GX: a GPU-native gyrokinetic turbulence code for tokamak and stellarator design

N.R. Mandell [1,2,†], **W. Dorland** [3,4], **I. Abel** [4], **R. Gaur** [4,5], **P. Kim** [3,5], **M. Martin** [6] **and T. Qian** [5]

[1]Princeton Plasma Physics Laboratory, Princeton, NJ 08543, USA

[2]Plasma Science and Fusion Center, Massachusetts Institute of Technology, Cambridge, MA, USA

[3]University of Maryland, College Park, MD 20742, USA

[4]Institute for Research in Electronics and Applied Physics, University of Maryland, College Park, MD 20742, USA

[5]Princeton University, Princeton, NJ 08543, USA

[6]Thea Energy, Princeton, NJ 08542, USA

GX is a code designed to solve the nonlinear gyrokinetic system for low-frequency turbulence in magnetized plasmas, particularly tokamaks and stellarators. In GX, our primary motivation and target is a fast gyrokinetic solver that can be used for fusion reactor design and optimization along with wide-ranging physics exploration. This has led to several code and algorithm design decisions, specifically chosen to prioritize time to solution. First, we have used a discretization algorithm that is pseudospectral in the entire phase space, including a Laguerre–Hermite pseudospectral formulation of velocity space, which allows for smooth interpolation between coarse gyrofluid-like resolutions and finer conventional gyrokinetic resolutions and efficient evaluation of a model collision operator. Additionally, we have built GX to natively target graphics processors (GPUs), which are among the fastest computational platforms available today. Finally, we have taken advantage of the reactor-relevant limit of small ρ_* by using the radially local flux-tube approach. In this paper we present details about the gyrokinetic system and the numerical algorithms used in GX to solve the system. We then present several numerical benchmarks against established gyrokinetic codes in both tokamak and stellarator magnetic geometries to verify that GX correctly simulates gyrokinetic turbulence in the small ρ_* limit. Moreover, we show that the convergence properties of the Laguerre–Hermite spectral velocity formulation are quite favourable for nonlinear problems of interest. Coupled with GPU acceleration, which we also investigate with scaling studies, this enables GX to be able to produce useful turbulence simulations in minutes on one (or a few) GPUs and higher fidelity results in a few hours using several GPUs. GX is open-source software that is ready for fusion reactor design studies.

Key words: plasma simulation, fusion plasma

---

† Email address for correspondence: noah.mandell@typeoneenergy.com

1. Introduction

Tokamaks and stellarators are fusion reactor concepts that use magnetic fields to reduce the rate of loss of particles, momentum and energy from the reactor's toroidal confinement volume. In general, the confining magnetic fields are produced both by magnets that surround the toroidal confinement region, and by currents flowing through the confined plasma. To a great extent, the detailed spatial patterns of magnetic fields determine whether the operating conditions of a tokamak or stellarator permit sustained thermonuclear fusion reactions. Consequently, a key step in the design of any tokamak or stellarator reactor is the evaluation of reactor performance as a function of magnetic geometry.

Turbulent transport is one major factor that limits reactor performance. Turbulence in the fuel column is driven by the steep gradients of density, velocity and temperature that exist between the hot, sometimes rapidly flowing, dense fuel core and the relatively cold, diffuse environment. Under optimal conditions, the turbulent fluctuations are small-amplitude perturbations of the steady-state plasma density, temperature, etc. but these fluctuations nonetheless enhance losses, reduce insulation (by rapidly mixing cold plasma from the outer region with hot plasma in the reactor core) and generally limit reactor performance (Abel *et al.* 2008). Decades of meticulous experiments and direct observations of plasma fluctuations in many tokamaks and stellarators have established a good understanding of when and where the turbulence arises, how intense it is, its dominant wavelengths and frequencies, and so on. The evaluation of the expected performance of any proposed tokamak or stellarator reactor design therefore essentially requires some way to estimate properties of the plasma turbulence that will arise as a result of the details of the proposed magnetic field geometry.

More than two decades of extensive validation campaigns have established that gyrokinetic theory quantitatively describes this turbulence. Many gyrokinetic simulation codes (Parker, Lee & Santoro 1993; Kotschenreuther, Rewoldt & Tang 1995; Dorland *et al.* 2000; Jenko 2000; Lin *et al.* 2000; Jost *et al.* 2001; Candy & Waltz 2003; Idomura, Tokuda & Kishimoto 2003; Watanabe & Sugama 2005; Wang *et al.* 2006; Jolliet *et al.* 2007; Idomura *et al.* 2008; Peeters *et al.* 2009; Candy, Belli & Bravenec 2016; Barnes, Parra & Landreman 2019; Lanti *et al.* 2020) have been used to study, test and ultimately validate our understanding of how plasma turbulence depends on the magnetic geometry (see e.g. Belli, Hammett & Dorland (2008), Marinoni *et al.* (2009), Mynick, Xanthopoulos & Boozer (2009) and White (2019)). Because there are many algorithmic trade-offs to consider when setting up a turbulence simulation code, no single gyrokinetic code is 'best' for these purposes. In this paper we present GX, a fast gyrokinetic solver where our primary motivations and targets are fusion reactor design, optimization and wide-ranging physics exploration at reactor scales.

What does this mean in practice? It means primarily that we have prioritized time to solution (together with quantitative uncertainty estimates) over comprehensiveness and full realism, and that we aim to provide simple runtime control parameters to shift smoothly from very fast but relatively less accurate calculations such as are of most value in 'outer loops' of a design activity, to more comprehensive, 'standard' results that are (much) more expensive to evaluate. Moreover, instead of retrofitting an existing CPU-based code, we started from scratch, with every algorithmic design decision made to target graphics processors (GPUs), which are among the fastest computational platforms available today, with NVIDIA or AMD GPUs providing most of the computer power for seven of the top 10 fastest supercomputers in the world (Strohmaier *et al.* 2022). Finally, since we are interested in reactors specifically, we focused on the reactor limit of small

$\rho_* \equiv \rho_i/R$, where ρ_i is the typical radius of gyration of an ion in the plasma, and R is the major radius of the device. In the small ρ_* limit, there are major simplifications to be had, corresponding physically to the opening of a wide 'gap' between typical radial correlation lengths $\lambda_{corr} \sim \rho_i$ of the turbulence and the radial gradient scale lengths $L \sim R$. In contrast, present-day devices are not in the small ρ_* limit, so turbulent eddies can be large enough to sample variations in the driving gradients and other relevant plasma properties. Additionally, because stellarators are non-axisymmetric, they have irreducible geometric variation within a given flux surface that makes the small ρ_* limit more challenging. The turbulent correlation length λ_{corr} must also be small compared with the distance that geometric differences within the flux surface change significantly. This limit exists for small enough ρ_*, but it is roughly more challenging by a factor of the aspect ratio R/a, where a is the minor radius of the torus. To account for the 'mesoscale' phenomena that may result and to complete the many quantitative validation studies that have been undertaken, many teams chose algorithms that handle radially non-local physics (often called 'global' codes). Codes that take advantage of the simpler, radially local, small ρ_* limit are known as 'flux-tube' codes, and our new code GX is in this category. Several flux-tube turbulence calculations can then be coupled together with a transport solver to obtain transport time scale profile evolution on the device scale (Barnes *et al.* 2009) or steady-state profiles (Candy *et al.* 2009). This multiscale approach is asymptotically valid in the limit of vanishing ρ_* (Abel *et al.* 2013). Recent work has also shown a novel approach to include radially global effects within the flux-tube formulation (Parra & Barnes 2015; St-Onge, Barnes & Parra 2022).

To verify that GX correctly simulates gyrokinetic turbulence in the small ρ_* limit, we present several tokamak and stellarator benchmarks of GX with established flux-tube gyrokinetic simulation codes. Note that while numerical verification (benchmarking) is code-specific, the extensive history of experimental validation of the gyrokinetic model is inherited by any verified gyrokinetic code and need not be repeated. GX is open-source scientific software[1] that is therefore ready for immediate design studies. Below, we show that GX can produce useful turbulence simulations in minutes using one (or a few) GPUs and higher fidelity results in a few hours using several GPUs.

Our algorithmic approach to solving the gyrokinetic system is based in pseudospectral methods. There is a long history of the use of Fourier pseudospectral methods in turbulence simulations to discretize the configuration space, especially in flux-tube gyrokinetic codes like GS2 (Kotschenreuther *et al.* 1995; Dorland *et al.* 2000), (non-global) GENE (Jenko & Dorland 2001), CGYRO (Candy *et al.* 2016) and stella (Barnes *et al.* 2019). These methods have the advantage of spectrally accurate evaluation of derivatives. Building on this, we have developed a pseudospectral formulation for discretizing the velocity space in the gyrokinetic system using a Laguerre–Hermite spectral basis (Mandell, Dorland & Landreman 2018). Projecting the (perturbed) gyrokinetic distribution function onto this basis produces spectral modes that correspond to (gyro)fluid moment quantities (like density, parallel momentum, etc.), so that projection of the gyrokinetic equation corresponds to a gyrofluid system with arbitrarily many moments. In the lowest-resolution limit the model retains critical conservation laws (density, momentum, energy and free energy) and corresponds precisely to the well-established gyrofluid models of Dorland & Hammett (1993), Beer & Hammett (1996) and Snyder & Hammett (2001) when their Landau-fluid closures are employed (Hammett & Perkins 1990; Beer & Hammett 1996). At moderate velocity-space resolution, one can

---

[1]The source code is available at https://bitbucket.org/gyrokinetics/gx. Documentation is available at https://gx.readthedocs.io/en/latest.

extend the Landau-fluid closure approach to accommodate more moments (Smith 1997) or use 'hypercollisions' to close the moment hierarchy. At high velocity-space resolution, the model can retain a similar number of degrees of freedom as typically used in conventional gyrokinetic Eulerian models. A key advantage of our approach is therefore the flexibility to interpolate between these limits, depending on the desired balance of accuracy and speed for a particular calculation. Additionally, our Fourier–Laguerre–Hermite pseudospectral algorithm is a good fit for modern GPU computing: the algorithm relies heavily on fast transform methods that are well-optimized on GPUs, and the memory requirements of spectral algorithms are low enough to fit a problem onto one or a few GPUs. Several gyrokinetic codes have been ported to GPUs in recent years to target modern heterogeneous computing platforms (Madduri *et al.* 2011; D'Azevedo *et al.* 2018; Sfiligoi, Candy & Kostuk 2018; Germaschewski *et al.* 2021; Ohana *et al.* 2021; Belli *et al.* 2022).

Similar works on Laguerre–Hermite spectral methods for drift kinetics (Jorge, Ricci & Loureiro 2017) and gyrokinetics (Jorge, Frei & Ricci 2019; Frei, Jorge & Ricci 2020; Frei *et al.* 2021; Frei, Ernst & Ricci 2022a; Frei, Hoffmann & Ricci 2022b; Hoffmann, Frei & Ricci 2023a,b) have shown promising results and can be viewed as complementary to our work. In particular, extensive linear benchmarks and convergence studies have been performed (Frei *et al.* 2022b, 2023), showing that the Laguerre–Hermite approach can be used successfully to model ion temperature gradient (ITG), trapped electron, kinetic ballooning and microtearing modes. Jorge *et al.* (2019), Frei *et al.* (2021, 2022a,b) develop, implement and study advanced gyrokinetic collision operators by leveraging the Laguerre–Hermite spectral approach. Nonlinear studies in Z-pinch (Hoffmann *et al.* 2023b) and full toroidal (Hoffmann *et al.* 2023a) geometry have also now been performed, showing that accurate nonlinear results, including the Dimits shift (Dimits *et al.* 2000), can be obtained with fewer basis functions than typically necessary for convergence of linear instabilities. Additionally, Frei *et al.* (2020) developed a Laguerre–Hermite projection of the full-f gyrokinetic system to target the periphery of tokamaks, and Frei, Mencke & Ricci (2024) implemented the full-f system in slab geometry for modelling linear devices.

The remainder of this paper is organized as follows. In § 2 we describe the gyrokinetic model. We describe the Fourier–Laguerre–Hermite pseudospectral formulation of the gyrokinetic system in § 4. Section 3 describes the magnetic geometry and coordinates used for tokamaks and stellarators. The time discretization scheme is described in § 5. In § 6, several benchmarks are presented comparing GX with established flux-tube gyrokinetic codes (GS2, GENE and stella) for linear and nonlinear cases in tokamak and stellarator geometries. Section 7 details the convergence properties of the Laguerre–Hermite basis for nonlinear calculations, along with details about GPU performance and multi-GPU scaling. Finally, we present the conclusions and future work in § 8.

2. Gyrokinetic model equations

In GX we solve the δf gyrokinetic system. The literature of gyrokinetic theory is vast; our notation and philosophy follow four particular references: Antonsen & Lane (1980), Frieman & Chen (1982), Barnes *et al.* (2010) and Abel *et al.* (2013). The velocity space in gyrokinetic theory is two-dimensional because the particles gyrate very rapidly around their guiding centre position, averaging over any field variations that are encountered in one gyration period so that we can ignore the instantaneous position of the particle around its gyro-orbit. We choose to write the gyrokinetic equation in $(v_\parallel, \mu)$ coordinates, where $v_\parallel$ is the speed in the direction of the magnetic field and $\mu \equiv v_\perp^2/2B$ is the magnetic moment, where $v_\perp$ is the speed perpendicular to the magnetic field and B is the local magnetic field strength. Using μ as a coordinate takes advantage of the fact that μ is a gyrokinetic adiabatic invariant. These coordinates are not optimal for resolving the boundary layers

in velocity that occur in linear gyrokinetic theory where particle orbits are 'trapped' in magnetic wells on one side of the boundary layer and 'passing' (untrapped) on the other. To resolve the sharp boundary layers it is better to use energy and pitch angle as one's velocity space coordinates (as is done in the GS2 code, for example, which optimizes its velocity space grid to enhance resolution near the trapped–passing boundary), but this leads to a considerably more complicated code implementation, particularly for stellarator applications, because of the need to carefully handle turning points in each magnetic well. However, in a sufficiently turbulent plasma, these boundary layers will be broadened, reducing the need for coordinates optimized for a sharp boundary. Thus, we have selected the coordinates that lead to simpler, more efficient code because we are particularly interested in calculating nonlinear turbulence properties (as opposed to linear stability conditions). These coordinates are also used in GENE and stella.

Following (144) of Abel *et al.* (2013), but expressing the equations in $(v_\parallel, \mu)$ velocity coordinates, neglecting strong equilibrium flows and normalizing all quantities to non-dimensional form (see appendix A), we have

$$
\frac{\partial h_s}{\partial t} + \left[v_{ts} v_\parallel \hat{\boldsymbol{b}} + \langle \boldsymbol{v}_\chi \rangle_{\boldsymbol{R}} + \frac{\tau_s}{Z_s} \boldsymbol{v}_d \right] \cdot \boldsymbol{\nabla} h_s - v_{ts} \mu (\hat{\boldsymbol{b}} \cdot \boldsymbol{\nabla} B) \frac{\partial h_s}{\partial v_\parallel}
$$
$$
= \frac{Z_s}{\tau_s} F_{Ms} \frac{\partial \langle \chi \rangle_{\boldsymbol{R}}}{\partial t} - \langle \boldsymbol{v}_\chi \rangle_{\boldsymbol{R}} \cdot \boldsymbol{\nabla}|_E F_{Ms} + C(h_s). \tag{2.1}
$$

Here, we have split the total distribution function as $F_s = F_{0s} + \delta f_s = F_{Ms}(1 - Z_s \Phi / \tau_s) + h_s$, with $(Z_s/\tau_s) F_{Ms} \Phi(\boldsymbol{r}, t)$ and $h_s(\boldsymbol{R}, v_\parallel, \mu, t)$ the Boltzmann and non-Boltzmann parts of the δf_s perturbation, respectively. The equilibrium distribution function is a Maxwellian in energy E (which we take to have no flows),

$$
F_{0s} = F_{Ms} = \frac{n_s}{(2\pi v_{ts}^2)^{3/2}} e^{-E} = \frac{n_s}{(2\pi v_{ts}^2)^{3/2}} e^{-v_\parallel^2/2 - \mu B}, \tag{2.2}
$$

and the fluctuating gyroaveraged distribution function satisfies $h_s \ll F_{Ms}$. Note that in (2.1), the gradient of F_M at constant energy is denoted $\boldsymbol{\nabla}|_E F_M$. We also have the following dimensionless species parameters: equilibrium density n_s; equilibrium temperature τ_s; mass m_s, charge Z_s; thermal velocity $v_{ts} = \sqrt{\tau_s/m_s}$.

The distribution function describes the probability of finding a particle of species s with gyrocentre position $\boldsymbol{R}$, velocity parallel to the magnetic field $v_\parallel$ and magnetic moment $\mu = v_\perp^2/2B$, where $v_\perp$ is the speed in the plane perpendicular to the magnetic field. The equilibrium magnetic field $\boldsymbol{B}$ has magnitude $B = B(\boldsymbol{r})$ and direction $\hat{\boldsymbol{b}} = \boldsymbol{B}/B$, and magnetic fluctuations are given by $\delta \boldsymbol{B} = \delta \boldsymbol{B}_\perp + \delta B_\parallel \hat{\boldsymbol{b}}$, where the perpendicular and parallel components can be expressed in terms of the vector potential $\boldsymbol{A} = \boldsymbol{A}_\perp + A_\parallel \hat{\boldsymbol{b}}$ via $\delta \boldsymbol{B}_\perp = [\boldsymbol{\nabla} \times \boldsymbol{A}]_\perp \simeq \boldsymbol{\nabla} A_\parallel \times \hat{\boldsymbol{b}}$ and $\delta B_\parallel = \hat{\boldsymbol{b}} \cdot (\boldsymbol{\nabla} \times \boldsymbol{A}_\perp)$, respectively. The gyrokinetic potential $\chi(\boldsymbol{r}, t)$ is composed of the electrostatic potential $\Phi(\boldsymbol{r}, t)$ and the vector potential $\boldsymbol{A}(\boldsymbol{r}, t)$,

$$
\chi(\boldsymbol{r}, \boldsymbol{v}, t) = \Phi(\boldsymbol{r}, t) - v_{ts} \boldsymbol{v} \cdot \boldsymbol{A}(\boldsymbol{r}, t), \tag{2.3}
$$

and is a function of particle position $\boldsymbol{r}$, where $\boldsymbol{r} = \boldsymbol{R} + \boldsymbol{\rho}$ so that $\boldsymbol{\rho}$ is the gyroradius vector that rotates at the gyrofrequency Ω and points from the gyrocentre (at $\boldsymbol{R}$) to the particle (at $\boldsymbol{r}$). In (2.1) gyroangle dependence is eliminated by gyroaveraging the potentials at fixed

gyrocentre position $\boldsymbol{R}$, denoted by

$$\langle \chi \rangle_{\boldsymbol{R}} = \langle \Phi \rangle_{\boldsymbol{R}} - v_{ts} v_{\parallel} \langle A_{\parallel} \rangle_{\boldsymbol{R}} - v_{ts} \langle \boldsymbol{v}_{\perp} \cdot \boldsymbol{A}_{\perp} \rangle_{\boldsymbol{R}}, \tag{2.4}$$

along with the corresponding gyroaveraged fluctuating velocity $\langle \boldsymbol{v}_{\chi} \rangle_{\boldsymbol{R}} = \hat{\boldsymbol{b}} \times \nabla \langle \chi \rangle_{\boldsymbol{R}}$. For most applications, the plasma β is low and the magnetic drift velocity is $\boldsymbol{v}_d = \hat{\boldsymbol{b}} \times (v_{\parallel}^2 \hat{\boldsymbol{b}} \cdot \nabla \hat{\boldsymbol{b}} + \mu \nabla B)/B$. GX normally uses this low-β approximation, but the full expression for the magnetic drifts can be selected at runtime.

The collision operator is denoted by $C(h)$. Here we will follow Mandell *et al.* (2018) and take the Dougherty model collision operator (Dougherty 1964). Expressed in Fourier space, the like-species Dougherty collision operator is given by

$$\begin{aligned} C_{ss}(h_s) = v_{ss} &\left\{ \left[\frac{\partial}{\partial v_{\parallel}} \left(\frac{\partial}{\partial v_{\parallel}} + v_{\parallel} \right) + 2 \frac{\partial}{\partial \mu} \left(\frac{\mu}{B} \frac{\partial}{\partial \mu} + \mu \right) + k_{\perp}^2 \rho_s^2 \right] h_s \right. \\ &\left. + (\bar{T}_s[(v_{\parallel}^2 - 1) + 2(\mu B - 1)]J_{0s} + [\bar{u}_{\parallel s} v_{\parallel} J_{0s} + \bar{u}_{\perp s} v_{\perp} J_{1s}])F_{Ms} \right\}, \end{aligned} \tag{2.5}$$

with the field-particle terms given by

$$\bar{u}_{\parallel s} \equiv \int \mathrm{d}^3 \boldsymbol{v} J_{0s} v_{\parallel} h_s, \tag{2.6}$$

$$\bar{u}_{\perp s} \equiv \int \mathrm{d}^3 \boldsymbol{v} J_{1s} v_{\perp} h_s, \tag{2.7}$$

$$\bar{T}_s = \frac{1}{3}(\bar{T}_{\parallel s} + 2\bar{T}_{\perp s}) \equiv \frac{1}{3} \int \mathrm{d}^3 \boldsymbol{v}[(v_{\parallel}^2 - 1) + 2(\mu B - 1)]J_{0s}h_s. \tag{2.8}$$

Here we have the Bessel functions $J_{0s} = J_0(\alpha_s)$ and $J_{1s} = J_1(\alpha_s)$ that result from the Fourier transform of gyroaverage operators, with $\alpha_s = \sqrt{2\mu B b_s}$, $b_s = k_{\perp}^2 \rho_s^2$ and $\rho_s = m_s v_{ts}/(Z_s B)$ is the normalized gyroradius for species s.

The Dougherty collision operator is a good physical model of like-particle collisions, which are important to gyrokinetic dynamics. It captures the physics of the collision operator presented in Abel *et al.* (2008), except that our collision frequency does not have velocity dependence, and there is no difference between the rates of pitch-angle scattering and slowing down. It is also attractive because its Laguerre–Hermite transform is sparse. More realistic collision operators could be included in our model in the future, such as those in Frei *et al.* (2021, 2022a), which employs a similar Laguerre–Hermite spectral velocity approach as ours. The Dougherty model operator can also be extended to multiple species (Francisquez *et al.* 2022).

The electromagnetic potentials are determined by the gyrokinetic equivalent of Maxwell's equations. The electrostatic potential is determined by the quasineutrality equation,

$$\sum_s \frac{Z_s^2 n_s}{\tau_s} \Phi = \sum_s Z_s n_s \int \mathrm{d}^3 \boldsymbol{v} \langle h_s \rangle_r, \tag{2.9}$$

where note that here the gyroaverage on the right-hand side is taken at constant particle position $\boldsymbol{r}$. The parallel magnetic vector potential is determined by the parallel component

of Ampère's law,

$$-\nabla_\perp^2 A_\parallel = \frac{\beta_{\mathrm{ref}}}{2} \sum_s Z_s n_s v_{ts} \int \mathrm{d}^3 \boldsymbol{v} \; v_\parallel \langle h_s \rangle_{\boldsymbol{r}}, \tag{2.10}$$

with $\beta_{\mathrm{ref}} = 8\pi n_{\mathrm{ref}} T_{\mathrm{ref}} / B_{\mathrm{N}}^2$ the plasma beta of the reference species. The perpendicular component of Ampère's law determines the perpendicular component of the vector potential, but is more conveniently expressed in terms of the parallel magnetic fluctuation, $\delta B_\parallel$,

$$\nabla_\perp^2 \delta B_\parallel = -\frac{\beta_{\mathrm{ref}}}{2B} \nabla_\perp \nabla_\perp : \sum_s n_s \tau_s \int \mathrm{d}^3 \boldsymbol{v} \; \langle \boldsymbol{v}_\perp \boldsymbol{v}_\perp h_s \rangle_r, \tag{2.11}$$

where the double dot product is defined as $A : B = \sum_{ij} A_{ij} B_{ij}$.

Finally, note that we can define an auxiliary distribution function

$$g_s = h_s - \frac{Z_s}{\tau_s} \langle \chi \rangle_{\boldsymbol{R}} F_{Ms}, \tag{2.12}$$

which eliminates the time derivative on the right-hand side of (2.1), resulting in

$$\frac{\partial g_s}{\partial t} + \left[v_{ts} v_\parallel \hat{\boldsymbol{b}} + \langle \boldsymbol{v}_\chi \rangle_{\boldsymbol{R}} + \frac{\tau_s}{Z_s} \boldsymbol{v}_d \right] \cdot \nabla h_s - v_{ts} \mu (\hat{\boldsymbol{b}} \cdot \nabla B) \frac{\partial h_s}{\partial v_\parallel} = -\langle \boldsymbol{v}_\chi \rangle_{\boldsymbol{R}} \cdot \nabla|_E F_{Ms} + C(h_s). \tag{2.13}$$

This is the form of the gyrokinetic equation that we will solve numerically.

3. Magnetic geometry and choice of coordinates

The geometric structure of a tokamak's magnetic field is axisymmetric, meaning there are no geometric variations that distinguish points at different toroidal angles. Most (but not all) gyrokinetic studies have focused on the axisymmetric limit. Stellarator magnetic fields are not axisymmetric. We have selected coordinates that are convenient for either case. Our spatial coordinates are aligned with the local background magnetic field so that we are able to efficiently resolve perturbations whose parallel wavelengths $\lambda_\parallel$ are $O(\rho_*^{-1})$ longer than the wavelengths $\lambda_\perp$, perpendicular to the local magnetic field; these field-line-following coordinates reduce the number of grid points required to resolve the turbulence by a factor of ρ_*. In the case of stellarators, the lack of axisymmetry implies that even when one can describe the turbulence successfully as radially local, it might happen that the geometric variations in the binormal direction (within the magnetic surface) are not numerically well-separated from $\lambda_\perp$, even for reactor-relevant values of ρ_*. We have chosen not to address this possibility; we assume that ρ_* is small enough that the turbulence within each flux tube in a given flux surface is not affected by turbulence occurring in a different flux tube within the same surface.

In this section, we will start with the general form of a divergence-free magnetic field, define key important quantities, various coordinate systems and briefly explain how they are used to obtain the geometry-dependent coefficients in the gyrokinetic model. Please refer to appendix E for specific mathematical details.

3.1. *Axisymmetric configurations (tokamaks)*

We start from the Clebsch form (D'haeseleer *et al.* 1991) for the equilibrium magnetic field:

$$\boldsymbol{B} = \nabla \alpha \times \nabla \psi. \tag{3.1}$$

We restrict our attention to solutions whose magnetic field lines lie on closed nested toroidal surfaces, known as flux surfaces. Here, flux surfaces are labelled by ψ, a

radial-like coordinate. On each surface, the binormal coordinate α is a field-line label such that a line of constant α gives the path of a magnetic field line.

For tokamaks, we choose to define the coordinate ψ to be the enclosed poloidal flux divided by 2π,

$$\psi = \frac{\Psi_{\text{pol}}}{2\pi} = \frac{1}{(2\pi)^2} \int_V d\tau \, \boldsymbol{B} \cdot \boldsymbol{\nabla}\theta, \tag{3.2}$$

and the binormal coordinate α is chosen to be

$$\alpha = \phi - q(\psi)\theta, \tag{3.3}$$

where ϕ and θ are 'straight-field-line' toroidal and poloidal angles, respectively, and

$$q(\psi) = \frac{1}{(2\pi)^2} \int_0^{2\pi} d\phi \int_0^{2\pi} d\theta \, \frac{\boldsymbol{B} \cdot \boldsymbol{\nabla}\phi}{\boldsymbol{B} \cdot \boldsymbol{\nabla}\theta}, \tag{3.4}$$

is the safety factor.

3.2. *Non-axisymmetric configurations (stellarators)*

For stellarators, we find it more convenient to write the magnetic field in Clebsch form as

$$\boldsymbol{B} = \boldsymbol{\nabla}\psi \times \boldsymbol{\nabla}\alpha, \tag{3.5}$$

with the toroidal flux chosen as the radial-like coordinate,

$$\psi = \frac{\Psi_{\text{tor}}}{2\pi} = \frac{1}{(2\pi)^2} \int_V d\tau \, \boldsymbol{B} \cdot \boldsymbol{\nabla}\phi, \tag{3.6}$$

and the binormal coordinate chosen to be

$$\alpha = \theta - \iota(\psi)\phi. \tag{3.7}$$

Here, $\iota(\psi) = 1/q(\psi)$ is the rotational transform, and once again ϕ and θ are 'straight-field-line' toroidal and poloidal angles, respectively.

3.3. *Field-aligned coordinate system*

Since GX is a local flux-tube code, it is advantageous to use a field-aligned coordinate system. Therefore, we introduce the field-aligned, flux-tube coordinates (x, y, z) used by Beer, Cowley & Hammett (1995). Here, $x = x(\psi)$ is a radial coordinate, $y = y(\alpha)$ is a binormal coordinate, and $z = z(\theta)$ is a field-line-following coordinate that parametrizes distance along the equilibrium magnetic field using the poloidal angle θ. The coordinates x and y are normalized forms of ψ and α such that

$$x = \frac{dx}{d\psi}(\psi - \psi_0), \tag{3.8}$$

$$y = \frac{dy}{d\alpha}(\alpha - \alpha_0), \tag{3.9}$$

where ψ_0 and α_0 are the values of equilibrium ψ and α at the centre of the flux tube. The coordinates (x, y, z) have units of length. In the local flux-tube, x and y vary on the length scale of the gyroradius whereas z varies on the length scale of the machine size. We require an equispaced coordinate z along the field line, which is needed to utilize the

Fourier spectral treatment of the parallel derivative terms in the gyrokinetic equation. The procedure for obtaining an equispaced, equal-arc z coordinate is provided in appendix E.

Upon defining the field-aligned coordinate system, we can calculate the geometric coefficients required to compute the various terms in the gyrokinetic equation, as described briefly in appendix E. For a three-dimensional equilibrium, GX can read the output data generated by the VMEC code (Hirshman & Whitson 1983). For a two-dimensional axisymmetric equilibrium, GX can use a local equilibrium defined using a Miller parametrization (Miller *et al.* 1998). The user also has the option to work with simpler geometries like a slab or a cylindrical Z-pinch. Besides these native geometry options, GX can also read geometry data generated by the geometry module in the GS2 code; for example, this enables the capability to use experimentally relevant geometry derived from an EFIT equilibrium reconstruction.

4. Fourier–Laguerre–Hermite pseudospectral formulation

Here we present the Fourier–Laguerre–Hermite formulation of the electromagnetic δf gyrokinetic system. Following Mandell *et al.* (2018), and extending to include electromagnetic fluctuations, we solve the gyrokinetic equation by taking a Fourier transform in the perpendicular directions (which we will call x and y; the parallel direction will be z), a Laguerre transform in μB, and a Hermite transform in $v_{\parallel}$. Note that there are no major disadvantages to the choice of μB versus μ as a coordinate, and our choice to use μB is natural because the Laguerre weight function becomes $\exp(-\mu B)$ just like in the Maxwellian distribution.

We define the Fourier–Laguerre–Hermite transform of the distribution function h_s for species s as

$$\mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{\boldsymbol{k}_\perp} h_s = 2\pi \int_0^\infty \mathrm{d}\mu B \psi_\ell(\mu B) \int_{-\infty}^\infty \mathrm{d}v_\parallel \phi_m(v_\parallel) \int \mathrm{d}x \,\mathrm{d}y \,\mathrm{e}^{-\mathrm{i}k_\perp \cdot R} h_s(x, y, z, v_\parallel, \mu)$$

$$\equiv H_{\ell,m}^s(\boldsymbol{k}_\perp, z), \tag{4.1}$$

$$H_{\ell,m}^s(\boldsymbol{k}_\perp, z) = 2\pi \int_0^\infty \mathrm{d}\mu B \psi_\ell(\mu B) \int_{-\infty}^\infty \mathrm{d}v_\parallel \phi_m(v_\parallel) \int \mathrm{d}x \,\mathrm{d}y \,\mathrm{e}^{-\mathrm{i}k_\perp \cdot R} h_s(x, y, z, v_\parallel, \mu), \tag{4.2}$$

where $\boldsymbol{k}_\perp = k_x \nabla x + k_y \nabla y$ is the perpendicular wavenumber vector, $\psi_\ell(\mu B) = (-1)^\ell \mathrm{L}_\ell(\mu B)$ with

$$\mathrm{L}_\ell(x) = \frac{\mathrm{e}^x}{\ell!} \frac{\mathrm{d}^\ell}{\mathrm{d}x^\ell} x^\ell \,\mathrm{e}^{-x}, \tag{4.3}$$

the Laguerre polynomials, and $\phi_m(v_\parallel) = \mathrm{He}_m(v_\parallel)/\sqrt{m!}$ with

$$\mathrm{He}_m(x) = (-1)^m \,\mathrm{e}^{x^2/2} \frac{\mathrm{d}^m}{\mathrm{d}x^m} \,\mathrm{e}^{-x^2/2}, \tag{4.4}$$

the probabilists' Hermite polynomials. In Fourier space, gyroaveraging operations (whether at constant $\boldsymbol{R}$ or constant $\boldsymbol{r}$) simply become multiplications by Bessel functions, so that the Fourier transform of the gyroaveraged potential is

$$\mathcal{F}_{k_\perp} \langle \chi \rangle_{\boldsymbol{R}} = J_{0s} \Phi(\boldsymbol{k}_\perp, z) - v_{ts} v_\parallel J_{0s} A_\parallel(\boldsymbol{k}_\perp, z) + \frac{\tau_s}{Z_s} 2\mu B \frac{J_{1s}}{\alpha_s} \frac{\delta B_\parallel(\boldsymbol{k}_\perp, z)}{B}. \tag{4.5}$$

Weighting the gyroaveraged potential by a Maxwellian and Laguerre–Hermite transforming results in

$$\mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{\mathbf{k}_\perp} \langle \chi \rangle_{\mathbf{R}} F_{Ms} = \mathcal{J}_\ell^s \Phi \delta_{m0} - v_{ts} \mathcal{J}_\ell^s A_\parallel \delta_{m1} + \frac{\tau_s}{Z_s} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) \frac{\delta B_\parallel}{B} \delta_{m0}, \qquad (4.6)$$

with

$$\mathcal{J}_\ell^s \equiv \int_0^\infty \mathrm{d}\mu B \psi_\ell J_{0s}(\sqrt{2\mu B b_s}) \, \mathrm{e}^{-\mu B} = \frac{1}{\ell!} \left(-\frac{b_s}{2} \right)^\ell \mathrm{e}^{-b_s/2} \quad (\ell \geq 0), \qquad (4.7)$$

and $\mathcal{J}_\ell^s \equiv 0$ when $\ell < 0$ or when $\ell \geq N_\ell$, with N_ℓ the number of evolved Laguerre moments. The Fourier–Laguerre–Hermite transform of the auxiliary distribution function g is then

$$\mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{\mathbf{k}_\perp} g_s \equiv G_{\ell,m}^s(\mathbf{k}_\perp, z) = H_{\ell,m}^s - \frac{Z_s}{\tau_s} \mathcal{J}_\ell^s \Phi \delta_{m0} + \frac{Z_s v_{ts}}{\tau_s} \mathcal{J}_\ell^s A_\parallel \delta_{m1} - (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) \frac{\delta B_\parallel}{B} \delta_{m0}. \qquad (4.8)$$

The Fourier–Laguerre–Hermite transform of (2.13) then gives

$$\begin{aligned}
\frac{\partial G_{\ell,m}^s}{\partial t} &+ \mathcal{N}_{\ell,m}^s + v_{ts} \nabla_\parallel (\sqrt{m+1} H_{\ell,m+1}^s + \sqrt{m} H_{\ell,m-1}^s) \\
&+ v_{ts} [-(\ell+1)\sqrt{m+1} H_{\ell,m+1}^s - \ell\sqrt{m+1} H_{\ell-1,m+1}^s \\
&+ \ell\sqrt{m} H_{\ell,m-1}^s + (\ell+1)\sqrt{m} H_{\ell+1,m-1}^s] \nabla_\parallel \ln B \\
&+ \mathrm{i}\frac{\tau_s}{Z_s} \omega_d^\kappa [\sqrt{(m+1)(m+2)} H_{\ell,m+2}^s + (2m+1) H_{\ell,m}^s + \sqrt{m(m-1)} H_{\ell,m-2}^s] \\
&+ \mathrm{i}\frac{\tau_s}{Z_s} \omega_d^{\nabla B} [(\ell+1) H_{\ell+1,m}^s + (2\ell+1) H_{\ell,m}^s + \ell H_{\ell-1,m}^s] \\
&= \mathcal{D}_{\ell,m}^s + \mathcal{C}_{\ell,m}^{ss}.
\end{aligned} \qquad (4.9)$$

Parallel convection, including bounce motion induced by magnetic trapping in the equilibrium magnetic field, is described by the terms proportional to $\nabla_\parallel \equiv \hat{\mathbf{b}} \cdot \nabla = (\hat{\mathbf{b}} \cdot \nabla z)\partial/\partial z$. The $\nabla_\parallel H_{\ell,m}^s$ terms are evaluated spectrally by Fourier transforming in z,

$$\nabla_\parallel H_{\ell,m}^s \equiv (\hat{\mathbf{b}} \cdot \nabla z)\mathcal{F}_{k_z}^{-1}[\mathrm{i}k_z \mathcal{F}_{k_z} H_{\ell,m}^s], \qquad (4.10)$$

with

$$\mathcal{F}_{k_z} H_{\ell,m}^s = \int \mathrm{d}z \, \mathrm{e}^{-\mathrm{i}k_z z} H_{\ell,m}^s(z). \qquad (4.11)$$

Note that a spectral evaluation of this term requires an equispaced grid in z, and we also require the z coordinate to be an equal-arclength coordinate so that $(\hat{\mathbf{b}} \cdot \nabla z)$ is a constant (see §E.1) to avoid a convolution in (4.10). Additional details about this operation related to boundary conditions are given in §4.2. Toroidicity gives rise to the terms proportional to the curvature drift operator $\mathrm{i}\omega_d^\kappa = (1/B)\hat{\mathbf{b}} \times (\hat{\mathbf{b}} \cdot \nabla \hat{\mathbf{b}}) \cdot \mathrm{i}\mathbf{k}_\perp$ and the ∇B drift operator $\mathrm{i}\omega_d^{\nabla B} = (1/B^2)\hat{\mathbf{b}} \times \nabla B \cdot \mathrm{i}\mathbf{k}_\perp$ is the ∇B. Drive terms from equilibrium gradients, denoted

by $\mathcal{D}_{\ell,m}$, are given by

$$\mathcal{D}_{\ell,m=0}^s = \mathrm{i}\omega_* \left[\frac{1}{L_{ns}} \mathcal{J}_\ell^s + \frac{1}{L_{Ts}} [\ell \mathcal{J}_{\ell-1}^s + 2\ell \mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s] \right] \Phi$$

$$+ \frac{\tau_s}{Z_s} \mathrm{i}\omega_* \left[\frac{1}{L_{ns}} [\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s] + \frac{1}{L_{Ts}} [\ell \mathcal{J}_{\ell-2}^s + 3\ell \mathcal{J}_{\ell-1}^s + (3\ell+1)\mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s] \right] \frac{\delta B_\parallel}{B},$$

$$\mathcal{D}_{\ell,m=1}^s = -v_{ts}\mathrm{i}\omega_* \left[\frac{1}{L_{ns}} \mathcal{J}_\ell^s + \frac{1}{L_{Ts}} [\ell \mathcal{J}_{\ell-1}^s + (2\ell+1)\mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s] \right] A_\parallel,$$

$$\mathcal{D}_{\ell,m=2}^s = \frac{1}{\sqrt{2}} \mathrm{i}\omega_* \frac{1}{L_{Ts}} \mathcal{J}_\ell^s \Phi + \frac{\tau_s}{Z_s} \frac{1}{\sqrt{2}} \mathrm{i}\omega_* \frac{1}{L_{Ts}} [\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s] \frac{\delta B_\parallel}{B},$$

$$\mathcal{D}_{\ell,m=3}^s = -v_{ts}\sqrt{\frac{3}{2}} \mathrm{i}\omega_* \frac{1}{L_{Ts}} \mathcal{J}_\ell^s A_\parallel,$$

$$\mathcal{D}_{\ell,m>3}^s = 0, \tag{4.12}$$

where $\mathrm{i}\omega_* \equiv -\nabla x \cdot \hat{\boldsymbol{b}} \times \mathrm{i}\boldsymbol{k}_\perp = \mathrm{i}k_y$, and L_{ns} and L_{Ts} are the normalized density and temperature gradient scale lengths, respectively.

The like-species collision terms are given by

$$\mathcal{C}_{\ell,0}^{ss} = -\nu_{ss}(b_s + 2\ell + 0)H_{\ell,0}^s$$

$$+ \nu_{ss}(\sqrt{b_s}\,(\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s)\bar{u}_{\perp s} + 2[\ell \mathcal{J}_{\ell-1}^s + 2\ell \mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s]\bar{T}_s),$$

$$\mathcal{C}_{\ell,1}^{ss} = -\nu_{ss}(b_s + 2\ell + 1)H_{\ell,1}^s + \nu_{ss}\mathcal{J}_\ell^s \bar{u}_{\parallel s},$$

$$\mathcal{C}_{\ell,2}^{ss} = -\nu_{ss}(b + 2\ell + 2)H_{\ell,2}^s + \nu_{ss}\sqrt{2}\mathcal{J}_\ell^s \bar{T}_s,$$

$$\mathcal{C}_{\ell,m}^{ss} = -\nu_{ss}(b_s + 2\ell + m)H_{\ell,m}^s, \quad (m > 2), \tag{4.13}$$

with the field-particle terms given by

$$\bar{u}_{\parallel s} = \int \mathrm{d}^3 \boldsymbol{v} J_{0s} v_\parallel h_s = \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,1}^s, \tag{4.14}$$

$$\bar{u}_{\perp s} = \int \mathrm{d}^3 \boldsymbol{v} J_{1s} v_\perp h_s = \sqrt{b_s} \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s)H_{\ell,0}^s, \tag{4.15}$$

$$\bar{T}_{\parallel s} = \int \mathrm{d}^3 \boldsymbol{v} J_{0s}(v_\parallel^2 - 1)h_s = \sqrt{2} \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,2}^s, \tag{4.16}$$

$$\bar{T}_{\perp s} = \int \mathrm{d}^3 \boldsymbol{v} J_{0s}(\mu B - 1)h_s = \sum_{\ell=0}^{N_\ell} [\ell \mathcal{J}_{\ell-1}^s + 2\ell \mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s]H_{\ell,0}^s, \tag{4.17}$$

$$\bar{T}_s = \frac{1}{3}(\bar{T}_\parallel + 2\bar{T}_\perp). \tag{4.18}$$

The nonlinear terms are evaluated pseudospectrally in $(x, y, z, \mu B, m)$ space to avoid convolutions in both Fourier and Laguerre coefficients as (Mandell *et al.* 2018)

$$\mathcal{N}_{\ell,m}^s = \mathcal{L}_\ell \mathcal{H}_m \mathcal{F}_{\boldsymbol{k}_\perp}[\langle \boldsymbol{v}_\chi \rangle_{\boldsymbol{R}} \cdot \nabla h_s], \tag{4.19}$$

where the complete pseudospectral expression for this term is given in appendix D.

Taking the Fourier–Laguerre–Hermite transform of the field equations, the quasineutrality equation, (2.9), becomes

$$\sum_s \frac{Z_s^2 n_s}{\tau_s} \Phi = \sum_s Z_s n_s \int d^3 \boldsymbol{v} J_{0s} h_s = \sum_s Z_s n_s \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,0}^s, \tag{4.20}$$

the parallel component of Ampère's law, (2.10), becomes

$$k_\perp^2 A_\parallel = \frac{\beta_{\text{ref}}}{2} \sum_s Z_s n_s v_{ts} \int d^3 \boldsymbol{v} v_\parallel J_{0s} h_s = \frac{\beta_{\text{ref}}}{2} \sum_s Z_s n_s v_{ts} \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,1}^s \tag{4.21}$$

and the perpendicular component of Ampère's law, (2.11), becomes

$$\frac{\delta B_\parallel}{B} = -\frac{\beta_{\text{ref}}}{2B^2} \sum_s n_s \tau_s \int d^3 \boldsymbol{v} 2\mu B \frac{J_{1s}}{\alpha_s} h_s = -\frac{\beta_{\text{ref}}}{2B^2} \sum_s n_s \tau_s \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) H_{\ell,0}^s. \tag{4.22}$$

To solve these field equations numerically, it is more convenient to express them in terms of G rather than H, which results in

$$\sum_s \frac{Z_s^2 n_s}{\tau_s} \left(1 - \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s)^2 \right) \Phi - \sum_s Z_s n_s \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) \frac{\delta B_\parallel}{B} = \sum_s Z_s n_s \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s G_{\ell,0}^s, \tag{4.23}$$

$$\left(k_\perp^2 + \frac{\beta_{\text{ref}}}{2} \sum_s \frac{Z_s^2 n_s}{m_s} \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s)^2 \right) A_\parallel = \frac{\beta_{\text{ref}}}{2} \sum_s Z_s n_s v_{ts} \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s G_{\ell,1}^s, \tag{4.24}$$

$$\frac{\beta_{\text{ref}}}{2B^2} \sum_s Z_s n_s \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) \Phi + \left(1 + \frac{\beta_{\text{ref}}}{2B^2} \sum_s n_s \tau_s \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s)^2 \right) \frac{\delta B_\parallel}{B}$$

$$= -\frac{\beta_{\text{ref}}}{2B^2} \sum_s n_s \tau_s \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) G_{\ell,0}^s. \tag{4.25}$$

Note that while the sum $\sum(\mathcal{J}_\ell^s)^2$ could be computed analytically in the $N_\ell \to \infty$ limit as $\Gamma_0(b) = I_0(b) e^{-b}$, with I_0 the modified Bessel function of the first kind, doing so breaks the energetic consistency of the equations (Mandell *et al.* 2018); thus these terms should be evaluated as truncated sums.

4.1. *Hyperdissipation and closure*

Grid-scale dissipation terms are used to ensure numerical stability and robustness. For dissipation in configuration space, we employ a hyperviscosity term of the standard form

$$\left(\frac{\partial G_{\ell,m}}{\partial t} \right)_{\text{hyperviscosity}} = -D \left(\frac{k_\perp^2}{k_{\perp,\text{max}}^2} \right)^n G_{\ell,m}, \tag{4.26}$$

where $k_{\perp,\text{max}}$ is the largest dealiased perpendicular wavenumber in the simulation, and typical values of the variable parameters for nonlinear simulations are $D = 0.05$ and

$n = 2$. This provides a sink at the shortest wavelengths in the domain, which can be useful in nonlinear simulations to avoid spectral pileup. Unlike in other Eulerian gyrokinetic codes which include dissipation along the parallel direction via upwinding, (Kotschenreuther *et al.* 1995; Jenko & Dorland 2001; Candy *et al.* 2016; Barnes *et al.* 2019), the spectral discretization scheme used for the parallel direction in GX does not itself introduce dissipation in the parallel direction. The hypercollision operator and the parallel boundary conditions discussed below introduce some parallel dissipation. A standard parallel hyperdissipation operator of the form $k_\parallel^{2n}$ could also be included, but we find that this is not necessary for the linear and nonlinear cases that we consider.

Fine-scale structure in velocity space must also be regulated. The Dougherty collision operator fulfils this purpose by acting increasingly strongly on higher Laguerre and Hermite moments, limiting their amplitude. Thus, for a given collisionality, there is a physical cutoff at some ℓ_{cut} and m_{cut} beyond which fine scales in velocity space are completely wiped out by collisions. At this point, closure of the Laguerre–Hermite moment series by simple truncation (setting $G_{\ell,m} = 0$ for $\ell \geq \ell_{cut}$ or $m \geq m_{cut}$) is justified. This is the simplest high-resolution closure, but not the only option. One can also obtain an asymptotically correct collisional closure by assuming the collision term becomes dominant in the unresolved moment equations (Zocco, Helander & Connor 2015; Loureiro *et al.* 2016; Jorge *et al.* 2017; Frei *et al.* 2022*b*).

In the limit of low collisionality or lower Laguerre–Hermite resolution, however, the closure situation is more complicated. Unresolved moments are not expected to be negligible at collisionalities of interest, so closure by truncation will generally give poor results. One possible approach is to follow the collisionless closure approach pioneered by Hammett & Perkins (1990), which was used and extended in the gyrofluid models of Dorland & Hammett (1993), Beer & Hammett (1996), Snyder & Hammett (2001) and Smith (1997) to model parallel, toroidal and nonlinear finite-Larmor-radius (FLR) phase mixing. However, the development of generalized closures in toroidal geometry for a system with an arbitrary number of moments is complicated and beyond the scope of the current work. Instead, we have found that employing a hypercollision operator that provides a sink at large m allows well-behaved results even with simple closure by truncation at relatively low resolution. The operator takes the generic form

$$\left(\frac{\partial G_{\ell,m}}{\partial t}\right)_{\text{hyp-coll}} = -\nu_{\text{hyp}} m^p G_{\ell,m} \quad (m > 2). \tag{4.27}$$

We ensure that the operator conserves density, momentum and energy by only operating on moments with $m > 2$. Similar hypercollision operators (in Hermite space only) have been used in other contexts (Parker 2015; Parker & Dellar 2015; Loureiro *et al.* 2016), and Parker (2015) also develops theory of a 'hypercollisional plateau' where the behaviour of the operator is insensitive to the choice of the variable parameters. In our operator, the coefficient ν_{hyp} is chosen to approximately remove the same amount of fluctuation energy from the largest ms as would be removed by parallel phase mixing. The details are provided in § B, and the result is

$$\nu_{\text{hyp}} = 2.5 f_{\text{hyp}} \frac{p + 1/2}{m_{\max}^{p+1/2}} |k_\parallel| v_{ts}, \tag{4.28}$$

where f_{hyp} is an adjustable coefficient that is typically set to unity. The $|k_\parallel|$ operator is evaluated spectrally by Fourier transforming in z. The exponent p is chosen to be $p = N_m/2$, which ensures that only the highest several Hermite modes are strongly damped, roughly independent of N_m. While one could consider a similar hyperdissipation

mechanism for large ℓ, we find that Laguerre hypercollisions are unnecessary even when using relatively low Laguerre resolution ($N_\ell \sim 4$) for problems of interest because phase mixing in $v_\parallel$ dominates.

4.2. *Boundary conditions*

In the flux tube approach we assume statistical periodicity in the perpendicular (x, y) plane. In the parallel direction, we use the 'twist-and-shift' boundary condition (Beer *et al.* 1995), which in the spectral representation can be accomplished by linking modes with different k_x values into an extended z domain, such that

$$H_{\ell,m}(k_x, k_y, z) = H_{\ell,m}(k_x + \delta k, k_y, z + 2\pi), \tag{4.29}$$

with $\delta k = 2\pi k_y \hat{s}$. We can then perform a Fourier transform on each extended z domain to evaluate the parallel streaming terms as given by (4.10).

A Fourier representation naturally enforces a periodic boundary condition on h at the ends of each extended domain, but this is not the physically correct boundary condition for non-zonal modes, which should have a zero incoming boundary condition in the infinite ballooning limit. Following Kotschenreuther *et al.* (1995), we would like to approximate this by enforcing a zero incoming boundary condition on h at the ends of the extended domain, but this boundary condition cannot be enforced directly in Hermite space. Instead, we use a wave-absorbing layer (see e.g. Durran (2010), and references within) near the ends of each extended z domain for the non-zonal components of h to damp out perturbations leaving and re-entering the domain due to the natural periodicity of the Fourier representation. Following Beer (1995), we use a damping filter that is zero in most of the domain and ramps up smoothly near the ends of the extended domain as $d(\hat{z}) = A[1 - 2\hat{z}^2/(1 + \hat{z}^4)]$, where $\hat{z} = (z - z_{\text{end}})/z_{\text{width}}$ is a normalized distance from the ends z_{end}. The width z_{width} and the scaling factor A are adjustable parameters; in appendix C we have performed a convergence study and found $z_{\text{width}} = L_z/8$ and $A = 0.1/\Delta t$ to work well, where L_z the total length of the extended domain. The wave-absorbing damping operation can then be expressed as

$$\left(\frac{\partial G_{\ell,m}}{\partial t}\right)_{\text{ends}} = -d(\hat{z})H_{\ell,m} \quad (|\hat{z}| < 1, \ k_y \neq 0). \tag{4.30}$$

The zonal modes ($k_y = 0$) are not filtered because these modes should be physically periodic, so the natural periodic boundary condition from the Fourier representation is the correct one. We have found this wave-absorbing layer to be particularly necessary for simulations with kinetic electrons.

The standard twist-and-shift boundary condition can be ill-suited for low magnetic shear regions, particularly in stellarators. Martin *et al.* (2018) have developed a generalization of the twist-and-shift boundary condition that alleviates some of these issues by using the integrated local shear (rather than global shear) in the 'twist'. This generalized twist-and-shift boundary condition can be used in GX for stellarator geometries with stellarator symmetry. Investigation of a non-twisting flux-tube boundary condition (Ball & Brunner 2021) using GX is also in progress.

Martin *et al.* (2018) prescribe methods to choose the parallel domain length to enforce exact periodicity or continuous magnetic drifts at the ends of the flux tube. In addition to these options, we adopt an additional criterion for choosing the parallel domain length: to enforce a desired aspect ratio for the perpendicular computational domain. When using the standard twist-and-shift boundary condition, the aspect ratio is quantized as $L_x/L_y = J/(2\pi|\hat{s}|)$, with J a non-zero integer. This results in $L_x \propto L_y/\hat{s}$, which makes resolution

requirements challenging for $\hat{s} \ll 1$. In the generalized twist-and-shift boundary condition of Martin *et al.* (2018), the perpendicular aspect ratio of the flux tube is given by

$$\frac{L_x}{L_y} = J \left[\frac{|\nabla x|^2}{2|\nabla x \cdot \nabla y|} \right]_{z_{\text{end}}}, \tag{4.31}$$

where z_{end} is the end of the flux tube and J is again a non-zero integer. In a stellarator, this quantity varies as a function of the parallel domain length (see figure 7 of Martin *et al.* (2018)). The generalized twist-and-shift boundary condition thus provides the freedom to choose z_{end} to obtain the desired aspect ratio (for some integer J). Using this prescription to enforce an aspect ratio of order unity can then alleviate the resolution requirements for geometries with small magnetic shear.

5. Timestepping schemes

GX uses explicit time integration methods to advance the system. Along with several standard Runge–Kutta (RK) and strong stability-preserving RK (SSP RK) schemes (Gottlieb, Shu & Tadmor 2001), we also provide an option for the SSP 10-stage fourth-order method of Ketcheson (2008) (a low-storage method ideal for the memory constraints of GPU computing). Additionally, the sparse banded structure of the linear system of (4.9) and the fact that the field equations only involve the lowest-order moments (instead of requiring integration over all velocity space) could allow the development of efficient implicit–explicit (IMEX) schemes for electron dynamics that avoid the timestep restriction of the fast electron motion; this will be a topic for future work. We have used the standard RK3 (third-order) scheme for all of the benchmark calculations presented in § 6.

In all of our timestepping schemes the timestep size Δt is constrained by the stability region of the scheme. While in principle the linear eigenvalues of the system could be computed exactly and used to constrain the timestep to ensure that all eigenvalues lie within the stability region, we instead estimate the maximum (linear and nonlinear) frequencies on the grid in each of the (x, y, z) directions via

$$\omega_{x,\text{max}} \approx \max \left(k_{x,\text{max}} \left(\frac{\tau_s}{Z_s} \boldsymbol{v}_d \cdot \nabla x \right)_{\text{max}}, \ k_{x,\text{max}} (\boldsymbol{v}_E \cdot \nabla x)_{\text{max}} \right), \tag{5.1}$$

$$\omega_{y,\text{max}} \approx \max \left(k_{y,\text{max}} \left(\frac{\tau_s}{Z_s} \boldsymbol{v}_d \cdot \nabla y \right)_{\text{max}}, \ k_{y,\text{max}} (\boldsymbol{v}_E \cdot \nabla y)_{\text{max}} \right), \tag{5.2}$$

$$\omega_{z,\text{max}} \approx \max((\hat{\boldsymbol{b}} \cdot \nabla z) k_{z,\text{max}} v_{\parallel,\text{max}} v_{t,\text{max}}, \omega_{A,\text{max}}), \tag{5.3}$$

with

$$\omega_{A,\text{max}} = \frac{(\hat{\boldsymbol{b}} \cdot \nabla z) k_{z,\text{max}} v_{\text{te}}}{\sqrt{\dfrac{\beta_{\text{ref}} n_e \tau_e}{2} \dfrac{m_i}{m_e} + k_{\perp,\text{min}}^2 \rho_s^2}}, \tag{5.4}$$

the maximum shear Alfvén frequency on the grid. In the electrostatic limit ($\beta_{\text{ref}} = 0$) this reduces to the high-frequency ω_H mode (Lee 1987). Here, $k_{z,\text{max}}$ is the largest parallel wavenumber in the simulation and $k_{\perp,\text{min}}$ is the smallest perpendicular wavenumber. To evaluate $v_{\parallel,\text{max}}$ and $(\mu B)_{\text{max}}$ we use the maximum value of the Gauss–Hermite and Gauss–Laguerre collocation grids, respectively, which depend on the number of spectral modes used in the calculation, N_m and N_ℓ.

The constraint on the maximum stable timestep is then approximately given by

$$(\omega_{x,\max} + \omega_{y,\max} + \omega_{z,\max})\Delta t \lesssim Cs, \qquad (5.5)$$

where s is dependent on the stability region for the particular timestepping scheme (for example, $s = 1.73$ for RK3, and $s = 2.82$ for RK4), and $0 < C \leq 1$ is a user-defined scaling factor. Formally, the stability constraint with $C = 1$ is only valid for linear modes that do not grow or decay; thus we typically use $C \sim 0.5$–0.9 to ensure stability in the full nonlinear system with a spectrum of growing and damped modes.

6. Numerical benchmarks

In this section we show a variety of linear and nonlinear benchmarks of GX by comparing results with widely benchmarked gyrokinetic codes (GS2, stella, GENE) in both tokamak and stellarator geometry. In all cases, both linear and nonlinear, an initial value problem is solved. The numerical resolution and associated parameters used in each case are given in appendix G.

6.1. *Tokamak benchmarks: Cyclone base case*

For benchmarks in tokamak geometry, we choose a configuration based on the 'Cyclone base case' (CBC), a widely used benchmark case with concentric circular flux surfaces (Dimits *et al.* 2000). For this we use a Miller local equilibrium (Miller *et al.* 1998) with $R_0/a = 2.78$, $r/a = 0.5$, $q = 1.4$, $\hat{s} = 0.8$, $\kappa = 1.0$, $\delta = 0.0$. Unless otherwise noted, the normalized gradient scale lengths are taken to be $a/L_{ni} = a/L_{ne} = 0.8$, $a/L_{Ti} = a/L_{Te} = 2.49$, and we also take $T_i = T_e$. In cases with kinetic electrons we take $m_i/m_e = 3670$ (deuterium ions). This series of benchmarks is inspired by the set of tokamak benchmarks chosen for stella by Barnes *et al.* (2019).

6.1.1. *Linear results*

We first make a linear comparison taking a Boltzmann (adiabatic) electron response, so that the quasineutrality equation (2.9) reduces to

$$\int d^3 \boldsymbol{v} \, \langle h_s \rangle_{\boldsymbol{r}} = \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell H_{\ell,0} = \frac{T_e}{T_i}[\Phi - \langle\langle\Phi\rangle\rangle] + \Phi, \qquad (6.1)$$

where $\langle\langle\Phi\rangle\rangle$ denotes the flux-surface average of Φ. In figure 1 we show normalized growth rates (figure 1a) and real frequencies (figure 1b) as a function of the normalized binormal wavenumber $k_y \rho_i$. GX results are shown with open blue circles connected by dashed lines, while results from GS2 are shown with yellow squares. The agreement between the two codes is very good across the range of unstable k_y values. Here we included a small number of collisions in both codes to provide velocity-space regularization, taking the normalized ion collision frequency to be $\nu_{ii} = 10^{-2} v_{ti}/a$. Hypercollisions are not used in GX in this case. We next make linear comparisons using kinetic electrons between GX, GS2 and stella. Figure 2(a) shows normalized growth rates and real frequencies as a function of $k_y \rho_i$ at ion scales in the electrostatic limit, with $\beta_{\mathrm{ref}} = 10^{-5}$ (we retain electromagnetic fluctuations even at low β to alleviate the timestep restriction from the electrostatic ω_H mode (Lee 1987)). In this case we take $\nu_{ee} = 10^{-2} v_{ti}/a$ and $\nu_{ii} = 1.65 \times 10^{-4} v_{ti}/a$; for best comparison with GX, we use the stella's Dougherty collision model option, even though stella also includes a more accurate Fokker–Planck
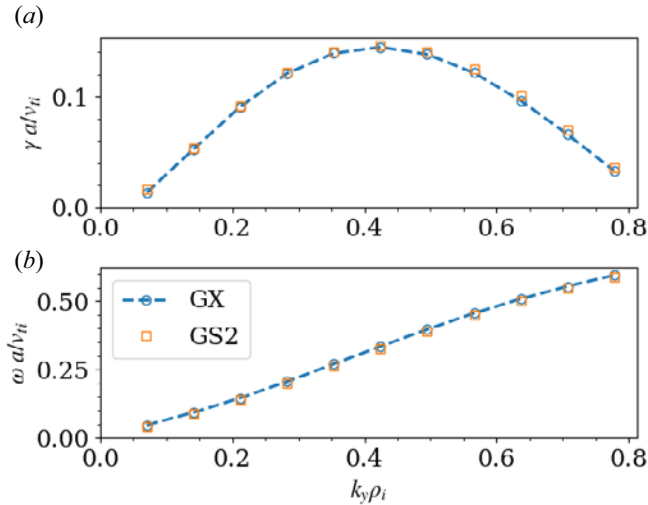
FIGURE 1. Normalized growth rates (*a*) and real frequencies (*b*) as a function of the normalized binormal wavenumber $k_y\rho_i$ from GX (blue circles and dotted lines) and GS2 (yellow squares) for 'Cyclone base case' (CBC) parameters using a Boltzmann electron response.

collision operator (which produces results that agree even more closely with GS2 in this case). Hypercollisions were again not used in GX. We observe excellent agreement between GX and stella for the ITG branch ($\omega > 0$), but the agreement is not as good for the trapped electron mode (TEM) branch ($\omega < 0$). If there is a weakness in our scheme it is accurately capturing linear TEM instability, which requires resolving sharp trapped–passing boundaries in velocity space (which is not optimal in $v_{\parallel}$, μ coordinates) and is sensitive to the details of the collision operator (our Dougherty model collision operator is not as accurate as the collision operators in other standard gyrokinetic codes like GS2). Indeed, even getting this level of agreement required 128 Hermite modes in GX (for more resolution details, see appendix G). A more detailed study of TEM modes is left to future work, where we will explore more accurate collision operators and methods to enhance convergence. Note that Frei *et al.* (2021, 2022*a*) have demonstrated the success of the Laguerre–Hermite method in resolving TEM modes with a more accurate collision operator. We also show growth rates and frequencies as a function of $k_y\rho_i$ at electron scales in figure 2(*b*), again in the electrostatic limit. There is once again strong agreement between GX and GS2 for these electron-temperature-gradient (ETG) modes.

We finally perform an electromagnetic linear benchmark of the transition from ITG instability to kinetic ballooning mode (KBM) instability. In this benchmark we include perpendicular magnetic fluctuations via finite $A_{\parallel}$ but we neglect parallel magnetic fluctuations ($\delta B_{\parallel} = 0$). Taking $k_y\rho_i = 0.3$, figure 3 shows normalized growth rates and real frequencies as a function of β_{ref}, the reference beta. As β_{ref} increases, both GX and GS2 agree well and show moderate stabilization of the ITG mode until the transition to KBM around $\beta_{\text{ref}} = 1.3\,\%$. Additional work studying reactor-relevant electromagnetic instabilities (e.g. microtearing and toroidal Alfvén eigenmodes) with GX will be reported in future publications; successful modelling of collisionless microtearing with a Laguerre–Hermite formulation has been shown by Frei *et al.* (2023).
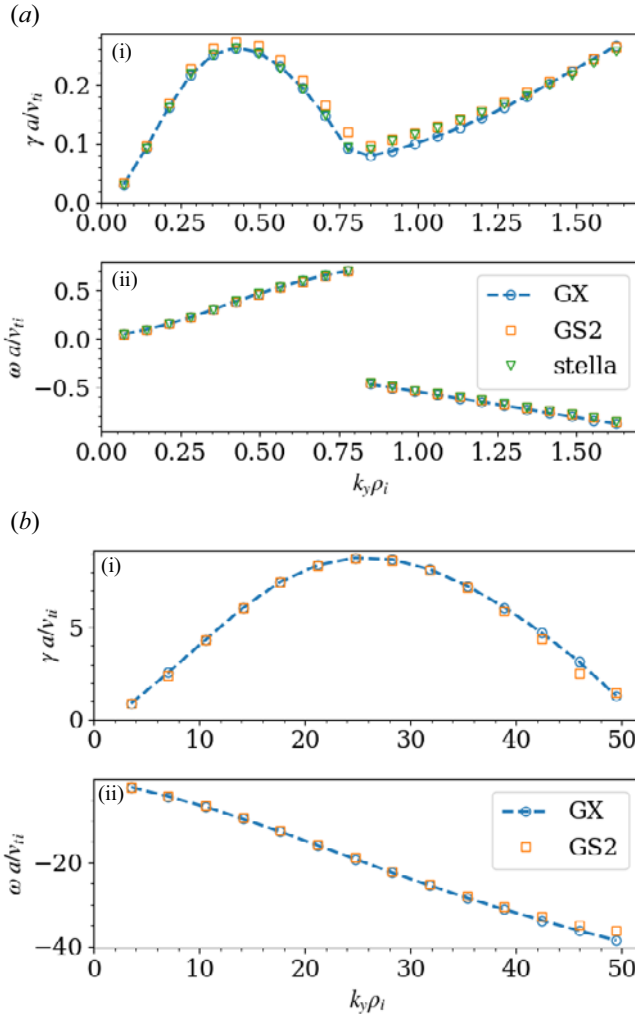
FIGURE 2. Normalized growth rates (i) and real frequencies (ii) as a function of the normalized binormal wavenumber $k_y \rho_i$ from GX (blue circles and dotted lines), GS2 (yellow squares) and stella (green inverted triangles) for CBC parameters using a kinetic electron response in the electrostatic limit ($\beta_{\text{ref}} = 10^{-5}$). Ion scales (ITG and TEM) are shown in (a) and electron scales (ETG) are shown in (b).

6.1.2. *Nonlinear results*

We first perform nonlinear CBC calculations using a Boltzmann electron response. Figure 4(a) shows time traces of the ion heat flux in gyro-Bohm units (computed using the expressions in appendix F), showing very good agreement in the time-averaged heat flux amongst GX and GS2. For GX, we show a case with moderate velocity resolution, $(N_\ell, N_m) = (8, 16)$, along with a coarse resolution case with $(N_\ell, N_m) = (4, 8)$. Additional details about velocity-space convergence for this case are given in § 7. We also show in figure 4(b) a scan of the ITG parameter, a/L_{Ti}, which again shows excellent agreement between GX and GS2 for all cases at both resolutions. The agreement at $a/L_T = 1.5$ is particularly notable because this is in the Dimits shift regime, where the system is linearly unstable to ITG but turbulence is suppressed by zonal flow dynamics. This regime was
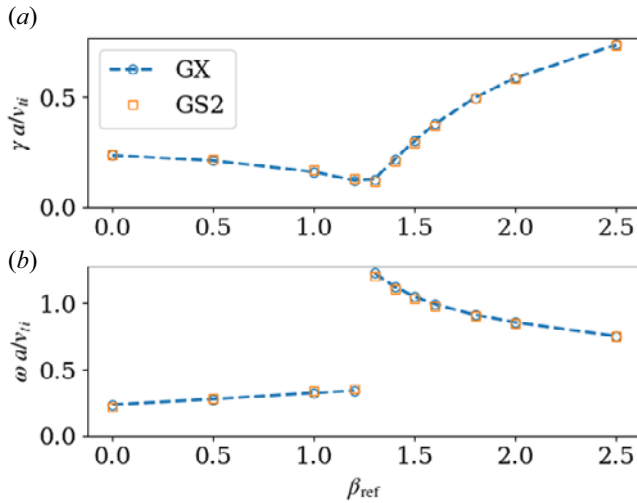
FIGURE 3. Normalized growth rates (*a*) and real frequencies (*b*) as a function of the reference beta β_{ref} from GX (blue circles and dotted lines) and GS2 (yellow squares) for CBC parameters using a kinetic electron response, taking $k_y \rho_i = 0.3$. Both codes show the transition from ITG instability at low β_{ref} to kinetic ballooning mode (KBM) instability at high β_{ref}.

especially troublesome for the Beer gyrofluid model (Beer & Hammett 1996; Dimits *et al.* 2000), but we see that by extending that model to more moments (and neglecting the collisionless closure schemes) via our Laguerre–Hermite approach we can recover the Dimits shift with $(N_\ell, N_m) = (4, 8)$. Hoffmann *et al.* (2023*a*) have studied the convergence of the Laguerre–Hermite basis for this case and also found that $(N_\ell, N_m) \geq (4, 8)$ is required to accurately capture the Dimits shift. Further from marginality, Hoffmann *et al.* (2023*a*) have found that even lower velocity resolution can be used with reasonable accuracy, which is consistent with the convergence study in § 7. Here, both GX and GS2 included a small number of collisions, taking the normalized ion collision frequency to be $\nu_{\text{ii}} = 10^{-2} \, \nu_{\text{ti}}/a$. Hypercollisions were included in the GX calculations, with details given in § G.3.

Nonlinear calculations with kinetic electrons are presented in figure 5, which shows excellent agreement amongst GX and GS2 for the time average of the heat flux in both the ion and electron channels. For GX, we show a case with moderate velocity resolution, $(N_\ell, N_m) = (4, 16)$, along with a higher resolution case with $(N_\ell, N_m) = (16, 32)$. Additional details about velocity-space convergence for this case are given in § 7. Here we took $\beta_{\text{ref}} = 10^{-3}$ along with $\nu_{\text{ee}} = 10^{-2} \nu_{\text{ti}}/a$ and $\nu_{\text{ii}} = 1.65 \times 10^{-4} \nu_{\text{ti}}/a$ for collisions in all codes, and hypercollisions were again used in GX (see § G.4).

6.2. *Stellarator benchmarks: W7-X*

For benchmarks in stellarator geometry, we choose a magnetic configuration from W7-X that has recently been used by González-Jerez *et al.* (2022) to benchmark stella and GENE. The numerical equilibrium is obtained from VMEC (Hirshman & Whitson 1983), and we choose the so-called bean flux tube with $\alpha_0 = 0$. To study ITG-driven instabilities and turbulence we take $a/L_{Ti} = 3$ and $a/L_n = 1$. The generalized twist-and-shift boundary condition of Martin *et al.* (2018) is used.
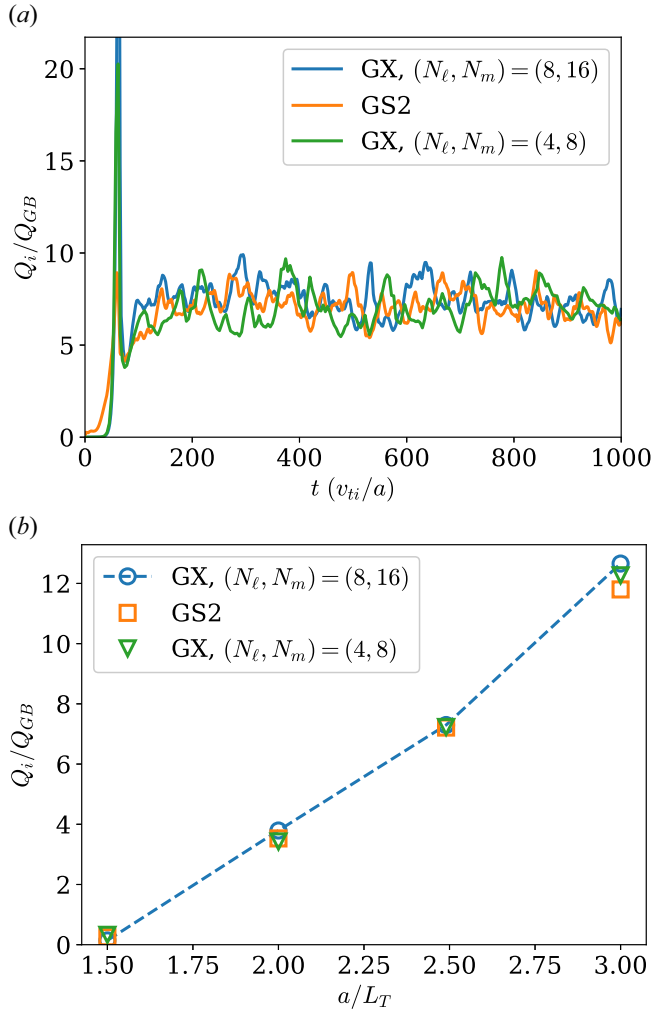
(*a*)

(*b*)

FIGURE 4. (*a*) Time traces of ion heat flux Q_i, normalized to gyro-Bohm units, $Q_{GB} = n_i T_i v_{ti} \rho_i^2 / a^2$, for the CBC with a Boltzmann electron response from GX with moderate (blue) and coarse (green) velocity resolution, and GS2 (yellow). (*b*) Time-averaged gyro-Bohm-normalized ion heat flux, Q_i/Q_{GB}, as a function of the normalized inverse ITG scale length, a/L_{Ti}.

6.2.1. *Linear results*

We first make a linear comparison between GX and stella, taking a Boltzmann electron response. Figure 6 shows the normalized growth rates (figure 6*a*) and real frequencies (figure 6*b*) for a range of k_y modes, with strong agreement between the two codes. Both codes use Dougherty model collisions with $v_{ii} = 0.01\, v_{ti}/a$. This test corresponds to figure 5 of González-Jerez *et al.* (2022). As observed in González-Jerez *et al.* (2022), the discontinuity in the frequency is due to a change in mode structure for two branches of ITG. Both branches have very similar growth rates at $k_y \rho_i = 1.0$, resulting in a small but likely inconsequential disagreement between the codes in the frequency of the fastest-growing mode.
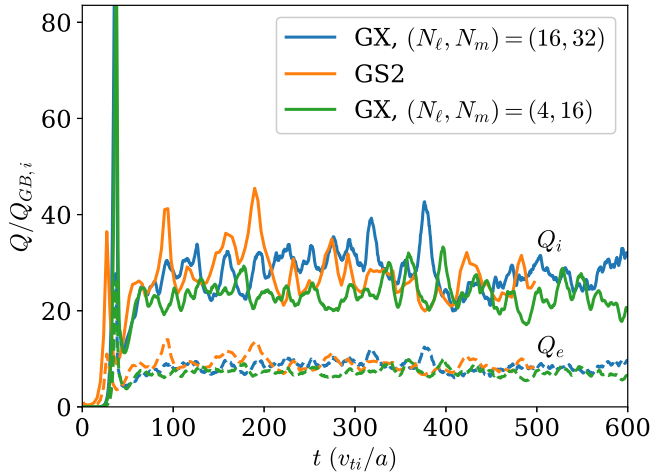
FIGURE 5. Time traces of ion (solid) and electron (dashed) heat flux, normalized to ion gyro-Bohm units, for the CBC with kinetic electrons from GX with high velocity resolution (blue) and moderate velocity resolution (green) and GS2 (yellow).
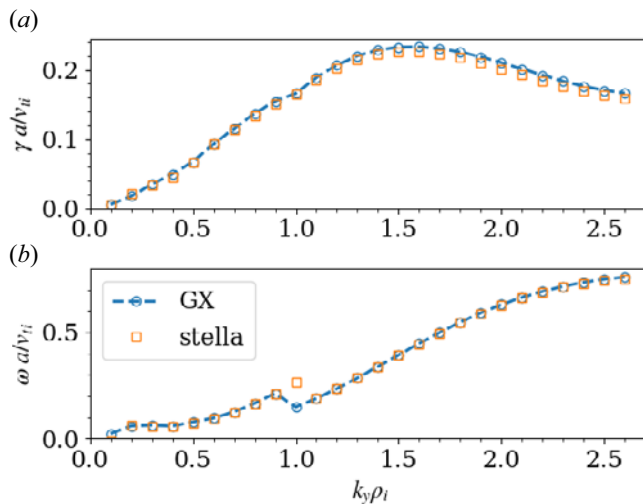

FIGURE 6. Normalized growth rates (*a*) and real frequencies (*b*) as a function of the normalized binormal wavenumber k_y from GX (blue circles and dotted lines) and stella (yellow squares) for W7-X bean flux-tube geometry and ITG parameters using a Boltzmann electron response.

6.2.2. *Nonlinear results*

Moving on to the nonlinear version of this test case, figure 7 shows time traces of ion heat flux from GX along with the time traces from stella and GENE, which were taken directly from the supplementary data made available by González-Jerez *et al.* (2022) (and corresponding to figure 12 in that work; however, note the factor of $\sqrt{2}$ difference in the definition of v_{ti}, and hence a factor of $2\sqrt{2}$ difference in Q_{GB}, in GX relative to the definitions used in stella and GENE). Both stella and GENE used rather high velocity resolution for this case, with $N_{v_{\parallel}} = 60$, $N_{\mu} = 24$. For GX we show a case with moderate velocity-space resolution, $(N_{\ell}, N_m) = (8, 16)$, and a coarse
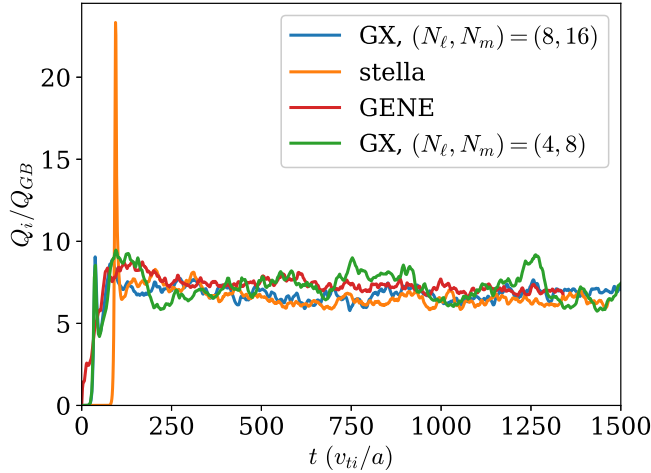
FIGURE 7. Time traces of ion heat flux Q_i, normalized to gyro-Bohm units, $Q_{GB} = n_i T_i v_{ti} \rho_i^2 / a^2$, for the W7-X bean flux-tube geometry with a Boltzmann electron response. Results are shown from GX with moderate (blue) and coarse (red) velocity resolution, stella (yellow) and GENE (green).

resolution case with $(N_\ell, N_m) = (4, 8)$. Both of these agree well with the higher-resolution cases from both stella and GENE, which indicates that the promising Laguerre–Hermite convergence observed in the tokamak benchmarks carries over to stellarators. Additionally, the wallclock time for the GX cases was 90 min (on four GPUs) in the moderate resolution case and 28 min (on one GPU) in the coarse case.

7. Velocity-space convergence and GPU performance

Part of the motivation for the Laguerre–Hermite pseudospectral approach is the ability to successfully run at low velocity-space resolution without uncontrolled approximations, since in the lowest-resolution limit the system corresponds to established gyrofluid models like the one of Beer & Hammett (1996), albeit without Beer & Hammett's collisionless closure schemes. Another motivation for the pseudospectral approach is its fit for GPU computing, since the spectral algorithm relies heavily on fast transform methods that are well-optimized on GPUs, and the memory requirements of the algorithm are low enough to fit a problem onto one or a few GPUs. We also make exclusive use of single-precision arithmetic, which is often sufficient for turbulence calculations (Maurer *et al.* 2020); there are no matrix inversions in our algorithm that would require higher precision, and we do not presently target simulations spanning both ion and electron scales that may stress the limits of single precision. The success of the benchmarks presented in the previous section is an additional indication that single-precision arithmetic is sufficient for a wide range of linear and nonlinear calculations.

In this section we examine the velocity-space convergence for the nonlinear CBC calculations with GX along with the performance of the calculations on one or several GPUs. Convergence and performance are complementary factors towards enabling first-principles transport calculations that are fast enough to be used for fusion reactor design along with wide-ranging parameter scans for physics discovery. In this section we will present results detailing convergence and performance for the nonlinear CBC benchmark cases from § 6.1.2.

N_ℓ	N_m	Q_i/Q_{GB}	wallclock (min)	time/step (s)	$\langle \Delta t \rangle$	N_{gpu}
16	32	7.8 ± 0.8	58	0.033	0.009	8
⋆ 8	16	7.3 ± 0.9	15	0.017	0.019	4
6	12	7.7 ± 0.8	11	0.017	0.026	2
⋆ 4	8	7.2 ± 0.8	5.5	0.014	0.041	1
4	6	7.0 ± 0.9	3.5	0.011	0.053	1
4	4	5.9 ± 0.7	2.1	0.0089	0.071	1
3	8	9.1 ± 0.8	4.3	0.011	0.042	1

TABLE 1. Velocity-space convergence and performance for the nonlinear CBC with Boltzmann electrons. Here N_m is the number of Hermite modes and N_ℓ is the number of Laguerre modes. Other resolution parameters were $N_x = 192$, $N_y = 64$, $N_z = 24$. Accurate ion heat flux calculations can be obtained with resolution as coarse as $(N_\ell, N_m) \geq (4, 6)$ (above the double bar). Each simulation was run to $t = 1000a/v_{ti}$, and we report the total wallclock time in minutes, the time per timestep in seconds, and the average timestep size $\langle \Delta t \rangle$ (normalized to a/v_{ti}), which changes with N_ℓ and N_m due to linear stability constraints. The number of NVIDIA A100 GPUs used for each calculation is listed in the final column. The resolutions shown in figure 4 are marked with a ⋆.

Starting with the case with Boltzmann electrons, table 1 shows the results of several GX calculations, with coarsening velocity resolution as we progress down the table. For this problem the convergence of the Laguerre–Hermite basis is quite remarkable, allowing accurate results with resolution as coarse as $(N_\ell, N_m) = (4, 6)$. This is still a factor of four more moments than used in the Beer & Hammett (1996) six-moment gyrofluid model, but it is quite coarse relative to typical gyrokinetic calculations with $O(10–100)$ velocity grid points. Thus, the flexibility of the Laguerre–Hermite representation to smoothly interpolate between gyrofluid and gyrokinetic resolution enables the full nonlinear CBC calculation to be run accurately in less than 4 min on a single GPU, a remarkable result.

When we look at the Laguerre and Hermite spectra for these cases shown in figure 8, where $W_{g,s}(\ell, m) = \int |G_{\ell,m}^s|^2 \, dx \, dy \, dz$ is the free energy in each moment, and $W_{g,s}(\ell) = \sum_m W_{g,s}(\ell, m)$ and $W_{g,s}(m) = \sum_\ell W_{g,s}(\ell, m)$, we can see why the convergence is very good: even for the coarsest resolution, the spectra at the scales that contribute dominantly to the heat flux (small ℓ and m) are nearly identical. Note that hypercollisions are being used in these calculations, and this is helping to enhance convergence. Without hypercollisions (not shown), the results are only accurate with $(N_\ell, N_m) \geq (6, 12)$. This means that hypercollisions enable a five-fold reduction in simulation (wallclock) time for this case.

We perform a similar convergence study for the CBC with kinetic electrons, as shown in table 2. As above, $N_\ell = 4$ is sufficient for accurate heat flux predictions in both ion and electron channels. However, Hermite convergence is not as strong here, with reasonable accuracy (within 15 % of the highest resolution case) requiring $N_m \gtrsim 16$. Examining the Laguerre and Hermite free energy spectra for these cases in figure 9, we see more structure in the Hermite spectra, especially for the electrons. The oscillatory nature of the electron spectrum, with higher amplitudes in even Hermite modes than odd, could be an indication that the toroidal drifts (which couple every other Hermite mode) are playing a strong role in the dynamics. This is consistent with the results of analytical and numerical convergence studies of the Laguerre–Hermite basis associated with toroidal drifts by Frei *et al.* (2022a, 2023). Despite this structure, reasonably accurate results can be achieved in 159 min on
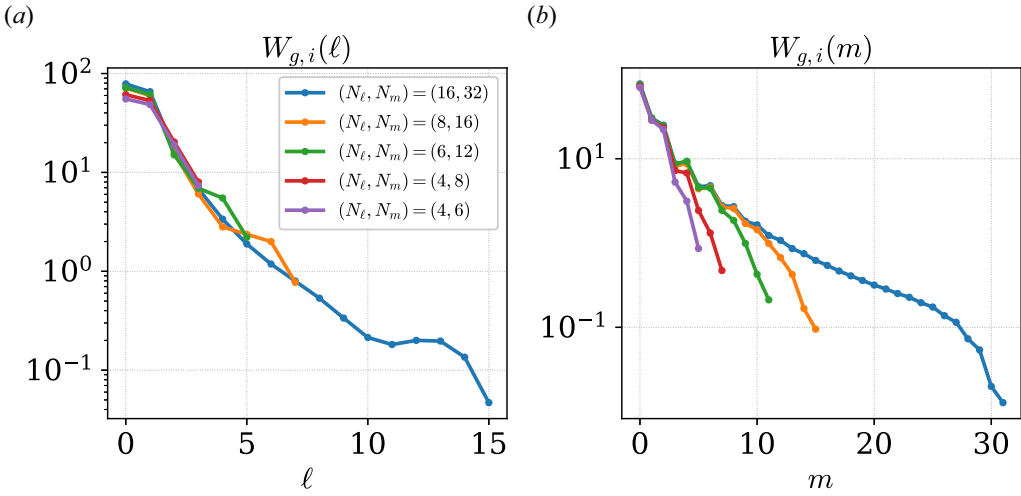
(*a*)

$W_{g,i}(\ell)$

(*b*)

$W_{g,i}(m)$

FIGURE 8. Laguerre (*a*) and Hermite (*b*) free energy spectra, W_g, for the nonlinear CBC with a Boltzmann electron response with varying velocity-space resolution. All the spectra are nearly identical at the scales that contribute dominantly to the heat flux (small ℓ and m), which is consistent with the excellent convergence observed in table 1.

N_ℓ	N_m	Q_i/Q_{GB}	Q_e/Q_{GB}	wallclock (min)	time/step (s)	$\langle\Delta t\rangle$	N_{GPU}
⋆ 16	32	27.8 ± 4.0	8.4 ± 1.2	900	0.068	0.00076	8
8	32	29.4 ± 3.3	8.8 ± 1.0	448	0.034	0.00076	8
4	32	24.8 ± 3.3	7.8 ± 1.0	255	0.020	0.00077	8
8	16	25.1 ± 3.0	7.6 ± 0.9	277	0.032	0.0012	4
⋆ 4	16	23.6 ± 3.2	7.3 ± 0.9	159	0.019	0.0012	4
8	8	18.6 ± 2.1	6.4 ± 0.7	159	0.03	0.0019	2
4	8	16.2 ± 2.2	5.7 ± 0.7	84	0.016	0.0019	2

TABLE 2. Velocity-space convergence and performance for the nonlinear CBC with kinetic electrons. Here N_m is the number of Hermite modes and N_ℓ is the number of Laguerre modes. Other resolution parameters were $N_x = 192$, $N_y = 64$, $N_z = 24$. Reasonably accurate heat flux calculations (within 15 % of the highest resolution case) can be obtained with resolution as coarse as $(N_\ell, N_m) \geq (4, 16)$ (above the double bar). Each simulation was run to $t = 600a/v_{ti}$, and we report the total wallclock time in minutes, the time per timestep in seconds, and the average timestep size $\langle\Delta t\rangle$ (normalized to a/v_{ti}). The number of NVIDIA A100 GPUs used for each calculation is also shown. The resolutions used for the GX calculations shown in figure 5 are marked with a ⋆.

four GPUs using $(N_\ell, N_m) = (4, 16)$. This is quite good for a nonlinear, electromagnetic gyrokinetic ITG simulation with real-mass-ratio kinetic electrons. Additional details about multi-GPU scaling are presented in § 7.1.

Note that using more Hermite modes not only adds to the size of the problem, but also makes the Courant–Friedrichs–Levy (CFL) constraint on the timestep more restrictive because the maximum parallel velocity $v_{\|max}$ on the Hermite collocation grid increases with the number of Hermite modes. In future work we will try to alleviate this issue by constraining the maximum velocity on the collocation grid. However, a commonly used
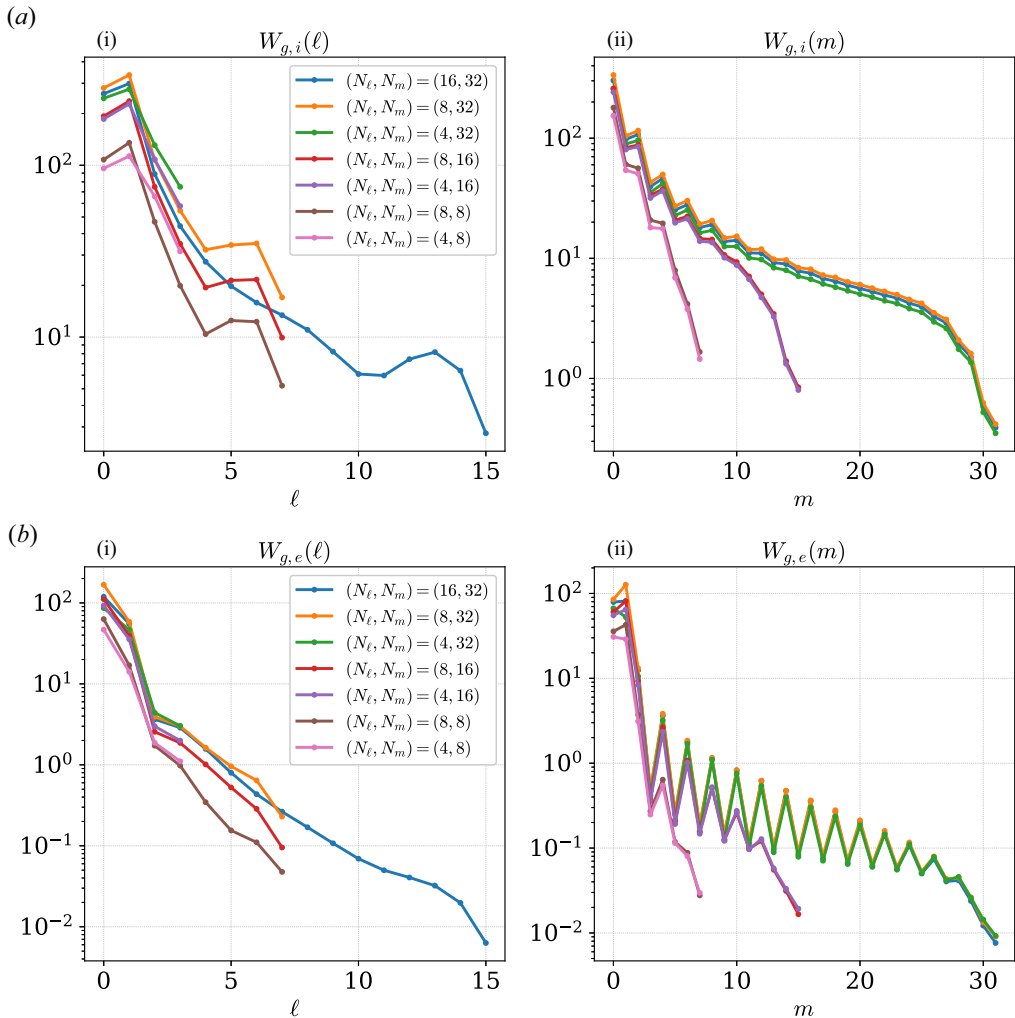
(a)

(b)

FIGURE 9. Laguerre (i) and Hermite (ii) spectra for ions (*a*) and electrons (*b*) for the nonlinear CBC with kinetic electrons with varying velocity-space resolution.

method of scaling the argument of the Hermite polynomials to reduce the extents of the collocation grid is ill-suited to our problem, since the resulting projection of a Maxwellian onto the scaled basis often diverges (Fok, Guo & Tang 2001). Instead, an approach similar to that used by Candy *et al.* (2016) to bound the energy grid in CGYRO could be used here.

7.1. *GPU scaling studies*

For single-GPU calculations, the GPU threading architecture effectively enables dynamic parallelism over the entire phase space. In figure 10, we show that for a nonlinear CBC calculation with Boltzmann electrons, the algorithm scales roughly linearly with the number of radial grid points, N_x. This is better than the $O(N \log N)$ cost that would be expected when the fast Fourier transform (FFT) in the pseudospectral algorithm dominates the cost. This is also better scaling than other gyrokinetic algorithms that require $O(N^2)$ operations due to matrix inversion.
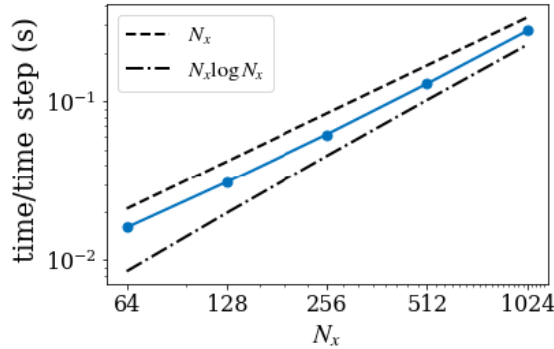
FIGURE 10. Scaling of time per timestep (in seconds) versus the number of radial grid points, N_x, for nonlinear CBC with Boltzmann electrons on a single GPU. Ideal linear scaling is shown with the dashed line, while the dot–dashed line shows $N_x \log N_x$ scaling (the expected scaling when FFTs dominate the algorithm).

When considering multi-GPU parallelization, the cost of inter-GPU memory transfers can be quite large compared with floating point operations. Thus, it is important to design the scheme to minimize memory transfers and take advantage of the capability of modern GPUs to overlap communication with computation. Thus, in GX we have chosen a targeted multi-GPU parallelization scheme that currently parallelizes only species and Hermite modes across GPUs. In this scheme, communication is required to compute the total charge density and currents in the field equations, which takes the form of an all-reduce operation on a configuration-space-sized ($N_{k_x} \times N_{k_y} \times N_z$) object. Additionally, since the equation for the Hermite mode m couples to modes $m-2$, $m-1$, $m+1$ and $m+2$, we use 'halo' (or 'ghost') modes on each GPU, similar to standard parallelization schemes for finite differencing. Halo exchange is thus the other dominant form of communication, requiring inter-GPU transfers of objects of size $N_{k_x} \times N_{k_y} \times N_z \times N_\ell \times 2$. We overlap the halo exchange with computation of most of the nonlinear terms (halo modes are only required for the nonlinear terms involving $A_\parallel$, and even for these the terms on the interior modes can be computed before the halo transfers have been completed).

In figure 11 we show a strong scaling study for nonlinear CBC cases with two kinetic species. In each case we fix resolution (shown in the legend) and plot the time per timestep (figure 11a) and scaling efficiency (figure 11b) as we increase the number of GPUs used for the calculation. These calculations have been performed on Perlmutter at NERSC, where each node contains four A100 GPUs. The three cases have the same configuration space resolution and number of kinetic species as the cases in table 2. For reference, the (16,64) case running on a single GPU consumes approximately 90 % of the available GPU memory on a 40GB A100 GPU. In each case, the scaling efficiency is above 75 % parallelizing across four GPUs. The efficiency also improves as the problem size increases, which is expected because there is more computing time relative to communication time.

We also show in figure 12 weak scaling studies for the nonlinear CBC with two kinetic species and the same configuration space resolution as the cases in table 2. Here, we increase the number of Hermite modes proportionally to the number of GPUs used, so that $N_m = 8N_{\text{gpu}}$ in each case. After parallelizing first over the two kinetic species, this results in 16 Hermite modes per GPU in all the calculations shown. We show the weak
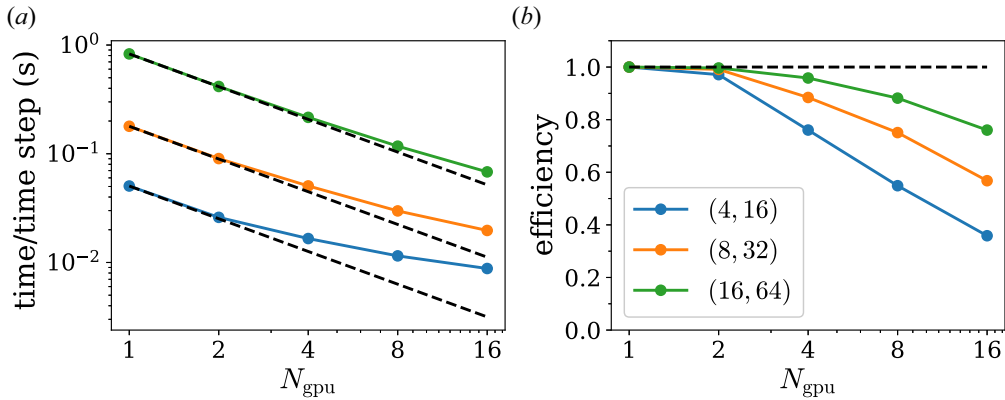
(a)

(b)

FIGURE 11. Strong scaling study for the nonlinear CBC with kinetic electrons showing the time per timestep (a) and the scaling efficiency (b) as a function of number of GPUs used, with fixed resolution for each curve. In each case the ideal scaling is shown with a dashed black line. The Laguerre–Hermite resolution parameters are listed in the legend as (N_ℓ, N_m). The other resolution parameters are $N_x = 192$, $N_y = 64$, $N_z = 24$, $N_{species} = 2$.
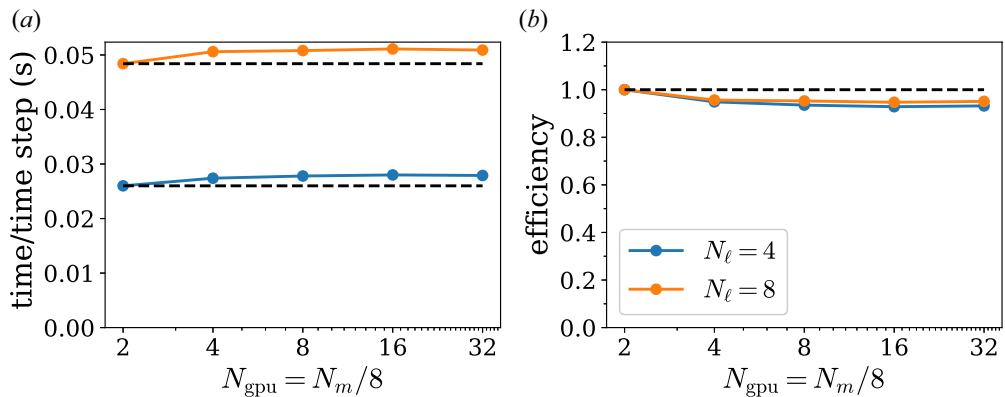
(a)

(b)

FIGURE 12. Weak scaling study for the nonlinear CBC with kinetic electrons showing the time per timestep (a) and the scaling efficiency (b) as a function of number of GPUs used, with the number of Hermite modes N_m scaling with the number of GPUs as $N_m = 8N_{gpu}$. We show results for both $N_\ell = 4$ (blue) and $N_\ell = 8$ (yellow). The ideal scaling is shown with dashed lines. The other resolution parameters are $N_x = 192$, $N_y = 64$, $N_z = 24$, $N_{species} = 2$.

scaling for both $N_\ell = 4$ (blue) and $N_\ell = 8$ (yellow). The weak scaling efficiency, shown in (figure 12b), is ideal in both cases up to 32 GPUs ($N_m = 256$), despite the fact that using more than four GPUs requires multiple nodes on Perlmutter, and communication between nodes can be slower than communication within a node due to the details of the interconnect hardware.

8. Conclusion and future opportunities

In this work we have presented GX, a GPU-native gyrokinetic code focused on tokamak and stellarator design at reactor scales. We have described the numerical algorithms that we use to solve the electromagnetic δf gyrokinetic system, in particular the discretization scheme that is pseudospectral in the entire five-dimensional phase

space, leveraging the Laguerre–Hermite velocity-space formulation developed by Mandell *et al.* (2018). We have shown several linear and nonlinear benchmarks against established flux-tube gyrokinetic codes in both tokamak and stellarator geometry, verifying that GX correctly simulates gyrokinetic turbulence in the small ρ_* limit. The combination of GPU acceleration and favourable convergence properties of the Laguerre–Hermite velocity-space basis for nonlinear problems enables useful turbulence simulations in minutes.

Additional work will also focus on improving the efficiency of simulations that include kinetic electrons. At present, the need to resolve the fast parallel electron motion results in an increase in simulation cost by a factor of roughly $2\sqrt{m_i/m_e} \approx 120$ compared with simulations with Boltzmann electrons (the factor of 2 is from the additional kinetic species, and the square root of the mass ratio comes from $v_{\text{te}}/v_{\text{ti}}$). We have also shown that kinetic electron cases may require more Hermite resolution, especially if resolving the trapped–passing boundary is essential. Exploration of different choices of velocity-space coordinates for the electrons or improvements to the basis functions used (for example, a method to bound the Hermite collocation points to constrain $v_{\parallel\text{max}}$, as in Candy *et al.* (2016), which could alleviate the CFL timestep restriction) could increase efficiency. In addition, the reduced electron models of Beer & Hammett (1996), Snyder & Hammett (2001) and Abel & Cowley (2013) provide insight into how to eliminate the fast time scales associated with parallel electron motion. A combination of these approaches with implicit timestepping schemes could enable kinetic electron simulations with nearly the same efficiency as the Boltzmann electron simulations we have presented. Additionally, we are investigating machine-learning methods for subgrid models and closures (Barbour *et al.* 2021, 2022), as well as methods for dynamically adapting the number of modes in the system.

An essential target for GX is transport time scale macroscale profile evolution via coupling to a transport solver like Trinity (Barnes *et al.* 2009; Qian *et al.* 2022) or a steady-state profile prediction solver like TGYRO (Candy *et al.* 2009) or PORTALS (Rodriguez-Fernandez, Howard & Candy 2022). This will be described in a forthcoming paper. In this case, several GX flux-tube calculations are run in parallel at various radii, and the resulting turbulent fluxes are used to advance the transport equations for the equilibrium profiles on transport time scales. The efficiency of GX makes these simulations quite tractable in comparison with previous efforts. Further, since transport in a fusion reactor is usually quite stiff, small changes in profile gradients result in large changes in transport fluxes; conversely, this means that errors of order 10 % are quite tolerable since they will result in relatively negligible differences in the final equilibrium profiles. Thus, we can also leverage the flexibility of GX to run at quite low velocity resolution with only 10 %–20 % errors in the turbulent fluxes, as shown in the convergence studies.

Finally, since GX is specialized for reactor design, using GX as a turbulence model in the inner loop of optimization frameworks is of great interest. Kim *et al.* (2024) have leveraged the DESC (Dudt *et al.* 2022) optimization framework to demonstrate that stellarators can be optimized for turbulence using nonlinear GX calculations in the optimization loop. Additionally, coupling of GX to the SIMSOPT stellarator optimization framework (Landreman *et al.* 2021) is in progress. Following the work of Highcock *et al.* (2018), optimization of core equilibrium profiles (and hence fusion power) is also tractable with a fast multiscale transport model composed of GX coupled to a transport solver.

length (m)	magnetic field (T)	temperature (K)	mass (kg)	number density (m^{-3})	charge (C)
a_{N}	B_{N}	T_{ref}	M_{ref}	n_0	q_{ref}

TABLE 3. Fundamental normalizing quantities.

k_x, k_y, x, y	∇, z	$v_\parallel$	μ	ω, t	ψ
ρ_{ref}	a_{N}	$v_{\mathrm{th,s}}$	$v_{\mathrm{th,s}}^2/B_{\mathrm{N}}$	$v_{\mathrm{th,ref}}/a_{\mathrm{N}}$	$a_{\mathrm{N}}^2 B_{\mathrm{N}}$

TABLE 4. Normalizing the quantities for the independent variables.

Appendix A. Normalization

We choose a set of normalizations, described in table 3, for the fundamental quantities.

Note that the specific definition of the normalizing quantities is arbitrary; for example, the normalizing length could be chosen to be the minor radius ($a_N = a$), the major radius ($a_N = R$) or some other length. Using the fundamental normalizations, we can define the reference thermal speed $v_{\mathrm{th,\,ref}} = \sqrt{T_{\mathrm{ref}}/M_{\mathrm{ref}}}$, the reference cyclotron frequency $\Omega_{\mathrm{ref}} = q_{\mathrm{ref}} B_{\mathrm{N}}/M_{\mathrm{ref}}$ and the reference gyroradius $\rho_{\mathrm{ref}} \equiv v_{\mathrm{th,ref}}/\Omega_{\mathrm{ref}}$. The ratio of the reference gyroradius ρ_{ref} and the normalizing length a_{N} is defined as

$$\rho_* \equiv \frac{\rho_{\mathrm{ref}}}{a_{\mathrm{N}}}, \tag{A1}$$

h	Φ	$A_\parallel$	$\delta B_\parallel$
$\rho_* n_0/v_{\mathrm{th,ref}}^3$	$\rho_* T_{\mathrm{ref}}/q_{\mathrm{ref}}$	$\rho_*(\rho_{\mathrm{ref}} B_{\mathrm{N}})$	$\rho_* B_{\mathrm{N}}$

TABLE 5. Normalizing quantities for dependent variables.

which is the fundamental small parameter in the gyrokinetic model. Using table 3, we can also normalize all the independent variables and operators as shown in table 4 and all the dependent variables as shown in table 5. Using the normalizations given in tables 4 and 5, we can normalize the gyrokinetic model. Note that the term ρ_* cancels out of all the equations once they are normalized. We also define the following non-dimensional species-specific parameters that appear throughout the normalized equations: $\tau_s = T_s/T_{\mathrm{ref}}$, $m_s = M_s/M_{\mathrm{ref}}$, $v_{ts} = v_{\mathrm{th},s}/v_{\mathrm{th,ref}}$, $Z_s = q_s/q_{\mathrm{ref}}$ and $\rho_s = (v_{\mathrm{th},s}/\Omega_s)/\rho_{\mathrm{ref}}$.

Appendix B. Derivation of hypercollision operator

Consider the simplified system where the distribution function g evolves due to only parallel free streaming. Fourier transforming in the parallel direction, a mode with parallel wavenumber k evolves via

$$\frac{\partial g}{\partial t} + \mathrm{i} k v_t v_\parallel g = 0. \tag{B1}$$

The solution is

$$g(k, v, t) \propto \mathrm{e}^{-\mathrm{i} k v_t v_\parallel t}. \tag{B2}$$

Free streaming thus causes energy transfer to smaller scales in velocity space; this is the parallel phase mixing behaviour that produces Landau damping, as is well known. In the Hermite representation results this results in transfer of energy to higher m, which can be seen by taking Hermite moments of the solution, giving

$$g_m(k, t) \propto t^m \, \mathrm{e}^{-k^2 v_t^2 t^2/2}. \tag{B3}$$

Thus, a pulse propagating in m reaches $m(t) = k^2 v_t^2 t^2$ at time t. By differentiating in time we can obtain the advection 'velocity' in m-space,

$$\frac{\mathrm{d}m}{\mathrm{d}t} = 2k^2 v_t^2 t = 2k v_t \sqrt{m}. \tag{B4}$$

Now consider the advection equation for the mode energy, $E(m) = |g_m|^2/2$,

$$\frac{\partial E}{\partial t} + \frac{\partial}{\partial m}\left[\frac{\mathrm{d}m}{\mathrm{d}t} E\right] = S + D. \tag{B5}$$

The source S only acts at low m. For dissipation, we will assume a hypercollision operator of the form

$$\left(\frac{\partial g_m}{\partial t}\right)_{\mathrm{hyp}} = -\nu_{\mathrm{hyp}} m^p g_m, \tag{B6}$$

which results in

$$D = -2\nu_{\mathrm{hyp}} m^p E. \tag{B7}$$

In steady state ($\partial/\partial t \to 0$) at large enough m to neglect the source, we have

$$\frac{\partial}{\partial m}\left[\frac{\mathrm{d}m}{\mathrm{d}t}E\right] = -2\nu_{\mathrm{hyp}}m^p E. \tag{B8}$$

Inserting the advection velocity from (B4), we obtain

$$\frac{\partial}{\partial m}[2kv_t\sqrt{m}E] = -2\nu_{\mathrm{hyp}}m^p E. \tag{B9}$$

The solution of this differential equation is of the form

$$E(m) = \frac{C}{\sqrt{m}}\exp\left(-\frac{\nu_{\mathrm{hyp}}m^{p+1/2}}{(p+1/2)kv_t}\right). \tag{B10}$$

Note that in the limit of no dissipation ($\nu_{\mathrm{hyp}} \to 0$), we should have a constant $m-$flux Γ, which restricts the constant of integration to $C = \Gamma/(2kv_t)$ and makes the full solution

$$E(m) = \frac{\Gamma}{2kv_t\sqrt{m}}\exp\left(-\frac{\nu_{\mathrm{hyp}}m^{p+1/2}}{(p+1/2)kv_t}\right). \tag{B11}$$

Now we would like to choose ν_{hyp} so that the energy is damped to some fraction f by the time the energy reaches the grid scale at $m = m_{\max}$. From this we obtain

$$\nu_{\mathrm{hyp}} = \ln(f^{-1})\frac{p+1/2}{m_{\max}^{p+1/2}}|k|v_t. \tag{B12}$$

Taking $f = 0.1$ results in the operator presented in §4.1. Note that undamped energy will be reflected and damped again as it travels from high m to low m, resulting in only $f^2 = 0.01$ of the initial energy returning to low m.

Appendix C. Convergence study for parallel boundary damping filter

In §4.2 we defined the parallel boundary damping filter using two adjustable parameters: the width of the damping region z_{width}, and the scaling factor A. Here we perform convergence studies to verify that we have chosen reasonable values of these parameters. We will use the kinetic electron CBC for these studies.

We first perform a scan of the scaling factor A. Figure 13 shows the ion and electron heat flux as a function of $A\Delta t$, with Δt the timestep size. For this scan we have used $z_{\mathrm{width}} = L_z/8$. We can see that the heat fluxes are converged when $A\Delta t \gtrsim 0.05$. We next perform a scan of the damping region width z_{width}, as shown in figure 14. For this scan we use $A\Delta t = 0.1$. By varying z_{width}/L_z for several values of N_z, we find that $z_{\mathrm{width}} = L_z/8$ produces accurate results for each choice of N_z.

Appendix D. Pseudospectral evaluation of nonlinear terms

The nonlinear terms are evaluated pseudospectrally in $(x, y, z, \mu B, m)$ space to avoid convolutions in both Fourier and Laguerre coefficients as (Mandell *et al.* 2018)

$$\mathcal{N}_{\ell,m}^s = \mathcal{L}_\ell\mathcal{H}_m\mathcal{F}_{k_\perp}[\langle\boldsymbol{v}_\chi\rangle_{\boldsymbol{R}}\cdot\boldsymbol{\nabla}h_s] = \mathcal{L}_\ell\mathcal{H}_m\mathcal{F}_{k_\perp}[\langle\boldsymbol{v}_\chi\rangle_{\boldsymbol{R}}\cdot\boldsymbol{\nabla}g_s]$$

$$= \mathcal{L}_\ell\mathcal{F}_{k_\perp}\left(\mathcal{F}_{k_\perp}^{-1}[ik_y J_{0s}\Phi]\mathcal{F}_{k_\perp}^{-1}\left[ik_x\sum_{\ell'}\psi^{\ell'}G_{\ell',m}^s\right]\right.$$
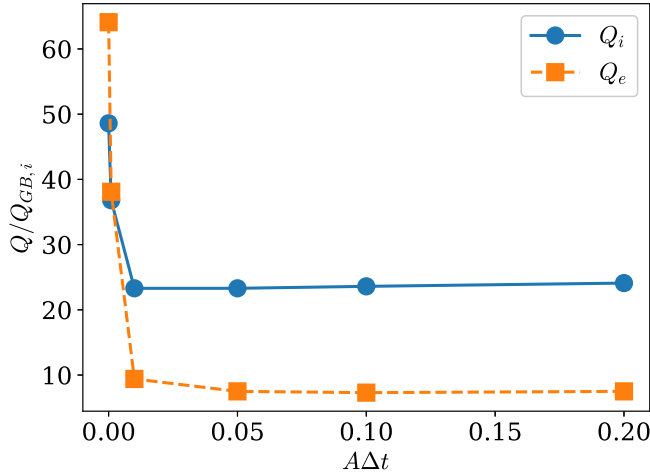
FIGURE 13. Time-averaged gyro-Bohm-normalized ion (solid blue circles) and electron (dashed yellow squares) heat fluxes as a function of $A\Delta t$, with A the scaling factor for the damping filter and Δt the timestep size.
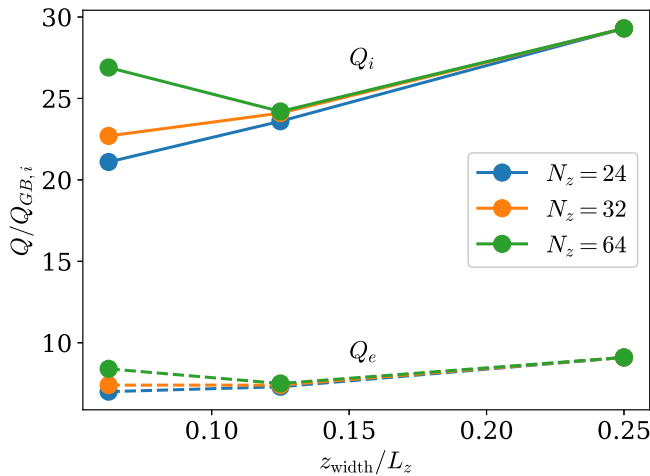

FIGURE 14. Time-averaged gyro-Bohm-normalized ion (solid) and electron (dashed) heat fluxes as a function of z_{width}/L_z, with z_{width} the width of the damping region and L_z the parallel length of the flux tube. Colours indicate different values of parallel resolution N_z.

$$
\begin{aligned}
&- \mathcal{F}_{\boldsymbol{k}_\perp}^{-1}[\mathrm{i}k_x J_{0s}\varPhi]\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}\left[\mathrm{i}k_y \sum_{\ell'}\psi^{\ell'}G_{\ell',m}^s\right]\Bigg) \\
&- v_{ts}\sqrt{m+1}\,\mathcal{L}_\ell \mathcal{F}_{\boldsymbol{k}_\perp}\left(\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}[\mathrm{i}k_y J_{0s}A_\parallel]\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}\left[\mathrm{i}k_x \sum_{\ell'}\psi^{\ell'}G_{\ell',m+1}^s\right]\right. \\
&- \mathcal{F}_{\boldsymbol{k}_\perp}^{-1}[\mathrm{i}k_x J_{0s}A_\parallel]\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}\left[\mathrm{i}k_y \sum_{\ell'}\psi^{\ell'}G_{\ell',m+1}^s\right]\Bigg) \\
&- v_{ts}\sqrt{m}\,\mathcal{L}_\ell \mathcal{F}_{\boldsymbol{k}_\perp}\left(\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}[\mathrm{i}k_y J_{0s}A_\parallel]\mathcal{F}_{\boldsymbol{k}_\perp}^{-1}\left[\mathrm{i}k_x \sum_{\ell'}\psi^{\ell'}G_{\ell',m-1}^s\right]\right.
\end{aligned}
$$

$$- \mathcal{F}_{k_\perp}^{-1}[ik_x J_{0s} A_\|] \mathcal{F}_{k_\perp}^{-1}\left[ik_y \sum_{\ell'} \psi^{\ell'} G_{\ell',m-1}^s \right] \Bigg)$$

$$+ \frac{\tau_s}{Z_s} \mathcal{L}_\ell \mathcal{F}_{k_\perp} \left(\mathcal{F}_{k_\perp}^{-1}\left[ik_y 2\mu B \frac{J_{1s}}{\alpha_s} \delta B_\| \right] \mathcal{F}_{k_\perp}^{-1}\left[ik_x \sum_{\ell'} \psi^{\ell'} G_{\ell',m}^s \right] \right.$$

$$\left. - \mathcal{F}_{k_\perp}^{-1}\left[ik_x 2\mu B \frac{J_{1s}}{\alpha_s} \delta B_\| \right] \mathcal{F}_{k_\perp}^{-1}\left[ik_y \sum_{\ell'} \psi^{\ell'} G_{\ell',m}^s \right] \right). \tag{D1}$$

This requires zero-padding of the arrays in the perpendicular and Laguerre dimensions to avoid aliasing, as described in Appendix D of Mandell *et al.* (2018). Fourier transforms are evaluated using the cuFFT library. The Laguerre transforms are expressed as matrix multiplications and implemented using the cuBLAS library.

Appendix E. Geometry details

In this appendix, we will describe how the geometry-related coefficients are calculated in the GX code. First, we define the different coordinate systems and relevant identities and briefly outline the process of calculating the geometry-related coefficients. Next, we list all the geometric quantities used by GX in the field-line following coordinate system. Finally, we describe the process of obtaining an equal-arc, equispaced coordinate from a general coordinate z.

Consider a magnetic coordinate system (ψ, ϕ, θ), where ψ is a flux surface label, ϕ is the cylindrical toroidal angle and θ is a generalized poloidal angle. A general three-dimensional equilibrium is a union of nested flux surfaces labelled with the flux surface label ψ, which is usually produced by an equilibrium code in the cylindrical coordinate (R, ϕ, Z). The first step is to calculate the Jacobian from the cylindrical to the curvilinear coordinates (ψ, ϕ, θ),

$$\mathcal{J} = \frac{1}{(\nabla\psi \times \nabla\phi) \cdot \nabla\theta} = \left(\frac{\partial \boldsymbol{R}}{\partial \psi} \times \frac{\partial \boldsymbol{R}}{\partial \phi} \right) \cdot \frac{\partial \boldsymbol{R}}{\partial \theta}, \tag{E1}$$

where the last equality in (E1) is a consequence of the duality relation (D'haeseleer *et al.* 1991); and

$$\nabla X_i \cdot \frac{\partial \boldsymbol{R}}{\partial X_j} = \delta_{ij}, \tag{E2}$$

where X_i and X_j are coordinates. Using the Jacobian and the duality relation we calculate the gradients $\nabla\psi, \nabla\phi, \nabla\theta$ from the gradients $\nabla R, \nabla\phi, \nabla Z$. For example, we calculate $\nabla\psi$ using the following relation:

$$\nabla\psi = \frac{\left(\dfrac{\partial \boldsymbol{R}}{\partial \theta} \times \dfrac{\partial \boldsymbol{R}}{\partial \phi} \right)}{\left(\dfrac{\partial \boldsymbol{R}}{\partial \psi} \times \dfrac{\partial \boldsymbol{R}}{\partial \phi} \right) \cdot \dfrac{\partial \boldsymbol{R}}{\partial \theta}}, \tag{E3}$$

where the vector $\boldsymbol{R} = (R, \phi, Z)$ is output from a typical equilibrium solver. Upon calculating the gradients of the coordinates (ψ, θ, ϕ), we can easily obtain the gradients of the field-line-following coordinate (x, y, z). In addition to that, an equilibrium code also outputs scalar-valued physical quantities like the plasma pressure $p(\psi)$, and the safety

factor $q(\psi)$ and vector-valued physical quantities like B, $\boldsymbol{B} \cdot \nabla\theta$, $\boldsymbol{B} \cdot \nabla\phi$. These quantities are then used to compute all the geometric coefficients defined below,

$$bmag = \frac{B}{B_{\mathrm{N}}}, \tag{E4}$$

$$gradpar = \boldsymbol{b} \cdot \nabla z, \tag{E5}$$

$$gds21 = \hat{s}(\nabla x \cdot \nabla y), \quad \hat{s} \equiv \frac{x}{q}\frac{\mathrm{d}q}{\mathrm{d}x}, \tag{E6}$$

$$gds22 = \hat{s}^2 |\nabla x|^2, \tag{E7}$$

$$gds2 = |\nabla y|^2, \tag{E8}$$

$$gbdrift = \frac{2}{bmag^2}(\boldsymbol{b} \times \nabla B) \cdot \nabla y, \tag{E9}$$

$$cvdrift = \frac{8\pi}{bmag^2}(\boldsymbol{b} \times \nabla y) \cdot \nabla p + gbdrift, \tag{E10}$$

$$cvdrift0 = gbdrift0 = \frac{2\hat{s}}{bmag^2}(\boldsymbol{b} \times \nabla B) \cdot \nabla x. \tag{E11}$$

In the local flux-tube limit we evaluate all of these quantities at $x = x_0$, with x_0 denoting the flux surface of interest. In non-axisymmetric configurations the coefficients also in general depend on the field line label y, so we also evaluate the coefficients at $y = y_0$ to select a particular field line of interest. This ensures that the geometric coefficients are only functions of the parallel coordinate z. These quantities form a complete set of geometry-related coefficients needed for a GX simulation. These quantities also coincide with the geometry definitions in GS2.

E.1. *Calculating an equispaced, equal-arc z*

Generally speaking, *gradpar* is a function of z. However, it is often convenient to transform to a z-grid where *gradpar* is a constant. This is strictly necessary in GX to allow Fourier spectral treatment of the parallel streaming term in the gyrokinetic equation. To do this, we seek a grid $\hat{z}$ such that

$$\boldsymbol{b} \cdot \nabla\hat{z} = C, \tag{E12}$$

where C is a constant. A differential element in the field-line following coordinates (x, y, z) can be written as

$$\mathrm{d}\boldsymbol{\ell} = \frac{\partial \boldsymbol{R}}{\partial x}\mathrm{d}x + \frac{\partial \boldsymbol{R}}{\partial y}\mathrm{d}y + \frac{\partial \boldsymbol{R}}{\partial z}\,\mathrm{d}z. \tag{E13}$$

On a given field line, $\mathrm{d}x = \mathrm{d}y = 0$. Using the duality relation (E2), we can write

$$\mathrm{d}\boldsymbol{\ell} = \frac{\nabla x \times \nabla y}{(\nabla x \times \nabla y) \cdot \nabla z}\,\mathrm{d}z. \tag{E14}$$

For a differential line element along the field line, we dot (E14) with the unit vector $\boldsymbol{b}$ to get

$$\mathrm{d}\ell = \frac{\mathrm{d}z}{\boldsymbol{b} \cdot \nabla z}. \tag{E15}$$

Repeating the same process for the coordinate system $(x, y, \hat{z})$ and using (E15)

$$\frac{\mathrm{d}z}{\boldsymbol{b} \cdot \nabla z} = \frac{\mathrm{d}\hat{z}}{\boldsymbol{b} \cdot \nabla \hat{z}}. \tag{E16}$$

Integrating (E16) first for one period in z (or $\hat{z}$) and using the fact that

$$\boldsymbol{b} \cdot \nabla z|_{z=0,L} = \boldsymbol{b} \cdot \nabla \hat{z}|_{\hat{z}=0,L}, \tag{E17}$$

we get

$$C = \boldsymbol{b} \cdot \nabla \hat{z} = L \left(\int_0^L \frac{\mathrm{d}z}{\boldsymbol{b} \cdot \nabla z} \right)^{-1}. \tag{E18}$$

Using C, we can now obtain $\hat{z}$ for each value of z by integrating (E16)

$$\hat{z} = C \int_0^L \frac{\mathrm{d}z}{\boldsymbol{b} \cdot \nabla z}. \tag{E19}$$

The resulting $\hat{z}$ coordinate is an equal-arc coordinate. Upon obtaining the equal-arc $\hat{z}$, we interpolate all geometric coefficients onto an equispaced grid $\tilde{z}$. This makes $\tilde{z}$ an equispaced, equal-arc coordinate. Note that this procedure can be used for any field-line-following coordinate.

Appendix F. Turbulent fluxes

Here we document expressions for turbulent fluxes in the Fourier–Laguerre–Hermite basis.

F.1. *Heat flux*

The normalized radial heat flux for species s is defined as

$$\begin{aligned}
\frac{Q_s}{n_s \tau_s} &= \int \frac{\mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \int \mathrm{d}^3 \boldsymbol{v} \langle \boldsymbol{v}_\chi \rangle_R \cdot \nabla x \left(\frac{1}{2} v_\parallel^2 + \mu B \right) h_s \\
&= \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \{ik_y \Phi(\boldsymbol{k}_\perp, z)\}^* \int \mathrm{d}^3 \boldsymbol{v} \left(\frac{1}{2} v_\parallel^2 + \mu B \right) J_{0s} h_s \\
&\quad - v_{ts} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \{ik_y A_\parallel(\boldsymbol{k}_\perp, z)\}^* \int \mathrm{d}^3 \boldsymbol{v} \left(\frac{1}{2} v_\parallel^2 + \mu B \right) J_{0s} v_\parallel h_s \\
&\quad + \frac{\tau_s}{Z_s} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \left\{ ik_y \frac{\delta B_\parallel(\boldsymbol{k}_\perp, z)}{B} \right\}^* \int \mathrm{d}^3 \boldsymbol{v} \left(\frac{1}{2} v_\parallel^2 + \mu B \right) 2\mu B \frac{J_{1s}}{\alpha_s} h_s \\
&= \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \{ik_y \Phi(\boldsymbol{k}_\perp, z)\}^* \frac{3}{2} \bar{p}_s - v_{ts} \\
&\quad \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \{ik_y A_\parallel(\boldsymbol{k}_\perp, z)\}^* \bar{q}_{\parallel s} \\
&\quad + \frac{\tau_s}{Z_s} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{k_\perp} \left\{ ik_y \frac{\delta B_\parallel(\boldsymbol{k}_\perp, z)}{B} \right\}^* \bar{q}_{\perp s},
\end{aligned} \tag{F1}$$

where $\{\cdots\}^*$ denotes a complex conjugate, and

$$\frac{3}{2}\bar{p}_s = \sum_{\ell=0}^{N_\ell} \left\{ \left[\ell \mathcal{J}_{\ell-1}^s + \left(2\ell + \frac{3}{2}\right)\mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s \right] H_{\ell,0}^s + \frac{1}{\sqrt{2}}\mathcal{J}_\ell^s H_{\ell,2}^s \right\}, \quad (\text{F2})$$

$$\bar{q}_{\|s} = \sum_{\ell=0}^{N_\ell} \left\{ \left[\ell \mathcal{J}_{\ell-1}^s + \left(2\ell + \frac{5}{2}\right)\mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s \right] H_{\ell,1}^s + \sqrt{\frac{3}{2}}\mathcal{J}_\ell^s H_{\ell,3}^s \right\}, \quad (\text{F3})$$

$$\bar{q}_\perp = \sum_{\ell=0}^{N_\ell} \left\{ \left[\ell \mathcal{J}_{\ell-2}^s + \left(3\ell + \frac{3}{2}\right)\mathcal{J}_{\ell-1}^s + \left(3\ell + \frac{5}{2}\right)\mathcal{J}_\ell^s + (\ell+1)\mathcal{J}_{\ell+1}^s \right] H_{\ell,0}^s \right.$$
$$\left. + \frac{1}{\sqrt{2}}(\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s)H_{\ell,2}^s \right\}. \quad (\text{F4})$$

Note that since $\boldsymbol{v}_\chi \cdot \nabla\chi = 0$, the convection of the integral of h at fixed $\boldsymbol{r}$ is equal to the convection of the integral of g at fixed $\boldsymbol{r}$. In physical units, the heat flux for species s is $Q_{\text{phys},s} = (n_0 v_{\text{th}} T)_{\text{ref}} \rho_*^2 Q_s \equiv Q_{GB,\text{ref}} Q_s$.

F.2. *Particle flux*

Similarly, the normalized radial particle flux for species s is defined as

$$\frac{\Gamma_s}{n_s} = \int \frac{\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \int \mathrm{d}^3 \boldsymbol{v} \langle \boldsymbol{v}_\chi \rangle_R \cdot \nabla x h$$

$$= \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \{ik_y \Phi(\boldsymbol{k}_\perp, z)\}^* \int \mathrm{d}^3 \boldsymbol{v} J_{0s} h_s$$

$$- v_{ts} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \{ik_y A_\|(\boldsymbol{k}_\perp, z)\}^* \int \mathrm{d}^3 \boldsymbol{v} J_{0s} v_\| h_s$$

$$+ \frac{\tau_s}{Z_s} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \left\{ik_y \frac{\delta B_\|(\boldsymbol{k}_\perp, z)}{B}\right\}^* \int \mathrm{d}^3 \boldsymbol{v}\, 2\mu B \frac{J_{1s}}{\alpha_s} h_s$$

$$= \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \{ik_y \Phi(\boldsymbol{k}_\perp, z)\}^* \bar{n}_s$$

$$- v_{ts} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \{ik_y A_\|(\boldsymbol{k}_\perp, z)\}^* \bar{u}_{\|s}$$

$$+ \frac{\tau_s}{Z_s} \int \frac{\mathrm{d}z}{\nabla x \times \nabla y \cdot \nabla z} \sum_{\boldsymbol{k}_\perp} \left\{ik_y \frac{\delta B_\|(\boldsymbol{k}_\perp, z)}{B}\right\}^* \frac{\bar{u}_{\perp s}}{\sqrt{b_s}}, \quad (\text{F5})$$

with

$$\bar{n}_s = \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,0}^s, \quad (\text{F6})$$

$$\bar{u}_{\parallel s} = \sum_{\ell=0}^{N_\ell} \mathcal{J}_\ell^s H_{\ell,1}^s, \tag{F7}$$

$$\frac{\bar{u}_{\perp s}}{\sqrt{b_s}} = \sum_{\ell=0}^{N_\ell} (\mathcal{J}_\ell^s + \mathcal{J}_{\ell-1}^s) H_{\ell,0}^s, \tag{F8}$$

as defined in §4. In physical units, the particle flux for species s is $\Gamma_{\text{phys},s} = (n_0 v_{\text{th}})_{\text{ref}} \rho_*^2 \Gamma_s \equiv \Gamma_{GB,\text{ref}} \Gamma_s$.

Appendix G. Numerical resolution for benchmark cases

Here we provide the numerical resolution and associated parameters used in each benchmark case from §6. In all cases, normalizing quantities below follow the GX conventions above (as opposed to the normalization conventions of the various other codes).

G.1. *Linear CBC, Boltzmann electrons (figure 1)*

GX calculations used: three 2π segments in an extended ballooning domain each with $N_z = 24$ parallel grid points; $N_m = 48$ Hermite modes; $N_\ell = 16$ Laguerre modes.

GS2 calculations used: three 2π segments in an extended ballooning domain each with $N_z = 25$ parallel grid points; 32 energy grid points; 33 pitch angles (20 in the untrapped region of velocity space and 13 in the trapped region).

G.2. *Linear CBC, kinetic electrons (figures 2 and 3)*

GX calculations used: three 2π segments in an extended ballooning domain each with $N_z = 24$ parallel grid points; $N_m = 128$ Hermite modes; $N_\ell = 16$ Laguerre modes.

GS2 calculations used: three 2π segments in an extended ballooning domain each with $N_z = 25$ parallel grid points; 32 energy grid points; 33 pitch angles (20 in the untrapped region of velocity space and 13 in the trapped region).

The stella calculations used: three 2π segments in an extended ballooning domain each with $N_z = 25$ parallel grid points; 48 $v_\parallel$ (evenly spaced) grid points with $v_{\parallel,\text{max}} = 3\sqrt{2}v_{\text{ti}}$; 16 μ grid points with $v_{\perp,\text{max}} = 3\sqrt{2}v_{\text{ti}}$.

G.3. *Nonlinear CBC, Boltzmann electrons (figure 4 and table 1)*

GX calculations used: one 2π ballooning domain segment with $N_z = 24$ parallel grid points; $N_x = 192$ radial grid points (which corresponds to $N_{k_x} = 127$ dealiased Fourier modes); $N_y = 64$ binormal grid points (which corresponds to $N_{k_y} = 22$ dealiased $k_y \geq 0$ Fourier modes, with the $k_y < 0$ modes determined by the reality condition). The velocity-space resolution varied by case and is provided in the main text. The perpendicular box dimensions were approximately $190\rho_i \times 190\rho_i$. The hyperdissipation parameters used were: hyperviscosity with $D = 0.05$ and $n = 4$; hypercollisions with $f_{\text{hyp}} = 1$; $p = N_m/2$.

GS2 calculations used: one 2π ballooning domain segment with $N_z = 24$ parallel grid points, $N_x = 192$ radial grid points ($N_{k_x} = 127$); $N_y = 64$ binormal grid points ($N_{k_y} = 22$); 16 energy grid points; 33 pitch angles (20 in the untrapped region of velocity space and 13 in the trapped region). The perpendicular box dimensions were approximately $190\rho_i \times 190\rho_i$.

G.4. *Nonlinear CBC, kinetic electrons (figure 5 and table 2)*

GX calculations used: one 2π ballooning domain segment with $N_z = 24$ parallel grid points; $N_x = 192$ radial grid points (which corresponds to $N_{k_x} = 127$ dealiased Fourier modes); $N_y = 64$ binormal grid points (which corresponds to $N_{k_y} = 22$ dealiased Fourier modes). The velocity-space resolution varied by case and is provided in the main text. The perpendicular box dimensions were approximately $190\rho_i \times 190\rho_i$. The hyperdissipation parameters used were: hyperviscosity with $D = 0.05$ and $n = 4$; hypercollisions with $f_{\text{hyp}} = 1$; $p = N_m/2$.

GS2 calculations used: one 2π ballooning domain segment with $N_z = 32$ parallel grid points, $N_x = 192$ radial grid points ($N_{k_x} = 127$); $N_y = 64$ binormal grid points ($N_{k_y} = 22$); 16 energy grid points; 36 pitch angles (19 in the untrapped region of velocity space and 17 in the trapped region). The perpendicular box dimensions were approximately $190\rho_i \times 190\rho_i$.

G.5. *Linear W7-X case (figure 6)*

GX calculations used: six 2π ballooning domain segments (corresponding to six poloidal turns) with $N_z = 256$ parallel grid points; $N_m = 16$ Hermite modes; $N_\ell = 8$ Laguerre modes.

The stella calculations used: 33 field periods with $N_z = 256$ parallel grid points; $N_{v_\parallel} = 32$; $N_\mu = 8$.

G.6. *Nonlinear W7-X case (figure 7)*

GX calculations used: one 2π ballooning domain segment (corresponding to a single poloidal turn) with $N_z = 128$ parallel grid points, $N_x = 64$ radial grid points (which corresponds to $N_{k_x} = 43$ dealiased Fourier modes); $N_y = 192$ binormal grid points (which corresponds to $N_{k_y} = 64$ dealiased Fourier modes). The velocity-space resolution varied by case and is provided in the main text. The perpendicular box dimensions were approximately $132\rho_i \times 89\rho_i$. The hyperdissipation parameters used were: hyperviscosity with $D = 0.05$ and $n = 4$; hypercollisions with $f_{\text{hyp}} = 1$ and $p = N_m/2$.

For stella and GENE calculation details, refer to Tables 2 and 3 (test 5) of González-Jerez *et al.* (2022).

REFERENCES

ABEL, I., BARNES, M., COWLEY, S., DORLAND, W. & SCHEKOCHIHIN, A. 2008 Linearized model Fokker-Planck collision operators for gyrokinetic simulations. I. Theory. *Phys. Plasmas* **15** (12), 122509.

ABEL, I. & COWLEY, S. 2013 Multiscale gyrokinetics for rotating tokamak plasmas: II. Reduced models for electron dynamics. *New J. Phys.* **15**.

ABEL, I., PLUNK, G., WANG, E., BARNES, M., COWLEY, S., DORLAND, W. & SCHEKOCHIHIN, A. 2013 Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. *Rep. Prog. Phys.* **76** (11).

ANTONSEN, T. & LANE, B. 1980 Kinetic equations for low frequency instabilities in inhomogeneous plasmas. *Phys. Fluids* **23** (6), 1205–1214.

BALL, J. & BRUNNER, S. 2021 A non-twisting flux tube for local gyrokinetic simulations. *Plasma Phys. Control. Fusion* **63** (6).

BARBOUR, N., DORLAND, W., BARBOUR, N. & DORLAND, W. 2021 Subgrid modeling of gyrokinetic turbulence using machine learning. In *Bulletin of the American Physical Society*.

BARBOUR, N., DORLAND, W., MANDELL, N., GAUR, R., MCGRAE-MENGE, M., PIERCE, J., ALMANZA, M., CHOU, J., ALVES, E., FIUZA, F. & LOUREIRO, N. 2022 Machine-learning

closures of the kinetic moment hierarchy in the context of Landau damping. In *Bulletin of the American Physical Society*.

BARNES, M., ABEL, I., DORLAND, W., ERNST, D., HAMMETT, G., RICCI, P., ROGERS, B., SCHEKOCHIHIN, A. & TATSUNO, T. 2009 Linearized model Fokker-Planck collision operators for gyrokinetic simulations. II. Numerical implementation and tests. *Phys. Plasmas* **16** (7), 72107.

BARNES, M., ABEL, I., DORLAND, W., GÖRLER, T., HAMMETT, G. & JENKO, F. 2010 Direct multiscale coupling of a transport code to gyrokinetic turbulence codes. *Phys. Plasmas* **17** (5), 969.

BARNES, M., PARRA, F. & LANDREMAN, M. 2019 stella: an operator-split, implicit–explicit δf-gyrokinetic code for general magnetic field configurations. *J. Comput. Phys.* **391**, 365–380.

BEER, M. 1995 Gyrofluid models of turbulent transport in tokamaks. PhD thesis, Princeton University.

BEER, M., COWLEY, S. & HAMMETT, G. 1995 Field–aligned coordinates for nonlinear simulations of tokamak turbulence. *Phys. Plasmas* **2** (7), 2687–2700.

BEER, M. & HAMMETT, G. 1996 Toroidal gyrofluid equations for simulations of tokamak turbulence. *Phys. Plasmas* **3** (11), 4046–4064.

BELLI, E., CANDY, J., SFILIGOI, I. & WÜRTHWEIN, F. 2022 Comparing single-node and multi-node performance of an important fusion HPC code benchmark. In *PEARC 2022 Conference Series - Practice and Experience in Advanced Research Computing 2022 - Revolutionary: Computing, Connections, You*. Association for Computing Machinery.

BELLI, E., HAMMETT, G. & DORLAND, W. 2008 Effects of plasma shaping on nonlinear gyrokinetic turbulence. *Phys. Plasmas* **15** (9), 92303.

CANDY, J., BELLI, E. & BRAVENEC, R. 2016 A high-accuracy Eulerian gyrokinetic solver for collisional plasmas. *J. Comput. Phys.* **324**, 73–93.

CANDY, J., HOLLAND, C., WALTZ, R., FAHEY, M. & BELLI, E. 2009 Tokamak profile prediction using direct gyrokinetic and neoclassical simulation. *Phys. Plasmas* **16** (6), 060704.

CANDY, J. & WALTZ, R. 2003 An Eulerian gyrokinetic-Maxwell solver. *J. Comput. Phys.* **186** (2), 545–581.

D'AZEVEDO, E., ABBOTT, S., KOSKELA, T., WORLEY, P., KU, S., ETHIER, S., YOON, E., SHEPHARD, M., HAGER, R., LANG, J., CHOI, J., PODHORSZKI, N., KLASKY, S., PARASHAR, M. & CHANG, C. 2018 The fusion code XGC. In *Exascale Scientific Applications*, pp. 529–552. Chapman and Hall/CRC.

D'HAESELEER, W., HITCHON, W., CALLEN, J. & SHOHET, J. 1991 *Flux Coordinates and Magnetic Field Structure*. Springer.

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

DORLAND, W. & HAMMETT, G. 1993 Gyrofluid turbulence models with kinetic effects. *Phys. Fluids* B **5** (3), 812–835.

DORLAND, W., JENKO, F., KOTSCHENREUTHER, M. & ROGERS, B. 2000 Electron temperature gradient turbulence. *Phys. Rev. Lett.* **85** (26), 5579–5582.

DOUGHERTY, J. 1964 Model Fokker-Planck equation for a plasma and its solution. *Phys. Fluids* **7** (11), 1788.

DUDT, D., CONLIN, R., PANICI, D. & KOLEMEN, E. 2022 The DESC stellarator code suite part III: quasi-symmetry optimization. *J. Plasma Phys.* **89** (2), 955890201.

DURRAN, D. 2010 Nonreflecting boundary conditions. In *Numerical Methods for Fluid Dynamics: With Applications to Geophysics*, pp. 453–495. Springer.

FOK, J., GUO, B. & TANG, T. 2001 Combined Hermite spectral-finite difference method for the Fokker-Planck equation. *Maths Comput.* **71** (240), 1497–1529.

FRANCISQUEZ, M., JUNO, J., HAKIM, A., HAMMETT, G. & ERNST, D. 2022 Improved multispecies Dougherty collisions. *J. Plasma Phys.* **88** (3), 905880303.

FREI, B., BALL, J., HOFFMANN, A., JORGE, R., RICCI, P. & STENGER, L. 2021 Development of advanced linearized gyrokinetic collision operators using a moment approach. *J. Plasma Phys.* **87** (5), 905870501.

FREI, B., ERNST, S. & RICCI, P. 2022*a* Numerical implementation of the improved Sugama collision operator using a moment approach. *Phys. Plasmas* **29** (9), 093902.

FREI, B., HOFFMANN, A. & RICCI, P. 2022*b* Local gyrokinetic collisional theory of the ion-temperature gradient mode. *J. Plasma Phys*. **88** (3), 905880304.

FREI, B., HOFFMANN, A., RICCI, P., BRUNNER, S. & TECCHIOLL, Z. 2023 Moment-based approach to the flux-tube linear gyrokinetic model. *J. Plasma Phys*. **89** (4), 905890414.

FREI, B., JORGE, R. & RICCI, P. 2020 A gyrokinetic model for the plasma periphery of tokamak devices. *J. Plasma Phys*. **86**.

FREI, B., MENCKE, J. & RICCI, P. 2024 Full-F turbulent simulation in a linear plasma device using a gyro-moment approach. *Phys. Plasmas* **31** (1), 012301.

FRIEMAN, E. & CHEN, L. 1982 Nonlinear gyrokinetic equations for low-frequency electromagnetic waves in general plasma equilibria. *Phys. Fluids* **25** (3), 502–508.

GERMASCHEWSKI, K., ALLEN, B., DANNERT, T., HRYWNIAK, M., DONAGHY, J., MERLO, G., ETHIER, S., D'AZEVEDO, E., JENKO, F. & BHATTACHARJEE, A. 2021 Toward exascale whole-device modeling of fusion devices: porting the GENE gyrokinetic microturbulence code to GPU. *Phys. Plasmas* **28** (6).

GONZÁLEZ-JEREZ, A., XANTHOPOULOS, P., GARCÍA-REGAÑA, J., CALVO, I., ALCUSÓN, J., BAÑÓN NAVARRO, A., BARNES, M., PARRA, F. & GEIGER, J. 2022 Electrostatic gyrokinetic simulations in Wendelstein 7-X geometry: benchmark between the codes stella and GENE. *J. Plasma Phys*. **88** (3), 905880310.

GOTTLIEB, S., SHU, C. & TADMOR, E. 2001 Strong stability–preserving high–order time discretization methods. *SIAM Rev*. **43** (1), 89–112.

HAMMETT, G. & PERKINS, F. 1990 Fluid moment models for Landau damping with application to the ion-temperature-gradient instability. *Phys. Rev. Lett*. **64** (25), 3019–3022.

HIGHCOCK, E., MANDELL, N., BARNES, M. & DORLAND, W. 2018 Optimisation of confinement in a fusion reactor using a nonlinear turbulence model. *J. Plasma Phys*. **84** (2), 905840208.

HIRSHMAN, S. & WHITSON, J. 1983 Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria. *Phys. Fluids* **26** (12), 3553–3568.

HOFFMANN, A., FREI, B. & RICCI, P. 2023*a* Gyrokinetic moment-based simulations of the Dimits shift. *J. Plasma Phys*. **89** (6), 905890611.

HOFFMANN, A., FREI, B. & RICCI, P. 2023*b* Gyrokinetic simulations of plasma turbulence in a Z-pinch configuration using a moment approach and advanced collision operators. *J. Plasma Phys*. **89** (2), 905890214.

IDOMURA, Y., IDA, M., KANO, T., AIBA, N. & TOKUDA, S. 2008 Conservative global gyrokinetic toroidal full-f five-dimensional Vlasov simulation. *Comput. Phys. Commun*. **179** (6), 391–403.

IDOMURA, Y., TOKUDA, S. & KISHIMOTO, Y. 2003 Global gyrokinetic simulation of ion temperature gradient driven turbulence in plasmas using a canonical Maxwellian distribution. *Nucl. Fusion* **43** (4), 234–243.

JENKO, F. 2000 Massively parallel Vlasov simulation of electromagnetic drift-wave turbulence. *Comput. Phys. Commun*. **125** (1), 196–209.

JENKO, F. & DORLAND, W. 2001 Nonlinear electromagnetic gyrokinetic simulations of tokamak plasmas. *Plasma Phys. Control. Fusion* **43** (12A), A141.

JOLLIET, S., BOTTINO, A., ANGELINO, P., HATZKY, R., TRAN, T., MCMILLAN, B., SAUTER, O., APPERT, K., IDOMURA, Y. & VILLARD, L. 2007 A global collisionless PIC code in magnetic coordinates. *Comput. Phys. Commun*. **177** (5), 409–425.

JORGE, R., FREI, B. & RICCI, P. 2019 Nonlinear gyrokinetic Coulomb collision operator. *J. Plasma Phys*. **85** (6), 905850604.

JORGE, R., RICCI, P. & LOUREIRO, N. 2017 A drift-kinetic analytical model for scrape-off layer plasma dynamics at arbitrary collisionality. *J. Plasma Phys*. **83** (6).

JOST, G., TRAN, T., COOPER, W., VILLARD, L. & APPERT, K. 2001 Global linear gyrokinetic simulations in quasi-symmetric configurations. *Phys. Plasmas* **8** (7), 3321–3333.

KETCHESON, D. 2008 Highly efficient strong stability-preserving Runge-Kutta methods with low-storage implementations. *SIAM J. Sci. Comput*. **30** (4), 2113–2136.

KIM, P., BULLER, S., CONLIN, R., DORLAND, W., DUDT, D., GAUR, R., JORGE, R., KOLEMEN, E., LANDREMAN, M., MANDELL, N. & PANICI, D. 2024 Optimization of nonlinear turbulence in stellarators. *J. Plasma Phys.* **90** (2), 905900210.

KOTSCHENREUTHER, M., REWOLDT, G. & TANG, W. 1995 Comparison of initial value and eigenvalue codes for kinetic toroidal plasma instabilities. *Comput. Phys. Commun.* **88** (2–3), 128–140.

LANDREMAN, M., MEDASANI, B., WECHSUNG, F., GIULIANI, A., JORGE, R. & ZHU, C. 2021 SIMSOPT: a flexible framework for stellarator optimization. *J. Open Source Softw.* **6** (65), 3525.

LANTI, E., OHANA, N., TRONKO, N., HAYWARD-SCHNEIDER, T., BOTTINO, A., MCMILLAN, B., MISHCHENKO, A., SCHEINBERG, A., BIANCALANI, A., ANGELINO, P. & BRUNNER, S. 2020 ORB5: a global electromagnetic gyrokinetic code using the PIC approach in toroidal geometry. *Comput. Phys. Commun.* **251**, 107072.

LEE, W. 1987 Gyrokinetic particle simulation model. *J. Comput. Phys.* **72** (1), 243–269.

LIN, Z., HAHM, T., LEE, W., TANG, W. & WHITE, R. 2000 Gyrokinetic simulations in general geometry and applications to collisional damping of zonal flows. *Phys. Plasmas* **7** (5), 1857–1862.

LOUREIRO, N., DORLAND, W., FAZENDEIRO, L., KANEKAR, A., MALLET, A., VILELAS, M. & ZOCCO, A. 2016 Viriato: a Fourier-Hermite spectral code for strongly magnetized fluid-kinetic plasma dynamics. *Comput. Phys. Commun.* **206**, 45–63.

MADDURI, K., IBRAHIM, K., WILLIAMS, S., IM, E., ETHIER, S., SHALF, J. & OLIKER, L. 2011 Gyrokinetic toroidal simulations on leading multi- and manycore HPC systems. In *Proceedings of 2011 SC - International Conference for High Performance Computing, Networking, Storage and Analysis*. Association for Computing Machinery.

MANDELL, N., DORLAND, W. & LANDREMAN, M. 2018 Laguerre-Hermite pseudo-spectral velocity formulation of gyrokinetics. *J. Plasma Phys.* **84** (1), 905840108.

MARINONI, A., BRUNNER, S., CAMENEN, Y., CODA, S., GRAVES, J., LAPILLONNE, X., POCHELON, A., SAUTER, O. & VILLARD, L. 2009 The effect of plasma triangularity on turbulent transport: modeling TCV experiments by linear and non-linear gyrokinetic simulations. *Plasma Phys. Control. Fusion* **51** (5), 55016.

MARTIN, M., LANDREMAN, M., XANTHOPOULOS, P., MANDELL, N. & DORLAND, W. 2018 The parallel boundary condition for turbulence simulations in low magnetic shear devices. *Plasma Phys. Control. Fusion* **60** (9), 95008.

MAURER, M., BAÑÓN NAVARRO, A., DANNERT, T., RESTELLI, M., HINDENLANG, F., GÖRLER, T., TOLD, D., JAREMA, D., MERLO, G. & JENKO, F. 2020 GENE-3D: a global gyrokinetic turbulence code for stellarators. *J. Comput. Phys.* **420**, 109694.

MILLER, R., CHU, M., GREENE, J., LIN-LIU, Y. & WALTZ, R. 1998 Noncircular, finite aspect ratio, local equilibrium model. *Phys. Plasmas* **5** (4), 973–978.

MYNICK, H., XANTHOPOULOS, P. & BOOZER, A. 2009 Geometry dependence of stellarator turbulence. *Phys. Plasmas* **16** (11), 110702.

OHANA, N., GHELLER, C., LANTI, E., JOCKSCH, A., BRUNNER, S. & VILLARD, L. 2021 Gyrokinetic simulations on many- and multi-core architectures with the global electromagnetic Particle-In-Cell Code ORB5. *Comput. Phys. Commun.* **262**, 107208.

PARKER, J. 2015 Gyrokinetic simulations of fusion plasmas using a spectral velocity space representation. PhD thesis, University of Oxford.

PARKER, J. & DELLAR, P. 2015 Fourier-Hermite spectral representation for the Vlasov-Poisson system in the weakly collisional limit. *J. Plasma Phys.* **81** (2).

PARKER, S., LEE, W. & SANTORO, R. 1993 Gyrokinetic simulation of ion temperature gradient driven turbulence in 3D toroidal geometry. *Phys. Rev. Lett.* **71** (13), 2042–2045.

PARRA, F. & BARNES, M. 2015 Equivalence of two different approaches to global δf gyrokinetic simulations. *Plasma Phys. Control. Fusion* **57** (5), 054003.

PEETERS, A., CAMENEN, Y., CASSON, F., HORNSBY, W., SNODIN, A., STRINTZI, D. & SZEPESI, G. 2009 The nonlinear gyro–kinetic flux tube code GKW. *Comput. Phys. Commun.* **180** (12), 2650–2672.

QIAN, T., BUCK, B., GAUR, R., MANDELL, N., KIM, P. & DORLAND, W. 2022 Stellarator profile predictions using Trinity3D and GX. In *Bulletin of the American Physical Society*. American Physical Society.

RODRIGUEZ-FERNANDEZ, P., HOWARD, N. & CANDY, J. 2022 Nonlinear gyrokinetic predictions of SPARC burning plasma profiles enabled by surrogate modeling. *Nucl. Fusion* **62** (7), 076036.

SFILIGOI, I., CANDY, J. & KOSTUK, M. 2018 CGYRO Performance on Power9 CPUs and Volta GPUs. In *International Conference in High Performance Computing*, Lecture Notes in Computer Science, vol. 11203, pp. 365–372. Springer.

SMITH, S. 1997 Dissipative closures for statistical moments, fluid moments, and subgrid scales in plasma turbulence. PhD thesis, Princeton University.

SNYDER, P. & HAMMETT, G. 2001 A Landau fluid model for electromagnetic plasma microturbulence. *Phys. Plasmas* **8** (7), 3199–3216.

ST-ONGE, D., BARNES, M. & PARRA, F. 2022 A novel approach to radially global gyrokinetic simulation using the flux-tube code stella. *J. Comput. Phys.* **468**, 111498.

STROHMAIER, E., DONGARRA, J., SIMON, H. & MEUER, M. 2022 60th edition of the TOP500 (November, 2022). Available at: https://www.top500.org/lists/top500/2022/11/.

WANG, W., LIN, Z., TANG, W., LEE, W., ETHIER, S., LEWANDOWSKI, J., REWOLDT, G., NAHM, T. & MANICKAM, J. 2006 Gyro-kinetic simulation of global turbulent transport properties in tokamak experiments. *Phys. Plasmas* **13** (9), 969.

WATANABE, T. & SUGAMA, H. 2005 Velocity–space structures of distribution function in toroidal ion temperature gradient turbulence. *Nucl. Fusion* **46** (1), 24.

WHITE, A. 2019 Validation of nonlinear gyrokinetic transport models using turbulence measurements. *J. Plasma Phys.* **85** (1).

ZOCCO, A., HELANDER, P. & CONNOR, J. 2015 Magnetic compressibility and ion-temperature-gradient-driven microinstabilities in magnetically confined plasmas. *Plasma Phys. Control. Fusion* **57** (8).