Towards optimal explicit time-stepping schemes for the gyrokinetic equations

H. Doerk*, F. Jenko
Max-Planck-Institut für Plasmaphysik, Boltzmannstraße 2, D-85748 Garching, Germany

ARTICLE INFO

Article history:
Received 24 October 2013
Received in revised form
11 March 2014
Accepted 24 March 2014
Available online 29 March 2014

Keywords:
Gyrokinetic simulation
Plasma turbulence
Optimized explicit Runge–Kutta schemes
Operator splitting
Eigenvalue computation
Spectral analysis

ABSTRACT

The nonlinear gyrokinetic equations describe plasma turbulence in laboratory and astrophysical plasmas. To solve these equations, massively parallel codes have been developed and run on present-day supercomputers. This paper describes measures to improve the efficiency of such computations, thereby making them more realistic. Explicit Runge–Kutta schemes are considered to be well suited for time-stepping. Although the numerical algorithms are often highly optimized, performance can still be improved by a suitable choice of the time-stepping scheme, based on the spectral analysis of the underlying operator. Here, an operator splitting technique is introduced to combine first-order Runge–Kutta–Chebychev schemes for the collision term with fourth-order schemes for the remaining terms. In the nonlinear regime, based on the observation of eigenvalue shifts due to the (generalized) $E \times B$ advection term, an accurate and robust estimate for the nonlinear timestep is developed. The presented techniques can reduce simulation times by factors of up to three in realistic cases. This substantial speedup encourages the use of similar timestep optimized explicit schemes not only for the gyrokinetic equation, but also for other applications with comparable properties.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

Gyrokinetic simulation codes are a common tool for obtaining \textit{ab-initio} predictions of turbulence properties in strongly magnetized high-temperature plasmas. \cite{1,2} Such plasmas are present in magnetic confinement fusion devices, and in astrophysics. Gyrokinetic theory describes the time evolution of each species’ particle distribution function $f$ in five-dimensional phase space (one velocity space variable, the gyro-angle, is averaged out). Obtaining a solution of this nonlinear partial integro-differential equation generally requires high-performance computing. In the past decades, gyrokinetic codes have become substantially more realistic by applying higher numerical resolution and by moving to more comprehensive physics models. For example, the effect of collisions is formally weak in dilute high-temperature plasmas and thus has often been neglected. Today, one realizes that including a suitable collision operator in gyrokinetic turbulence is not only required for a physically correct entropy balance, \cite{3} but can also greatly influence the turbulence level – through damping of zonal flows – or even change the turbulence regime by modifying the growth rate of certain types of microinstabilities \cite{4–8}. Since more realistic physics models require increased computational effort, progress is enabled by the availability of more powerful computers and by the use of advanced algorithms, the importance of the latter often being underestimated.

Three classes of gyrokinetic turbulence codes (particle-in-cell, semi-Lagrangian, and Eulerian) exist. Here, the Eulerian approach, which became popular approximately fifteen years ago, is considered. Several major code projects exist in this area, for instance GENE \cite{9–12}, GS2 \cite{13,14}, GYRO \cite{15,16}, GKW \cite{17}, and AstroGK \cite{18}. The common basic procedure is the so-called method of lines: after discretizing phase space on a fixed grid, the resulting large system of ordinary differential equations is evolved with a time integration scheme. However, the choice of algorithms can differ substantially. Besides various possible choices for phase space grids and the representation of derivatives on those grids, time discretization is performed in several ways; see Ref. \cite{19} for a useful overview. Operator splitting techniques for the collisional term are used in GYRO, GS2 and AstroGK. Some codes (like GS2) even choose to split off the nonlinear term from linear dynamics, while others avoid splitting to treat these terms on an equal level. Moreover, implicit as well as explicit schemes are applied. While GS2 (and AstroGK) treat all linear terms implicitly, the GYRO algorithm splits off fast linear terms (the parallel electron dynamics) in an implicit–explicit (IMEX) fashion. Here, we focus on fully explicit
time integration, as employed in GENE and GKw, for example. Explicit methods offer the advantages of an excellent performance on massively parallel systems and the straightforward implementation of nonlinear terms. The drawback is a strict stability limit that is set on the timestep $\Delta t$, which depends on the fastest dynamics in the system. A major advance from the gyrokinetic theory is to analytically remove extremely fast timescales like compressional Alfvén waves or particle gyromotion, leaving only relevant dynamics and enabling an explicit treatment. One of the fastest remaining terms is then given by the (generalized) nonlinear drift velocity $v_x = E_x \times B$ that combines electric and magnetic field fluctuations. When this nonlinear advection limits the timestep according to a Courant–Friedrichs–Lewy (CFL) relation $\Delta t \lesssim \Delta x / v_x$ [20], fully explicit schemes are likely to be the more efficient choice (particularly in view of increasing problem size) [9].

It is sometimes stated that collisions require an implicit treatment, since the explicit diffusive timestep limit would be too strict [21]. However, we find severe restrictions only for rather large collision frequencies (in the tokamak edge, for example) or for very high velocity resolution. In this work, we introduce a splitting scheme involving Runge–Kutta–Chebyshev (RKC) schemes with extended real stability boundary [22,23], which enables an explicit treatment of a sophisticated collision operator even in these extreme cases. Partitioned RKC schemes have recently been developed which are also stable for advective terms, involving, however, a larger number of operator evaluations per step [24,25].

In principle, accuracy limits can also be imposed on the timestep. In this context, we note that the overall numerical accuracy of gyrokinetic simulations is generally strongly restricted by the grid resolution in five-dimensional phase space. A relative error tolerance of approximately $10^{-3}$ is already considered to be sufficient, even for linear simulations. Nonlinear simulations are subject to statistical errors of the order of 10%, underlining the fact that long simulation times rather than highly accurate steps are needed. In consequence, the use of low-order time integration schemes is well justified to speed up computations.

In this paper, a detailed analysis of the spectral properties of the discretized system allows us to identify a class of highly efficient first-order explicit schemes (with largely extended stability boundaries), which we apply to the gyrokinetic code GENE. The remainder of this paper is organized as follows. The relevant physics is discussed in Section 2. The gyrokinetic equations are summarized in Section 2, and timestep limiting results in a large system of ordinary differential equations for fluid dynamics, such as spectral methods, finite differencing, and the GENET code, it may be necessary to add hyperdiffusion terms to $L[g]$ that remove unphysical grid-size oscillations in some phase space directions [28,29].

One way of solving this space-discretized system is to perform initial value computations, for which we consider Runge–Kutta (RK) schemes here. In the nonlinear case, we desire to find a statistically stationary turbulent state. Linear initial value computations yield the fastest growing solution (sometimes referred to as a mode), which constitute the driving force for plasma turbulence and are thus of great interest. Typical growth rates and frequencies are of the order of $c_s / L_{ref}$, where $c_s = (T_e / m_i)^{1/2}$ denotes the ion sound speed and $L_{ref}$ is a typical macroscopic scale length, often set to the tokamak major radius. Additionally, the linearized system can be formulated as an eigenvalue problem. In this context, GENE features the use of optimized iterative algorithms provided by the SLEPc package [30–34], which select a subset of eigenvector–eigenvalue pairs $(\lambda, \mathbf{g})$ that fulfill some user-specified criteria. For convenience, we split the complex eigenvalue $\lambda = \gamma + i \omega$ into a growth rate $\gamma$ and a frequency $\omega$. The eigenvalues of largest magnitude $|\lambda|$ are quickly found (for example by Krylov–Schur subspace iteration), which proves extremely useful for the exact computation of the maximum stable timestep for initial value simulations. Due to the shape of the spectrum, obtaining the fastest growing solution with SLEPc is more cumbersome, but can still be faster than a corresponding initial value simulation. Moreover, subdominant and marginally stable solutions only become accessible by such eigenvalue computations. Finally, GENE can also

### 2. The gyrokinetic equations

The gyrokinetic equation

$$\partial_t g = \mathcal{G}[g] = N[\tilde{\chi} \cdot g] + L[g] + C[g]$$

(1)

describes the time evolution of the (modified) perturbed gyrocenter distribution $g$ for each plasma species in $(x, y, z, v_\parallel, \mu)$ phase space. The notation

$$\tilde{\chi} = \tilde{\phi}_t - v_\parallel \tilde{A}_\parallel + \mu \tilde{B}_\parallel$$

$$f = g + v_\parallel \tilde{A}_\parallel F_0$$

$$F_0 = \frac{n_0}{\pi^{1/2} v_0^3} \exp\left(\frac{v_0^2 + \mu B_0}{T_0}\right)$$

introduces the fluctuating potential $\tilde{\chi}$, consisting of electrostatic perturbations $\phi_t$ and magnetic perturbations $\tilde{A}_\parallel$ and $\tilde{B}_\parallel$, where the overbar denotes a gyroaverage. The gyrocenter distribution is split into a background (Maxwellian) distribution $F_0$ and a small fluctuating part $f$. The background magnetic field is $B_0$ and the background density $n_0$, temperature $T_0$, thermal velocity $v_\parallel = (2m_0 / m)^{1/2}$, and particle mass $m$ are given for each plasma species. The gyrokinetic version of Maxwell’s equations is used to compute a self-consistent fluctuating potential from $g$, which closes the system of equations. We refer to Refs. [2,10,12] for a detailed description and derivation.

Eq. (1) is symbolically written as the sum of three integro-differential operators whose physical meaning is briefly discussed in the following. The linear terms $L[g]$ contain parallel advection along the magnetic field lines, as well as the perpendicular drifts such as curvature and $\nabla \times B$ drifts, and temperature and density gradient terms. The nonlinear term $N[\tilde{\chi} \cdot g]$ describes turbulent redistribution of free energy due to perpendicular $E \times B$ advection, where the generalized fluctuating field is defined as $E_x = -\nabla \times \tilde{\chi}$. Finally, the linearized Landau–Boltzmann collision operator $C[g]$ describes diffusion and dynamical friction in velocity space, including back-reaction terms that ensure the conservation of particles, momentum, and energy. Details of the implementation of the collision operator in GENE can be found in Refs. [26,27].

For a numerical solution, Eq. (1) is discretized on a fixed grid in phase space, where common techniques from computational fluid dynamics, such as spectral methods, finite differencing, finite element, and finite volume schemes, can be used. This results in a large system of ordinary differential equations for the time evolution of the state vector $g$. When non-dissipative differencing schemes are employed, as is the case with the GENE code, it may be necessary to add hyperdiffusion terms to $L[g]$ that remove unphysical grid-size oscillations in some phase space directions [28,29].

One way of solving this space-discretized system is to perform initial value computations, for which we consider Runge–Kutta (RK) schemes here. In the nonlinear case, we desire to find a statistically stationary turbulent state. Linear initial value computations yield the fastest growing solution (sometimes referred to as a mode), which constitute the driving force for plasma turbulence and are thus of great interest. Typical growth rates and frequencies are of the order of $c_s / L_{ref}$, where $c_s = (T_e / m_i)^{1/2}$ denotes the ion sound speed and $L_{ref}$ is a typical macroscopic scale length, often set to the tokamak major radius. Additionally, the linearized system can be formulated as an eigenvalue problem. In this context, GENE features the use of optimized iterative algorithms provided by the SLEPc package [30–34], which select a subset of eigenvector–eigenvalue pairs $(\lambda, \mathbf{g})$ that fulfill some user-specified criteria. For convenience, we split the complex eigenvalue $\lambda = \gamma + i \omega$ into a growth rate $\gamma$ and a frequency $\omega$. The eigenvalues of largest magnitude $|\lambda|$ are quickly found (for example by Krylov–Schur subspace iteration), which proves extremely useful for the exact computation of the maximum stable timestep for initial value simulations. Due to the shape of the spectrum, obtaining the fastest growing solution with SLEPc is more cumbersome, but can still be faster than a corresponding initial value simulation. Moreover, subdominant and marginally stable solutions only become accessible by such eigenvalue computations. Finally, GENE can also...
compute the full spectrum (using ScALAPACK routines), but this is only feasible for small problems.

As we will see in Section 3, the maximum stable timestep for Runge–Kutta methods is determined by the spectral properties of the underlying operator. Focusing on the linear case first, either the fastest oscillating $\omega_{\text{max}}$ or the most damped $\gamma_{\text{min}} = \min[\text{Re}(\lambda)]$ solutions are typically most restrictive. Let us briefly summarize physical mechanisms behind these extreme eigenvalues. Importantly, the integro-differential character of the parallel advection term $\delta g = v \phi g + v_x \phi$ does not allow for a rigorous CFL approach of the form $\lambda_{\text{max}} = k_{\text{max}} v$, since $\phi$ is computed from $g$ integrals. Here, $v$ is an advection velocity and $k_{\text{max}}$ is the largest wavenumber in the system. A popular example for the origin of very high-frequency (and timestep limiting) solutions are kinetic shear Alfvén waves. In simplified slab geometry (and in the relevant low-$\beta_e$ limit), the dispersion relation reads

$$\omega_{\text{RSA}}^2 = \frac{1}{\mu_e (k^2 + k^2)} + \beta_e k^2 \Gamma_c^2,$$

(2)

where $k_\perp$ is a perpendicular wavenumber and $k_\parallel$ is a parallel wavenumber [35,36]. Here, the electron to ion mass ratio $\mu_e = m_e/m_i$, the electron sound speed $\beta_e = 4\pi neT_e/\rho_i$, the ion sound gyroradius $\rho_i = c_i/\Omega_i$, and the ion cyclotron frequency $\Omega_i = (\epsilon_B) / (mc_i)$ are introduced. The $\beta_e$ parameter controls the response in Ampère’s law, whereas electrostatic models use $\beta_e = 0$. We observe that as $\beta_e$ approaches zero, the frequency $\omega_{\text{RSA}}$ approaches $k_\parallel (k_\perp \rho_i)$ becomes very large. Indeed, setting $k_\parallel \rho_i \rightarrow \pi / \Delta k \rightarrow 10, k_\parallel \rho_i \sim 0.05, \mu_e = 1 / 3600$ and $\beta_e = 0$ we obtain $\omega_{\text{RSA}} \approx 4000 (c_i / L_i)$, about four orders of magnitude larger than the typical values for growth or damping rates. Fortunately, even small values of $\beta_e$ prevent the divergence of $\omega_{\text{RSA}}$, so that it can be beneficial to include electromagnetic effects for kinetic electron simulations, even if the dominant physics is of electrostatic nature. In the opposite limit of $\beta_e \gg (k_\parallel \rho_i)^2$, which is more relevant to actual fusion plasmas, Eq. (2) transitions into the classical Alfvén wave dispersion relation $\omega_A^2 = v_A^2 k^2$ with $v_A \equiv c_i/\rho_i^{1/2}$ denoting the Alfvén velocity. Also the parallel streaming of electrons is often relevant, even if field-aligned coordinates are used. The characteristic frequency is given as $\omega_A \approx k \rho_i v_A$, which can be linked to a CFL condition. In typical fusion experiments, the electron thermal velocity is larger or comparable to the Alfvén velocity ($v_A \rho_i \sim v_e \rho_i \sim (\beta_e / \mu_e)^{1/2} \sim 1$), so that the CFL condition for kinetic electrons is usually more restrictive than the limit due to Alfvén waves.

A third notable source of high frequency solutions is linked to magnetic curvature and $VB$ drifts, which are (roughly) proportional to particle energy $\epsilon = m_e v^2/2$ and perpendicular wavenumber $k_\perp$. Thus, if either highly energetic particles or very high wavenumbers are involved, these drifts are expected to play a relevant role.

Of course, the above considerations are based on simplified versions of the gyrokinetic equation, or even on single terms. In the general, more comprehensive case, all these terms are coupled and one has to numerically compute the spectrum. The result (using ScALAPACK) is shown in Fig. 1(a) for the discretized, collisionless linear gyrokinetic operator. Indeed, the eigenvalue spectrum is stretched along the imaginary axis. One generally finds the highest frequency either at the smallest or largest wavenumber $k_\parallel$, which is consistent with the phenomena discussed above. In the appropriate limits, the magnitude of the numerically computed frequency scales as expected. While hyperdiffusion terms can be necessary to stabilize spurious grid-size oscillations, their effect on the eigenvalue spectrum is just a comparably small shift along the negative real axis, which has no big impact on stability considerations. It is important to note that upwind (or other dissipative) discretization methods can strongly distort the spectrum, and that dissipation on low $k_\parallel$ potential fluctuations

\[ N[\dot{x}, g] = \frac{c}{E_y B_0 r^2} \left( \partial_{\dot{x}} \dot{x} \partial_{\dot{g}} \partial_{\dot{x}} - \partial_{\dot{x}} \dot{g} \partial_{\dot{x}} \right) \]

(3)

prohibits a direct eigenvalue computation, since the advection velocities $\nu_x^x$ and $\nu_x^i$ are computed self-consistently from $g$. In such cases, a common technique is to freeze $v_x$ at time $t_n$ (and maximize it in space) for computing a timestep estimate. Details on this procedure and the combination of nonlinear and linear stability limits are given in Section 5. Due to the advective character of $N[\dot{x}, g]$, a frequency shift along the imaginary axis is found.

---

\footnote{Imaginary parts that arise from discretization and boundary conditions are negligible.}
3. Stability properties of relevant explicit Runge–Kutta methods

According to the method of lines the nonlinear, time-dependent operator $G(t, g)$ is discretized on a fixed grid in phase space, which turns Eq. (1) into a large set of first-order ordinary differential equations for the time evolution of the state vector $g$. In this section, explicit RK methods are considered to advance $g_n$ at time $t_n$ to $g_{n+1}$ at time $t_{n+1}$ with the timestep $\Delta t = t_{n+1} - t_n$. We focus on explicit RK schemes of the diagonal form

$$g_{n+1} = g_n + \Delta t \sum_{j=1}^{s} b_j k_j$$

where $s$ is the number of stages and the coefficients fulfill $\sum_{j=1}^{s} b_j = 1$ as well as $a_1 = 0$. In this simplified scheme, only $k_1$ is used for computing $\delta g$; while in general, all $k_j$ with $j < s$ can be allowed to contribute. Obviously, this procedure is memory efficient, since only up to three additional vectors of the size of $g$ have to be stored. The order of consistency $p$ is determined by comparing Eq. (4) with a Taylor expansion

$$g_{n+1} = g_n + \Delta t \sum_{j=1}^{s} c_j \frac{d^j g}{dt^j} \bigg|_{t_n}$$

$$= g_n + \Delta t \sum_{j=1}^{s} \frac{c_j}{j!} \frac{d^j g}{dt^j} \bigg|_{t_n} + \sum_{j=p+1}^{s} \frac{c_j}{j!} \frac{d^j g}{dt^j} \bigg|_{t_n}$$

of $g$ about $t_n$, where the $\{c_j\}$ can be computed from the coefficients ($a_1, b_j$) in Eq. (4). The required order of consistency $p \leq s$ thus imposes constraints on the $a_1, b_j$. For the linear problem, one inserts the eigenvalue equation $\lambda \frac{d g}{d t} = \lambda g$ into the Eq. (5) to obtain the stability polynomial

$$P_\lambda(\lambda) = 1 + \sum_{j=1}^{s} c_j (\lambda \Delta t)^j,$$

allowing to write down the RK stability condition

$$\vert P_\lambda(\lambda \Delta t) \vert \leq 1$$

that must be fulfilled for all $\lambda$ in the left complex half-plane ($\gamma \leq 0$) to ensure that these actually stable or damped solutions are not artificially destabilized by the explicit scheme. Thus, a sufficiently small timestep must be chosen. We define $\beta_{\text{mag}}(\beta_{\text{rel}})$ to be the extent of the stability boundary along the imaginary (negative real) axis, i.e., $\vert P_\lambda(-\beta_{\text{rel}}) \vert = \vert P_\lambda(\beta_{\text{mag}}) \vert = 1$. For instance, in the case of a simple advection problem $\partial_t g = v \partial_x g$, the eigenvalues $\lambda = ik_0 v t_n$ are imaginary, and the maximum timestep is $\Delta t = \beta_{\text{mag}}/\omega_{\text{max}}$. For maximum time order schemes ($p = s$), the total stability region is completely determined by the number of stages. Increasing the number of stages above $p$ requires additional evaluations of $G[g]$, but adds free parameters for shaping the stability polynomial to lower timestep constraints and reduce the overall computational cost.

Internal stability is found to become increasingly important at a large number of stages [22,23]. Since we use diagonal methods with no more than six stages, no restrictions are found in practice.

Among other choices, GENE features the use of standard $s = p$ second-order (RK2) and fourth-order (RK4) schemes, as well as an optimized fourth-order scheme (RK4M) with six stages, following Ref. [39]. The corresponding coefficients are given in Table 1 and the stability boundaries are depicted in Fig. 1(c). Additionally, a class of $s$-stage Runge–Kutta–Chebychev (RKC) schemes is considered, which are unconditionally unstable for (undamped) waves, but are powerful in the case of the real spectrum of the collision operator. The Chebychev polynomials are defined as

$$T_0(x) = 1 \quad T_1(x) = x$$

$$T_j(x) = 2x T_{j-1}(x) - T_{j-2}(x), \quad j \geq 2.$$
The eigenvalues of the collision operator \( C(g) \), on the other hand, extend along the negative real axis. In this case, the RKC methods discussed in Section 3 possess a near-optimal stability polynomial with a computational efficiency \( \beta_{\text{real}}/s \approx 1.93s \) that increases linearly in the number of stages.

Although it is possible to include the collision operator in \( L \) and perform time integration with a RK4 method, the strong benefits of RKC methods can only be exploited when an operator splitting technique is applied. In exponential notation, it can easily be shown that the symmetric (Strang) splitting
\[
g_{n+1}^{L} = e^{\Delta t g_{n}} + O(\Delta t^{2})
\]
is second-order accurate in \( \Delta t \) [41]. In contrast, the non-symmetric splitting
\[
g_{n+1}^{L} = e^{\Delta t g_{N+1}}g_{n} = e^{\Delta t g_{N+1}}g_{n} + O(\Delta t^{2})
\]
is formally only first-order accurate. However, when the propagation of \( g \) with the first operator \( C \) does not change the second operator \( N + L \), the second half-step can be combined with the first half-step of the next time iteration. In this case, both of the above splitting schemes are of second order [42]. In gyrokinetics, this argument holds for the linear computations only. In the nonlinear case, applying collisions on \( g \) does generally alter the self-consistent potentials computed with Maxwell’s equations, which in turn changes the nonlinear operator \( N \). In consequence, second-order accuracy in nonlinear simulations is expected only for the symmetric splitting. Since the RKC schemes that we consider here are only first-order accurate anyway, we can choose the simple approach
\[
g_{n+1}^{L} = R \mathcal{K}^{L} \{ I[g_{n}] + N[\chi_{n}, g_{n}], \Delta t \}
\]
\[
g_{n+1}^{RKC} = g_{n+1}^{L} + R \mathcal{K}^{RKC} \{ C[g_{n+1}], \Delta t \}
\]
(8)
of alternating propagation with the collisionless (Vlasov) operator \((L + N)\) and the collision operator \( C \), using a common timestep. The time-stepping schemes \( R \mathcal{K}^{L} \{ \cdot, \Delta t \} \) and \( R \mathcal{K}^{RKC} \{ \cdot, \Delta t \} \) can now be chosen individually. Scanning the timestep for different choices with the GENE code, Fig. 2 confirms the above considerations on a simple test case of an ion temperature gradient driven (ITG) mode, with kinetic electrons and \( \beta_{e} = 0.1\%. \) The frequency error \( \Delta \omega = \omega - \omega_{\text{comp}} \) compared to the converged RK4 result is measured with a precision of \( 5 \times 10^{-11} \) in this example. For an understanding of the results it is important to note that a pth-order scheme has an \( O(\Delta t^{p+1}) \) error for computing \( g_{n+1}^{L} \) from \( g_{n} \). Since the eigenvalue is basically determined by fitting an exponential as
\[
\lambda + \Delta \lambda = \frac{g_{n+1}^{L} - g_{n} + O(\Delta t^{p+1})}{\Delta t}
\]
we expect pth-order convergence for the frequency error \( \Delta \omega \). As a side note, also the global error for reaching a fixed simulation time \( t \) is \( O(\Delta t^{p}) \), since choosing a smaller value for \( \Delta t \) requires an accordingly larger number of timesteps to be computed.

In summary, using RKC schemes for collisions brings us back to first-order in time, as expected. However, the observed prefactor of the order of \( 10^{-2} \) in Fig. 2 is relatively small, so that an acceptable accuracy of at least \( 10^{-4} \) is obtained. Thus, in practice, no effect on the accuracy of the physically relevant solutions is visible with respect to higher-order methods, which has been confirmed for a large number of linear and nonlinear cases.

Two main advantages of applying this operator splitting are identified: (i) in cases of moderate collisionality, the one stage RKC1 scheme is sufficient for collisions, which saves three (five) calls of \( C(g) \), compared to including it in the RK4(RK4M) scheme. This remains valid in nonlinear computations, although the total fraction of CPU time spent to compute collisions will be smaller. (ii) In strongly collisional cases, the timestep is restricted by strongly damped collisional eigenvalues. Computation time can then be saved by adding RKC stages, which allows larger timesteps. In nonlinear simulations the timestep is rarely dominated by collisions, even if it is in the corresponding linear case.

Thus, it is reasonable to determine the optimal number of RKC stages dynamically, so that the maximum stable RKC timestep \( \Delta t_{\text{RKC}} \) is always somewhat larger than \( \Delta t_{\text{RK4}} \), the maximum stable RK4 timestep. In this way, collisions never restrict the timestep. This makes sense, as long as evaluating the collisionless part dominates the computational cost, which is generally the case. For evaluating \( \Delta t_{\text{RK4}} \) and \( \Delta t_{\text{RKC}} \), a small set of most restrictive eigenvalues of \( L \) and \( C \), \( \lambda_{c} \) and \( \lambda_{\ell} \), are pre-computed with fast largest-magnitude SLEPc algorithms. When multiple Fourier modes \( k_{r} \) are present, we make use of the fact that \( L \) and \( C \) are block-diagonal in this dimension. In this way, only a very low percentage of the following initial value computation (\( \leq 1\% \) of a linear run) is needed for this step.

We note, however, that the equations implemented in GENE require the computation of \( f(\chi) \) before every call of \( C \), because
these fields have to be kept consistent with \( g \). This produces an additional overhead of the splitting scheme. In rare cases, the computation of \( f \) and \( \chi \) is found to be relatively costly, but in general this is easily over-compensated by the gain in timestep or the less frequent calls of \( f \) itself. A positive side effect is related to the fact that in the GENE code the velocity space dimension \( N \) is stored in the last index of \( g \) and therefore is widely spread in the system memory. Since only collisions require ghost-cells in this dimension, which are exchanged via the message passing interface library, the reduction in number of calls of \( f \) improve the parallelization efficiency.

The efficiency of the adaptive RK operator splitting methods in combination with RK4 and RK4M is demonstrated in the following. We first focus on linear physics and compute the fastest growing solution with the initial value solver. Two typical cases that are sensitive to collisions are chosen. One is a trapped electron mode problem (TEM) in circular model geometry. The other is a microtearing mode problem (MTM) for physics parameters of the ASDEX Upgrade discharge 27 963 at the radial position \( \rho_{ref} = 0.85 \). The eigenvalues computed by the GENE code are \( \gamma_{TEM} = 0.2610 c_s/\rho_{ref} \), \( \omega_{TEM} = -0.6380 c_s/\rho_{ref} \) for the TEM case, and \( \gamma_{MTM} = 0.181 c_s/\rho_{ref} \), \( \omega_{MTM} = -1.332 c_s/\rho_{ref} \) for the MTM case. They coincide for all schemes considered up to the given convergence accuracy of 10^{-4}. Table 3 summarizes the results for code efficiency. In both cases, operator splitting leads to a strong increase in efficiency. In the TEM case with \( v_{nd}/\omega = 0.38 \), the timestep is not limited by collisions. Here, less frequent calls of the collision operator lead to shorter runtime and thus RK1 is most efficient. In the MTM case, the collisionality is larger (\( v_{nd}/\omega = 1.2 \)), so that adding up to two or three RK stages increases efficiency due to a gain in timestep. Interestingly, the RK1 splitting method requires a smaller timestep (which is explained by the lower stability boundary) but still is slightly more efficient due to the reduced number of calls of the collision operator. As expected, for both cases the runtime increases as soon as the optimal number of RK stages is exceeded.

The benefits of operator splitting with RKC schemes become even more striking when replacing the RK4 method with the optimized six-stage RK4M method, that is more efficient for large imaginary eigenvalues. Again, this is attributed to the use of a larger timestep, which also results in less overall evaluations of the collision operator. In consequence, the combination of RK4M and RKC1 leads to the lowest runtime for these linear runs. Compared to the standard RK4 scheme, the speedup is a factor of two to three, depending on the parameter set. This is significant when comprehensive physics models are employed to perform large (multidimensional) parameter studies, as is routinely done in quasilinear transport predictions for fusion plasmas (see [11,43–45], for example).

Additionally, we have compared the time traces of a nonlinear simulation of the TEM case with and without operator splitting using up to two stages. The simulation times are given in Table 4.

Here, the timestep is set by linear physics of the high wavenumbers even in the nonlinear simulation. Due to reduced computational time per step, the RK1 operator splitting method in combination with the RK4M time scheme has the largest speedup with respect to the RK4 scheme without operator splitting. We note that the physical results are identical within the statistical error bars inherent to nonlinear turbulence simulations. For the present MTM case the nonlinear terms dominate the timestep limit, as shown in the next section, similarly to previously published MTM simulations in a slightly different parameter regime [46]. In such cases, the RKC1 collision scheme is most efficient.

### 5. Timestep optimization in nonlinear simulations

This section addresses the modification of the maximum stable timestep by the nonlinear term Eq. (3). For simplicity, in the present section we will include collisions in the linear operator \( L \), unless \( C \) is separated from the RK4 scheme, as described in the previous section. In this sense, we denote by \( \Delta t \) the timestep obtained by considering only the linear terms. The nonlinear term will be analyzed in a linearized form \( N[\chi_0,g] \), which is obtained by freezing the potential to a snapshot of \( \chi_0 = \chi(t_r) \) taken at the current time \( t_r \). This procedure allows us to access the largest magnitude eigenvalues \( \lambda_{\text{lev}} \) of the combined operator \( L[g] + N[\chi_0,g] \), from which the maximum timestep \( \Delta t_{\text{lev}} \) can be computed exactly from Eq. (6). Since the drift velocities \( v^0 \) and \( v^1 \) do not change much during one step, \( \Delta t_{\text{lev}} \) is expected to accurately describe the stability limit of the RK scheme in the nonlinear regime. We test our hypothesis by performing three simulations with fixed timestep and identical initial condition. As illustrated in Fig. 3, the scheme becomes unstable, when \( \Delta t_{\text{lev}} \) sinks below the simulation timestep. Up to this point the time-traces of physical quantities are identical, reflecting the fact that the results are converged with respect to the timestep. Thus, it seems that we have a powerful tool at hand to exactly compute the maximum stable timestep, even in nonlinear simulations. However, performing this kind of "nonlinear eigenvalue computation" is not feasible at every timestep. In the
Eq. (11) requires $\Delta t_{\text{cfl}}$ to be set smaller than one in order to

would intuitively imply an eigenvalue shift along the negative real axis. Indeed, nonlinear $E_x \times B$ advection mediates dissipation by transporting fluctuation energy from driven phase space regions into dissipative regions. While this mechanism is important for the formation of a statistically stationary turbulent state, $N$ is purely advective at each point in time, which is relevant for the stability of time integration. The third condition (iii) holds, because $\lambda_{x,\max}$ is a global maximum over phase space and thus overestimates the actual stability restriction of the nonlinear term. Also the interplay with linear terms has been observed to lower the stability constraint in some cases. However, the theoretical maximum can closely be reached (cases with dominant $A_{||}$ fluctuations and fine $k_x$ resolution show this behavior). Nevertheless, we allow $\Delta t_{x,\max}$ to be multiplied with a user-specified constant $C_{\text{ef}}$, which can be set larger than unity in most cases. Robustness for the general case is obtained for $C_{\text{ef}} = 0.95$ to compensate the threshold mentioned above. It should be noted that a non-spectral treatment of the $x(y)$ dimension demands a correction factor to $k_x$ ($k_y$) in Eq. (9), which can be deduced from the corresponding differencing scheme.

The shift of the eigenvalues due to the nonlinearity as well as our model are illustrated in Fig. 4 for three distinct cases with increasing nonlinear contribution. The linear eigenvalues are shifted along the imaginary axis, as expected from the advective character of $N$. Using the same three parameter sets, we show in Fig. 5 that the sum of the maximum linear eigenvalue and the maximum nonlinear shift $\lambda_{L} + \lambda_{x,\max}$ is indeed always overestimating the exact result $\lambda_{\text{lev}}$. In the third case, the exact result is only overestimated by about 20% and also other cases showed almost no difference between modeled and real timestep limits. As mentioned above, other cases allow a $C_{\text{ef}}$ prefactor to $\Delta t_{x,\max}$ to be set larger than one. While this increases the efficiency, general robustness is lost. A less strict, yet robust automatic adaptation scheme would of course be desirable, but this seems to be impossible to find without performing actual expensive eigenvalue computations.

However, we want to point out that the presented estimate of Eq. (10) is already superior to the previously used CFL method

Fig. 3. This figure demonstrates that the eigenvalues with linearized $v_r$ terms are meaningful for stability of the RK4 scheme. As an example, an electromagnetic ITG/TEM simulation in ASDEX Upgrade outer core geometry is chosen. The timesteps $\Delta t_{\text{cfl}}$ and $\Delta t_{\text{lev}}$ are computed by inserting $\lambda_{L}$ and $\lambda_{\text{lev}}$ into the stability condition equation (6), respectively. The linear timestep is $\Delta t_{\text{cfl}} = 0.00518$ $\lambda_{\text{lev}}/c_s$ in this case. While the simulation is stable when the timestep is just below $\Delta t_{\text{lev}}$, a numerical instability is detected in two other simulations (marked by dashed lines ending in red crosses), when $\Delta t_{\text{cfl}}$ exceeds $\Delta t_{\text{lev}}$.
Fig. 4. This figure illustrates the position of the most timestep-relevant eigenvalues $\lambda_L$ and $\lambda_{nlev}$ at one point in time. The sum of $\lambda_L$ and the maximum nonlinear shift $\lambda_{x,\text{max}}$ is also depicted, accounting for the sign of $\lambda_L$. Three parameter sets with increasing nonlinear contribution are chosen. (a) is an ITG/TEM case in circular geometry; (b) and (c) use realistic tokamak geometry of the ASDEX Upgrade discharges #26459 and #27963, respectively. The shift of $\lambda_{nlev}$ with respect to $\lambda_L$ is along the imaginary axis, as expected. Note that Eq. (11) uses an effective eigenvalue $\lambda_{x,\text{max}}/c_{\text{cfl}}$. Setting $c_{\text{cfl}} = 0.3$ to a typical value, this lies (far) outside the figure in cases (b) and (c), which demonstrates the improvement of Eq. (10).

Fig. 5. Time traces of linear, estimated nonlinear, and exact nonlinear eigenvalues for the three sets of parameters of Fig. 4. The sum $\lambda_L + \lambda_{x,\text{max}}$ overestimates $\lambda_{nlev}$ at any time, leading to a robust timestep limit. The corresponding time trace of $1/\Delta t$ does not exactly match the estimate due to built-in thresholds for timestep changes, but it stays quite close. Here, the $\lambda_{nlev}$ are only computed each 1000th timestep.

6. Conclusions

In summary, we presented two methods for increasing the efficiency of gyrokinetic simulations and applied these to the plasma turbulence code GENE. First, we matched individual explicit Runge–Kutta schemes to the properties of individual parts of the equation by applying an operator splitting technique. For the collisionless part we chose classical and advanced fourth-order schemes. For collisions, we restricted ourselves to first-order Runge–Kutta–Chebychev schemes, since they possess optimal stability properties, while higher-order schemes offer much smaller efficiency gains. Thereby, we reached an increased timestep and/or fewer evaluations of the collision operator, resulting in a speedup by a factor of up to three both in strongly and weakly collisional cases. A possible application is given by extremely large (multi-dimensional) parameter studies, which are, for example, needed in quasilinear transport modeling of tokamak plasmas. Time savings due to our method are striking especially in the tokamak edge, where the collisionality is increased.

Second, we investigated the impact of nonlinear advection on the timestep. Based on the observation of a frequency-shift in the eigenvalue spectrum due to the $E \times B$ advection velocity (which is an interesting topic in itself), we developed an improved and robust timestep estimate for nonlinear simulations. Beyond the examples shown in this paper, the new adaptation scheme has been successfully applied to a large number of simulations. Avoiding unnecessarily small timesteps, a speedup of up to a factor of two to three is realized for realistic problems. This is particularly important for large simulations including comprehensive physics, experimental plasma shaping, kinetic electrons, multiple scales, and possibly also profile variations. Constituting the high-end of fusion plasma modeling, such simulations yield the most accurate description of plasma turbulence currently available, but they are expensive: one run can consume millions of CPU hours on present-day supercomputers.

Since the choice of algorithms and their implementation are already highly optimized in GENE (as in other state-of-the-art codes), this further increase of efficiency is really significant. The techniques discussed in this work can prove extremely useful, also for other simulation codes with similar numerical schemes.

Acknowledgments

The authors would like to thank J. Abiteboul, T. Dannert, T. Görler, D. R. Hatch, F. Merz, E. Sonnendrücker, and D. Told for fruitful discussions and M. Dunne for extracting code input from the ASDEX Upgrade database. This work was supported by the NuFuSE project which is funded through G8 Multilateral Research by Funding Nu-FuSE Grant JE 520/4-1. The research leading to these results has received funding from the European Research Council under the European Union’s Seventh Framework Programme (FP7/2007–2013)/ERC Grant Agreement No. 277870. The numerical results presented in this work were carried out using the HELIOS supercomputer system at the Computational Simulation