considering are effectively described in the spherical coordinates of (p, ξ, θ) , with p the normalized momentum magnitude, $\xi = p_{\parallel}/p$ the pitch, and θ the azimuthal angle. Here p is dimensionless with normalization by $m_e c$, with m_e the rest mass of an electron and c the speed of light, and ξ is the cosine of the polar angle in the standard spherical coordinate. The parallel direction, denoted by subscript $\parallel$, aligns with the magnetic field. We further follow the guiding center model and assume the solution has azimuthal symmetry, which reduces the momentum space to two dimensions. This corresponds to a momentum-space domain of $[0, +\infty) \times [-1, 1]$. The normalized relativistic Fokker–Planck–Boltzmann (RFP) equation to describe runaway electron distribution, $f(t, p, \xi)$, in a slab geometry, is given by

$$\frac{\partial f}{\partial t} - E \left(\xi \frac{\partial f}{\partial p} + \frac{1 - \xi^2}{p} \frac{\partial f}{\partial \xi} \right) = C(f) + \alpha R(f) + S(f), \quad (1)$$

where E stands for the normalized electric field parallel to the magnetic field in a tokamak and its normalization scale is $E_c := m_e c / e \tau_c$ with e the elementary charge and τ_c the time scale. Here the time scale of the equation, τ_c , is the relativistic electron collision time given by

$$\tau_c := \frac{4\pi\epsilon_0^2 m_e^2 c^3}{e^4 n_e \ln \Lambda}, \quad (2)$$

where n_e is the (background) thermal electron density, ϵ_0 is the electrical permittivity, and $\ln \Lambda$ is the Coulomb logarithm. The so-called Connor–Hastie E_c sets the critical value for runaway electron generation [32]. In the current study all those values are given as constant based on practical devices such as ITER. On the right-hand side are several secondary effects due to different types of collisions and synchrotron radiation (defined in detail later), and α is introduced to describe the intensity of radiation damping. All the physical quantities appearing in the current work have been summarized in Appendix B.

This problem has been well studied by plasma physicists [3–6,33], commonly using conservative finite difference or finite volume schemes on a stretched grid. The dynamics of the equation is primarily dominated by the advection term due to the strong electric field. The distribution function has a sharp boundary layer at $p = p_{\min}$ due to the boundary condition (defined in detail later); and as time evolves, the distribution accumulates around $\xi = -1$ if $E > 0$ (or $\xi = 1$ if $E < 0$) due to the advection. The distribution is further impacted by the collisions and the source term in the large p region and eventually forms a nontrivial tail structure that is of the greatest interest. A majority of the previous work uses the stretched grid to resolve the localized distribution and tails. Adaptive mesh refinement has not been explored on this problem to the best of our knowledge.

We note that in [6] a different coordinate of $(p_{\parallel}, p_{\perp})$ is used. Choosing the coordinate of (p, ξ) for the RFP equation has several advantages. The solution is highly localized, as found in [3–6,33]. It is more efficient to resolve the same structure in (p, ξ) while the

coordinate of $(p_{\parallel}, p_{\perp})$ leads to half of the domain resolving a trivial distribution tail. Note that the physics of interest in this study is the small population of runaway electrons. It is therefore much easier to resolve this small population tail in (p, ξ) by chopping off the large Maxwellian bulk close to $p = 0$. We often set $p_{\min}$ to correspond to a few times the thermal energy of the background electrons, and we impose an approximate Dirichlet boundary condition given by the Maxwellian bulk, which is commonly adopted in the physics codes [3–6,33] and will be detailed later. We also provide a test that extends the boundary condition to $p = 0$ through a compatible boundary condition.

2.1. Coulomb collision

In this study the bulk distribution is considered to be close to a Maxwellian with a thermal speed much less than the light speed, namely, $v_t/c \ll 1$, where v_t stands for the thermal velocity of the background electrons. The (small-angle) Coulomb collision of the runaway electrons, which are of much lower density, with the background electrons, $C(f)$, can be approximated by the so-called test-particle collision operator [3,4]. We use the test-particle collision operator proposed in [34] that is applicable to both thermal ($p \sim v_t/c \ll 1$) and relativistic ($p > 1$) electrons. This Coulomb collision operator has the form

$$C(f) = \frac{1}{p^2} \frac{\partial}{\partial p} \left[p^2 \left(C_F f + C_A \frac{\partial f}{\partial p} \right) \right] + \frac{C_B}{p^2} \frac{\partial}{\partial \xi} \left[(1 - \xi^2) \frac{\partial f}{\partial \xi} \right], \quad (3)$$

where all the collision coefficients are p -dependent and given by

$$\begin{aligned} C_F &= \frac{2c^2}{v_t^2} \Psi(x), \\ C_A &= \frac{\sqrt{1+p^2}}{p} \Psi(x), \\ C_B &= \frac{\sqrt{1+p^2}}{2p} \left(Z + \operatorname{erf}(x) - \Psi(x) + \frac{v_t^2}{2c^2} \frac{p^2}{1+p^2} \right). \end{aligned}$$

Here x is p -dependent,

$$x = \frac{c}{v_t} \frac{p}{\sqrt{1+p^2}},$$

Z stands for a charge number, and $\Psi(x) = [\operatorname{erf}(x) - x \operatorname{erf}'(x)]/(2x^2)$ is the Chandrasekhar function. A study of the collision coefficients in a transition from nonrelativistic to relativistic regimes is presented in [3], where the authors show that the collision coefficients at small p are several orders of magnitude larger than those at larger p and thus the diffusion becomes dominant compared with the advection field. Therefore, a fully implicit treatment of the collision operator is necessary. When needed, the collision coefficients are also modified to include the partial screening effect [5], which addresses the role of bound electrons in collisions between fast electrons and partially ionized impurities. Mathematically, it corresponds to modifying the C_B and the Coulomb logarithm $\ln \Lambda$, but it does not involve additional numerical challenges. For more details, see [5,35].

2.2. Synchrotron radiation

When there is an acceleration of a charged particle, there is an energy loss through the emitted electromagnetic radiation. In a dominating magnetic field where a charged particle follows cyclotron motion, relativistic electrons can emit a significant amount of synchrotron radiation. The resulting radiation damping operator satisfies

$$R(f) = \frac{1}{p^2} \frac{\partial}{\partial p} \left[p^3 \gamma (1 - \xi^2) f \right] - \frac{\partial}{\partial \xi} \left[\frac{\xi(1 - \xi^2)}{\gamma} f \right], \quad (4)$$

where $\gamma = \sqrt{1+p^2}$ is the Lorentz factor. The dimensionless parameter for the intensity of damping is formally defined as $\alpha := \tau_c/\tau_s$, a ratio between the collision time scale and the synchrotron radiation time scale, where $\tau_s := 6\pi\epsilon_0 m_e^3 c^3 / e^4 B^2$ and B is the magnitude of the magnetic field. Typically it is about 0.1 to 0.3 for ITER-like tokamaks and 0.001 to 0.05 for a smaller device, which is the range we use in our simulations.

2.3. Knock-on collision

Another important source in runaway electron generations is the avalanche mechanism due to knock-on collisions between runaway electrons and the background thermal electrons [36,37]. The knock-on source considers the secondary large-angle collisions in addition to the primary small-angle collisions of the Fokker–Planck operator [1,2]. The knock-on source is

$$S(p, \xi) = S_1(p, \xi) + S_2(p, \xi), \quad (5)$$

which consists of the secondary electron source term

$$S_1(p, \xi) = \int_{-1}^1 \int_0^\infty \Pi(\gamma', \xi'; \gamma, \xi) f(p', \xi') 2\pi p'^2 dp' d\xi', \quad (6)$$

and the annihilation term

$$S_2(p, \xi) = -f(p, \xi) \int_{-1}^1 \int_0^\infty \Pi(\gamma, \xi; \gamma', \xi') 2\pi p'^2 dp' d\xi', \quad (7)$$

with $\gamma = \sqrt{1 + p^2}$ and $\gamma' = \sqrt{1 + p'^2}$.

In this work we use the Chiu model in [37] and approximate the scattering rate function in S_1 by

$$\Pi(\gamma', \xi'; \gamma, \xi) \approx \frac{1}{\ln \Lambda} \frac{p'^2}{2\pi p^2 |\xi|} \frac{d\sigma(\gamma', \gamma)}{dp} \delta(p' - p^*), \quad (8)$$

where $\delta(\cdot)$ is a Dirac delta function that indicates only electrons with particular energy at p^* can generate the required secondary electron at a given point (p, ξ) with $p^* = \sqrt{(\gamma^*)^2 - 1}$ and $\gamma^* = \frac{(\gamma+1)/(\gamma-1)\xi^2+1}{(\gamma+1)/(\gamma-1)\xi^2-1}$. The Møller cross section for large-angle collisions is used,

$$\frac{d\sigma(\gamma', \gamma)}{dp} = \frac{p}{\gamma} \frac{2\pi\gamma'^2}{(\gamma' - 1)^3(\gamma' + 1)} \left[x^2 - 3x + \left(\frac{\gamma' - 1}{\gamma'} \right)^2 (1 + x) \right], \quad (9)$$

where

$$x = \frac{1}{v(1-v)}, \quad \text{and} \quad v = \frac{\gamma - 1}{\gamma' - 1}.$$

Note that the Møller cross section is turned on only if $\gamma' \geq 2\gamma - 1$; otherwise S_1 is set to 0, which is commonly imposed (see [38], for instance). One can easily show that the source term S_1 is positive only in a thin region of $\xi \in [-\sqrt{\gamma/(\gamma+1)}, -p/(\gamma+1)]$. An adaptive mesh is beneficial to resolve this thin region. The annihilation term is integrated in the range of $\gamma' \in [\gamma_0, \frac{\gamma+1}{2}]$, where γ_0 is computed from the computation domain, namely, $\gamma_0 = \sqrt{1 + p_{\min}^2}$. The Møller cross section is well defined in such a range. This leads to

$$S_2(p, \xi) = -\frac{1}{\ln \Lambda} f(p, \xi) \sigma(\gamma, \gamma_0), \quad (10)$$

where $\sigma(\gamma, \gamma_0)$ is the integrated form of the Møller cross section, given by

$$\sigma(\gamma, \gamma_0) = \frac{2\pi}{\gamma^2 - 1} \left[\frac{\gamma + 1}{2} - \gamma_0 - \gamma^2 \left(\frac{1}{\gamma - \gamma_0} - \frac{1}{\gamma_0 - 1} \right) + \frac{2\gamma - 1}{\gamma - 1} \ln \frac{\gamma_0 - 1}{\gamma - \gamma_0} \right]$$

for $\gamma \geq 2\gamma_0 - 1$ or otherwise $\sigma(\gamma, \gamma_0) = 0$.

2.4. Computational domain and boundary conditions

Since the focus of this study is on relativistic runaway electrons where the bulk of the distribution is still assumed to be a Maxwellian, it is reasonable to use a computational domain: $[p_{\min}, p_{\max}] \times [-1, 1]$, where $p_{\min}$ is chosen as being thermal (e.g., let $p_{\min} = v_t/c$ or a few times of v_t/c). One can further assume that the left boundary condition at $p = p_{\min}$ is Dirichlet and given by the Maxwell–Jüttner distribution, which is uniform in ξ . At the right boundary of $p = p_{\max}$, a Neumann boundary condition $\partial f / \partial n = 0$ is prescribed, so one must check the solution to make sure that $p_{\max}$ is high enough that f is negligibly small there. At the top and bottom boundaries, because of the involvement of the coefficient $1 - \xi^2$, all the fluxes along ξ become 0, which is in fact the natural outcome of the coordinate chosen. Therefore, Neumann boundary conditions are used at the top and bottom boundaries.

2.5. Conservation and an alternative RFP form

The determinant of the Jacobian corresponding to the transformation between $(p_{\parallel}, \mathbf{p}_{\perp})$ to (p, ξ) , under azimuthal symmetry, is given by $J = p^2$ (see Appendix A). Thus, the definition of the divergence of a vector $\mathbf{v} = (v_1, v_2)$ in (p, ξ) becomes

$$\nabla \cdot \mathbf{v} := \frac{1}{p^2} \left[\frac{\partial}{\partial p} (p^2 v_1) + \frac{\partial}{\partial \xi} (p^2 v_2) \right].$$

One can show that the electric field “advection” term in (1), $(-E\xi, -E(1 - \xi^2)/p)$, is divergence-free. This is in fact more obvious in the coordinate of $(p_{\parallel}, \mathbf{p}_{\perp})$, as the RFP equation in $(p_{\parallel}, \mathbf{p}_{\perp})$ becomes

$$\frac{\partial f}{\partial t} - E \frac{\partial f}{\partial p_{\parallel}} = C(f) + \alpha R(f) + S(f)$$

and E is a constant. Moreover, fp^2 is a conserved value in the coordinate of (p, ξ) , which is consistent with the fact that the distribution function is conservative in $(p_{\parallel}, \mathbf{p}_{\perp})$. As a result, a conservative finite difference method is chosen in this study to satisfy the conservation of the distribution.

Following the previous work [3], we choose to solve $\tilde{f} := fp$. Then we can rewrite the RFP equation (1) as

$$\frac{\partial \tilde{f}}{\partial t} + \frac{1}{p} \frac{\partial \Gamma_p(\tilde{f})}{\partial p} + \frac{\partial \Gamma_{\xi}(\tilde{f})}{\partial \xi} = S(\tilde{f}), \quad (11)$$

where the fluxes are

$$\Gamma_p(\tilde{f}) = -p \left[E\xi + \alpha p \gamma (1 - \xi^2) \right] \tilde{f} - (pC_F - C_A) \tilde{f} - pC_A \frac{\partial \tilde{f}}{\partial p}, \quad (12)$$

$$\Gamma_{\xi}(\tilde{f}) = -(1 - \xi^2) \left[\left(\frac{E}{p} - \frac{\alpha \xi}{\gamma} \right) \tilde{f} + \frac{C_B}{p^2} \frac{\partial \tilde{f}}{\partial \xi} \right]. \quad (13)$$

This conservative form is a choice of our implementation, which eases the head-to-head comparison between the current implementation with our previous one [3]. This choice is not critical for the success of the proposed AMR algorithm.

3. Adaptive mesh refinement: indicators and their prediction in time

Indicators are spatial fields that determine at each grid point whether to refine or coarsen the mesh. This section presents the indicators for adaptivity and shows how indicators are propagated forward in time ahead of the simulation in order to predict the regions of the mesh that need to be refined and derefined. We consider an abstract setting; hence the discussion is general and applicable beyond the RFP equation, which is otherwise the focus of this work. First, we present how the indicators are computed, and then we propose an approach for the prediction of indicators.

3.1. AMR indicators for coarsening and refinement

We define an indicator as a functional, χ , that maps a function, say f , to a single positive scalar value for each mesh cell Ω_h within the entire domain Ω . Hence we write

$$\chi(f; \Omega_h) \in \mathbb{R}_+ \quad \text{for each } \Omega_h \subset \Omega.$$

In the following, the dependence on Ω_h is assumed implicitly such that we can use the brief notation $\chi(f) = \chi(f; \Omega_h)$. The result is then used to evaluate the criteria for AMR. A mesh cell is coarsened if the indicator is below a threshold

$$\chi(f) < \chi_{\min},$$

and it is refined if the indicator is above a threshold

$$\chi_{\max} < \chi(f),$$

where $0 \leq \chi_{\min} < \chi_{\max}$ are (prescribed) constants that stay fixed during the simulation. These criteria for coarsening and refinement are carried out uniformly for every cell of the mesh. This process yields the requirement that $\chi(f)$ has to produce values that are independent of any scale, because the function f is likely to vary significantly. Indeed, large variations of f are the reason to utilize AMR in the first place.

To proceed to defining AMR indicators, we introduce the following preliminary definitions and observations. Let $f : \Omega \rightarrow \mathbb{R}_+$ be a positive¹ and continuously differentiable function, $f \in C^1(\Omega)$, over an open domain $\Omega \in \mathbb{R}^d$. We define the gradient scale of a function $f > 0$ as the nondimensionalized quantity $(\nabla f)/f$, and as a result the scale of f is neutralized. It satisfies the *gradient-scale identity*

$$\frac{\nabla f(x)}{f(x)} = \nabla (\log f(x)), \quad \text{for } x \in \Omega. \quad (14)$$

Since we are concerned with meshes of a finite resolution, we denote $\Omega_h \subset \Omega$ to be a mesh cell with a *characteristic size* $h > 0$, and we let $\overline{\Omega}_h$ be its closure. Then we define a *discrete local gradient magnitude*, which is motivated by finite difference gradients,

$$G_h[f](x) := \max_{y \in \overline{\Omega}_h} \frac{|f(x) - f(y)|}{h}, \quad \text{for } x \in \Omega_h, \quad (15)$$

where $|\cdot|$ denotes the absolute value. Note that for (15) to be well defined, it is only required for f to be Lipschitz continuous.

We define two AMR indicators based on the gradient scale (GS) in two different ways, therefore making use of both sides of the identity (14).

¹ To simplify the notation, we assume the function f to have positive values. One could, however, remove the assumption on positivity and, in place of f , use $|f| + \varepsilon$ with a chosen $0 < \varepsilon \ll 1$.

Definition 3.1 (Gradient-scale indicator). Let $f > 0$ be Lipschitz continuous, and let $\Omega_h \subset \Omega$ be a cell of the mesh. The first version of the GS indicator is

$$\chi_{GS}(f) := h \max_{x \in \Omega_h} \left(\frac{G_h[f](x)}{f(x)} \right). \quad (16)$$

The second version is

$$\chi_{LGS}(f) := h \max_{x \in \Omega_h} \left(G_h[\log f](x) \right). \quad (17)$$

The indicators in Definition 3.1 are derived from a discrete version of the gradient-scale identity with an additional multiplication by the characteristic size h of the cell. The multiplication by h reduces the value of the indicator as the mesh is refined, and it also neutralizes division by h appearing in the discrete gradient (15). Hence, the indicator becomes nondimensionalized regarding the cell size. Overall, because we eliminated the scale of the function f and the scale of the discrete gradient h , the resulting indicators χ_{GS} and χ_{LGS} are nondimensional.

An alternative indicator based on the dynamic ratio (i.e., maximum value divided by minimum value of f in $\bar{\Omega}_h$) is defined next. Subsequently, a relation between the two definitions of indicators is established in Proposition 3.3.

Definition 3.2 (Log-DR indicator). We define an AMR indicator based on the logarithm of the dynamic ratio (DR) within a cell Ω_h

$$\chi_{LDR}(f) := \log \left(\frac{\max_{x \in \bar{\Omega}_h} f(x)}{\min_{x \in \bar{\Omega}_h} f(x)} \right). \quad (18)$$

Proposition 3.3 (Equivalence between gradient scale and log-DR indicators). Given a Lipschitz continuous function $f > 0$, let χ_{LGS} be the gradient-scale AMR indicator defined in (17) and χ_{LDR} be the log-DR indicator defined in (18). Then the two indicators satisfy

$$\chi_{LGS}(f) = \chi_{LDR}(f). \quad (19)$$

Proof. Starting from the definition (17), we obtain

$$\begin{aligned} \chi_{LGS}(f) &= h \max_{x \in \Omega_h} \left(G_h[\log f](x) \right) \\ &= h \max_{x, y \in \Omega_h} \left(\frac{|\log f(x) - \log f(y)|}{h} \right) \\ &= \max_{x, y \in \bar{\Omega}_h} \left| \log \left(\frac{f(x)}{f(y)} \right) \right|. \end{aligned}$$

Utilizing the monotonicity of the logarithm and $f > 0$, we continue

$$\begin{aligned} \max_{x, y \in \bar{\Omega}_h} \left| \log \left(\frac{f(x)}{f(y)} \right) \right| &= \max \left\{ \max_{x, y \in \bar{\Omega}_h} \log \left(\frac{f(x)}{f(y)} \right), -\min_{x, y \in \bar{\Omega}_h} \log \left(\frac{f(x)}{f(y)} \right) \right\} \\ &= \max \left\{ \log \left(\max_{x, y \in \bar{\Omega}_h} \frac{f(x)}{f(y)} \right), -\log \left(\min_{x, y \in \bar{\Omega}_h} \frac{f(x)}{f(y)} \right) \right\} \\ &= \max \left\{ \log \left(\frac{\max_{x \in \bar{\Omega}_h} f(x)}{\min_{y \in \bar{\Omega}_h} f(y)} \right), -\log \left(\frac{\min_{x \in \bar{\Omega}_h} f(x)}{\max_{y \in \bar{\Omega}_h} f(y)} \right) \right\}. \end{aligned}$$

Additionally, the second argument in the outermost maximum is

$$-\log \left(\frac{\min_{x \in \bar{\Omega}_h} f(x)}{\max_{y \in \bar{\Omega}_h} f(y)} \right) = \log \left(\frac{\min_{x \in \bar{\Omega}_h} f(x)}{\max_{y \in \bar{\Omega}_h} f(y)} \right)^{-1}.$$

Therefore we can simplify

$$\max_{x, y \in \bar{\Omega}_h} \left| \log \left(\frac{f(x)}{f(y)} \right) \right| = \log \left(\frac{\max_{x \in \bar{\Omega}_h} f(x)}{\min_{y \in \bar{\Omega}_h} f(y)} \right),$$

which shows the claim of the proposition. $\square$

Remark 3.4. The indicator χ_{LGS} can be seen in relation to the Lipschitz constant. Let us assume that K_h is a discrete version of the Lipschitz constant such that $|\log f(x) - \log f(y)| \leq K_h |x - y|$ for all $x, y \in \bar{\Omega}_h$ with $h \leq |x - y|$. Then the following estimate holds:

$$\max_{\substack{x, y \in \Omega_h \\ h \leq |x-y|}} \frac{|\log f(x) - \log f(y)|}{|x-y|} \leq \max_{x, y \in \Omega_h} \frac{|\log f(x) - \log f(y)|}{h} = \frac{\chi_{LGS}(f)}{h}.$$

Thus the indicator computes an upper bound on the discrete Lipschitz constant scaled by the characteristic cell size: $hK_h \leq \chi_{LGS}(f)$.

Remark 3.5. In order to compute the indicators (16), (17), and (18), a minimum and/or maximum needs to be computed over a mesh cell. In practice this can be done by looping over the nodes or degrees of freedom of the particular discretization. Furthermore, it is sufficient to approximate the discrete local gradient magnitude (15) by utilizing the routines for gradient computation that are native to the discretization.

For computing the log-DR indicator in (18) from a discretized f that has been polluted by errors due to finite precision arithmetic or solver truncations, which are unavoidable in practice, we recommend computing $\chi_{LDR}(f + \epsilon)$ with a small constant $\epsilon > 0$. This ensures numerically that the argument of χ_{LDR} is positive.

Proposition 3.3 shows that the gradient-scale and the log-DR indicators can be used interchangeably. Specifically, the simple-to-implement and computationally cheaper log-DR indicator quantifies features in the solution equally well as the (slightly) more complex gradient-scale indicator. To decide which formulation of indicator to use depends on a particular discretization method and the quantities that are computed in the implementation. If, for example, a gradient already needs to be computed, then the gradient-scale indicator is a cheap byproduct of this computation. For the finite difference-based discretization that we consider in this work, the log-DR indicator presents the computationally cheapest option. This is why we will utilize log-DR for our adaptivity criterion in the numerical experiments that follow in Section 6.

3.2. Dynamic AMR and indicator prediction in time

This section is concerned with adapting a mesh dynamically to a function $f = f(t)$ that is evolving in time. Dynamic mesh adaptivity is a requirement that is imposed by the PDEs we aim to solve (see Section 2). These PDEs are advection-dominated in some parts of the domain and diffusion-dominated in other parts.

The mesh can be adapted in between time steps of the implicit time integration scheme. To track the variations of the solution as accurately as possible, one would ideally adapt the mesh in between every time step. However, this will increase the computational cost of a parallel implicit solver, because adapting the mesh induces additional computational costs, such as redistributing the mesh cells across compute cores, interpolation between coarse and fine cells, and reallocating and setting up of solver components. If the mesh remains unchanged for too many time steps, on the other hand, accuracy may suffer, and fine-scale features can become insufficiently resolved because they will migrate from fine mesh cells to coarse ones. Therefore dynamic AMR faces two competing constraints: (i) how long can the time intervals be between adapting the mesh in order to adequately track the quantity of interest and (ii) how large is the additional computational cost that each change of the mesh induces.

The goal in this section is to introduce a new technique to balance these two constraints. We want to reduce the frequency of changing the mesh and at the same time provide sufficient resolution where it is needed (within a certain time interval). We also want to avoid overly fine meshes where resolution is not needed. It therefore is desired to “predict” AMR indicators in a time evolution simulation. We will show that such a prediction is possible in an advection-dominated setting.

We denote Δt as the length of the time step of the time evolution scheme. We define a number $n_{\text{adapt}} \in \mathbb{N}_+$ (i.e., positive integer) that determines after how many time steps Δt adaptivity is triggered; therefore the mesh remains fixed during the interval $\Delta t_{\text{adapt}} := n_{\text{adapt}} \Delta t$. This implies that within the n_{adapt} time steps the features of a function $f(t)$, which we want to resolve, will migrate. This is what we want to address with AMR prediction. The superscript notation $f^{(A)}(t)$ indicates that $f(t)$ is discretized (in space) on a particular mesh A .

The diagram in Fig. 1 illustrates how dynamic AMR is carried out between time steps without AMR prediction. For simplicity of the diagram, we set $n_{\text{adapt}} = 1$, and hence $\Delta t_{\text{adapt}} = \Delta t$. The diagram shows the sequence of mesh adaptivity, where interpolation from mesh A to mesh B is performed, alternating with the time evolution scheme, where $f^{(B)}(t)$ evolves to $f^{(B)}(t + \Delta t)$.

With AMR prediction, we aim to avoid a loss of accuracy when fine-scale features of $f(t)$ migrate away from fine to coarser regions of the mesh. We propose to predict the “path of refinement” by evolving the AMR indicators “ahead” of the quantity of interest, $f(t)$, by intervals of at most Δt_{adapt} . In effect, we are leapfrogging the AMR indicators forward in time relative to the (regular) simulation. Dynamic AMR with prediction is illustrated in the diagram in Fig. 2. For simplicity of the diagram, we set $n_{\text{adapt}} = n_{\text{pred}} = 1$, and hence $\Delta t_{\text{adapt}} = \Delta t_{\text{pred}} = \Delta t$. First, the indicator χ is computed as in the previous case without prediction; however, χ is subsequently evolved forward in time by Δt_{pred} ahead of the simulation. To keep the computational costs low, we propose to use an explicit time integration scheme, which should be significantly faster to carry out compared with the implicit scheme used to evolve f . As χ evolves, it is simultaneously reduced into a time-independent indicator χ_{pred} . To adapt the mesh, we then use χ_{pred} , which predicts the path of adaptivity.

Algorithm 1 describes in more detail how dynamic AMR with prediction is carried out during the time evolution of a function f . Specifically, the reduction of the leapfrogged indicator χ is performed by taking the maximum of the evolving indicator across all time steps.

In the general setting, as it has been considered in this section, we have not made assumptions about the time-stepping methods used to evolve $f(t)$ and to leap-frog χ . The choice of the pairing of these two time-stepping methods is important when considering

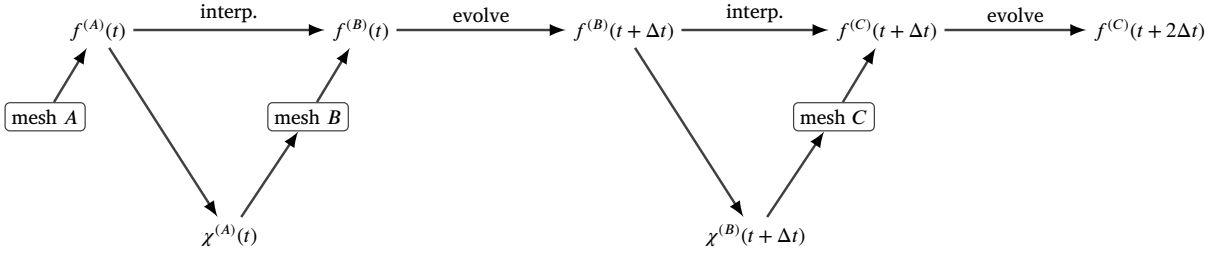


Fig. 1. Diagram of **dynamic AMR without prediction**. In the “interp.” (interpolation) step, the indicator χ is computed from f , and this indicator is used to adapt the mesh (e.g., from mesh A to mesh B). The “evolve” step advances f forward in time by $\Delta t_{\text{adapt}} = \Delta t$.

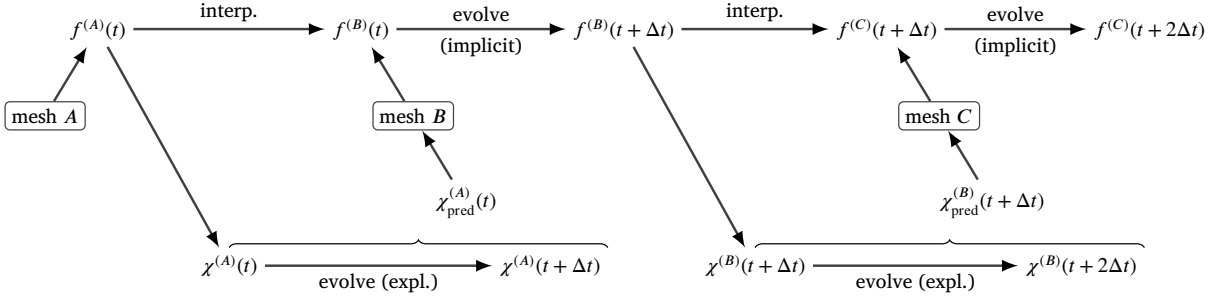


Fig. 2. Diagram of **dynamic AMR with prediction**. The indicator χ is first computed from f and subsequently evolved forward in time by $\Delta t_{\text{pred}} = \Delta t$ ahead of the simulation, thus predicting the path of AMR. The evolution of χ is reduced into one time-independent indicator χ_{pred} . In the “interp.” (interpolation) step, χ_{pred} is used to adapt the mesh (e.g., from mesh A to mesh B). The “evolve” step advances f forward in time by $\Delta t_{\text{adapt}} = \Delta t$.

Algorithm 1 AMR prediction by leapfrogging AMR indicator.

Given: Mesh A and function $f^{(A)}(t)$ supported on mesh A at time t ; time step length $\Delta t > 0$, number of steps $n_{\text{adapt}} \in \mathbb{N}_+$ for adaptivity intervals, and number of steps $n_{\text{pred}} \in \mathbb{N}_+$ for prediction intervals.

- 1: **while** t has not reached final time **do**
- 2: $\Delta t_{\text{pred}} = n_{\text{pred}} \Delta t$
- 3: $\Delta t_{\text{adapt}} = n_{\text{adapt}} \Delta t$
- 4: compute indicator $\chi^{(A)}(t)$ from function $f^{(A)}(t)$ on mesh A
- 5: leapfrog $\chi^{(A)}(t)$ until $\chi^{(A)}(t + \Delta t_{\text{pred}})$ ▷ (“fast” scheme)
- 6: reduce $\chi^{(A)}$ to predictive indicator $\chi_{\text{pred}}^{(A)}(t) = \max_{\tau \in [t, t + \Delta t_{\text{pred}}]} \chi^{(A)}(\tau)$
- 7: create mesh B by adapting mesh A based on $\chi_{\text{pred}}^{(A)}(t)$ ▷ (AMR)
- 8: interpolate $f^{(A)}(t)$ to $f^{(B)}(t)$ onto mesh B
- 9: evolve $f^{(B)}(t)$ until $f^{(B)}(t + \Delta t_{\text{adapt}})$ ▷ (“accurate” scheme)
- 10: update time step length Δt ▷ (adaptive time stepping)
- 11: **end while**

the computational complexity of the overall simulation with dynamic AMR and AMR prediction. In the context of the governing equations we are aiming to solve (Section 2), we are dealing with an advection–diffusion PDE. The PDE will require implicit time-stepping schemes, which have a significantly higher computational complexity when compared with explicit schemes. The leapfrogging of the indicators χ , on the other hand, has to be performed only locally in time: for short time intervals and with a new initial condition at each invocation of AMR prediction. Therefore we can relax the requirement for accuracy, and we can approximate local short-time behavior by eliminating the diffusion term of the PDE. Hence, the equation for AMR prediction will be an advection-only PDE, and we will use the same advection coefficient as for the governing equations of the physical models (see Section 2).

In Section 6.2 we discuss the increase in accuracy of the numerical solution while keeping the computational costs low, which is achieved due to AMR prediction.

4. Relativistic electron drift-kinetic solver based on PETSc and p4est

DMBF (Data Management for Block-structured Forest-of-trees) is a new data management object in PETSc that we have developed and implemented along with a new relativistic electron drift-kinetic solver. We first outline the capabilities of the implementation that are incorporated into PETSc; subsequently, we give a high-level description of the new application code for relativistic electrons.

4.1. DMBF: new PETSc data management for p4est-based AMR solvers

Our overall goal is to discretize a computational domain with locally adaptively refined quadrilateral meshes in two dimensions. In particular for relativistic electrons, extreme local refinement is critical for resolving the rapidly changing solution, which can vary 15–20 orders of magnitude, while over regions where the solution is relatively flat, coarser meshes can help reduce the computational cost. Therefore, a key requirement of our solver is adaptive mesh refinement.

The dynamic mesh adaptivity in parallel is, at its lowest level, enabled by the p4est library [21,39]. This library implements hierarchically refined quadrilateral and hexahedral meshes utilizing forest-of-octree algorithms and space-filling curves [40]. Mesh refinement and coarsening are efficient in parallel because p4est performs these tasks locally on each compute core without communication. The representation of the mesh as a quadtree/octree topology enables efficient 2:1 mesh balancing in parallel as well as repartitioning of the mesh cells across compute cores, for both of which communication is necessary. Space-filling curves transform a two- or three-dimensional space that is subdivided in cells into a (one-dimensional) sequence of cells. This sequence is used for an efficient partitioning of mesh cells in parallel with the desirable property of keeping spatially neighboring cells nearby each other in the sequence; hence exhibiting memory locality. This property of space-filling curves is responsible for keeping the amount of communication low. Overall, the algorithms of p4est have demonstrated scalability up to $O(100,000)$ of processes on distributed-memory CPU-based systems [41,42]; moreover, scalability to $O(1,000,000)$ of CPU cores have been achieved for complex implicit PDE solvers [43–45]. Generally, tree-based AMR algorithms have shown impressive scalability results [46,47].

DMBF is a new type of data management in PETSc. It provides interfaces for block-structured forest-of-trees meshes. The block structure is understood in the sense that leaf elements of the trees can be equipped with blocks of uniformly refined cells. DMBF implements PETSc functions that give direct access to the functionality of p4est with the least amount of overhead compared with existing PETSc data management approaches. In addition, DMBF provides cellwise data management for storing any kind of data that is mapped to the cells of the p4est mesh. Because DMBF associates cellwise data directly with the p4est mesh, the cell data managed by DMBF is created and destroyed when p4est cells are refined and coarsened; and DMBF-managed cell data migrates across processes in parallel as the p4est mesh is repartitioned. In the context of discretizations of PDEs, it is essential that data from neighboring cells be shared, which in a distributed-memory setting is handled via point-to-point communication between processes in order to communicate data of a ghost layer. DMBF supports the communication of DMBF-managed cell data within a ghost layer.

The focus of DMBF in managing general data objects enables scientific applications that use the DMBF interface to store any kind of cell-dependent data, for instance, the data required for finite element, finite volume, or finite difference discretizations. The shape and size of the cell data are left entirely up to the application using DMBF and are not imposed by DMBF's interface. This approach allows for greater flexibility, because more applications can utilize DMBF. At the same time, it requires the application to implement its own discretization routines (or rely on other libraries). We have taken this approach of separating concerns. We have implemented a new relativistic electron drift-kinetic solver that is supported by the DMBF interfaces (see Section 4.2).

To perform computations with the DMBF-managed cell data, the DMBF interface provides iterator functions that call a user-specified function for each cell. Additional functions are capable of iterating over the edges/faces of cells. The adaptation of a mesh is handled with similar user-provided functions that act on the data of each cell individually. The DMBF interface is completed by allowing the generation of PETSc vectors and matrices. Therefore, calculations that utilize DMBF leverage the linear and nonlinear solvers of PETSc as well as PETSc's time-stepping functionality.

4.2. Relativistic electron drift-kinetic solver based on DMBF

This section gives a brief overview of the relativistic drift-kinetic Fokker–Planck solver that is a new application code building on the DMBF data structure.

The spatial discretization chooses a finite volume scheme, MUSCL (Monotonic Upstream-centered Scheme for Conservation Law) [20,21], or a conservative finite difference scheme, QUICK (Quadratic Upstream Interpolation for Convective Kinematics) [22], for the advection operators. Both schemes are second-order accurate in space on a uniform mesh for time-dependent advection problems. The collision operator is discretized by a central scheme using a three-point stencil that also gives second-order accuracy. Note that the second-order central scheme for the collision operator can be interpreted as finite difference or finite volume [22]. We support both finite difference and finite volume schemes through associating corresponding degrees of freedom to each cell. An adaptive quadtree-based quadrilateral mesh is used. Thus our solver supports the handling of hanging nodes and can be nonsymmetric in a single cell. More details on the AMR and the implementation can be found in Section 4.1.

Numerical solutions throughout the current paper are supported on nodes at cell centers. Each cell of the adapted mesh is further subdivided into four uniform cells, which constitute the degrees of freedom (DOFs) of a solution's discretization. To distinguish between the two notions of cells, we use the term *mesh cells*, denoting the cells of the adapted p4est-based mesh, and we use the term *cells* for the four uniformly refined cells inside a mesh cell. Additional to the cells inside each mesh cell, the numerical schemes need to access DOFs from neighboring cells. To this end, the DMBF cell data includes a layer—also referred to as *guard layer*—of two cells around a mesh cell. This outer layer of DOFs requires point-to-point communication, if neighboring cells reside on different parallel processes.

The size of the guard layer is determined from the required DOFs of a numerical scheme. Here, we employ the second-order QUICK and MUSCL schemes, which need DOFs from up to two neighboring cells. Hence, each mesh cell has a guard layer of two (discretization) cells in each of the horizontal p - and vertical ξ -directions. Since we choose the adaptive mesh to be block-structured by incorporating four cells of uniform refinement into each mesh cell, the parallel communication of ghost cells is feasible with

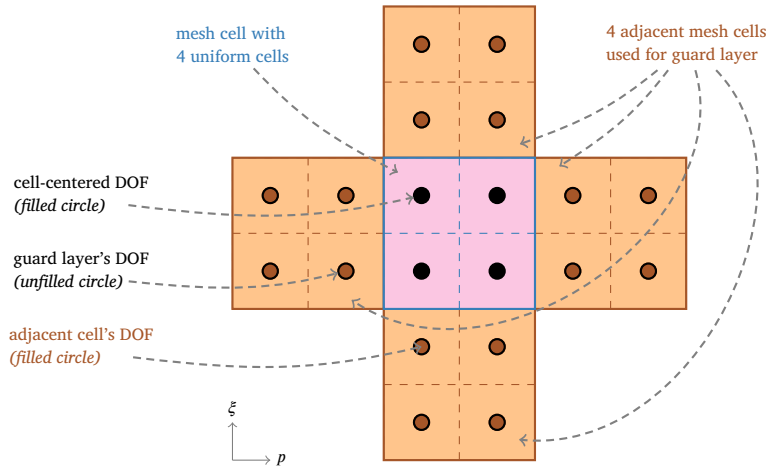


Fig. 3. A mesh cell (blue square) is subdivided into four cells with one cell-centered DOF (back filled circle). The guard layer of this mesh cell is comprised of four adjacent mesh cells (brown squares), if the neighboring cells have the same level of refinement. Filling the guard layer's DOFs (black unfilled circles) of the (blue) mesh cell is done by copying the DOFs from adjacent cells (brown filled circles) in the case of uniform level of refinement.

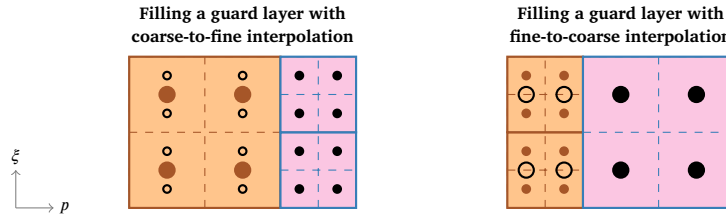


Fig. 4. Filling a guard layer's DOFs (black unfilled circles) in the case of adaptive refinement requires interpolating the DOFs of the adjacent mesh cell(s) (brown filled circles) to the locations at the guard layer's DOFs, while the latter are positioned to resemble a stretched, uniformly refined grid for the DOF of the mesh cell(s) (black filled circles).

standard point-to-point communication of adjacent neighbors of mesh cells (as opposed to needing ghost layers over two mesh cells). Filling a guard layer, if the neighboring mesh cells have the same level of refinement, is straightforward: assuming a ghost layer communication has taken place, the values of the DOFs from neighbors can be copied into the guard layer of a mesh cell; this is depicted in Fig. 3. When the level of refinement differs, then it will differ only by one level, because we enforce 2:1-balancing supported by p4est. This leaves two cases to be discussed: when a guard layer needs to be filled with DOFs from one coarser mesh cell or from two mesh cells of one level finer. We will present the example of the left edge of a mesh cell, while all other edges are treated analogously.

First, the interpolation from one coarse mesh cell to the guard layers of two fine mesh cells involves the DOFs shown in Fig. 4, left, where the figure depicts the coarse-fine interface along the left edge of two fine mesh cells. The guard layers' (fine) DOFs have coordinates that are horizontally aligned with the coarse DOFs; this allows to interpolate only along the vertical ξ -dimension. We employ linear projections, because they are more stable than higher-order projections with respect to extrapolation, which affects the fine DOFs at the top and bottom of the guard layer (see Fig. 4, left); and we enforce nonnegativity after projection. Second, the interpolation from two fine mesh cells to the guard layer of one coarse mesh cell is depicted in Fig. 4, right, for the fine-coarse interface along the left edge of a mesh cell. The guard layer's (coarse) DOFs are horizontally aligned with the fine DOFs (i.e., similarly to the previously presented fine guard layers). A linear interpolation only along the ξ -direction is necessary; and no extrapolation is performed, because coarse DOFs will always be surrounded by fine DOFs.

The spatial discretization is coupled with explicit or implicit time stepping. To ease the mesh adaptivity, we choose single-step integrators. For the explicit case a SSP Runge-Kutta is chosen, while for the implicit case a fully implicit DIRK or ESDIRK time integrator is chosen. Here the explicit integrators are chosen only for testing purposes. All the production runs use implicit integrators because we are interested in physics of the collisional time scale.

Depending on the choice of the advection scheme, the problem can be nonlinear (in the case of MUSCL) or linear (in the case of QUICK). For nonlinear problems, the JFNK algorithm is the primary solver, and it is further preconditioned with an approximate Jacobian. The Jacobian is computed through finite difference coloring provided by PETSc. A GMRES iterative solver is used to invert the linearized system (and also the linear system arising from QUICK). A hypre [30] algebraic multigrid (BoomerAMG [48]) solver is used as the preconditioner. Those choices of the algorithm are state of the art for large-scale nonlinear simulations.

Two commonly used operators on adaptive meshes are interpolations and integrations, both of which need to support handling of hanging nodes/edges. For interpolation from coarse to fine and reverse, we use tensor product interpolation matrices because of their lower memory usage per generated Flops. When needed, a postprocessing step is performed after interpolation to preserve

the positivity and mass conservation of the algorithm. Computations of fluxes across hanging edges require tailored algorithms for filling the guard cells; these are derived from the interpolation algorithms and inherit their properties (tensor product formulation, positivity, and mass conservation). A line integral along ξ direction is performed to compute runaway electrons. A simple quadrature rule is used in each cell, and the positivity can be enforced at the price of reducing the spatial accuracy to first-order.

5. Experimental setup

This section focuses on the derivation of manufactured solutions, the setup of numerical experiments of relativistic electron drift-kinetic simulations, and the experimental hardware and software environment. The goal of this section is to ensure the reproducibility of the work and to introduce the computational studies in the numerical results of Section 6.

5.1. Manufactured solution

To verify the accuracy of our implementation, we rely on the method of manufactured solutions. Several exact solutions are constructed for the RFP equation (1). Consider the original RFP model for the distribution function f ,

$$\frac{\partial f}{\partial t} - E\xi \frac{\partial f}{\partial p} - E \frac{1-\xi^2}{p} \frac{\partial f}{\partial \xi} = C(f), \quad (20)$$

where $C(f)$ represents a general collision operator. To determine an exact solution, we convert the RFP equation back to the coordinates of $(p_{\parallel}, \mathbf{p}_{\perp})$, giving

$$\frac{\partial f}{\partial t} - E \frac{\partial f}{\partial p_{\parallel}} = C(f). \quad (21)$$

Note that without the collision operator the exact solution to (21) satisfies the following identity between $t = T$ and $t = 0$:

$$f(T, p_{\parallel}, \mathbf{p}_{\perp}) = f(0, p_{\parallel} + ET, \mathbf{p}_{\perp}). \quad (22)$$

Therefore, exact solutions can be constructed by following characteristics. For instance, the exact solutions to the pure advection equation can be

- exponential solutions, such as

$$\begin{aligned} f(t, p, \xi) &= \exp \left[-p_{\perp}^2 - (p_{\parallel} + Et)^2 \right] \\ &= \exp \left[-p^2 - 2p\xi Et - (Et)^2 \right]. \end{aligned}$$

(we note that this exact solution decays properly at $p_{\max}$), or

- sinusoidal solutions, such as

$$f(t, p, \xi) = \sin(p\xi + Et), \quad \text{or} \quad f(t, p, \xi) = \cos(p\xi + Et)^2.$$

Now we consider the RFP equation with a simplified collision operator:

$$\frac{\partial f}{\partial t} - E\xi \frac{\partial f}{\partial p} - E \frac{1-\xi^2}{p} \frac{\partial f}{\partial \xi} = \frac{\epsilon}{p^2} \frac{\partial}{\partial p} \left[p^2 \frac{\partial f}{\partial p} \right] + \frac{\epsilon}{p^2} \frac{\partial}{\partial \xi} \left[(1-\xi^2) \frac{\partial f}{\partial \xi} \right].$$

One nontrivial exact solution can be derived as

$$f(t, p, \xi) = \sin(p\xi + Et) \exp(-\epsilon t) \quad (23)$$

through separation of variables [49]. Note that this collision operator corresponds to a special case of $C_F = 0$, $C_A = C_B = \epsilon$ in the general Fokker-Planck collision operator. These exact solutions can be used to verify numerical solutions with or without collision operators.

Boundary conditions associated with the manufactured solutions For practical problems, the RFP equation is discretized in the domain of $[p_{\min}, p_{\max}] \times [-1, 1]$ with $p_{\min} > 0$ to avoid the singularity of the factor $1/p$. We note, however, that the cited exact solutions are valid in the entire domain of $[0, +\infty) \times [-1, 1]$. Therefore, the time-dependent Dirichlet boundary condition is used in the p direction at both $p_{\min}$ and $p_{\max}$. Because of the coefficients involving the term $1-\xi^2$, no boundary conditions are needed at $\xi = \pm 1$. Alternatively, one can enforce a Dirichlet boundary condition at $\xi = \pm 1$ for testing purposes.

5.2. Setup of relativistic electron drift-kinetic simulations

The RFP model considered in the current work has rich physics that are often used to describe runaway electrons, which are one of the major sources of tokamak disruptions. The current work focuses on examining the capabilities and performance of the

proposed RFP solver; further advancement of physics studies will be carried out in follow-up work. This section lays out details of the RFP model in order to introduce the numerical results in the next section. It is of interest to numerically analyze the linear and nonlinear solvers in practical physics regimes, examine the efficiency gained by AMR, and demonstrate the parallel scalability of RFP simulations.

The RFP model is considered in a domain of $[0.3, 60] \times [-1, 1]$ for $[p_{\min}, p_{\max}] \times [\xi_{\min}, \xi_{\max}]$. The model's dynamics are dominated primarily by the interplay of the electric field, the damping term, and the Fokker–Planck collision operator. We aim to demonstrate in which regime the solvers and the preconditioners converge robustly. It is well known that an algebraic multigrid preconditioner deteriorates in efficacy when the advection–diffusion operator becomes increasingly nonsymmetric [50]. In the RFP model, this is possible when the electric field E becomes large. On the other hand, there is a well-known threshold that E has to surpass for a runaway avalanche to happen [51]. The solver is desired to be robust at or above this threshold for E .

Additionally, we are concerned with the numerical performance in the presence of the radiation damping term, $R(f)$ in (1). This term is close to a friction force, which is effectively another advection term, and it mainly impacts the high-energy region (large p). This operator interplays with the electric field in the moderately high energy region, and thus it is critical for the forming of the runaway tail. We are also interested in determining the range of α , which is the scaling factor of $R(f)$ in (1), where the solver exhibits a robust performance.

Another important value of interest is the runaway electron population along the p -direction. This is computed through a line integration of the distribution along the ξ direction with a metric that will be defined explicitly in the numerical section. Note that computing a line integral over an adaptively refined mesh becomes a nontrivial task. In this work we focus only on the time evolution of this important population. One immediate next goal would be to couple this population with the field solver so that it leads to a self-consistent model that is capable of describing the impact of the runaway electrons on the tokamak field.

Furthermore, we are interested in evolving the runaway electron distribution under the impact of the secondary knock-on source as described in Section 2.3. Recall that we have shown that the dominant term of the Chiu knock-on source (5) is nonzero in a very thin region. The major difficulty in evaluating (5) is thus twofold: (i) the thin region needs to be sufficiently well resolved; and (ii) the evaluation of the source terms requires another line integral along ξ direction, and in a parallel context the resulting integral needs to be communicated to each process participating in the distributed parallel simulation. To address (i), we rely on AMR to resolve such a region dynamically in time. Therefore, this case demonstrates another aspect for the need of an adaptive solver. To address (ii), we design a positivity-preserving interpolation approach for performing the line integral on an octree-based adaptively refined mesh, and we implement MPI communication routines to distribute the integral information across processes.

Unless otherwise noted, adaptivity criteria for mesh refinement and coarsening are carried out with AMR indicator prediction as proposed in Section 3.2. Other details of AMR will be described in each numerical example.

5.3. Hardware and software environments

The hardware and software environments are presented in this section for the purpose of reproducibility of numerical experiments. The development of the solvers and the studies of the algorithmic performance were carried out on three different HPC platforms: Argonne National Laboratory's Bebop cluster, Los Alamos National Laboratory's Chicoma system, and the Cori system at the National Energy Research Scientific Computing Center. The computations pertaining to performance, specifically to demonstrate the parallel scalability of the solvers, were carried out on the Frontera system at the Texas Advanced Computing Center (TACC). Table 1 lists the hardware specifications of Frontera. The software environment consists of Intel C/C++ compilers, the Intel MPI (Message Passing Interface) library, and Intel MKL, as well as the libraries PETSc, p4est, and hypre. The versions of the software utilized for parallel scalability studies are listed at the bottom of Table 1. The PETSc library has been extended with additional code that implements the data management functionalities for adaptive meshes based on p4est. This work relies on these extensions (currently residing in a public branch of the PETSc repository²), and they will be integrated into a future release version of PETSc.

6. Numerical results

Numerical experiments were performed in order to demonstrate the convergence of the AMR-based numerical schemes and solvers, the efficiency gained by AMR with prediction, the robustness of the solvers for practical physics regimes, and the algorithmic as well as parallel scalability of RFP simulations.

6.1. Manufactured solution results: convergence of algorithms

This section employs the manufactured solution derived in Section 5.1 to demonstrate a convergence analysis of the overall solver, including the numerical scheme and handling of the finite volume discretization at hanging faces of neighboring cells with different levels of refinement.

The manufactured solution is approximated in the domain $[0.3, 60] \times [-1, 1]$ for $[p_{\min}, p_{\max}] \times [\xi_{\min}, \xi_{\max}]$. The parameter $E = 0.5$ is used in the model (23); the simulation of the manufactured solution takes place in the time interval $t = 0, \dots, 10$; and the MUSCL scheme is used. We utilize the 3rd-order explicit Runge–Kutta time integrator of PETSc with a constant time step length (varying

² Link to DMBF code (branch in PETSc repository): <https://gitlab.com/petsc/petsc/-/tree/johann/jcp2023/>.

Table 1

Hardware specifications of the TACC Frontera system based on the CPU Intel Xeon Platinum 8280 (“Cascade Lake”) and the Mellanox HDR interconnect technology. The bottom rows list the software environment utilized for our parallel scalability studies.

Racks (total)	101
Nodes (total)	8,368
Cores (total)	468,608
CPU cores per node	2
CPU cores per node	56
Hardware threads per core	1
CPU clock rate (nominal)	2.7 GHz
Memory per node	192 GB
C/C++ compiler	Intel 19.1.1
MPI library	Intel MPI 19.0.9
BLAS library	Intel MKL 2020.1
PETSc base version	3.16.4
hypre version	2.23.0
p4est version	2.0.7

Table 2

Numerical results pertaining to the manufactured solution (23), where rows show simulations with increasing levels of refinement. An order of convergence of ~ 2 is asymptotically reached.

Number of cells (per refinement level)								#Cells total	Δt	Time steps	Refine freq.	Error	Conv. order
level 2	level 3	level 4	level 5	level 6	level 7	level 8	level 9						
8	324	1,648	–	–	–	–	–	1,980	0.08	125	16	6.90e-1	–
–	164	1,304	4,448	–	–	–	–	5,916	0.04	250	32	4.55e-1	0.60
–	–	1,296	3,520	14,336	–	–	–	19,152	0.01	1,000	128	2.90e-1	0.65
–	–	–	7,268	8,056	48,096	–	–	63,420	0.0025	4,000	512	9.92e-2	1.55
–	–	–	–	33,832	20,452	163,312	–	217,596	0.000625	16,000	2048	2.80e-2	1.82
–	–	–	–	–	146,276	57,632	574,784	778,692	0.00015625	64,000	8192	7.05e-3	1.99

sizes determined by spatial refinement). The range of levels of mesh refinement, in between which our AMR refinement criteria from Section 3.2 are adapting the mesh dynamically, is varied. We prescribe the range of levels to be three and shift the minimum and maximum permitted levels up by one in each simulation setup. Consequently, the coarsest setup has refinement levels (2, 3, 4); the next finer setup has levels (3, 4, 5), etc.; and finally we reach levels (7, 8, 9). Table 2 lists the number of cells that are generated by the AMR algorithm with prediction (see Section 3.2) at each refinement level.

The increasingly finer mesh resolution requires a shorter time step length that is selected to satisfy the CFL condition; see column Δt in Table 2. Because of the changing time step length, we also are adjusting the frequency of time steps at which mesh refinement is performed; the corresponding column in Table 2 is called *Refine freq.* The refinement frequency determines after how many time steps the mesh is updated to track the dynamically changing solution. We increase the refinement frequency proportionally to the number of time steps, hence the total number of AMR operations stays constant across all experiments.

The errors of the numerical solution relative to the known exact solution (23) and the order of convergence of these errors are listed in the two rightmost columns of Table 2. The error initially decreases with an order that is around 0.6 at coarser levels of refinement (first three rows of Table 2). As the levels of refinement increase, the order of convergence also increases, and it reaches a value of around two. From this we observe an asymptotic order of convergence of ~ 2 for our numerical scheme with adaptive meshes. The lower order of error reduction that is present at coarser meshes is likely due to the specific function that represents the manufactured solution. A sufficiently refined mesh may be necessary to capture certain oscillations of that sinusoidal function before the desired asymptotic order of convergence can be observed.

6.2. AMR prediction results: higher accuracy with low computational overhead

This section demonstrates numerically the benefit of adding the proposed prediction capabilities to AMR as described in Section 3.2. The benefit is that the solution function is resolved with increased accuracy with a comparatively small increase in computational costs. We focus on an experimental setup where the frequency of mesh refinement (and coarsening) varies such that AMR is performed each 32, 16, 8, and 4 time step iterations, where the time step length is adapted based on a local truncation error. For brevity, we refer to these varying frequencies of mesh refinement as RF32, RF16, RF8, and RF4, respectively. These refinement frequencies are used in experiments with standard mesh refinement without prediction and AMR with prediction. The adaptive levels for mesh refinement range from 2 to 6, hence capturing a level contrast of 4, where a uniform mesh of level 2 would contain 192 cells and each additional level subdivides a parent cell into 4 child cells.

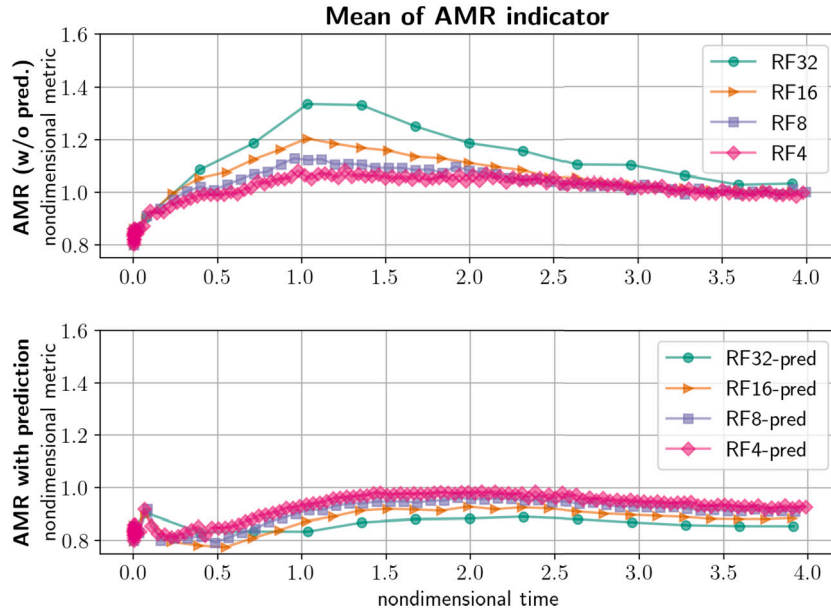


Fig. 5. Spatial mean of the AMR indicator LogDR (18) evolving in time: AMR without prediction (top) and AMR with prediction (bottom). Decreasing refinement frequencies (RF32, RF16, RF8, RF4) are lowering the LogDR metric, whereas with prediction the metric remains at low levels for any refinement frequency.

The RFP model parameters for this study are $E = 0.5$ and $\alpha = 0.1$; the time interval of the simulation is $t = 0, \dots, 4$; and the MUSCL scheme is used.

The AMR indicator LogDR (18) is a metric based on features of the numerical solution that quantifies the smoothness of the solution with nondimensional values. When the solution exhibits artifacts or oscillations arising from too coarse meshes, the metric increases to values larger than one. Therefore, we use this metric to evaluate how well the solution is resolved by a dynamically adapted mesh, and we compare the differences as the refinement frequency varies and for AMR without and with prediction.

Initially, we observe how the mean of the AMR indicator evolves in time. In Fig. 5 the top graphs correspond to standard AMR (without prediction), and the bottom graphs show AMR with prediction. The colors of each graph indicate different refinement frequencies. The top of the figure shows how more frequent mesh adaptation is resulting in a reduction of the spatial mean of the metric and how the metric is approaching unity. These results stand in contrast to the bottom of the figure, where the AMR indicators are propagated ahead of the solution, which results in the metric being significantly lower, by about 10–40%. Additionally, the curves corresponding to different refinement frequencies are more clustered together and below unity, showing that less frequent mesh adaptation must not result in poorer-resolved solutions as in the case without AMR prediction. The trend as refinement frequency increases (RF32-pred ... RF4-pred) is reversed compared with the top curves (RF32 ... RF4), because the metric associated with RF4-pred (pink color) is above the one for RF32-pred (green color). The reason is that the prediction intervals are longer for RF32-pred and therefore more cells of the mesh are refined along the predicted path of the solution.

Complementary to the evolution of the mean of the metric, Fig. 6 plots the evolution of the mesh size. This figure shows how many more mesh cells are being generated due to the prediction of AMR indicators. The increase in the number of cells is $<10\%$ when comparing RF4 and RF4-pred and can reach an increase of $\sim 20\%$ when comparing RF32 and RF32-pred. The larger mesh sizes of RF32-pred compared with RF4-pred also support the above observations about the mean of the metric (i.e., RF32-pred's metric is below RF4-pred). While the discussion has been highlighting the extreme cases RF32-pred and RF4-pred, the intermediate setups RF16-pred and RF8-pred show only moderate increases in mesh sizes (orange and purple curves in Fig. 6) while at the same time keeping the metric uniformly bounded below one (orange and purple curves in Fig. 5). This shows that a balance between mesh size, implying computational cost, and boundedness of the metric, implying a well-resolved solution, is possible in practice.

Going beyond observations of the spatial mean of the metric as it evolves in time, we also give a sense of the distribution of the metric at time instances. Fig. 7 shows the envelope of one standard deviation about the mean for each of the experiments discussed previously. The bottom of the figure shows how AMR with prediction is keeping steady control of the distribution of the AMR indicators, because the curves are staying predominantly flat. AMR without prediction, however, shown at the top of the figure, demonstrates that large values of the metric are being created for lower frequencies of refinement (see RF32, green color). To reduce the spread of values, one has to refine more frequently (see RF4, pink color), which consequently comes at an increased computational cost (see Table 3 discussed below).

The most extreme discrepancy between AMR with and without prediction is demonstrated in Fig. 8. The graphs in this figure show envelopes between the spatial mean and maximum values of the metric as it evolves in time. That is, these graphs are showing a contrast of the metric between mean and maximum (note that the spatial minimum of the metric is zero for all experiments and hence cannot be used to define a contrast). The differences between AMR with and without prediction are most pronounced in these figures, with a factor of ~ 3 difference between the top and bottom graphs of Fig. 8. Note that in Fig. 8 the initial mesh at $t = 0$ is the

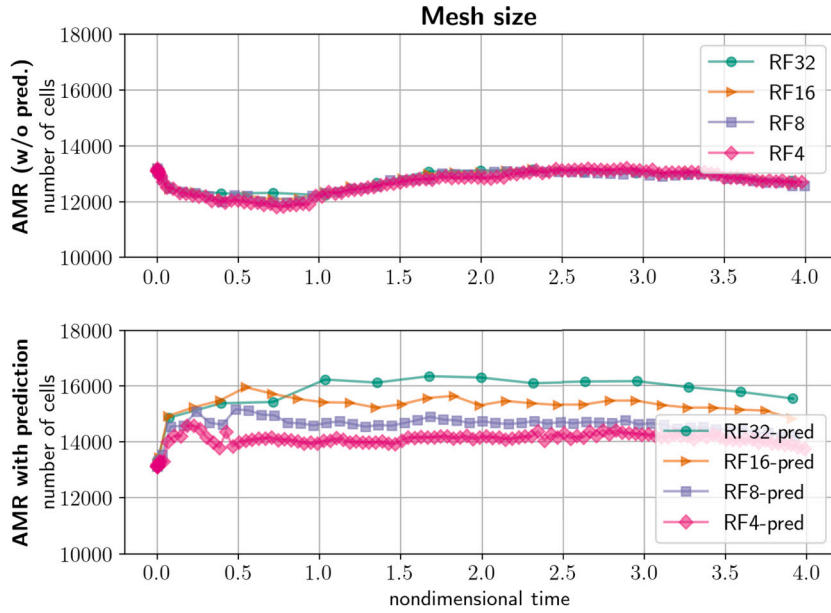


Fig. 6. Temporal evolution of the mesh size reported as number of cells. The *top* and *bottom* graphs and different *colors* correspond to the metric shown in Fig. 5.

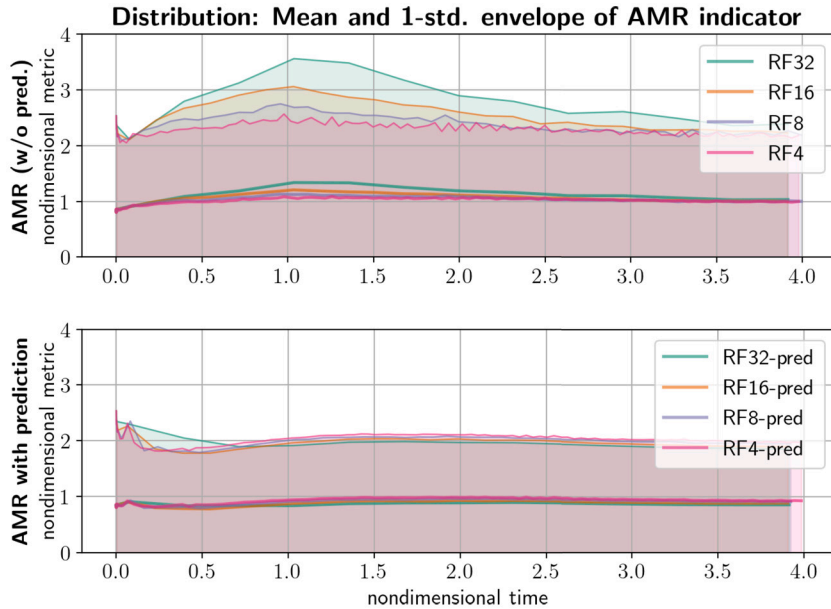


Fig. 7. Spatial distribution of the AMR indicator around the mean from Fig. 5. The distribution is shown as an envelope of one standard deviation above and below the mean curve. (Note: The vertical axis is clipped at zero because the AMR indicator is nonnegative.) AMR without prediction is shown at the *top* and AMR with prediction at the *bottom*.

same for all the cases. Thus a few mesh adaptation steps are needed until a more steady state of the maximum metric is reached, and it takes longer for RF32-pred, where AMR is performed less frequently.

We summarize the preceding observations in Table 3. This table lists summary statistics of the AMR indicator by taking the time average of the spatial mean, denoted by $\langle \text{mean} \rangle_t$, and the time averages of the standard deviation and the maximum, denoted by $\langle \text{std} \rangle_t$ and $\langle \text{max} \rangle_t$, respectively. Further, the table lists a time-averaged mesh size and the maximum mesh size over all time steps. Complementing the earlier figures, the table presents data about the computational costs in terms of the time steps used until final time (these values are closely clustered), the total number of Newton iterations across all time steps (these increase with frequency of AMR), and the total number of GMRES iterations used (these follow a trend similar to the Newton iterations). The last column of the table gives the run time of the simulation overall, which was obtained on one node with 56 CPU cores of the Frontera

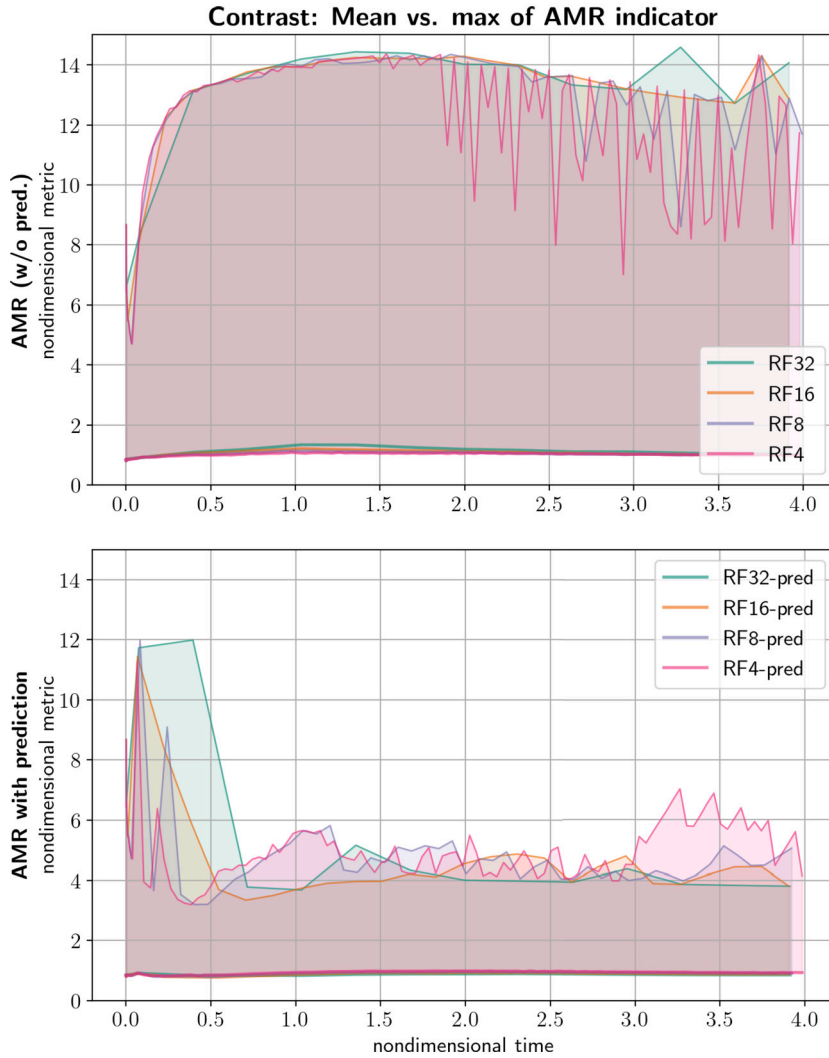


Fig. 8. Maximum value of the AMR indicator relative to the mean from Fig. 5. The maximum value is shown above the mean curve. AMR without prediction is at the top and AMR with prediction is at the bottom.

supercomputer (details in Table 1). The run times show, as expected, that higher frequencies of mesh refinement result in larger run times. Additionally the run times for AMR with prediction are slightly larger than for AMR without prediction, while AMR frequency is the same. However, we have demonstrated above that the accuracy of the solution is significantly improved by AMR prediction. The table shows that the run times are, for instance, $\sim 15\%$ larger for RF16 vs. RF16-pred (and the same for RF8 vs. RF8-pred). This is a modest increase in computational cost but a dramatic increase in accuracy as measured by the AMR indicator (shown in the second column).

We further want to illustrate the quantitative observations with qualitative figures of the numerical solution. The effects of the different AMR settings can be observed qualitatively by visualizing the solution and the associated mesh. This is done in Fig. 9, where the plots visualize the solution at time $t = 1.99$. Looking at the edges of the solution's distribution, we can clearly see artifacts for nonpredictive AMR in the left column of the figure, when the refinement lags behind the solution. Frequency RF32 (top left) shows the most severe artifacts. These artifacts can be ameliorated by faster refinement frequencies, RF32, RF16, RF8, and RF4 (top left to bottom left, respectively). On the other hand, in the right column of Fig. 9 when AMR prediction is active, the solution's features are resolved as well as or better than was the case for RF4 (bottom left) before. This result holds for all plots: RF32-pred, RF16-pred, RF8-pred, and RF4-pred (top right to bottom right, respectively). The additional mesh cells that AMR prediction generates are most pronounced for RF32-pred (top right in Fig. 9), where the prediction window is the longest. This illustrates the additional computational cost that was observed before (e.g., in Table 3). Reducing the prediction window to RF16-pred or RF8-pred results in a better balance between extra generated cells and resolving the numerical solution.

Table 3

Comparison of AMR without prediction (*top four rows*) and AMR with prediction (*bottom four rows*). The table summarizes Figs. 5–8 by reporting temporal averages of the AMR indicator and mesh sizes, where $\langle \cdot \rangle_t$ denotes a temporal average for the time interval $0.5 < t \leq 4$. The computational cost is given in terms of iterations of Newton's method, the iterations of GMRES for the linearized systems of Newton, right-hand side (RHS) evaluations, and run time.

Refinement frequency	AMR indicator $\langle \text{mean} \rangle_t; \langle \text{std} \rangle_t; \langle \text{max} \rangle_t$	Number of cells average ; max	Time steps	Newton iterations	GMRES iterations	RHS eval's	Run time [seconds]
RF32	1.16 ; 1.70 ; 13.88	12,884 ; 13,596	456	1,559	4,896	59,853	21.02
RF16	1.09 ; 1.49 ; 13.68	12,836 ; 13,596	459	1,711	5,449	68,163	23.58
RF8	1.05 ; 1.35 ; 13.32	12,760 ; 13,596	458	1,871	6,006	75,902	26.42
RF4	1.03 ; 1.28 ; 12.48	12,740 ; 13,596	454	1,934	6,262	79,222	28.12
RF32-pred	0.86 ; 1.06 ; 4.07	15,400 ; 16,356	456	1,586	4,938	62,916	26.41
RF16-pred	0.89 ; 1.07 ; 4.15	15,028 ; 15,960	457	1,630	5,158	66,400	26.24
RF8-pred	0.92 ; 1.08 ; 4.61	14,444 ; 15,168	454	1,694	5,480	69,627	32.84
RF4-pred	0.95 ; 1.09 ; 4.99	14,004 ; 14,592	454	2,034	6,661	85,296	32.84

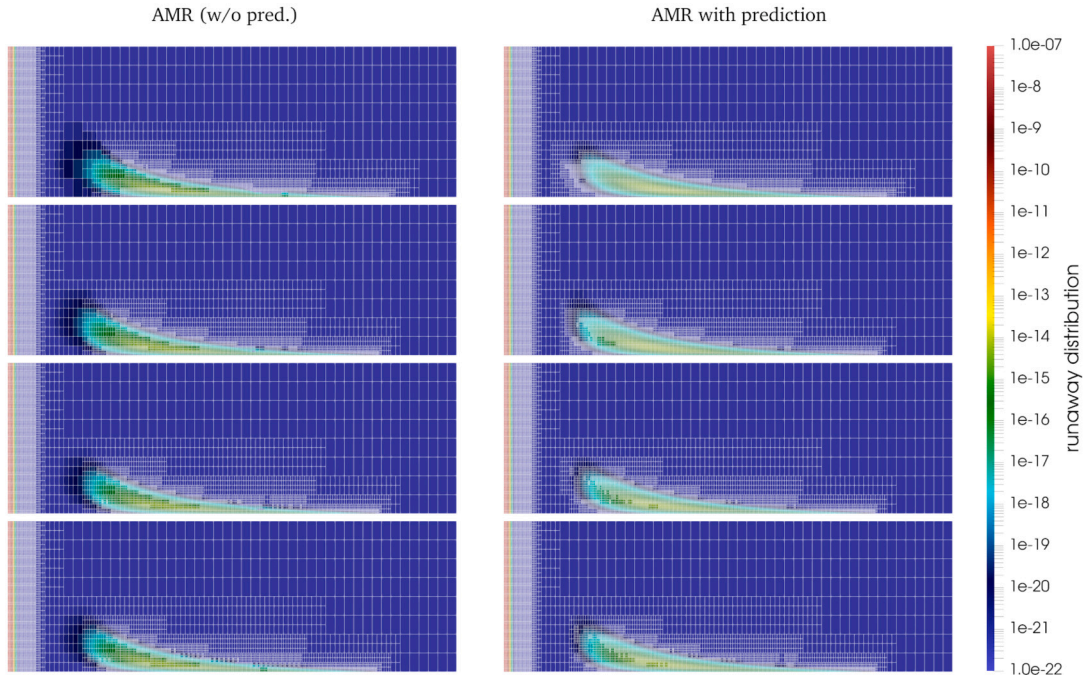


Fig. 9. Visualization of the refinement levels of the dynamically adapted mesh (white lines) at $t = 1.99$, without prediction (left column) vs. AMR with prediction (right column), while showing the numerical solution in colors. The four rows of plots correspond, from top to bottom, to RF32, RF16, RF8, and RF4.

6.3. Algorithmic robustness under different damping coefficients and electric fields

The first practical study focuses on the impact of two important coefficients, the damping coefficient α , and the electric field E . In this study, the Fokker–Planck collision is turned on as usual while the knock-on source is turned off. The runs here all use the same initial condition, which is a Maxwellian with a small perturbation in the tail region. We vary two coefficients in a range close to practice. The AMR algorithm in this example uses a base mesh of 48×8 for a domain of $[0.3, 60] \times [-1, 1]$ and a total of 7 levels of refinement.

Distribution functions at $T = 1$ with different damping coefficients are presented in Fig. 10. One can see that a larger damping term leads to the initial high-energy perturbation moving close to the low-energy region (low p). Meanwhile, Maxwellian bulks at the low-energy region are similar under different α , as expected. Note that solutions in Fig. 10 are presented in log scale, which shows the distribution function spreads more than 15 orders of magnitude.

Distribution functions at $T = 1$ with different electric fields are presented in Fig. 11. We find that a larger E field will push the entire distribution to the high-energy region, as expected. The expansion of the Maxwellian is found to be faster in a larger E case. In practice, however, the actual distribution will be a result of the interplay of many factors such as the damping force, the electric field, and the Fokker–Planck collision. Therefore, the distribution will not move all the way to the high p region but instead form an interesting fat tail in a moderate high p region. This effect will be studied more carefully later. In addition, we find that AMR can capture both the Maxwellian expansion and the movement of the tail very well (not presented here).

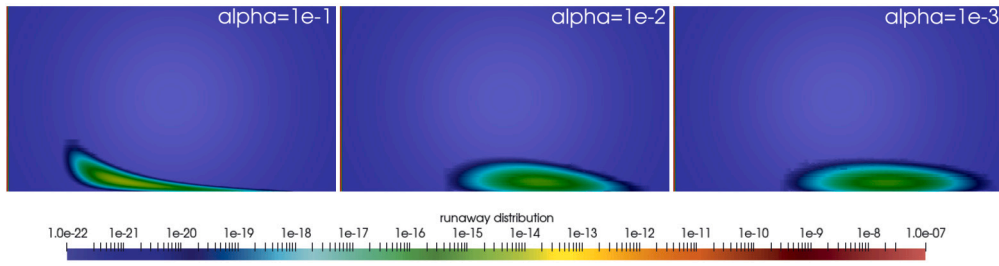


Fig. 10. Comparison of RFP solutions under different damping coefficients α . The field is fixed as $E = 0.5$, and the final time is $T = 1$. The knock-on source is turned off. Note that the solutions are presented in the log scale and spread over more than 15 orders of magnitude.

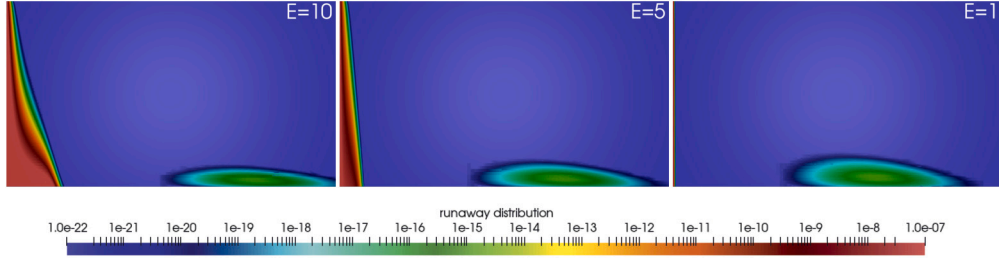


Fig. 11. Comparison of RFP solutions under different electric fields E . The damping coefficient is fixed as $\alpha = 0.01$, and the final time is $T = 1$. The knock-on source is turned off. Note that the solutions are presented in log scale and spread over more than 15 orders of magnitude.

Table 4

Comparison of RFP solver performance under different damping coefficients α . The field is fixed as $E = 2$, and the final time is $T = 1$.

α	Averaged Δt	Averaged DOFs	RHS eval./solve	GMRES it./solve
1e-1	0.00351	132358	42.85	4.49
5e-2	0.00420	120592	43.02	4.52
2e-2	0.00441	114166	43.17	4.67
1e-2	0.00457	111694	43.22	4.66
5e-3	0.00463	110459	43.02	4.71

Table 5

Comparison of RFP solver performance under different electric fields E . The damping coefficient is fixed as $\alpha = 0.1$, and the final time is $T = 1$.

E	Averaged Δt	Averaged DOFs	RHS eval./solve	GMRES it./solve
2	0.00351	132358	42.85	4.49
5	0.00244	130881	43.02	4.50
10	0.00192	131288	43.12	3.96
15	0.00183	132139	41.82	3.70
20	0.00179	132657	41.77	3.51

The second focus of this example is to study the performance of the numerical algorithms under different coefficients. Such a study for our solver is not easy because the time steps and details of adaptive meshes change dynamically throughout the simulations. (Here we use ESDIRK and estimate the time step using local truncation error estimators.) In addition, when the coefficients change, the numerical difficulty of the model also changes. To have a fair comparison, we run all the simulations with the same final time $T = 1$ and average all the important algorithm quantities over the number of time steps.

We start with testing the damping coefficients. Recall that the practical range of α is 0.001 to 0.3. We test several α and present the corresponding results in Table 4. The electric field is fixed as $E = 2$, which is above the so-called avalanche threshold (see Figure 7 in [51]). Here we use hypre's algebraic multigrid solver (BoomerAMG) as the preconditioner for the linearized system. We further choose hypre's parallel ILU solver as the multigrid smoother (Euclid) and use default values in the rest hypre options. Table 4 shows that the algorithm performs well for different α . We also note that as α increases, the averaged time step becomes smaller, and the averaged AMR degrees of freedom become larger, which all indicate that the problem becomes harder for large α . Nevertheless, the solver still performs well. We also present the averaged RHS evaluation per solve. The majority of those evaluations come from evaluating the Jacobian through finite difference coloring.

Table 6

Algorithmic scalability results show the number of GMRES iterations per time step either as a range (*min* ... *max* value) over runs with increasing core counts or as a single number (when *min* is equal to *max*). Columns show increasing problem sizes as number of cells, and rows show decreasing refinement frequencies. The number of GMRES iterations needed per solve is not increasing as the mesh becomes finer, which means the solver achieves optimal algorithmic scalability.

Refinement frequency	1.4 M cells (112 to 7,168 cores)	2.9 M cells (224 to 14,336 cores)	4.4 M cells (448 to 28,672 cores)
RF256	12.7 ... 17.1	6.0	6.0
RF128	11.2 ... 14.5	6.0	6.0
RF64	13.6 ... 14.1	6.0	6.0
RF32	13.6 ... 14.6	6.0	6.0

We then consider the impact of the electric field. A reasonable region for the normalized electric field from practice would be the interval $[0, 20]$. We test several E and present the corresponding results in Table 5. We find that the linear and nonlinear solvers all perform well. We also note that the averaged time steps become smaller for larger E , which is expected as the problem becomes stiffer.

6.4. Algorithmic and parallel scalability

This section presents the algorithmic as well as parallel scalability of the overall numerical simulations. One aspect of scalability is algorithmic scalability, which is the dependence of Newton and/or Krylov iterations on the spatial and temporal resolutions. The second aspect is parallel scalability of the implementation, which is the runtime measured on increasing numbers of compute cores. Studying both aspects is required to fully assess the performance of a solver at scale. The runtimes measuring parallel scalability include the AMR algorithms, the numerical scheme, the linear solver, and the setup of the AMG-based preconditioner. The computations are carried out on the Frontera system, a CPU-based platform detailed in Section 5.3.

In order to obtain strong scalability results, the parameters of the problem, and the problem size specifically, remain the same for each run while the number of compute cores increases by a factor of 2. The setup of weak scalability runs entails the challenge to control the increase of the problem size proportional to the number of compute cores, because the mesh is adaptively refined in a dynamic fashion depending on properties of the solution.

We generate different experiment configurations that exhibit increasing problem sizes. Each of these configurations provides a new baseline for showing strong scalability, where the baseline core counts are increasing with larger problem sizes: (i) problem size with 1.4 million (time-averaged) cells is scaled from 112 to 7,168 cores; (ii) problem size with 2.9 million (time-averaged) cells is scaled from 224 to 14,336 cores; and (iii) problem size with 4.4 million (time-averaged) cells is scaled from 448 to 28,672 cores. These three different mesh sizes feature increasingly aggressive mesh adaptivity, which is measured by the difference between finest mesh level and coarsest mesh level of refinement; we call this the *mesh level contrast*. The mesh level contrast for the above described problem sizes is (i) 8 levels contrast, (ii) 9 levels contrast, and (iii) 10 levels contrast. In addition to the different mesh configurations, we alter the frequency of mesh adaptation analogous to Section 6.2. As before, the label RF32 stands for performing mesh adaptivity after every 32 time steps, and additionally reported refinement frequencies are RF64, RF128, and RF256. The AMR indicators are predicted, as proposed in Section 3.2, for all of these frequencies.

The RFP model parameters for this study are $E = 5$ and $\alpha = 0.1$, which correspond to a realistic parameterization (see Section 6.3). The Fokker–Planck collision is turned on, while the knock-on source is turned off. To demonstrate scalability, we prescribe 512 time steps of the implicit second-order ESDIRK time integrator of PETSc with a constant time step length that decreases along with finer mesh resolutions. We use the QUICK scheme.

Table 6 demonstrates the algorithmic scalability of the overall solver. It shows along the columns the three increasing problem sizes from 1.4 million to 4.4 million cells; and it also lists the varying refinement frequencies, RF256 to RF32, in each row. The number of GMRES iterations per time step is given either as a range, to present its variation with increasing core counts, or as a single number if there are no variations. Recall that the number of time steps is prescribed and constant. Moreover, the stopping criterion for GMRES is the tolerance 10^{-6} for relative residual reduction. We can see from the table that the number of GMRES iterations per time step is around 14 for the smallest problem size (1.4 M cells), and it is six for the larger problem sizes (2.9 and 4.4 M cells). This demonstrates optimal algorithmic scalability, because the number of iterations is not increasing as the mesh is refined. The larger GMRES iteration count of ~ 14 for 1.4 M cells can be explained by slightly worse conditioning of the discretized PDE operator, which can be caused by large variations in the coefficients relative to the resolution of the mesh.

Implicit solvers with variable coefficients and discretized on adaptive meshes are well known to be challenging to scale strongly in parallel requiring significant efforts (e.g., see [43] for variable coefficient Poisson and Stokes solvers). The following two figures show strong scalability for the previously defined problem configurations: Fig. 12 corresponds to problem size (i) 1.4 million and Fig. 13 corresponds to problem size (iii) 4.4 million (we omit the presentation for problem size (ii) 2.9 million for brevity). Each figure also shows four differently colored curves, which we use to distinguish between refinement frequencies. Every figure shows the number of cores on the horizontal axis. Below the core count, the average number cells per core is reported in parentheses, which tells how large the portion of the distributed problem is for each core. Mainly because of the reduction in problem size per core, the

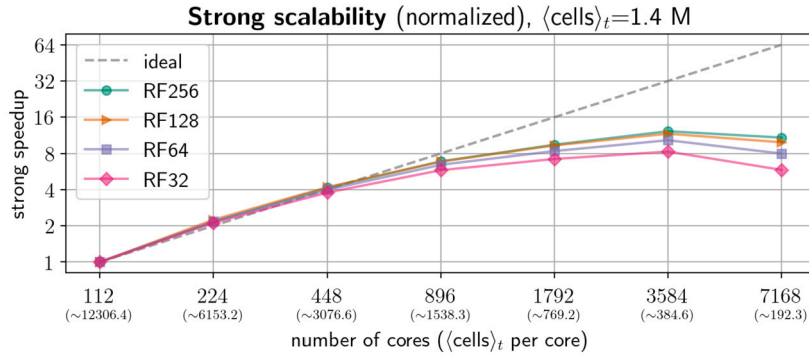


Fig. 12. Strong scalability results on Frontera from 2 to 128 nodes, where the baseline problem has 1.4 million cells and a range of 8 levels of refinement; $\Delta t = 10^{-5}$.

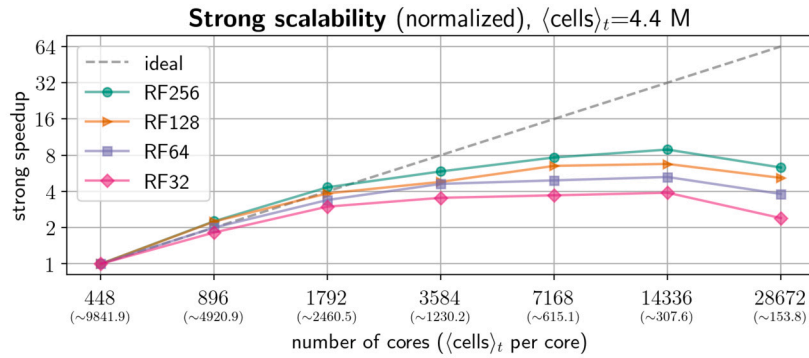


Fig. 13. Strong scalability results on Frontera from 8 to 512 nodes, where the baseline problem has 4.4 million cells and a range of 10 levels of refinement; $\Delta t = 10^{-6}$.

communication starts to dominate the simulation's run time, because the fewer degrees of freedom that reside at a compute unit's local memory, the more communication needs to be performed via MPI. The vertical axis shows the speedup in run time, which is calculated as the quotient of the baseline run time over the run time with associated core count. We report normalized speedup, where the run time is normalized with respect to the number of GMRES iterations per time step, because we document parallel scalability separately from algorithmic scalability. The speedup depicted by the gray dashed line is the idealized speedup, which, for implicit solvers, is well known to be out of reach.

Each of the Figs. 12 and 13 shows a general trend of speedup that is unavoidable when strongly scaling an implicit method: at first, the speedup stays close to the ideal speedup with initial increases in cores; this is where the run time is dominated by computations and/or communication is overlapped with computations. Speedup then gradually deviates from ideal and starts to flatten up to a certain increase in cores. After the flattening, the speedup decreases because the time used for communication dominates compared with the time used for computations. Our strong scalability results follow this general trend. Specifically, they demonstrate that the highest achievable speedup is 12.2 (green curve at 3,584 cores in Fig. 12). The highest speedup is slightly lower in Fig. 13 (green curve reaches speedup of 8). We show the scalability of different refinement frequencies because this work focuses on the AMR algorithms as well as AMR-induced overheads. The more often the mesh is refined dynamically (i.e., the lower the refinement frequency), the wider the gap is between the case with only two refinements during the whole simulation (green curve for RF256) and the frequencies RF128 (orange), RF64 (purple), and RF32 (pink). This implies that to achieve better scalability in our solver, mesh adaptation would need to happen less often. Note that as discussed in Section 4.1, the AMR algorithm based on p4est is highly scalable. The inefficiency comes from the overhead in rebuilding various operators when the mesh is updated. However, less frequent adaptation of the mesh is detrimental for the accuracy of the solution, as demonstrated in Section 6.2. These conflicting objectives need to be balanced, which was the main motivation in proposing indicator prediction for AMR, which achieves control over higher accuracies while keeping the impact on computational overheads, and hence on scalability, at a minimum. With the demonstrated improvements in accuracy from Section 6.2 due to AMR prediction, we can expect that each of the refinement frequencies in Figs. 12 and 13 resolves the solution similarly well. Therefore, one can choose the frequency that is most advantageous for scalability without sacrificing accuracy.

The strong scalability results of Fig. 13 are complemented by weak scalability results in Fig. 14. We observe a nearly ideal speedup up to 7,168 cores and see a slight deviation from the ideal for the maximum number of cores. Because of the dynamic adaptivity, the problem sizes are not perfectly proportionally increasing with core counts; see the fewer cells per core shown below the numbers of cores in brackets. This is a contributing reason that the weak scalability at 28 thousand cores performs slightly worse. The trend of fewer refinements with RF256 (green curve) performing better than the other refinement frequencies is similarly observed for weak

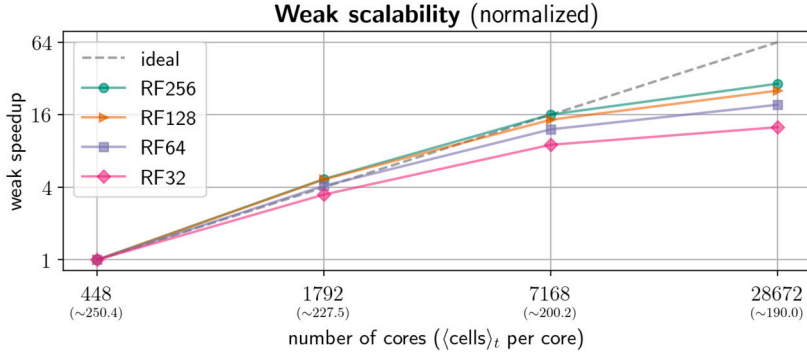


Fig. 14. Weak scalability results on Frontera from 8 to 512 nodes, where the problem sizes range from 112 thousand (at 448 cores) to 5.4 million cells (at 28,672 cores).

scalability as it was for strong scalability. Note that we also observe better than ideal speedups for 1,792 cores, which is a known phenomenon for time dependent solvers.

6.5. Benchmark with nonlinear PDE

One issue of interest is how our scalable implicit solver and dynamic AMR methods would perform for nonlinear PDEs as opposed to the linear relativistic kinetic equation shown in this paper. In the context of runaway electron dynamics in a magnetized plasma, the nonlinear coupling is through the inductive parallel electric field, which is described by a modified Ohm's law in which the plasma current density is replaced by the difference between total current density and the runaway current density. The leading order physics has a dynamically changing parallel inductive electric field that is independent of spatial position, so the nonlinear coupling via the inductive electric field is numerically straightforward, and the implementation of which is trivially simple compared with the AMR kinetic solver in momentum space. To give a more reasonable test of the AMR machinery for nonlinear problems, we design new numerical experiments and implement them within our solver codes. To this end we take a two-dimensional nonlinear convection–diffusion equation as a test problem because of its common use for numerical benchmarks.

The domain supporting the PDE is a channel of extensions $x \in [0, 6]$ horizontally and $y \in [-1, 1]$ vertically. We seek the solution $f(t, x, y)$ of the following PDE written in conservative form

$$\frac{\partial f}{\partial t} + n_x \frac{\partial}{\partial x} \left(\frac{1}{2} f^2 \right) + n_y \frac{\partial}{\partial y} \left(\frac{1}{2} f^2 \right) = \nu \frac{\partial^2 f}{\partial x^2} + \nu \frac{\partial^2 f}{\partial y^2}, \quad (24)$$

for $t > 0, x \in (0, 6), y \in (-1, 1)$, where $\mathbf{n} = (n_x, n_y)$ is a given unit vector, and $\nu > 0$ is a given diffusion coefficient. Equation (24) is complemented with Neumann boundary conditions and we propagate an initial Gaussian distribution $f(0, x, y) = \exp(-(x-5)^2/0.16) + \exp(-y^2/0.16)$ with the direction of advection $\mathbf{n} = (-1, 0)$ and viscosity $\nu = 10^{-4}$ until final time $T_{\text{final}} = 20$. A one-dimensional version of (24) was discussed in [52], and in the case $\nu = 0$, the PDE is commonly referred to as Burgers' equation. Due to the nonlinear nature of the PDE (24), the changing advective speed results in the formation of discontinuous solutions. This behavior is similar to shock waves, and the shock dissipates as the solution travels because of the viscosity.

Dynamic AMR is performed every 100th time step with prediction of adaptive refinement using a linear approximation to (24) and $\nu = 0$, such that a computationally cheap explicit scheme can be employed for propagation of AMR indicators. The overall levels of mesh refinement range from two to eight, resulting in a mesh level contrast of six. Fig. 15 presents the snapshots of the numerical solution at times $t = 0, 5, 10, 20$. We observe the formation of a shock discontinuity at the left tip of the solution. AMR is facilitating a higher resolution of the solution's sharp features as is depicted with a mesh wireframe overlaying the solution function at final time $t = 20$.

6.6. Compatible boundary condition at $p = 0$ with AMR

In this test, we provide an extension of the boundary up to $p = 0$. This study uses a primary runaway electron test, in which the initial condition is given as

$$f_0(p, \xi) = \frac{1}{\hat{v}_t^3 \pi^{\frac{3}{2}}} \exp \left(-\frac{1 - \sqrt{1 + p^2}}{\hat{v}_t^2 / 2} \right),$$

where the normalized thermal velocity is taken as $\hat{v}_t = 0.1$. The runaway electron in the tail region will grow exponentially due to the driving electric field. We select an electric field of $E = 3$, which is well above the runaway electron threshold. We extended the idea given in [53] to handle the boundary condition at $p = 0$ with AMR. Using the expansion of $f(p, \xi)$ with azimuthal symmetry, it can be derived that the compatible boundary condition for distribution at $p = 0$ should be

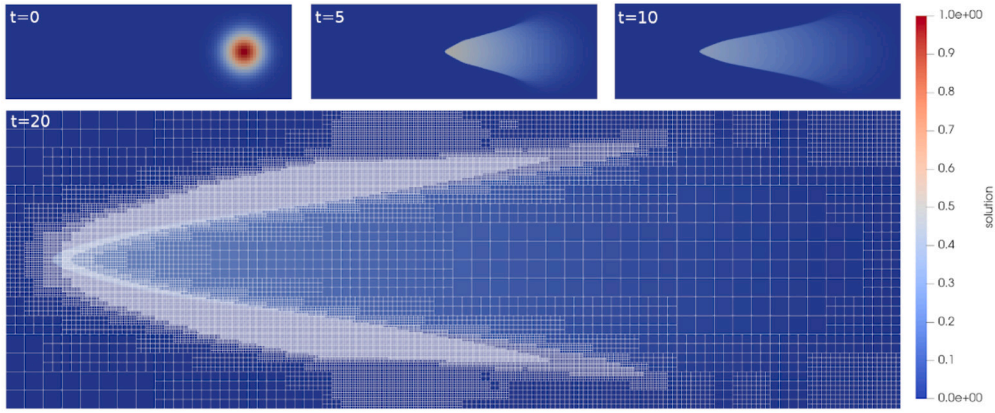


Fig. 15. The numerical solution of the nonlinear PDE (24) shown as snapshots over time. At final time $t = 20$, the solution is overlaid with the mesh wireframe to show the adaptive refinement resolving sharp features of the solution.

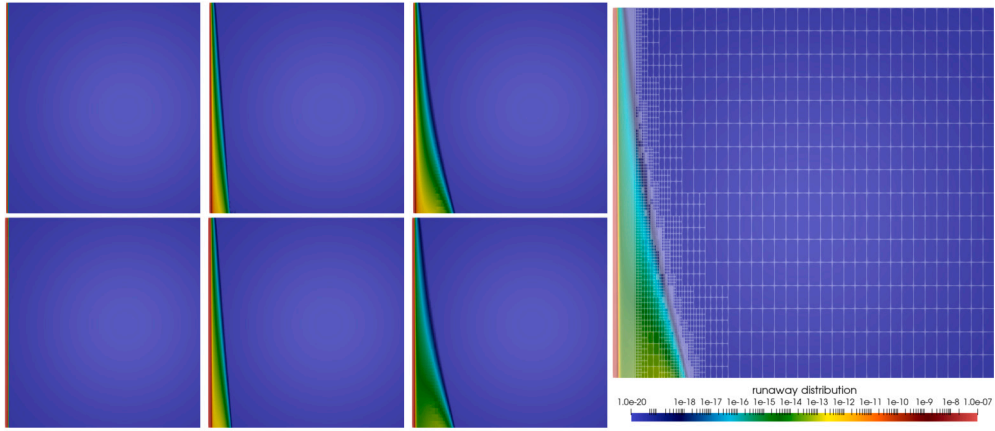


Fig. 16. Left: Distribution functions at $t = 0, 2$ and 4 . The top row uses the Dirichlet boundary condition at $p = 3\hat{v}_t$, and the bottom row uses the compatible boundary condition at $p = 0$. Right: The adaptive mesh at the final time.

$$f(-p, -\xi) = f(p, \xi). \quad (25)$$

This condition is used to fill in the ghost point at the negative p location. We utilize the connectivity of trees of the p4est library to connect two coarse quadrants in ξ -direction along the $p = 0$ quadrant face; and we also flip the orientation of this connectivity to account for the alternating sign of the ξ argument in (25). Therefore, the mesh adaptivity along the $p = 0$ boundary can be treated consistently with all other cell boundaries with varying levels of refinement. The results of two different types of boundary conditions are presented in Fig. 16. The physical locations are perfectly aligned in the figure for easy comparison. It is clear to see that the boundary layer of $p = 0$ is thicker than that of $p = 3\hat{v}_t$, but they match well starting from $p = 3\hat{v}_t$, despite different types of boundary conditions being used. It is also observed that the shapes of runaway tails match well throughout the entire run. At the final time, it is observed that the details of runaway tails are slightly different, which is due to the impact of different boundary conditions. In our forthcoming physics study, we will document the specifics of runaway growth rates in the two different setups.

6.7. Interaction between a Maxwellian and a runaway tail

We next present a physics-motivated example to demonstrate the AMR capability for resolving both of a bulk Maxwellian and a runaway distribution tail. Our aim is to study the algorithm performance in a practical setting, and thus we design a test that is still relatively simple. The problem is initialized with a Maxwellian uniformly in ξ under a small perturbation in the tail, which is centered around $(p, \xi) = (40, -0.9)$,

$$f_0(p, \xi) = \frac{1}{\hat{v}_t^3 \pi^{\frac{3}{2}}} \exp\left(-\frac{1 - \sqrt{1 + p^2}}{\hat{v}_t^2/2}\right) + 10^{-15} \exp\left(-\frac{(p - 40)^2}{25}\right) \exp\left(-\frac{-(\xi + 0.9)^2}{0.0025}\right),$$

where the normalized thermal velocity is taken as $\hat{v}_t = 0.1$. A Neumann boundary condition is applied at the right boundary, while the left boundary uses a Dirichlet boundary condition from the initial Maxwellian. The computational domain is $[0.3, 60] \times [-1, 1]$

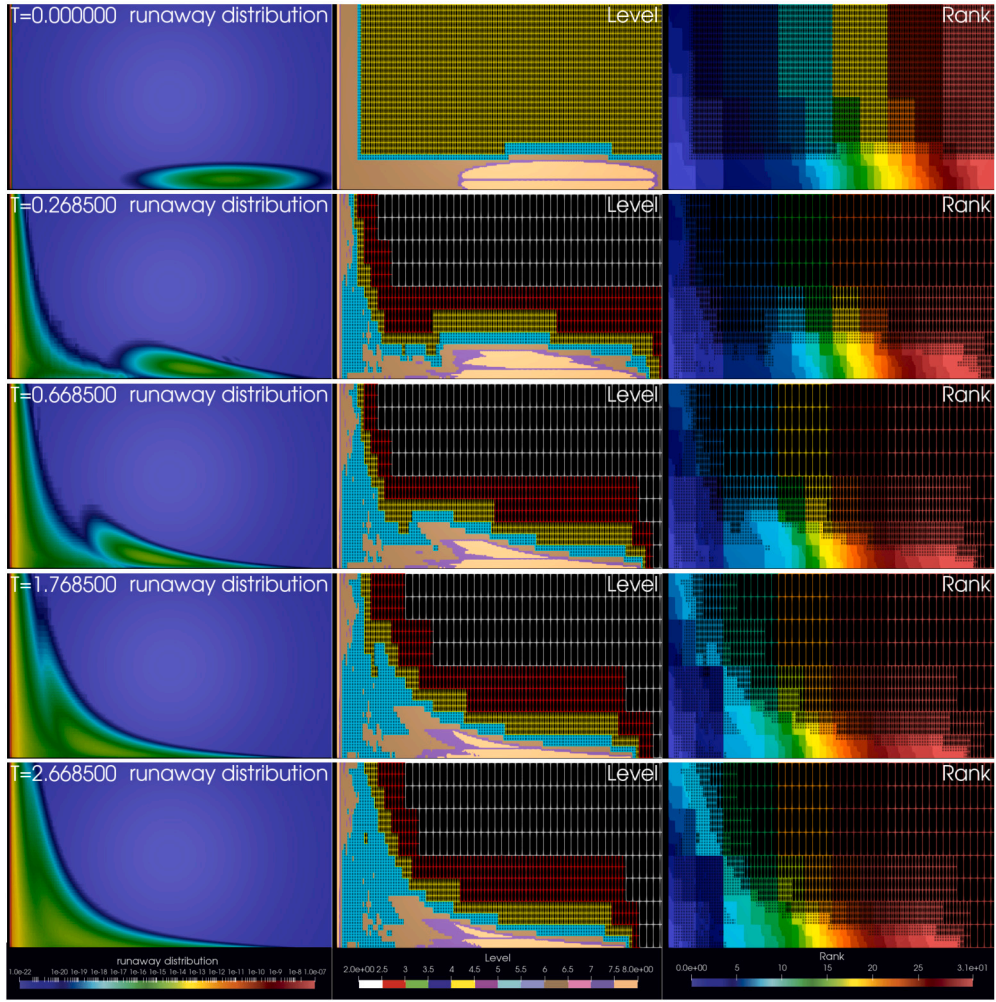


Fig. 17. Distribution functions (left), mesh and AMR levels (middle), and MPI ranks (right) over time.

as we let $p_{\min} = 3 \hat{v}_t$ for which the Dirichlet is still a good approximation. We choose a numerical scheme based on the QUICK finite difference scheme and an implicit second-order ESDIRK integrator through PETSc's TS interface.

The proposed AMR algorithm and indicators are tested using this problem. The coarsest mesh in the AMR algorithm is chosen to be 48×8 , and we use a maximum of 6 more levels of refinement. Initially, we enforce the AMR algorithm to refine around the left and bottom boundaries for better resolving the Maxwellian bulk and the large portion of the perturbation (see the top row in Fig. 17). After the initial indicator is computed from the given solution, the proposed AMR indicator prediction is used to evolve those indicator values over time. The mesh is checked and updated every 6 time steps. The averaged number of DOFs in the whole simulation is only 71,156, which amounts to a reduction in DOFs of $\sim 4.5\%$ when compared to a uniformly refined mesh with the same finest level as the adaptive mesh. The previous work [3] needed 0.62 million DOFs through a structured stretched grid to have a comparable resolution, which again highlights the computational savings obtained with AMR.

The numerical results are presented in Fig. 17. The distribution solutions along with the adaptive meshes are shown. We note that the Maxwellian bulk and the tail perturbation are slowly merging into each other due to the interaction between the driven electric field, the relativistic Fokker–Planck collision, synchrotron radiation, and the secondary knock-on source. Here we let the damping strength be $\alpha = 0.1$ and turn on the partial screening effects. We note that our AMR algorithm captures the interesting features in the solution very well. The finest mesh is around the region where it has a large gradient, which is either near the Maxwellian boundary layer or near the runaway tail. It is also critical for this test that the solver be capable of resolving a large variation of the solution. The solution varies from 10^{-2} (the bulk Maxwellian boundary layer) to 10^{-20} (the runaway tail). In order to demonstrate the large variation, the distribution function in Fig. 17 is presented in the log scale, which shows that our adaptive solver is capable of resolving both regions. To demonstrate the dynamic load-balancing, we include the MPI ranks in the figure. Note that this test is designed to be relatively simple so that 32 processors are sufficient.

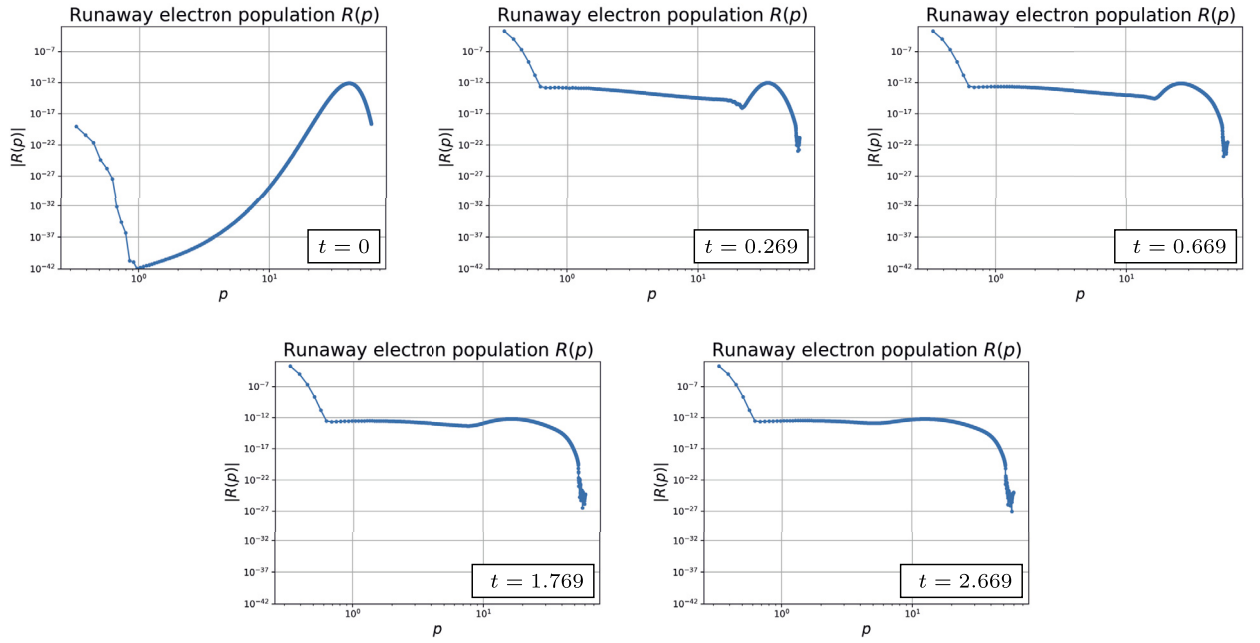
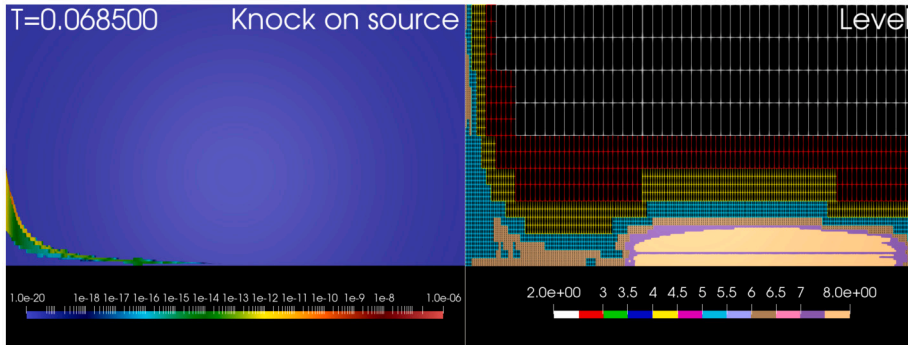
Fig. 18. Runaway electron distributions in p over time.

Fig. 19. Knock-on electron source at the first checkpoint.

In addition to the distribution, we evaluate the so-called runaway electron population $R(p, t)$,

$$R(p, t) := \int_{\xi=-1}^{\xi=1} f(p, \xi, t) v_{\parallel} 2\pi p^2 d\xi,$$

where the parallel velocity is defined as $v_{\parallel} = p\xi/\gamma m_e$. This is a distribution in the p direction to measure the runaway electron strength, which is a good indicator of the number of high-energy runaway electrons in the system. Fig. 18 shows that initially there is a large portion of runaway electrons due to the perturbation and that they are slowly reduced and merged into the bulk.

We further present the knock-on electron source term in Fig. 19. We note that the knock-on source is exactly in the range of $\xi \in [-\sqrt{\gamma/(\gamma+1)}, -p/(\gamma+1)]$ as we indicated in Section 2.3. The mesh plot of Fig. 19 shows that the AMR algorithm is able to capture the source term in this narrow region through adjusting the mesh on the fly.

7. Conclusion

In this work we designed and developed adaptive, scalable, fully implicit solvers for the relativistic Fokker-Planck equation in phase space. One key goal of the work is to develop a scalable and efficient dynamic AMR. In practice, one needs to determine the requirements on scalability and then adjust the parameters of the AMR algorithms in terms of refinement frequencies and the AMR prediction horizon. The present work proposes a new prediction strategy for refinement indicators, giving a potential tool to exert control over accuracy at computational overheads that were demonstrated to be low. Numerical experiments quantify the predictive approach behaving better than conventional indicators, which lack prediction, in terms of resolving features in the solutions. When

the need for scalability is of higher importance to the application, one can use the proposed AMR prediction at the expense of further increases of computational overheads from finer meshes.

We contribute a new scalable implementation of the proposed algorithm using the p4est and PETSc frameworks. Although the focus of the current work is to develop a solver aiming specifically at runaway electrons during tokamak disruptions, a large portion of our proposed algorithm is general and thus can be potentially applied to solving many other time-dependent PDEs with dynamic mesh adaptivity. Several numerical examples are presented in order to verify the solver's accuracy, scalability, and efficiency in adaptivity. These numerical results confirm the parallel and algorithmic scalabilities and accuracy of the schemes. A significant improvement of computational cost owing to the AMR algorithm was demonstrated using a practical disruption study.

Future work includes generalizing the fixed external field in the current model to a self-consistent model that evolves the electric field through involving the runaway current.

CRedit authorship contribution statement

Johann Rudi: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization. **Max Heldman:** Writing – review & editing, Validation, Software, Methodology, Investigation. **Emil M. Constantinescu:** Writing – review & editing, Supervision, Project administration, Conceptualization. **Qi Tang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Xian-Zhu Tang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgement

This research used resources provided by the Los Alamos National Laboratory Institutional Computing Program, which is supported by the U.S. Department of Energy's National Nuclear Security Administration under Contract No. 89233218CNA000001, and The National Energy Research Scientific Computing Center (NERSC), a U.S. Department of Energy Office of Science User Facility located at Lawrence Berkeley National Laboratory, operated under Contract No. DE-AC02-05CH11231 using NERSC award FES-ERCAP0021219.

This research was funded in part and used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract No. DE-AC02-06CH11357.

The Texas Advanced Computing Center (TACC) at The University of Texas at Austin provided HPC resources that have contributed to the research results reported within this paper.

Appendix A. Guiding center coordinate transformation

The coordinate transformation from $(\mathbf{X}, p_{\parallel}, \mathbf{p}_{\perp})$ to $(\mathbf{X}, p, \xi)$ is described in this section. The discussion focuses on the phase space. The differential element for momentum space integration is

$$d^3\mathbf{p} = dp_{\parallel} d^2\mathbf{p}_{\perp} = dp_{\parallel} (p_{\perp} dp_{\perp} d\theta). \quad (\text{A.1})$$

The guiding center model assumes azimuthal symmetry. Thus one can integrate over $\theta \in (0, 2\pi)$ (here $p_{\perp}$ is positive), giving

$$\int_{\theta} d^3\mathbf{p} = 2\pi p_{\perp} dp_{\parallel} dp_{\perp}. \quad (\text{A.2})$$

The transformation from $(p_{\parallel}, p_{\perp})$ to (p, ξ) coordinate is given by

$$p = \sqrt{p_{\parallel}^2 + p_{\perp}^2} \quad (\text{A.3})$$

$$\xi = \frac{p_{\parallel}}{p} = \frac{p_{\parallel}}{\sqrt{p_{\parallel}^2 + p_{\perp}^2}}. \quad (\text{A.4})$$

The Jacobian of this transformation is

Table B.7
Physical quantities in the RFP model and the notations used in the current work.

Description	Symbol	Description	Symbol
time	t	electron charge	e
configuration space	$\mathbf{x}$	electron mass	m_e
phase space	$\mathbf{p}$	speed of light	c
phase space parallel to $\mathbf{B}$	$p_{\parallel}$	Lorentz factor	γ
phase space perpendicular to $\mathbf{B}$	$p_{\perp}$	electrical permittivity	ϵ_0
magnitude of the momentum	$p = \ \mathbf{p}\ $	Coulomb logarithm	$\ln \Lambda$
pitch angle	$\xi = p_{\parallel}/p$	thermal velocity	v_i
runaway electron density	f	effective charge number	Z
energy flux	Γ_p	collision time scale	τ_c
pitch angle flux	Γ_{ξ}	synchrotron time scale	τ_s
magnetic field	$\mathbf{B}$	intensity of damping	$\alpha = \tau_c/\tau_s$
induced electric field parallel to $\mathbf{B}$	E	Møller cross section	$d\sigma/dp$

$$\frac{\partial(p_{\parallel}, p_{\perp})}{\partial(p, \xi)} = \left| \frac{\partial p_{\parallel}}{\partial p} \frac{\partial p_{\perp}}{\partial \xi} - \frac{\partial p_{\perp}}{\partial \xi} \frac{\partial p_{\parallel}}{\partial p} \right| = \left| -\xi \frac{p_{\xi}}{\sqrt{1-\xi^2}} - p\sqrt{1-\xi^2} \right| = \frac{p}{\sqrt{1-\xi^2}}. \quad (\text{A.5})$$

One then has

$$\int_{\theta} d^3\mathbf{p} = 2\pi p_{\perp} dp_{\parallel} dp_{\perp} = 2\pi p \sqrt{1-\xi^2} \frac{p}{\sqrt{1-\xi^2}} dp d\xi = 2\pi p^2 dp d\xi. \quad (\text{A.6})$$

Noting that $\xi \in (-1, 1)$, the integration over the entire momentum space is simply $(4\pi/3)p^3$, as expected. This also indicates that the overall Jacobian of (p, ξ, θ) coordinates is p^2 , for gyro-angle independent problems, and the overall Jacobian for the (p, ξ) coordinate is $2\pi p^2$. For convenience, we will be using $J = p^2$ for the (p, ξ) coordinates.

Appendix B. Summary of physical quantities in RFP

Table B.7 summarizes some physical quantities and the corresponding notations in this paper.

References

- [1] A.H. Boozer, Theory of runaway electrons in ITER: equations, important parameters, and implications for mitigation, *Phys. Plasmas* 22 (3) (2015) 032504.
- [2] B.N. Breizman, P. Aleynikov, E.M. Hollmann, M. Lehnen, Physics of runaway electrons in tokamaks, *Nucl. Fusion* 59 (8) (2019) 083001.
- [3] Z. Guo, C.J. McDevitt, X.-Z. Tang, Phase-space dynamics of runaway electrons in magnetic fields, *Plasma Phys. Control. Fusion* 59 (4) (2017) 044003.
- [4] A. Stahl, M. Landreman, O. Embr us, T. F l p, NORSE: a solver for the relativistic non-linear Fokker–Planck equation for electrons in a homogeneous plasma, *Comput. Phys. Commun.* 212 (2017) 269–279.
- [5] L. Hesslow, O. Embr us, G.J. Wilkie, G. Papp, T. F l p, Effect of partially ionized impurities and radiation on the effective critical electric field for runaway generation, *Plasma Phys. Control. Fusion* 60 (7) (2018) 074010.
- [6] D. Daniel, W.T. Taitano, L. Chac n, A fully implicit, scalable, conservative nonlinear relativistic Fokker–Planck 0D-2P solver for runaway electrons, *Comput. Phys. Commun.* 254 (2020) 107361.
- [7] H. Strauss, D. Longcope, An adaptive finite element method for magnetohydrodynamics, *J. Comput. Phys.* 147 (2) (1998) 318–336.
- [8] B. Philip, L. Chac n, M. Pernice, Implicit adaptive mesh refinement for 2D reduced resistive magnetohydrodynamics, *J. Comput. Phys.* 227 (20) (2008) 8855–8874.
- [9] H. Baty, FINMHD: an adaptive finite-element code for magnetic reconnection and formation of plasmoid chains in magnetohydrodynamics, *Astrophys. J. Suppl. Ser.* 243 (2) (2019) 23.
- [10] Z. Peng, Q. Tang, X.-Z. Tang, An adaptive discontinuous Petrov–Galerkin method for the Grad–Shafranov equation, *SIAM J. Sci. Comput.* 42 (5) (2020) B1227–B1249.
- [11] Q. Tang, L. Chac n, T.V. Kolev, J.N. Shadid, X.-Z. Tang, An adaptive scalable fully implicit algorithm based on stabilized finite element for reduced visco-resistive MHD, *J. Comput. Phys.* 454 (2022) 110967.
- [12] J.A. Hittinger, J.W. Banks, Block-structured adaptive mesh refinement algorithms for Vlasov simulation, *J. Comput. Phys.* 241 (2013) 118–140.
- [13] M.F. Adams, E. Hirvijoki, M.G. Knepley, J. Brown, T. Isaac, R. Mills, Landau collision integral solver with adaptive mesh refinement on emerging architectures, *SIAM J. Sci. Comput.* 39 (6) (2017) C452–C465.
- [14] B.S. Wettervik, T.C. DuBois, E. Siminos, T. F l p, Relativistic Vlasov–Maxwell modelling using finite volumes and adaptive mesh refinement, *Eur. Phys. J. D* 71 (6) (2017) 1–14.
- [15] R.R. Arslanbekov, V.I. Kolobov, A.A. Frolova, Kinetic solvers with adaptive mesh in phase space, *Phys. Rev. E* 88 (6) (2013) 063301.
- [16] V. Kolobov, R. Arslanbekov, D. Levko, Boltzmann-Fokker-Planck Kinetic Solver with Adaptive Mesh in Phase Space, *AIP Conference Proceedings*, vol. 2132, AIP Publishing, 2019.
- [17] M.J. Berger, P. Colella, Local adaptive mesh refinement for shock hydrodynamics, *J. Comput. Phys.* 82 (1) (1989) 64–84.
- [18] E. Constantinescu, A. Sandu, G. Carmichael, Modeling atmospheric chemistry and transport with dynamic adaptive resolution, *Comput. Geosci.* 12 (2) (2008) 133–151.
- [19] N. Offermans, D. Massaro, A. Peplinski, P. Schlatter, Error-driven adaptive mesh refinement for unsteady turbulent flows in spectral-element simulations, *Comput. Fluids* (2022) 105736.
- [20] W. Guo, Y. Cheng, An adaptive multiresolution discontinuous Galerkin method for time-dependent transport equations in multidimensions, *SIAM J. Sci. Comput.* 39 (6) (2017) A2962–A2992.
- [21] C. Burstedde, L.C. Wilcox, O. Ghattas, p4est: scalable algorithms for parallel adaptive mesh refinement on forests of octrees, *SIAM J. Sci. Comput.* 33 (3) (2011) 1103–1133.

- [22] R.J. LeVeque, Finite Volume Methods for Hyperbolic Problems, Cambridge Texts in Applied Mathematics, Cambridge University Press, 2002.
- [23] E.F. Toro, Riemann Solvers and Numerical Methods for Fluid Dynamics, Springer, Berlin, Heidelberg, 2009.
- [24] B.P. Leonard, A stable and accurate convective modelling procedure based on quadratic upstream interpolation, *Comput. Methods Appl. Mech. Eng.* 19 (1) (1979) 59–98.
- [25] S. Balay, S. Abhyankar, M.F. Adams, S. Benson, J. Brown, P. Brune, K. Buschelman, E. Constantinescu, L. Dalcin, A. Dener, V. Eijkhout, J. Faibussowitsch, W.D. Gropp, V. Hapla, T. Isaac, P. Jolivet, D. Karpeev, D. Kaushik, M.G. Knepley, F. Kong, S. Kruger, D.A. May, L.C. McInnes, R.T. Mills, L. Mitchell, T. Munson, J.E. Roman, K. Rupp, P. Sanan, J. Sarich, B.F. Smith, S. Zampini, H. Zhang, H. Zhang, J. Zhang, PETSc/TAO users manual, Tech. Rep. ANL-21/39 – Revision 3.18, Argonne National Laboratory, 2022.
- [26] R. Alexander, Diagonally implicit Runge–Kutta methods for stiff ODE's, *SIAM J. Numer. Anal.* 14 (6) (1977) 1006–1021.
- [27] F.X. Giraldo, J.F. Kelly, E.M. Constantinescu, Implicit-explicit formulations of a three-dimensional nonhydrostatic unified model of the atmosphere (numa), *SIAM J. Sci. Comput.* 35 (5) (2013) B1162–B1194.
- [28] D.A. Knoll, D.E. Keyes, Jacobian-free Newton–Krylov methods: a survey of approaches and applications, *J. Comput. Phys.* 193 (2) (2004) 357–397.
- [29] *hypre: high performance preconditioners*, <https://llnl.gov/casc/hypre>, <https://github.com/hypre-space/hypre>.
- [30] R.D. Falgout, U.M. Yang, hypre: a library of high performance preconditioners, in: P.M.A. Sloot, A.G. Hoekstra, C.J.K. Tan, J.J. Dongarra (Eds.), *Computational Science — ICCS 2002*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2002, pp. 632–641.
- [31] A.J. Brizard, A.A. Chan, Nonlinear relativistic gyrokinetic Vlasov–Maxwell equations, *Phys. Plasmas* 6 (12) (1999) 4548–4558.
- [32] J. Connor, R. Hastie, Relativistic limitations on runaway electrons, *Nucl. Fusion* 15 (3) (1975) 415.
- [33] Z. Guo, C. Mcdevitt, X. Tang, Toroidal effect on runaway vortex and avalanche growth rate, *Phys. Plasmas* 26 (8) (2019) 082503.
- [34] G. Papp, M. Drevlak, T. Fülöp, P. Helander, Runaway electron drift orbits in magnetostatic perturbed fields, *Nucl. Fusion* 51 (4) (2011) 043004.
- [35] L. Hesslow, O. Embréus, A. Stahl, T.C. DuBois, G. Papp, S.L. Newton, T. Fülöp, Effect of partially screened nuclei on fast-electron dynamics, *Phys. Rev. Lett.* 118 (25) (2017) 255001.
- [36] M. Rosenbluth, S. Putvinski, Theory for avalanche of runaway electrons in tokamaks, *Nucl. Fusion* 37 (10) (1997) 1355.
- [37] S. Chiu, M. Rosenbluth, R. Harvey, V. Chan, Fokker–Planck simulations mylb of knock-on electron runaway avalanche and bursts in tokamaks, *Nucl. Fusion* 38 (11) (1998) 1711.
- [38] C.J. McDevitt, Z. Guo, X.-Z. Tang, Avalanche mechanism for runaway electron amplification in a tokamak plasma, *Plasma Phys. Control. Fusion* 61 (5) (2019) 054008.
- [39] T. Isaac, C. Burstedde, L.C. Wilcox, O. Ghattas, Recursive algorithms for distributed forests of octrees, *SIAM J. Sci. Comput.* 37 (5) (2015) C497–C531.
- [40] H. Sundar, R.S. Sampath, G. Biros, Bottom-up construction and 2:1 balance refinement of linear octrees in parallel, *SIAM J. Sci. Comput.* 30 (5) (2008) 2675–2708.
- [41] C. Burstedde, O. Ghattas, M. Gurnis, T. Isaac, G. Stadler, T. Warburton, L.C. Wilcox, Extreme-scale AMR, in: *SC10: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ACM/IEEE, 2010.
- [42] H. Sundar, G. Biros, C. Burstedde, J. Rudi, O. Ghattas, G. Stadler, Parallel geometric-algebraic multigrid on unstructured forests of octrees, in: *SC12: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ACM/IEEE, 2012.
- [43] J. Rudi, A.C.I. Malossi, T. Isaac, G. Stadler, M. Gurnis, P.W. Staar, Y. Ineichen, C. Bekas, A. Curioni, O. Ghattas, An extreme-scale implicit solver for complex PDEs: highly heterogeneous flow in Earth's mantle, in: *SC15: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, ACM, 2015, pp. 1–12.
- [44] J. Rudi, G. Stadler, O. Ghattas, Weighted BFBT preconditioner for Stokes flow problems with highly heterogeneous viscosity, *SIAM J. Sci. Comput.* 39 (5) (2017) S272–S297.
- [45] J. Rudi, Y.-H. Shih, G. Stadler, Advanced Newton methods for geodynamical models of Stokes flow with viscoplastic rheologies, *Geochem. Geophys. Geosyst.* 21 (9) (2020).
- [46] T. Weinzierl, The Peano software—parallel, automaton-based, dynamically adaptive grid traversals, *ACM Trans. Math. Softw.* 45 (2) (2019) 1–41.
- [47] M. Fernando, D. Neilsen, H. Lim, E. Hirschmann, H. Sundar, Massively parallel simulations of binary black hole intermediate-mass-ratio inspirals, *SIAM J. Sci. Comput.* 41 (2) (2019) C97–C138.
- [48] V.E. Henson, U.M. Yang, BoomerAMG: a parallel algebraic multigrid solver and preconditioner, *Appl. Numer. Math.* 41 (1) (2002) 155–177.
- [49] J.W. Banks, W.D. Henshaw, D.W. Schwendeman, Q. Tang, A stable partitioned FSI algorithm for rigid bodies and incompressible flow. Part I: model problem analysis, *J. Comput. Phys.* 343 (2017) 432–468.
- [50] T.A. Manteuffel, J. Ruge, B.S. Southworth, Nonsymmetric algebraic multigrid based on local approximate ideal restriction (ℓ AIR), *SIAM J. Sci. Comput.* 40 (6) (2018) A4105–A4130.
- [51] C.J. McDevitt, Z. Guo, X.-Z. Tang, Relation of the runaway avalanche threshold to momentum space topology, *Plasma Phys. Control. Fusion* 60 (2) (2018) 024004.
- [52] J. Burgers, A Mathematical Model Illustrating the Theory of Turbulence, *Advances in Applied Mechanics*, vol. 1, Elsevier, 1948, pp. 171–199.
- [53] K. Mohseni, T. Colonius, Numerical treatment of polar coordinate singularities, *J. Comput. Phys.* 157 (2) (2000) 787–795.