

Spectrally accurate, reverse-mode differentiable bounce-averaging algorithm and its applications

K. Unalmis^{1†}, R. Gaur², R. Conlin³, D. Panici², and E. Kolemen^{2,4,5‡}

¹Electrical and Computer Engineering, Princeton University, New Jersey, USA

²Mechanical and Aerospace Engineering, Princeton University, New Jersey, USA

³IREEAP, University of Maryland, College Park, Maryland, USA

⁴Andlinger Center for Energy and the Environment, Princeton University, New Jersey, USA

⁵Princeton Plasma Physics Laboratory, Princeton, New Jersey, USA

(Received xx; revised xx; accepted xx)

We present a fast, spectrally accurate, automatically differentiable bounce-averaging algorithm implemented in the DESC stellarator optimization suite. Using this algorithm, we can perform efficient optimization of many objectives to improve stellarator performance, such as the ϵ_{eff} proxy for the neoclassical transport coefficient in the low collisionality regime. By employing this differentiable approximation, for the first time, we optimize a finite-beta stellarator to directly reduce neoclassical ripple transport using reverse-mode differentiation. This ensures the cost of differentiation is independent of the number of controllable parameters.

1. Introduction

Stellarators, first conceived by Spitzer Jr (1958), represent a distinct approach to magnetic confinement fusion that offers unique advantages over tokamaks. These toroidal devices achieve plasma confinement through external magnetic fields rather than through plasma current, providing greater design flexibility and operational stability. The absence of a continuous toroidal symmetry allows for magnetic field optimization through boundary shaping, which helps minimize the net toroidal current and thereby avoids current-driven plasma disruptions that plague tokamak operation (Helander 2014).

The design of optimal stellarator configurations is a complex optimization problem involving hundreds of degrees of freedom. Traditional optimization approaches have evolved significantly over the past decades. VMEC (Variational Moments Equilibrium Code), developed by Hirshman & Whitson (1983), served as the foundation for stellarator optimization. Building upon VMEC, several frameworks have emerged: STELLOPT (Lazerson *et al.* 2020; Spong *et al.* 1998), which implements a suite of physics-based optimization criteria, ROSE (Drevlak *et al.* 2018), which focuses on coil optimization and engineering constraints, and more recently, SIMSOPT (Landreman *et al.* 2021). In DESC (Dudt *et al.* 2025; Dudt & Kolemen 2020; Panici *et al.* 2023; Conlin *et al.* 2023; Dudt *et al.* 2023), unlike previous optimizers, it is not necessary to re-solve the magnetohydrodynamic (MHD) force balance equation at each optimization step (Conlin *et al.* 2024). Additional objectives that depend on equilibrium force balance can be optimized simultaneously while ensuring force balance.

In general, traditional approaches to stellarator optimization rely on finite difference techniques. Such techniques yield low-order accurate estimates of derivatives that can

† Email address for correspondence: kunalmis@alumni.princeton.edu

‡ Email address for correspondence: ekolemen@princeton.edu

hinder the ability of the optimizer to find good solutions. Furthermore, finite difference techniques require computing the objective function multiple times to estimate the derivative in the direction of each optimizable parameter; this is infeasible when the number of parameters is large. In contrast, automatic differentiation can compute derivatives of the objective with respect to all input parameters in a single computational pass.

We present a fast, automatically differentiable bounce-averaging algorithm that is used to simplify kinetic models such as drift and gyrokinetics that study phenomena at timescales longer than the bounce orbit time. This algorithm has been implemented in the DESC optimization suite. Previous works (Matsuda & Stewart 1986; Nemov *et al.* 1999, 2008; Kernbichler *et al.* 2016; Petrov & Harvey 2016; Lazerson *et al.* 2020; Velasco *et al.* 2020) have used bounce-averaging to accelerate solution of Fokker-Planck equations. However, such works are not compatible with automatically differentiable optimization. Moreover, while they employ respectable numerical techniques, their computation is discretized with lower-order accuracy than in this work. They may also use an adaptive quadrature, which would be expensive to automatically differentiate. This work enables automatically differentiable optimization to improve stellarator performance with exponential accuracy.

In section 2, we present an application of bounce-averaging to compute the neoclassical transport coefficient in the low collisionality regime where the transport coefficients increase with decreasing collision frequency. Then, in section 3, we describe our numerical methods to compute bounce-averaged objectives for optimization. In section 4, we apply this framework to optimize against neoclassical transport. In section 5, we conclude this work and explain how it can be extended.

2. Neoclassical model of plasma

Our study concerns configurations where magnetic field lines lie on closed, nested toroidal surfaces, known as flux surfaces. We label these surfaces with their enclosed toroidal flux ψ . Such a divergence-free magnetic field may be written in the Clebsch form (D'haeseleer *et al.* 2012), showing that curves of constant (ψ, α) trace field lines.

$$\mathbf{B} = \nabla\psi \times \nabla\alpha \quad (2.1)$$

The dynamics of a magnetized hot plasma differ significantly from that of an unmagnetized fluid. Unlike isotropic hard-sphere collisions that govern the behavior of an uncharged fluid, a plasma behaves differently in directions perpendicular and parallel to the magnetic field lines because of Coulomb collisions. In magnetized plasmas, particles traverse helical trajectories, gyrating around magnetic field lines and drifting across them. The classical transport model assumes a simplistic view of particle collisions and does not adequately incorporate the effects of these drifts. To properly account for these drifts, trapped and passing particles, and the magnetic geometry, we use the neoclassical transport theory.

There are three fundamental length and time scales relevant to magnetized plasmas. The time scales correspond to the particle transit frequency $v_{\text{th},s}/L_B$, where $v_{\text{th},s} = (2T_s/m_s)^{1/2}$ is the thermal speed, the Coulomb collision frequency $\nu_{ss'} \propto T^{3/2}$, and the gyration frequency $\Omega_s = Z_s e |B| / (m_s c)$ where s, s' are the species of interest, $Z_s e$ is the charge, m_s is the mass, and c is the speed of light. For each time scale, the corresponding length scales are the gradient scale length of the magnetic field L_B , the mean free path λ_{mfp} , and the gyroradius $r_{\text{gyro},s} = v_{\text{th},s}/\Omega_s$, respectively. In a magnetized plasma,

$$\nu_{ss'} \sim \frac{v_{\text{th},s}}{L_B} \ll \Omega_s \quad (2.2)$$

$$\lambda_{\text{mfp}} \sim L_B \gg r_{\text{gyro},s}. \quad (2.3)$$

Using a random walk estimate, we can calculate the classical heat transport coefficient in the perpendicular direction as $D_{\perp} \sim \nu_{ss'} r_{\text{gyro},s}^2 \sim T^{5/2}$ (Helander & Sigmar 2005) whereas, using neoclassical theory, we have $\Delta r \sim r_{\text{gyro},s} |B| / |B|_{\text{poloidal}}$ with $|B|$ and $|B|_{\text{poloidal}}$ given by the total and poloidal magnetic field strength, respectively. The transport coefficient is then $D_{\perp} \sim \nu_{ss'} r_{\text{gyro},s}^2 |B|^2 / |B|_{\text{poloidal}}^2 \sim T^{5/2} |B|^2 / |B|_{\text{poloidal}}^2$. Note the ratio $|B| / |B|_{\text{poloidal}}$ strongly depends on the magnetic field geometry and significantly affects the regime of neoclassical transport.

Magnetized plasmas can be weakly or strongly collisional. This is defined by the collisionality $\nu_* = L_B / \lambda_{\text{mfp}}$. In a strongly collisional plasma, particles undergo frequent collisions without covering a significant distance along a magnetic field line, *i.e.*, $\nu_* \gg 1$. Conversely, in a weakly collisional plasma, particles can traverse significant distance before colliding, *i.e.*, $\nu_* \ll 1$. Stellarator plasmas in practical applications tend to be weakly collisional.

Based on the stellarator geometry, the weak collisionality regime can be further divided into the banana or plateau regime depending on the reciprocal of the aspect ratio $\epsilon \sim \iota^{-1} |B|_{\text{poloidal}} / |B|$ where ι is the rotational transform (Helander 2014). Most stellarators lie in the regime where the collisionality $\nu_* \ll \epsilon^{3/2}$. This categorization is illustrated in figure 1.

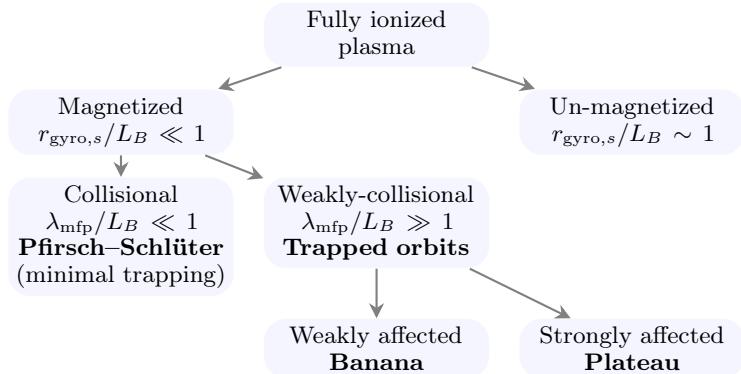


Figure 1: A schematic categorizing neoclassical transport is shown.

The standard neoclassical theory first enabled computation of the neoclassical transport coefficients in the low collisionality regime for a simplified model of the magnetic field. This analysis was later extended to stellarator magnetic fields (Kovrizhnykh 1984; Ochs 2025). The following section outlines this process in one regime of interest to stellarator equilibrium optimization.

2.1. Effective ripple

In the low collision limit $\nu_* \ll \epsilon^{3/2}$, the neoclassical model studies the plasma distribution f determined by a simplified Boltzmann equation known as the drift-kinetic equation. For a particle with mass m , let \mathbf{v}_{\parallel} and \mathbf{v}_{\perp} be the velocity parallel and orthogonal, respectively, to the unit vector magnetic field \mathbf{b} . In the drift-kinetic equation, the velocity space may be parameterized with three independent coordinates: the total energy E , the magnetic moment $\mu = m|\mathbf{v}_{\perp}|^2/(2|B|)$, and the gyrophase angle. In this treatment, the equation is averaged over the gyrophase angle. We seek a steady-state solution and linearize the distribution of guiding centers $f = f_0 + f_1$ into a background f_0 that is Maxwellian in velocity and a higher order correction f_1 . Thus, the background is

parameterized in velocity space with E and the higher order correction with (E, μ) . The linearized drift-kinetic equation reduces to the following PDE (Abel *et al.* 2013).

$$\mathcal{C}[f] = \mathbf{v}_{Ds} \cdot \nabla f_0 + |v_{\parallel}| \mathbf{b} \cdot \nabla f_1 \quad (2.4)$$

$$\mathbf{v}_{Ds} = \frac{|v_{\parallel}|^2}{\Omega_s} \mathbf{b} \times (\mathbf{b} \cdot \nabla \mathbf{b}) + \frac{|v_{\perp}|^2}{2\Omega_s} \frac{\mathbf{b} \times \nabla |B|}{|B|} + \mathbf{v}_{Baños} \quad (2.5)$$

The electric field is neglected in this section as we focus on the $1/\nu$ collisionality regime.

To reduce neoclassical transport, one may minimize the radial particle flux.

$$\Gamma = \int d^3v f_1 \mathbf{v}_{Ds} \cdot \nabla \psi \quad (2.6)$$

Appendix B shows a derivation to extract a dimensionless quantity Γ_0 (2.8) for the optimization objective which is proportional to the flux surface average $\langle \Gamma \rangle$ (B13).

$$\langle \Gamma \rangle = \Gamma_0 \frac{2^{3/2} \pi c^2}{3^2 e^2 m^{3/2}} \int_0^\infty dE \frac{E^{5/2}}{\nu} \frac{\partial f_0}{\partial \psi} \quad (2.7)$$

$$\Gamma_0 = \left(\int_0^{2\pi} d\alpha \int_{|B|_{\min}}^{|B|_{\max}} \frac{d\varrho}{\varrho^3} \sum_w \frac{I_1^2}{I_2} \right) \left(\int_0^{2\pi} d\alpha \int_{\zeta_1}^{\zeta_2} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} \right)^{-1} \quad (2.8)$$

$$I_1(\psi, \alpha, \varrho, w) = \int_{\zeta_1(w)}^{\zeta_2(w)} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} (1 - |B|/\varrho)^{1/2} (4\varrho/|B| - 1) |\nabla \psi| \kappa_G \quad (2.9)$$

$$I_2(\psi, \alpha, \varrho, w) = \int_{\zeta_1(w)}^{\zeta_2(w)} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} (1 - |B|/\varrho)^{1/2} \quad (2.10)$$

The quantity κ_G is the geodesic curvature of the field line (B5), and the velocity space coordinate ϱ is defined as

$$\varrho = E/\mu. \quad (2.11)$$

The number w indexes the well with boundaries $\zeta_1(w)$ and $\zeta_2(w)$ where a bouncing particle is trapped. These boundaries are referred to as bounce points. Only the particles which are trapped within the interval $[\zeta_1, \zeta_2]$ are considered so that $\zeta_1 \leq \min_w \zeta_1(w)$ and $\max_w \zeta_2(w) \leq \zeta_2$. An illustration is shown in figure 2.

In an axisymmetric configuration, integration along the field line for a single poloidal transit between two global maxima of $|B|$ is sufficient for convergence of Γ_0 . On an irrational magnetic surface, it is sufficient to integrate along a single field line (D'haeseleer *et al.* 2012, section 4.9). On a rational or near-rational surface in a non-axisymmetric configuration, it is necessary to integrate along multiple field lines until the surface is covered sufficiently.

The effective ripple modulation amplitude ϵ_{eff} is related to Γ_0 as follows.

$$\epsilon_{\text{eff}}^{3/2} = \frac{\pi}{2^{7/2}} \frac{(B_0 R_0)^2}{\langle |\nabla \psi| \rangle^2} \Gamma_0 \quad (2.12)$$

B_0 is a background magnetic field typically chosen to be $|B|_{\max}$. R_0 is the average major radius of the stellarator. A reason ϵ_{eff} is preferred to Γ_0 as an optimization objective is that the latter vanishes near the magnetic axis, which reduces the ability to distinguish between good and bad configurations. Since ϵ_{eff} depends only on geometry, reducing it by varying the plasma boundary can reduce the radial neoclassical loss of trapped particles.

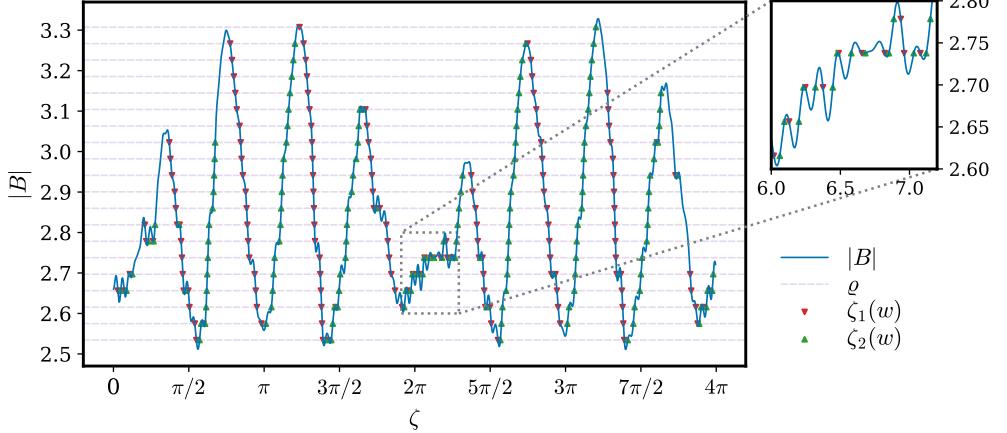


Figure 2: This figure shows bounce points within $(\zeta_1, \zeta_2) = (0, 4\pi)$ on the field line $(\psi, \alpha) = (1, 0)$ for a mesh of ϱ values on a W7-X stellarator. For a given ϱ marked by a horizontal line, $|v_{\parallel}| = 0$ at the bounce points marked by triangles. The plasma distribution vanishes in the hypograph of $|B|$.

3. Algorithm

We briefly describe a few fundamental parts of our algorithm. Section 3.1 discusses the bounce integral in more detail. In section 3.2, we describe efficient quadrature used for these integrals. Section 3.3 discusses our inverse method to solve the ideal MHD equation. In sections 3.4, 3.5, and 3.6 we describe our approach to obtain data along field lines.

To motivate the need for an efficient algorithm, let us estimate the cost of bounce-averaging with a blunt approach to the computation. After discretizing to N_s field lines, where each field line is followed over N_w magnetic wells for each of N_{ϱ} pitch angles, there will be $\mathcal{O}(N_s N_w N_{\varrho})$ bounce integrals. With N_q quadrature points each, the integrand is evaluated at $\mathcal{O}(N_s N_w N_{\varrho} N_q) \sim 10^8$ points. Furthermore, the path of integration is unknown *a priori* because the field lines move during optimization. Finding the position of the field lines on a known grid may involve N_i Newton iterations for each point. With N_c spectral coefficients used to approximate the map on which that root-finding is done, the cost grows to $\mathcal{O}(N_c N_i N_s N_w N_{\varrho} N_q)$. Moreover, the memory required to reverse-mode differentiate the objective grows linearly with the problem size.

3.1. Bounce integral

The bounce integral of x may be written as a time-weighted integral over the trajectory of the particle along its bounce orbit (Mackenbach *et al.* 2023b, section 2). Since the dynamics parallel to the field lines dominate, the particle trajectory may be approximated to follow field lines by parameterizing time t as the distance along a field-line following coordinate. Since the magnetic moment is an adiabatic invariant for which the gyro-average of $d\mu/dt$ is approximately zero, the pitch angle of a bouncing particle stays nearly constant over the timescale to complete bounce orbits when energy is conserved. Labeling the boundaries $\zeta_1(w)$ and $\zeta_2(w)$ of magnetic well w where the parallel velocity vanishes, using the streamline property in curvilinear coordinates

$$|v_{\parallel}|dt = \frac{d\zeta}{\mathbf{b} \cdot \nabla \zeta}, \quad (3.1)$$

and $|v_{\parallel}|^2 = (2E/m)(1 - |B|/\varrho)$ then allows writing the integral as follows.

$$\bar{x}(\psi, \alpha, \varrho, w) = \frac{m^{1/2}}{(2E)^{1/2}} \int_{\zeta_1(w)}^{\zeta_2(w)} \frac{d\zeta}{\mathbf{b} \cdot \nabla \zeta} (1 - |B|/\varrho)^{-1/2} x \quad (3.2)$$

More generally, integrals between bounce points involve a map g , smooth in ζ , weighted by a non-smooth map with behavior matching $|v_{\parallel}|^\eta$ near the bounce points.

$$\int_{\zeta_1(w)}^{\zeta_2(w)} d\zeta |v_{\parallel}|^\eta g(\psi, \alpha, \zeta, \varrho) \quad \eta \in \{-1, 1\} \quad (3.3)$$

3.2. Quadrature

Gaussian quadrature approximates (3.4) for some weight ς positive and continuous

$$\int_{-1}^1 d\zeta \varsigma g(\zeta) \approx \sum_{i=1}^{N_q} \sigma_i g(\zeta_i) \quad (3.4)$$

in the interior by approximating g with its Hermite interpolation polynomial and choosing σ_i, ζ_i to avoid evaluating the derivative (Süli & Mayers 2003). For integrable (3.3), we can construct such a quadrature for ς matching the non-polynomial behavior of $|v_{\parallel}|^\eta$ or, more generally, employ a change of variable whose Jacobian vanishes slowly near singularities such that the integrand can then be approximated by a low-degree polynomial. In the latter approach, the transformation should also be mild enough to prevent unnecessary clustering of quadrature points that would increase the condition number of the problem.

A good transformation for bounce integrals defines z such that $a_1(w, a_2[z]) = \zeta$.

$$a_1: \begin{cases} \mathbb{N} \times [-1, 1] \rightarrow \mathbb{R} \\ w, z \mapsto (z + 1)[\zeta_2(w) - \zeta_1(w)]/2 + \zeta_1(w) \end{cases} \quad (3.5)$$

$$a_2: \begin{cases} [-1, 1] \rightarrow [-1, 1] \\ z \mapsto \sin(\pi z/2) \end{cases} \quad (3.6)$$

$$\int_{\zeta_1(w)}^{\zeta_2(w)} d\zeta |v_{\parallel}|^\eta g(\zeta) \approx \frac{\zeta_2(w) - \zeta_1(w)}{2} \sum_{i=1}^{N_q} \sigma_i |v_{\parallel}|^\eta g(a_1(w, a_2[z_i])) \quad (3.7)$$

When neither bounce point is on a local maximum of $|B|$, the midpoint scheme in z (3.8) has spectral (exponential) convergence (Süli & Mayers 2003, 213).

$$\sigma_i = \pi \sin[\pi(2i - 1)/(2N_q)]/N_q, \quad a_2[z_i] = \cos[\pi(2i - 1)/(2N_q)] \quad (3.8)$$

If, in addition, $\eta = 1$, then (3.9) is preferred as it converges faster.

$$\sigma_i = \pi \sin[\pi i/(N_q + 1)]/(N_q + 1), \quad a_2[z_i] = \cos[\pi i/(N_q + 1)] \quad (3.9)$$

In general, Gauss-Legendre quadrature in z is spectrally accurate if (3.3) integrable. Figures 3, 4, 5, and 6 illustrate the convergence.

It is often of interest to integrate a nonlinear combination of bounce integrals over ϱ . Such integrands can be non-smooth in ϱ due to the logarithmic divergence (Calvo et al. 2017, section 4) of (3.2) as ϱ approaches the value of $|B|$ at any local maxima within or at the bounce points. The Alpert (1999) quadratures are arbitrarily high-order accurate and practical to implement for these singularities.

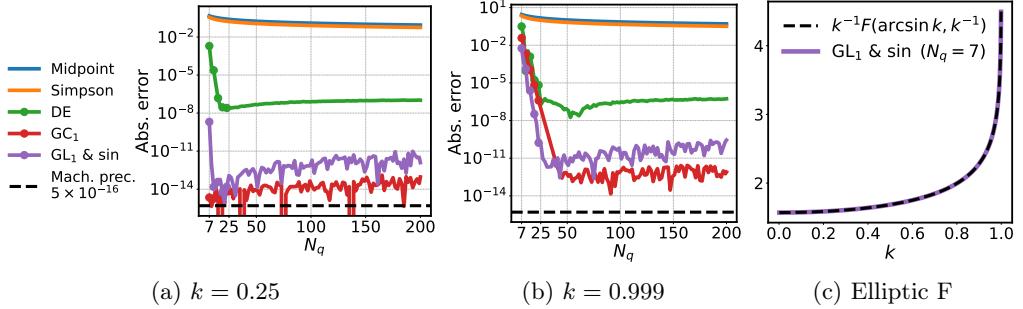


Figure 3: Convergence of quadratures for (3.10). GC₁ and GL₁ show spectral convergence whereas midpoint, Simpson, and double exponential quadratures hit floating point plateaus early.

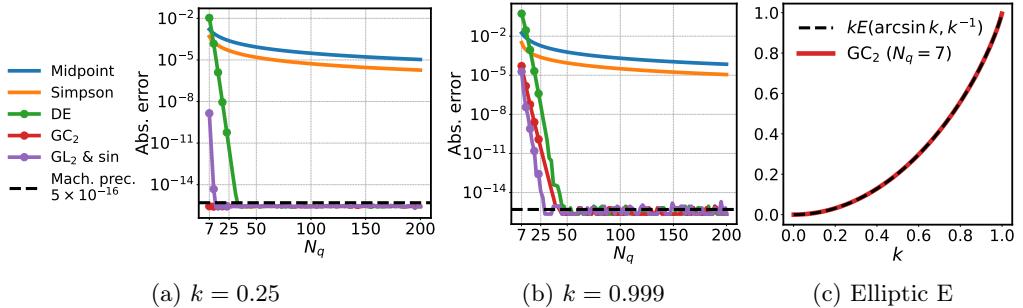


Figure 4: Convergence of quadratures for (3.11). GC₂, GL₂, and double exponential quadratures show spectral convergence.

We compare the following quadratures in their ability to compute elliptic integrals (3.10), (3.11), which are similar to bounce integrals in a simple stellarator geometry.

- (i) Midpoint scheme.
- (ii) Simpson's 1/3 in the interior completed by a midpoint scheme.
- (iii) Double exponential (DE) tanh – sinh.
- (iv) Implicitly weighted Gauss–Chebyshev of the first (GC₁) (3.8) or second kind (GC₂) (3.9).
- (v) Gauss–Legendre (GL₁) or Gauss–Lobatto–Legendre (GL₂) each composed with the sin transformation in (3.6). Compared to the Chebyshev quadrature, this quadrature offers more resolution near the boundary and less in the interior.

$$\int_{-\arcsin k}^{\arcsin k} d\zeta (k^2 - \sin^2 \zeta)^{-1/2} \underset{(3.10)}{=} 2K(k)$$

$$\int_{-\arcsin k}^{\arcsin k} d\zeta (k^2 - \sin^2 \zeta)^{1/2} \underset{(3.11)}{=} 2E(k) + 2(k^2 - 1)K(k)$$

To further benchmark the quadratures in a magnetic field with ripples, we show two more cases that model particles trapped in the following wells. The quantity labeled by f on the right is integrated.

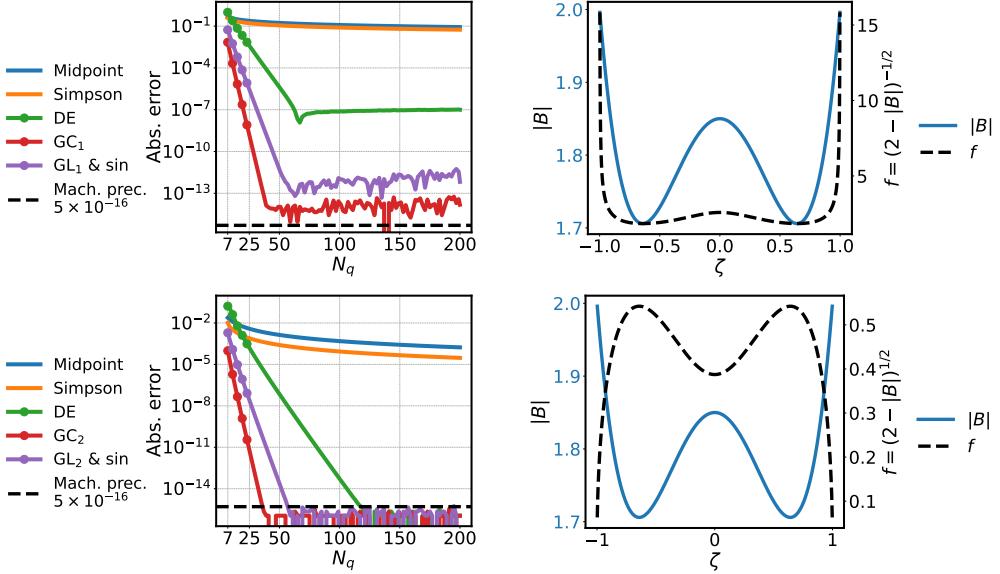


Figure 5: Convergence of quadratures for the well on the right is shown.

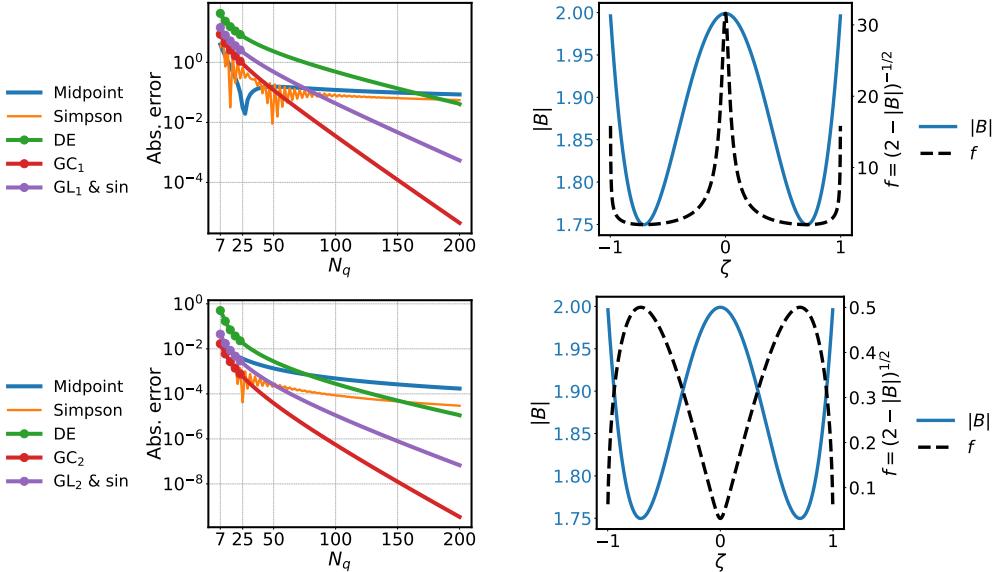


Figure 6: Convergence of quadratures for the well on the right is shown. In the top row, the integrand becomes nearly non-integrable as the parallel velocity nearly vanishes at $\zeta = 0$. In either case, splitting the quadrature there recovers fast convergence.

3.3. Inverse method

In this section, we briefly discuss how we find stellarator equilibria. At static equilibrium, the ideal MHD equations that approximate the behavior of the plasma reduce to

$$\mathbf{B} \cdot \nabla \mathbf{B} = \nabla(p + |B|^2/2) \quad (3.12)$$

$$\nabla \cdot \mathbf{B} = 0, \quad (3.13)$$

which describes a balance between the plasma pressure p , magnetic field pressure $|B|^2$ and the effect of field line curvature $\mathbf{B} \cdot \nabla \mathbf{B}$. We solve the ideal MHD equation using an inverse method. The computational domain is a solid torus in curvilinear flux coordinates (ρ, θ, ζ) where $\rho = (\psi/\psi_{\text{plasma boundary}})^{1/2}$, and (θ, ζ) are angles on a doubly-periodic surface. Λ and ω are to be determined maps that relate the angles (θ, ζ) that parameterize a given plasma boundary to the Clebsch angle.

$$\alpha = \theta + \Lambda - \iota(\zeta + \omega) \quad (3.14)$$

Fourier–Zernike series parameterized in flux coordinates (ρ, θ, ζ) are chosen to approximate Λ , ω , and the map to a cylindrical coordinate system (R, ϕ, Z) in the lab frame.

$$R(\rho, \theta, \zeta) \hat{\mathbf{R}}(\phi) + Z(\rho, \theta, \zeta) \hat{\mathbf{Z}} \quad (3.15)$$

It can be shown from (2.1) that $\nabla \theta \times \nabla \zeta \neq \mathbf{0}$ implies (3.16) and (3.17).

$$\mathbf{B} \cdot \nabla \theta = -[\nabla \psi \cdot (\nabla \theta \times \nabla \zeta)] \left(\frac{\partial \alpha}{\partial \zeta} \right)_{\psi, \theta} \quad (3.16)$$

$$\mathbf{B} \cdot \nabla \zeta = +[\nabla \psi \cdot (\nabla \theta \times \nabla \zeta)] \left(\frac{\partial \alpha}{\partial \theta} \right)_{\psi, \zeta} \quad (3.17)$$

Thus, we find equilibria by searching for a combination (R, Z, Λ, ω) to reduce the force balance error (3.12) at a set of collocation points using pseudo-spectral methods, subject to constraints on the pressure profile and the rotational transform or toroidal current profile. This boundary value problem is then solved as a minimization problem using a trust-region method. In an optimization constrained by force balance, varying (R, Z, Λ, ω) changes the magnetic field and (3.15) such that (3.14) remains valid.

Two advantages of this inverse approach for optimization with bounce-averaged objectives are stated below.

- (i) The variables (θ, ζ) on the boundary surface may be constructed so that maps parameterized in these coordinates are spectrally condensed (Hirshman & Breslau 1998; Hindenlang *et al.* 2025). Consequently, maps parameterized in (ρ, θ, ζ) in the plasma volume tend to have spectral expansions that converge more rapidly.
- (ii) Force balance and other geometric objectives are best computed on a particular grid in (ρ, θ, ζ) which is fixed throughout optimization. This ensures the spectral basis for (R, Z, Λ, ω) can be precomputed, avoiding “off-grid” interpolation of a three-dimensional basis that bottlenecks pseudo-spectral codes (Boyd 2013, section 10.7). Furthermore, if the coordinate system varied throughout the optimization, then so does the optimal grid for interpolation and quadrature. To preserve spectral accuracy, a pseudo-spectral code must first find this optimal grid and compute the basis there. This “moving-grid” interpolation is doubly expensive in optimization because the mentioned tasks must also be differentiated, which consumes significant memory.

These qualities enable faster generation of magnetic field data, which we discuss in the following section.

3.4. Fast interpolation

In this section, we outline our method for fast interpolation.

The Zernike basis concentrates the frequency transform of smooth maps on discs at lower frequencies than geometry-agnostic tensor-product bases. Boyd & Yu show the required number of spectral coefficients is typically half that of Fourier–Chebyshev (Boyd & Yu 2011). This ensures an optimization that varies a finite number of coefficients in the Fourier–Zernike series for (R, Z, Λ, ω) at a time has more freedom compared to expansions in other bases. However, the Zernike basis is expensive to evaluate.

Our algorithm computes the Fourier–Zernike basis once prior to optimization on a uniform $K_\theta \times K_\zeta$ grid in $(\theta, N_{\text{FP}}\zeta) \in [0, 2\pi]^2$ on each surface. Any smooth, periodic map g required by the objective is computed from (R, Z, Λ, ω) on this grid and interpolated with a fast Fourier transform (FFT). The resulting Fourier series are evaluated using type 2 non-uniform FFTs with computational cost that is linearithmic in $K_\theta K_\zeta$ plus linear in the number of points to evaluate (Barnett et al. 2019; Barnett 2021; huan Shih et al. 2021).

$$g_{k_\theta k_\zeta} = \frac{c_{k_\theta}}{4\pi^2} \iint_{[0, 2\pi)^2} d\theta d(N_{\text{FP}}\zeta) g(\theta, N_{\text{FP}}\zeta) e^{-ik_\theta \theta} e^{-ik_\zeta N_{\text{FP}}\zeta} \quad (3.18)$$

$$g(\alpha, \zeta) = \sum_{k_\theta=0}^{\lfloor K_\theta/2 \rfloor} \sum_{k_\zeta=-\lfloor K_\zeta/2 \rfloor}^{\lceil K_\zeta/2 \rceil - 1} \text{Real}\left(g_{k_\theta k_\zeta} e^{ik_\theta \theta(\alpha, \zeta)} e^{ik_\zeta N_{\text{FP}}\zeta}\right) \quad (3.19)$$

$$c_{k_\theta} = 1 \text{ if } k_\theta \in \{0, K_\theta/2\} \text{ else } 2$$

3.5. Map to the mesh of field lines

Evaluating maps along field lines requires finding the position of the field lines on some grid. To identify the coordinate θ at a given point (α, ζ) one may solve equation (3.14). To avoid repeating that inversion everywhere our objective demands, we compute the spectral projection $\{a_{xy}\}$ of the map $\alpha, \zeta \mapsto \theta - \alpha$ onto a tensor-product basis $\{b_{xy}\}$ that is orthogonal with respect to some weight ς .

$$a_{xy} \sim \iint d\alpha d\zeta (\theta - \alpha) \varsigma b_{xy}^*(\alpha, \zeta) \quad (3.20)$$

The Fourier–Chebyshev basis defined on the field period $(\alpha, N_{\text{FP}}\zeta) \in [0, 2\pi]^2$ is chosen for reasons discussed in (Mason & Handscomb 2002, section 5.5, 5.6, 6.3.4), (Boyd 2013, section 4.5).

$$b_{xy}(\alpha, \zeta) = e^{ix\alpha} \cos(y \arccos[N_{\text{FP}}\zeta/\pi - 1]) \quad (3.21)$$

On each flux surface, equation (3.14) is solved on a tensor-product grid of size $X \times Y$ on the Fourier nodes across field lines and the Chebyshev nodes along field lines using Newton iteration with a backtracking line search. The series (3.20) is computed by interpolating $\theta - \alpha$ on that grid with a discrete cosine transform along field lines, followed by an FFT across field lines. The convergence of the series is illustrated in figure 7.

To extend the map beyond a single field period, we use

$$\theta \equiv \alpha_{\text{mod}} + \sum_{x=0}^{\lfloor X/2 \rfloor} \sum_{y=0}^{Y-1} \text{Real}(a_{xy} b_{xy}(\alpha_{\text{mod}}, \zeta_{\text{mod}})) \pmod{2\pi} \quad (3.22)$$

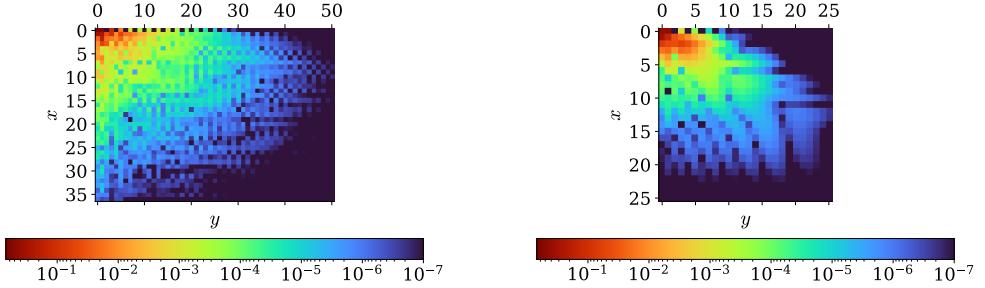
$$\alpha_{\text{mod}} = \alpha_{\text{shift}} \pmod{2\pi}$$

$$\alpha_{\text{shift}} = \alpha + \iota \lfloor N_{\text{FP}}\zeta/(2\pi) \rfloor 2\pi/N_{\text{FP}}$$

$$\zeta_{\text{mod}} = \zeta \pmod{2\pi/N_{\text{FP}}}.$$

The equivalence (3.22) is due to $\alpha + \iota\zeta = \alpha_{\text{shift}} + \iota\zeta_{\text{mod}}$ and uniqueness of solutions to (3.14). Figure 8 shows an illustration. The construction and evaluation of this series is accelerated with partial summation (Boyd 2013, section 10).

This approach avoids issues that result from changing the basis for (R, Z, Λ, ω) at each optimization step (appendix D).



(a) $|a_{xy}|$ on the plasma boundary of an NCSX stellarator with $N_{\text{FP}} = 3$. (b) $|a_{xy}|$ on the plasma boundary of a Heliotron stellarator with $N_{\text{FP}} = 19$.

Figure 7: These figures show the convergence of the spectral projection of $\alpha, \zeta \mapsto \theta - \alpha$ onto the Fourier–Chebyshev basis (3.22). Equation (3.14) was solved to error $\leq 10^{-7}$. Note that if $\omega \rightarrow \Lambda/\iota$ then $\theta - \alpha \rightarrow \iota\zeta$, so the spectral width reduces to one parameter. Thus, if the optimizer is motivated to match higher frequency spectral coefficients of ω with Λ/ι , then field lines can be tracked at lower resolution.

3.6. Jacobian of the map to the mesh of field lines

In this section, we explain how we accelerate the iterative solve discussed in the previous section throughout optimization. To bypass differentiating the iterative solve, we write the tangent and adjoint methods directly (Sapienza *et al.* 2025, section 3.3.3, 3.9.2). For this task, we leverage the implicit function theorem to differentiate solutions θ^* to (3.14) with respect to the optimizable parameters, denoted here with \mathbf{x}_{opt} . Define

$$f: \mathbf{x}_{\text{opt}}, \theta \mapsto \theta + \Lambda - \iota(\zeta + \omega) - \alpha.$$

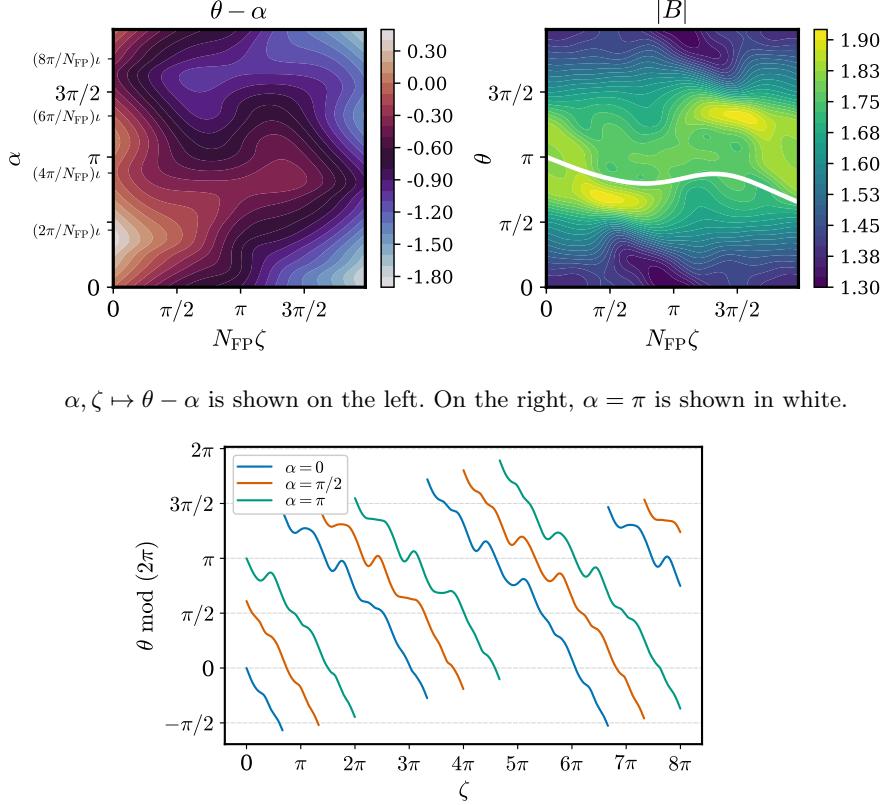
Let $(\mathbf{x}_{\text{opt}}^*, \theta^*)$ satisfy $f(\mathbf{x}_{\text{opt}}^*, \theta^*) = 0$.

$$\frac{\partial f}{\partial \theta}(\mathbf{x}_{\text{opt}}^*, \theta^*) = 1 + \frac{\partial(\Lambda - \iota\omega)}{\partial \theta}(\mathbf{x}_{\text{opt}}^*, \theta^*) \underset{(3.14)}{=} \left(\frac{\partial \alpha}{\partial \theta} \right)_{\psi, \zeta} (\mathbf{x}_{\text{opt}}^*, \theta^*) \quad (3.23)$$

In the (ψ, α, ζ) covariant basis, the only nonzero component of the non-vanishing magnetic field is (3.17), so the derivative (3.23) is invertible. By the implicit function theorem, θ^* is a continuously differentiable map of \mathbf{x}_{opt} and $f(\mathbf{x}_{\text{opt}}, \theta^*(\mathbf{x}_{\text{opt}})) = 0$ near $\mathbf{x}_{\text{opt}}^*$. Moreover,

$$\frac{\partial \theta^*}{\partial \mathbf{x}_{\text{opt}}}(\mathbf{x}_{\text{opt}}) = - \left[\frac{\partial f}{\partial \theta}(\mathbf{x}_{\text{opt}}, \theta^*(\mathbf{x}_{\text{opt}})) \right]^{-1} \frac{\partial f}{\partial \mathbf{x}_{\text{opt}}}(\mathbf{x}_{\text{opt}}, \theta^*(\mathbf{x}_{\text{opt}})). \quad (3.24)$$

Thus, we differentiate directly through the solution θ^* . Likewise, after updating \mathbf{x}_{opt} , we use (3.24) to warm start the next Newton iteration at an initial value that is correct to first order.



$\alpha, \zeta \mapsto \theta - \alpha$ is shown on the left. On the right, $\alpha = \pi$ is shown in white.

Figure 8: θ on the plasma boundary of an NCSX stellarator with $N_{\text{FP}} = 3$.

4. Optimization for reduced neoclassical transport

We present an optimization starting from a finite-beta helically omnigenous (OH) equilibrium. Finite-beta refers to the nonzero ratio of plasma pressure and magnetic pressure. We target flux surfaces near the boundary to reduce the effective ripple while maintaining reasonable elongation and curvature. With weights, w_A, w_C, w_E, w_O, w_R , the objective (4.1) is minimized while ensuring ideal MHD force balance (3.12) is maintained.

$$w_A f_{\text{aspect}}^2 + w_C f_{\text{curv}}^2 + w_E f_{\text{elongation}}^2 + w_O f_{\text{omni}}^2 + w_R f_{\text{ripple}}^2 \quad (4.1)$$

The initial equilibrium along with the definitions of the curvature and elongation objectives are provided in Gaur *et al.* (2025a); Gaur (2024). The results are presented in figure 9. The optimization took less than two hours with a GPU (NVIDIA Corporation 2020).

The omnigenicity objective is based on the work of Dudit *et al.* (2024), where it was shown that optimizing for omnigenicity can in turn reduce the effective ripple. Directly optimizing to reduce the effective ripple instead has the advantage that the optimizer is not biased toward a user-specified omnigenous field. For example in Gaur *et al.* (2025b), we used this property to optimize for an umbilic boundary while maintaining a low effective ripple without biasing the optimizer toward an omnigenous field with a specific helicity.

It should be noted that the assumptions used to derive the effective ripple increase in validity as the magnetic field becomes more omnigenous. For example, the bounce-averaged approach presented here assumes the radial orbit width is small compared to the magnetic field gradient scale length $\Delta r \ll L_B$. When finite orbit width effects

become dominant, particles may traverse “potato” orbits requiring a more global treatment ([Satake *et al.* 2002](#)). Hence, there is utility in optimization that uses both objectives, either simultaneously or in succession.

5. Conclusions

In this work, we optimized a finite-beta configuration to directly reduce neoclassical transport using reverse-mode differentiation. More generally, we upgraded the DESC stellarator optimization suite for fast, accurate, automatically differentiable bounce-averaging. We discussed how we perform moving-grid interpolation without sacrificing spectral accuracy. This accuracy ensures that changes in the objective due to small changes in controllable parameters reflect genuine improvement or degradation rather than noise due to error. Therefore, optimization is more likely to be successful.

Our algorithm enables optimization for many objectives to improve stellarator performance. These include maximization of the second adiabatic invariant ([Helander 2014](#), section 3.7), ([Rodríguez *et al.* 2024](#)), energetic particle confinement ([Nemov *et al.* 2008; Velasco *et al.* 2021](#)), and proxies for gyrokinetic turbulence such as the available energy ([Mackenbach *et al.* 2022, 2023a](#)). We have currently added all but the latter as objectives to DESC. Some of these objectives have previously had limited use in optimization due to expensive computational requirements or difficulty finding desirable configurations. Further demonstration of optimization with them remains as future work.

Acknowledgments

This work is funded through the SciDAC program by the US Department of Energy, Office of Fusion Energy Science, and Office of Advanced Scientific Computing Research under contract numbers DE-AC02-09CH11466, DE-SC0022005, and Field Work Proposal No. 1019. This work was also funded by the Peter B. Lewis Fund for Student Innovation in Energy and the Environment. This research used the resources of the Della computing cluster at Princeton University.

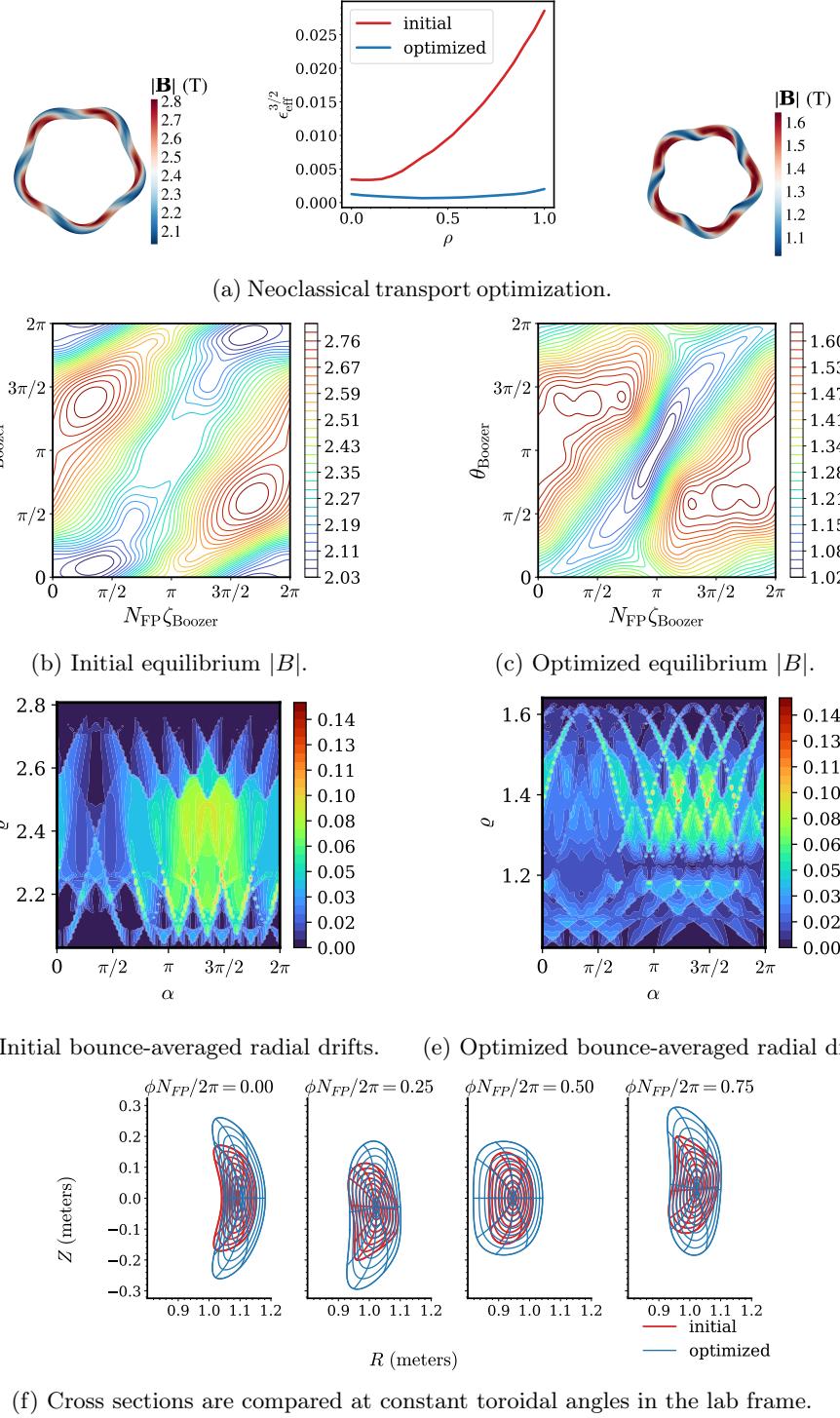


Figure 9: An OH transport optimization. Figures 9b, 9c are shown in Boozer coordinates (D’haeseleer *et al.* 2012). Figures 9d, 9e show bounce-averaged radial drifts, summed in magnitude over $\zeta \in (0, 2\pi)$. The size of the region with large drifts appears reduced.

Appendix A. Open source code

The implementation and supplementary information describing how correctness of automatic differentiation is enforced may be found in the `DESC` repository ([Dudt et al. 2025](#)). The implementation uses accelerated linear algebra `XLA` and Google’s `JAX` library ([Bradbury et al. 2018](#)). JIT compilation in `JAX` compiles the code at the start of an optimization. Optimization may be accelerated on CPUs, GPUs, and TPUs.

Appendix B. Effective ripple

We explain a short derivation of ϵ_{eff} , similar to the one used in [Nemov et al. \(1999\)](#).

To obtain an explicit expression for f_1 from (2.4), we will bounce-integrate the drift-kinetic equation. Applying this operator to the drift-kinetic equation discretizes the spatial coordinate ζ to a set of integral equations labeled by the magnetic well index w .

The collision operator in (2.4) is chosen to capture pitch angle scattering.

$$\mathcal{C}[f] = \nu m \frac{|v_{\parallel}|}{|B|} \frac{\partial}{\partial \mu} \left(\mu |v_{\parallel}| \frac{\partial f}{\partial \mu} \right) \quad (\text{B } 1)$$

These derivatives are at fixed position and energy. The collision frequency ν depends only on the energy of the particle. The velocity ratio is $|v_{\parallel}|/|v| = (1 - |B|/\varrho)^{1/2}$. The nullspace of this collision operator contains velocity-isotropic distributions, so $\mathcal{C}[f_0 + f_1] = \mathcal{C}[f_1]$. In this form, (2.4) is the linearized Lorentz-gas Fokker-Planck equation ([Goldston & Rutherford 1995](#), section 13).

In weakly collisional plasmas, the collision frequency is small compared to the particle bounce frequency. Consequently, fluctuations due to collisions homogenize along field lines rapidly, implying that the spatial variation in the plasma distribution along field lines in any particular magnetic well is small. Therefore, we approximate f_0 and f_1 to be spatially uniform along field lines in any particular magnetic well.

$$\nabla f = \left(\frac{\partial f}{\partial \psi} \right)_{\alpha, \zeta, E, \mu} \nabla \psi + \left(\frac{\partial f}{\partial \alpha} \right)_{\psi, \zeta, E, \mu} \nabla \alpha + \left(\frac{\partial f}{\partial \zeta} \right)_{\psi, \alpha, E, \mu} \nabla \zeta \quad (\text{B } 2)$$

$$|\nabla f| \gg |(\partial f / \partial \zeta) \nabla \zeta| \quad (\text{B } 3)$$

Nested flux surfaces (2.1) then imply the parallel drift $\mathbf{v}_{\text{Baños}}$ and the parallel spatial derivative of f_1 will be negligible in the bounce-integrated drift-kinetic equation.

$$\begin{aligned} \overline{\mathcal{C}[f_1]} &= \nu m \frac{\partial}{\partial \mu} \mu \int \frac{d\zeta}{\mathbf{b} \cdot \nabla \zeta} \frac{|v_{\parallel}|}{|B|} \frac{\partial f_1}{\partial \mu} \\ &= \nu m \frac{\partial}{\partial \mu} \mu \frac{\partial f_1}{\partial \mu} \frac{|v_{\parallel}|^2 / |B|}{|v_{\parallel}|^2 / |B|} \\ &= \frac{\partial f_0}{\partial \psi} \overline{\mathbf{v}_{\text{Ds}} \cdot \nabla \psi} \end{aligned} \quad (\text{B } 4)$$

To write the last relation (B4), we assume there are sufficiently many passing particles so that f_0 is independent of α .[†] We proceed to invert the collision operator. First label the geodesic curvature of the field line.

$$\kappa_G = [\mathbf{b} \times (\mathbf{b} \cdot \nabla \mathbf{b})] \cdot \frac{\nabla \psi}{|\nabla \psi|} = \frac{\mathbf{b} \times \nabla |B|}{|B|} \cdot \frac{\nabla \psi}{|\nabla \psi|} \quad (\text{B } 5)$$

[†] The claim $|\overline{(\partial f_0 / \partial \alpha) \mathbf{v}_{\text{Ds}} \cdot \nabla \alpha}| \ll |\overline{\mathbf{v}_{\text{Ds}} \cdot \nabla f_0}|$ requires care because $|\nabla \alpha|$ grows unbounded when the magnetic shear is nonzero. If the distribution has variation across field lines, we assume it is captured by the higher order correction f_1 .

The second equality is a consequence of ideal MHD force balance (3.12). Now the primitive with respect to μ of the bounce-integrated radial drift velocity is identified as follows.

$$\frac{\partial}{\partial \mu} |v_{\parallel}| \beta = \overline{\mathbf{v}_{Ds} \cdot \nabla \psi} \quad (\text{B } 6)$$

$$\frac{\partial \beta}{\partial \mu} = \frac{\mathbf{v}_{Ds} \cdot \nabla \psi}{|v_{\parallel}|} = (|v|^2 |v_{\parallel}|^{-1} + |v_{\parallel}|) \frac{|\nabla \psi| \kappa_G}{2 \Omega_s} \quad (\text{B } 7)$$

$$\beta = -(3|v|^2 |v_{\parallel}| + |v_{\parallel}|^3) \frac{m |\nabla \psi| \kappa_G}{6 \Omega_s |B|} \quad (\text{B } 8)$$

Inverting the μ derivative in equation (B 4) completes the inversion of the collision operator.

$$\begin{aligned} \nu m \frac{\partial}{\partial \mu} \left(\mu \frac{\partial f_1}{\partial \mu} \overline{|v_{\parallel}|^2 / |B|} \right) &= \frac{\partial}{\partial \mu} \left(\frac{\partial f_0}{\partial \psi} \overline{|v_{\parallel}| \beta} \right) \\ \frac{\partial f_1}{\partial \mu} &= \frac{\partial f_0}{\partial \psi} \frac{\overline{|v_{\parallel}| \beta}}{\nu m \mu \overline{|v_{\parallel}|^2 / |B|}} \end{aligned} \quad (\text{B } 9)$$

To compute (2.6) we will use the (E, μ) parameterization of velocity space.

$$\int d^3 \mathbf{v} = \frac{2\pi}{m^2} |B| \int_0^\infty dE \int_0^{E/|B|} \frac{d\mu}{|v_{\parallel}|} \quad (\text{B } 10)$$

$$= \frac{2^{1/2}\pi}{m^{3/2}} |B| \int_0^\infty dE E^{1/2} \int_\infty^{|B|} \frac{d\varrho}{\varrho^2 (1 - |B|/\varrho)^{1/2}} \quad (\text{B } 11)$$

The plasma distribution vanishes where $\mu \geq E/|B|$, so the integration region was truncated. Using (B 10), applying integration by parts in the μ coordinate, and enforcing the boundary condition $\lim_{\mu \rightarrow 0} f_1 = 0$ at fixed energy, the radial particle flux (2.6) can be written in terms of known quantities as follows.

$$\Gamma = - \int d^3 \mathbf{v} |v_{\parallel}| \beta \frac{\partial f_1}{\partial \mu} \quad (\text{B } 12)$$

To make optimization efficient, the flux surface average of the radial particle flux is of interest to minimize. This is the average on an infinitesimal volume covering the surface.

$$\langle \Gamma \rangle = \left(\int \frac{ds}{|\nabla \psi|} \Gamma \right) \left(\int \frac{ds}{|\nabla \psi|} \right)^{-1} \quad (\text{B } 13)$$

Here ds is the differential surface area Jacobian. As equation (B 12) enables computing the radial particle flux through a quotient of bounce integrals along the magnetic field line (B 9), it is more tractable to also compute the flux surface average along the field line.

$$\langle \Gamma \rangle = \left(\int_0^{2\pi} d\alpha \int_{\mathbb{R}} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} \Gamma \right) \left(\int_0^{2\pi} d\alpha \int_{\mathbb{R}} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} \right)^{-1} \quad (\text{B } 14)$$

We proceed to extract a dimensionless quantity Γ_0 for the optimization objective. First we use (B 11) and (B 14) to remove the spatial dependence in the boundary of the velocity

integral.

$$\begin{aligned} \langle I \rangle &= \frac{2\pi}{m^2} \left(\int_0^{2\pi} d\alpha \int_0^\infty dE E \int_{\mathbb{R}} \frac{d\zeta}{\mathbf{b} \cdot \nabla \zeta} \int_{|B|}^\infty \frac{d\varrho}{\varrho^2} \beta \frac{\partial f_1}{\partial \mu} \right) \left(\int_0^{2\pi} d\alpha \int_{\mathbb{R}} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} \right)^{-1} \\ &= \frac{2\pi}{m^2} \left(\int_0^{2\pi} d\alpha \int_0^\infty dE E \int_{|B|_{\min}}^{|B|_{\max}} \frac{d\varrho}{\varrho^2} \sum_w \overline{|v_\parallel| \beta} \frac{\partial f_1}{\partial \mu} \right) \left(\int_0^{2\pi} d\alpha \int_{\mathbb{R}} \frac{d\zeta}{\mathbf{B} \cdot \nabla \zeta} \right)^{-1} \end{aligned} \quad (\text{B } 15)$$

Here $|B|_{\min}$ and $|B|_{\max}$ are the min and max values on the flux surface. The integral was truncated at $|B|_{\max}$ as $f_1 = 0$ for passing particles. Now changing coordinates in (B 8)

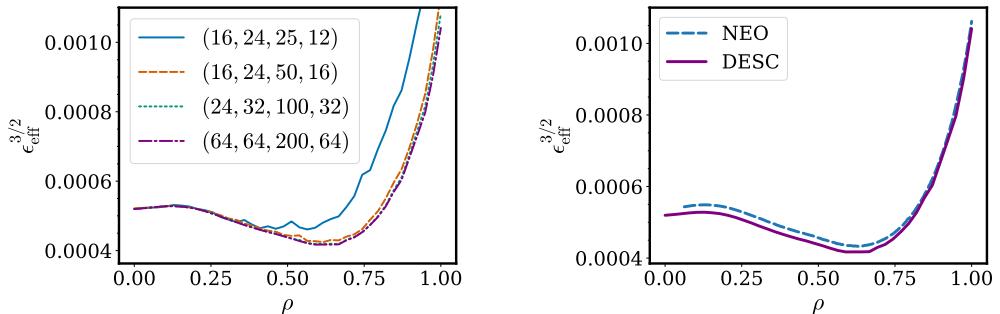
$$\beta = -\frac{(2mE^3)^{1/2}c}{3\varrho e|B|}(1-|B|/\varrho)^{1/2}(4\varrho/|B|-1)|\nabla\psi|\kappa_G \quad (\text{B } 16)$$

and using the new partition for the velocity integral (B 15), the expression (B 14) may be approximated using a sum over all wells in the interval $[\zeta_1, \zeta_2]$ (2.7).

B.1. Resolution scan for the neoclassical transport coefficient

Figure 10a presents a resolution scan for ϵ_{eff} . Figure 10b compares the result to the NEO code (Nemov *et al.* 1999), which uses a finite difference technique and requires transforming to Boozer coordinates. For the singular integrals, NEO uses a Runge-Kutta scheme equivalent to the Simpson method.

We mention some performance benchmarking below. Computing ϵ_{eff} and its derivative with respect to all parameters, on ten flux surfaces, following each field line for 75 field periods, with resolution $(K_\theta, K_\zeta, X, Y, N_\varrho, N_q) = (32, 32, 32, 32, 100, 32)$ was profiled to take less than one and ten seconds, respectively, with a CPU (Intel Corporation 2019). These computations are at least an order of magnitude faster with a GPU.



(a) Five field lines are followed for 100 field periods. The legend shows the resolution (X, Y, N_ϱ, N_q) . $K_\theta = K_\zeta = 33$.

(b) NEO-DESC comparison.

Figure 10: Resolution scan for ϵ_{eff} on the W7-X equilibrium in the DESC repository.

Appendix C. Bounce-averaged drifts in a shifted-circle model

In a shifted-circle model for plasma equilibrium, we can obtain analytical expressions for bounce-averaged drifts. We further verify our algorithm with this model in figure 11.

In this model, the magnetic field can be written

$$\mathbf{B} = \nabla\alpha \times \nabla\chi = F\nabla\phi + \frac{d\chi}{d\rho} \frac{\rho}{R_0} \nabla\vartheta \quad (\text{C } 1)$$

where $\alpha = \phi - \iota^{-1}\vartheta$, χ is the poloidal flux, F is the enclosed poloidal current, R_0 is the average major radius, and ρ is a radial coordinate. To lowest order, the Grad-Shafranov equation has the constant solution $F = F_0$. To the next order, the pressure gradient is

$$dp/d\rho = -F_0 R^{-2} (dF/d\rho). \quad (\text{C } 2)$$

To first order, the poloidal magnetic field can be ignored so that the field satisfies $|B| = B_0(1 - \epsilon \cos \vartheta)$ and $\mathbf{b} \cdot \nabla\vartheta = G_0(1 - \epsilon \cos \vartheta)$ where $\epsilon \ll 1$ is the reciprocal of the aspect ratio. B_0 and G_0 are constants. In this model, the global shear \hat{s} , normalized pressure gradient, and integrated local shear are

$$\hat{s} = -\rho\iota^{-1}(d\iota/d\rho) \quad (\text{C } 3)$$

$$\alpha_{\text{MHD}} = -2^{-1}\iota^{-2}(dp/d\rho) \quad (\text{C } 4)$$

$$\text{gds21} = (d\chi/d\rho)(d\iota^{-1}/d\rho)\nabla\chi \cdot \nabla\alpha = -\hat{s}(\hat{s}\vartheta - |B|^{-4}\alpha_{\text{MHD}} \sin \vartheta) + \mathcal{O}(\epsilon). \quad (\text{C } 5)$$

The binormal, geometric part of the $\nabla|B|$ drift is

$$(\nabla|B|)_{\text{drift}} = |B|^{-3}(\mathbf{B} \times \nabla|B|) \cdot \nabla\alpha = f_2(-\hat{s} + \cos \vartheta - \hat{s}^{-1}\text{gds21} \sin \vartheta). \quad (\text{C } 6)$$

The binormal, geometric part of the curvature drift is

$$\begin{aligned} \text{cvdrift} &= |B|^{-3}[\mathbf{B} \times \nabla(p + |B|^2/2)] \cdot \nabla\alpha \\ &= (\nabla|B|)_{\text{drift}} + f_3|B|^{-3}(dp/d\rho) \\ &= f_2(-\hat{s} + \cos \vartheta + \hat{s}\vartheta \sin \vartheta - B_0^{-4}\alpha_{\text{MHD}} \sin^2 \vartheta) + f_3B_0^{-2}\alpha_{\text{MHD}} + \mathcal{O}(\epsilon). \end{aligned} \quad (\text{C } 7)$$

The scalars f_2 and f_3 contain some constants. The bounce-averaged drift is

$$\langle v_D \rangle = \left(\int_{\vartheta_1}^{\vartheta_2} \frac{d\vartheta}{\mathbf{b} \cdot \nabla\vartheta} |v_\parallel|^{-1} \right)^{-1} \int_{\vartheta_1}^{\vartheta_2} \frac{d\vartheta}{\mathbf{b} \cdot \nabla\vartheta} \left[|v_\parallel| \text{cvdrift} + \frac{|v_\perp|^2}{2|v_\parallel|} (\nabla|B|)_{\text{drift}} \right]. \quad (\text{C } 8)$$

As used in Connor *et al.* (1983) and shown by Hegna (2015), in the limit of a large aspect ratio shifted-circle model, the parallel speed of a particle with a fixed energy is $|v_\parallel| = (2E/m)^{1/2}(2\epsilon\lambda B_0)^{1/2}(k^2 - \sin^2(\vartheta/2))^{1/2}$ where

$$k^2 = 2^{-1}[(1 - \lambda B_0)(\epsilon\lambda B_0)^{-1} + 1] \quad (\text{C } 9)$$

parametrizes the pitch angle λ . Using these simplifications and $|v_\perp|^2/2 = E - |v_\parallel|^2/2$,

$$\begin{aligned} \langle v_D \rangle &= \left(\int_{-2 \arcsin k}^{2 \arcsin k} \frac{d\vartheta}{\mathbf{b} \cdot \nabla\vartheta} (2\epsilon\lambda B_0)^{-1/2}(k^2 - \sin^2(\vartheta/2))^{-1/2} \right)^{-1} \\ &\quad \int_{-2 \arcsin k}^{2 \arcsin k} \frac{d\vartheta}{\mathbf{b} \cdot \nabla\vartheta} \left[(2\epsilon\lambda B_0)^{1/2}(k^2 - \sin^2(\vartheta/2))^{1/2} \text{cvdrift} \right. \\ &\quad \left. - 2^{-1/2}(\epsilon\lambda B_0)^{1/2}(k^2 - \sin^2(\vartheta/2))^{1/2} (\nabla|B|)_{\text{drift}} \right. \\ &\quad \left. + 2^{-3/2}(\epsilon\lambda B_0)^{-1/2}(k^2 - \sin^2(\vartheta/2))^{-1/2} (\nabla|B|)_{\text{drift}} \right]. \end{aligned} \quad (\text{C } 10)$$

The following identities simplify (C 10). The incomplete elliptic integrals are converted to complete elliptic integrals using the Reciprocal-Modulus transformation in (C 11) and (C 12) (Olver *et al.* 2024). K and E are complete elliptic integrals of the first and second kind, respectively.

$$I_0 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{-1/2} = 4K(k) \quad (\text{C } 11)$$

$$I_1 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{1/2} = 4 [E(k) + (k^2 - 1)K(k)] \quad (\text{C } 12)$$

$$I_2 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{-1/2} \vartheta \sin(\vartheta) = 16 [E(k) + (k^2 - 1)K(k)] \quad (\text{C } 13)$$

$$I_3 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{1/2} \vartheta \sin(\vartheta) = \frac{32}{9} [E + (k^2 - 1)^2 K] \quad (\text{C } 14)$$

$$I_4 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{-1/2} \sin^2(\vartheta) = \frac{16}{3} [(2k^2 - 1)E + (1 - k^2)K] \quad (\text{C } 15)$$

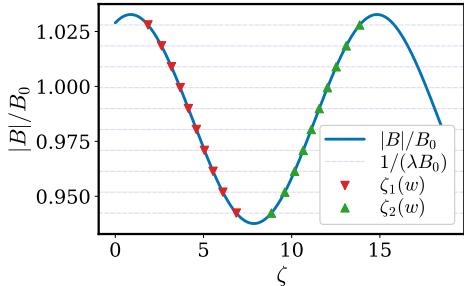
$$\begin{aligned} I_5 &= \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{1/2} \sin^2(\vartheta) \\ &= \frac{32}{30} [2(1 - k^2 + k^4)(E - (k^2 - 1)K) - (1 - 3k^2 + 2k^4)k^2 K] \end{aligned} \quad (\text{C } 16)$$

$$I_6 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{-1/2} \cos(\vartheta) = 8E - 4K \quad (\text{C } 17)$$

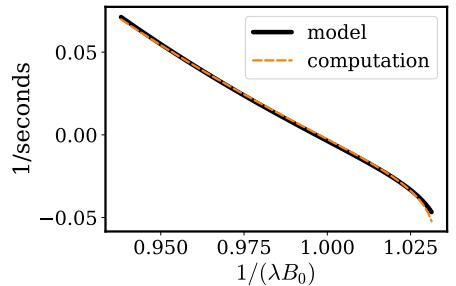
$$I_7 = \int_{-2 \arcsin k}^{2 \arcsin k} d\vartheta (k^2 - \sin^2(\vartheta/2))^{1/2} \cos(\vartheta) = \frac{4}{3} [(2k^2 - 1)E - (k^2 - 1)K] \quad (\text{C } 18)$$

Using these formulae, to lowest order, the bounce-averaged drift is

$$\langle v_D \rangle = \frac{1}{G_0} \left[f_3 \frac{\alpha_{\text{MHD}}}{B_0^2} I_1 - \frac{f_2}{2} \left(\hat{s} (I_0 + I_1 - I_2 - I_3) + \frac{\alpha_{\text{MHD}}}{B_0^4} (I_4 + I_5) - (I_6 + I_7) \right) \right]. \quad (\text{C } 19)$$



(a) For a pitch marked by a horizontal line, $|v_\parallel| = 0$ at the points marked by triangles.

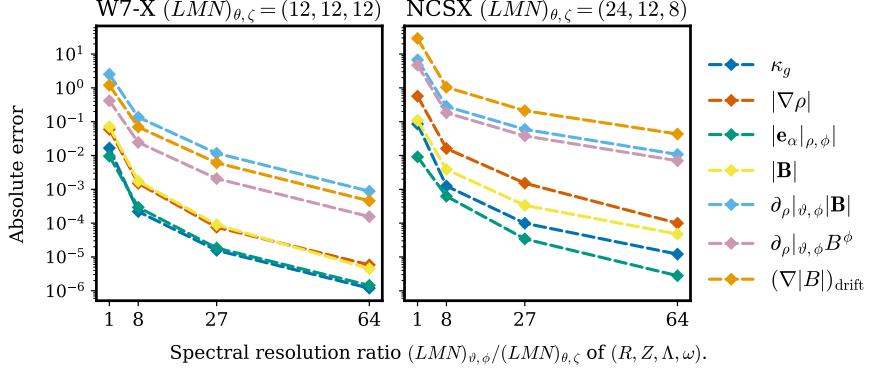


(b) The bounce-averaged binormal drift in the configuration in figure 11a is compared.

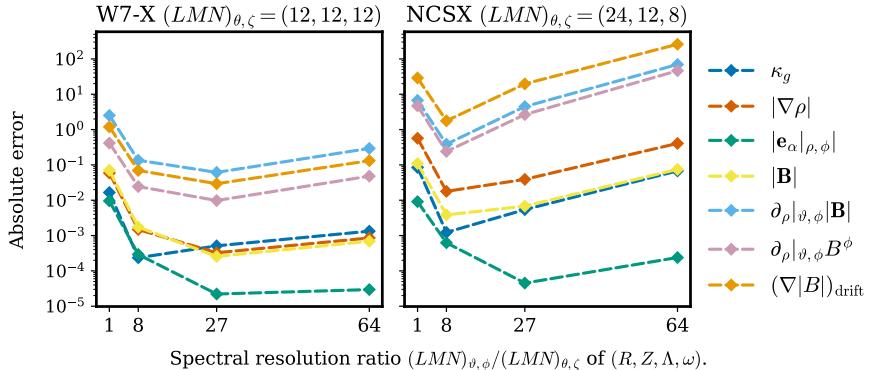
Figure 11: This figure compares our shifted-circle model for the binormal drift to the result computed by our algorithm. The minor difference in figure 11b is because the shifted-circle model is accurate to $\mathcal{O}(\epsilon^2)$.

Appendix D. Issues with changing the spectral basis to straight field line coordinates

In figure 12, we show that parameterizing our basis for (R, Z, Λ, ω) in straight field line coordinates $(\vartheta, \phi) = (\theta + \Lambda, \zeta + \omega)$ (Grimm et al. 2012) is inefficient and ill-conditioned. Therefore, we use the approach discussed in the main text instead.



(a) Equation (3.14) was solved to error $\leq 10^{-10}$ on the tensor-product of the optimal concentric sampling grid (Ramos-López et al. 2016) in $(\rho, \vartheta) \in [0, 1] \times [0, 2\pi]$ and a uniform grid in $\phi \in [0, 2\pi/N_{FP}]$. (R, Z, Λ, ω) were interpolated to a Fourier-Zernike series in (ρ, ϑ, ϕ) with maximum mode numbers $(L, M, N)_{\vartheta, \phi}$ on this grid. The interpolation used a $1.5 \times$ oversampled, in both ρ and ϑ , weighted least-squares fit to improve conditioning for the Zernike series, followed by an FFT in ϕ . (The optimal grid for interpolation to a Zernike series does not coincide with the optimal grid for quadrature to project onto the Zernike basis because the Zernike basis is not a tensor-product basis. Interpolation with a weighted least-squares fit was chosen because the interpolation grid is sparser than the quadrature grid.) Each quantity was then computed on a uniform grid in (ρ, ϑ, ϕ) . Quadrature required to compute a quantity in the plot was done on an over-sampled grid to account for nonlinearity in the computation from (R, Z, Λ, ω) . Zernike polynomials were evaluated with stable Jacobi polynomial recurrence relations using the algorithm in Elmacioglu et al. (2025).



(b) This is the same demonstration as figure 12a except equation (3.14) is solved to error 10^{-7} .

Figure 12: This figure shows the error induced by changing the Fourier-Zernike basis for (R, Z, Λ, ω) from flux coordinates (θ, ζ) to the straight field line coordinates (ϑ, ϕ) . Fitting at the resolution that obtains the error of 10^{-4} Tesla in $|B|$ on the NCSX stellarator in figure 12a took 10 minutes with a CPU (Intel Corporation 2019).

REFERENCES

- ABEL, I G, PLUNK, G G, WANG, E, BARNES, M, COWLEY, S C, DORLAND, W & SCHEKOCHIHIHIN, A A 2013 Multiscale gyrokinetics for rotating tokamak plasmas: fluctuations, transport and energy flows. *Reports on Progress in Physics* **76** (11), 116201.
- ALPERT, BRADLEY K. 1999 Hybrid gauss-trapezoidal quadrature rules. *SIAM Journal on Scientific Computing* **20** (5), 1551–1584.
- BARNETT, ALEX H. 2021 Aliasing error of the $\exp(\beta\sqrt{1-z^2})$ kernel in the nonuniform fast fourier transform. *Applied and Computational Harmonic Analysis* **51**, 1–16.
- BARNETT, ALEXANDER H., MAGLAND, JEREMY & AF KLINTEBERG, LUDVIG 2019 A parallel nonuniform fast fourier transform library based on an “exponential of semicircle” kernel. *SIAM Journal on Scientific Computing* **41** (5), C479–C504, arXiv: <https://doi.org/10.1137/18M120885X>.
- BOYD, J.P. 2013 *Chebyshev and Fourier Spectral Methods: Second Revised Edition*. Dover Publications.
- BOYD, JOHN P. & YU, FU 2011 Comparing seven spectral methods for interpolation and for solving the poisson equation in a disk: Zernike polynomials, logan–shepp ridge polynomials, chebyshev–fourier series, cylindrical robert functions, bessel–fourier expansions, square-to-disk conformal mapping and radial basis functions. *Journal of Computational Physics* **230** (4), 1408–1438.
- BRADBURY, JAMES, FROSTIG, ROY, HAWKINS, PETER, JOHNSON, MATTHEW JAMES, LEARY, CHRIS, MACLAURIN, DOUGAL, NECULA, GEORGE, PASZKE, ADAM, VANDERPLAS, JAKE, WANDERMAN-MILNE, SKYE & ZHANG, QIAO 2018 JAX v0.7.2: composable transformations of Python+NumPy programs.
- CALVO, IVÁN, PARRA, FELIX I, VELASCO, JOSÉ LUIS & ALONSO, J ARTURO 2017 The effect of tangential drifts on neoclassical transport in stellarators close to omnigeneity. *Plasma Physics and Controlled Fusion* **59** (5), 055014.
- CONLIN, RORY, DUDT, DANIEL W, PANICI, DARIO & KOLEMEN, EGEMEN 2023 The DESC stellarator code suite. Part 2. Perturbation and continuation methods. *Journal of Plasma Physics* **89**, 955890305.
- CONLIN, RORY, KIM, PATRICK, DUDT, DANIEL W., PANICI, DARIO & KOLEMEN, EGEMEN 2024 Stellarator optimization with constraints. *Journal of Plasma Physics* **90** (5), 905900501.
- CONNOR, J.W., HASTIE, R.J. & MARTIN, T.J. 1983 Effect of pressure gradients on the bounce-averaged particle drifts in a tokamak. *Nuclear Fusion* **23** (12), 1702.
- D’HAESELEER, W. D., HITCHON, W. N. G., CALLEN, J. D. & SHOHET, J. L. 2012 *Flux coordinates and magnetic field structure: a guide to a fundamental tool of plasma theory*. Springer Science & Business Media.
- DREVLAK, M., BEIDLER, C. D., GEIGER, J., HELANDER, P. & TURKIN, Y. 2018 Optimisation of stellarator equilibria with rose. *Nuclear Fusion* **59**, 016010.
- DUDT, D.W., CONLIN, R., PANICI, D. & KOLEMEN, E. 2023 The desc stellarator code suite part 3: Quasi-symmetry optimization. *Journal of Plasma Physics* **89** (2), 955890201.
- DUDT, DANIEL, CONLIN, RORY, PANICI, DARIO, UNALMIS, KAYA, ELMACIOGLU, YIGIT GUNSUR, GAUR, RAHUL, KIM, PATRICK & KOLEMEN, EGEMEN 2025 DESC. <https://github.com/PlasmaControl/DESC>. <https://doi.org/10.5281/zenodo.4876504>.
- DUDT, DW & KOLEMEN, E 2020 DESC: A stellarator equilibrium solver. *Physics of Plasmas* **27** (10).
- DUDT, DANIEL W, GOODMAN, ALAN G, CONLIN, RORY, PANICI, DARIO & KOLEMEN, EGEMEN 2024 Magnetic fields with general omnigenity. *Journal of Plasma Physics* **90** (1), 905900120.
- ELMACIOGLU, YIGIT GUNSUR, CONLIN, RORY, DUDT, DANIEL W., PANICI, DARIO & KOLEMEN, EGEMEN 2025 Zernipax: A fast and accurate zernike polynomial calculator in python. *Applied Mathematics and Computation* **505**, 129534.
- GAUR, RAHUL 2024 Omnigenous equilibria with enhanced stability: Dataset and analysis files. <https://doi.org/10.5281/zenodo.13887566>.
- GAUR, R, CONLIN, R, DICKINSON, D, PARISI, J F, PANICI, D, DUDT, D, KIM, P, UNALMIS, K, DORLAND, W D & KOLEMEN, E 2025a Omnigenous stellarators with improved ideal and kinetic ballooning stability. *Plasma Physics and Controlled Fusion* **67** (12), 125015.
- GAUR, R., PANICI, D.G., ELDER, T.M., LANDREMAN, M., UNALMIS, K., ELMACIOGLU, Y.G.,

- DUDT, D.W., CONLIN, R. & KOLEMEN, E. 2025b Omnigenous umbilic stellarators. *Journal of Plasma Physics* **91** (6), E151.
- GOLDSTON, R. J. & RUTHERFORD, P. H. 1995 *Introduction to Plasma Physics*. Taylor & Francis.
- GRIMM, R. C., GREENE, J. M. & JOHNSON, J. L. 2012 Computation of the magnetohydrodynamic spectrum in axisymmetric toroidal confinement systems. *Methods in Computational Physics* **16**, 253.
- HEGNA, C. C. 2015 The effect of three-dimensional fields on bounce averaged particle drifts in a tokamak. *Physics of Plasmas* **22** (7), 072510.
- HELANDER, PER 2014 Theory of plasma confinement in non-axisymmetric magnetic fields. *Reports on Progress in Physics* **77** (8), 087001.
- HELANDER, PER & SIGMAR, DIETER J 2005 *Collisional transport in magnetized plasmas*, , vol. 4. Cambridge University Press.
- HINDENLANG, FLORIAN, PLUNK, GABRIEL G & MAJ, OMAR 2025 Computing mhd equilibria of stellarators with a flexible coordinate frame. *Plasma Physics and Controlled Fusion* **67** (4), 045002.
- HIRSHMAN, S. P. & BRESLAU, J. 1998 Explicit spectrally optimized fourier series for nested magnetic surfaces. *Physics of Plasmas* **5** (7), 2664–2675.
- HIRSHMAN, S. P. & WHITSON, J. C. 1983 Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria. *The Physics of fluids* **26**, 3553.
- INTEL CORPORATION 2019 Intel core i7-9750h.
- KERNBICHLER, W., KASILOV, S V, KAPPER, G, MARTITSCH, A F, NEMOV, V V, ALBERT, C & HEYN, M F 2016 Solution of drift kinetic equation in stellarators and tokamaks with broken symmetry using the code neo-2. *Plasma Physics and Controlled Fusion* **58** (10), 104001.
- KOVRIZHNYKH, L.M. 1984 Neoclassical theory of transport processes in toroidal magnetic confinement systems, with emphasis on non-axisymmetric configurations. *Nuclear Fusion* **24** (7), 851.
- LANDREMAN, M., MEDASANI, B., WECHSUNG, F., GIULIANI, A., JORGE, R. & ZHU, C. 2021 SIMSOPT: A flexible framework for stellarator optimization. *Journal of Open Source Software* **6**, 3525.
- LAZERSON, SAMUEL, SCHMITT, JOHN, ZHU, CAOXIANG, BRESLAU, JOSHUA & STELLOPT DEVELOPERS, ALL 2020 Stellopt.
- MACKENBACH, R.J.J., PROLL, J.H.E., WAKELKAMP, R. & HELANDER, P. 2023a The available energy of trapped electrons: a nonlinear measure for turbulent transport. *Journal of Plasma Physics* **89** (5), 905890513.
- MACKENBACH, RJJ, PROLL, JOSEFINE HE & HELANDER, P 2022 Available energy of trapped electrons and its relation to turbulent transport. *Physical Review Letters* **128**, 175001.
- MACKENBACH, R. J. J., DUFF, J. M., GERARD, M. J., PROLL, J. H. E., HELANDER, P. & HEGNA, C. C. 2023b Bounce-averaged drifts: Equivalent definitions, numerical implementations, and example cases. *Physics of Plasmas* **30** (9), 093901.
- MASON, J.C. & HANDSCOMB, DAVID C. 2002 *Chebyshev Polynomials*. New York: Chapman and Hall/CRC.
- MATSUDA, Y & STEWART, J.J 1986 A relativistic multiregion bounce-averaged fokker-planck code for mirror plasmas. *Journal of Computational Physics* **66** (1), 197–217.
- NEMOV, VV, KASILOV, SV, KERNBICHLER, W & HEYN, MF 1999 Evaluation of $1/\nu$ neoclassical transport in stellarators. *Physics of plasmas* **6** (12), 4622–4632.
- NEMOV, V. V., KASILOV, S. V., KERNBICHLER, W. & LEITOLD, G. O. 2008 Poloidal motion of trapped particle orbits in real-space coordinates. *Physics of Plasmas* **15** (5), 052501.
- NVIDIA CORPORATION 2020 Nvidia A100 tensor core gpu.
- OCHS, IAN E. 2025 Bounce-averaged theory in arbitrary multi-well plasmas: solution domains and the graph structure of their connections. *Journal of Plasma Physics* **91** (4), E123.
- OLVER, F. W. J., OLDE DAALHUIS, A. B., LOZIER, D. W., SCHNEIDER, B. I., BOISVERT, R. F., CLARK, C. W., MILLER, B. R., B. V. SAUNDERS, H. S. COHL & M. A. McCORMICK, EDS. 2024 *NIST digital library of mathematical functions*.
- PANICI, DARIO, CONLIN, RORY, DUDT, DANIEL W, UNALMIS, KAYA & KOLEMEN, EGEMEN

- 2023 *The DESC stellarator code suite. Part 1. Quick and accurate equilibria computations*. *Journal of Plasma Physics* **89**, 955890303.
- PETROV, YU V & HARVEY, R W 2016 *A fully-neoclassical finite-orbit-width version of the cql3d fokker-planck code*. *Plasma Physics and Controlled Fusion* **58** (11), 115001.
- RAMOS-LÓPEZ, D., SÁNCHEZ-GRANERO, M.A., FERNÁNDEZ-MARTÍNEZ, M. & MARTÍNEZ-FINKELSTEIN, A. 2016 *Optimal sampling patterns for zernike polynomials*. *Applied Mathematics and Computation* **274**, 247–257.
- RODRÍGUEZ, E., HELANDER, P. & GOODMAN, A.G. 2024 *The maximum- j property in quasi-isodynamic stellarators*. *Journal of Plasma Physics* **90** (2), 905900212.
- SAPIENZA, FACUNDO, BOLIBAR, JORDI, SCHÄFER, FRANK, GROENKE, BRIAN, PAL, AVIK, BOUSSANGE, VICTOR, HEIMBACH, PATRICK, HOOKER, GILES, PÉREZ, FERNANDO, PERSSON, PER-OLOF & RACKAUCKAS, CHRISTOPHER 2025 *Differentiable programming for differential equations: A review*, arXiv: 2406.09699.
- SATAKE, SHINSUKE, OKAMOTO, MASAO & SUGAMA, HIDEO 2002 *Lagrangian neoclassical transport theory applied to the region near the magnetic axis*. *Physics of Plasmas* **9** (9), 3946–3960.
- HSUAN SHIH, YU, WRIGHT, GARRETT, ANDÉN, JOAKIM, BLASCHKE, JOHANNES & BARNETT, ALEX H. 2021 *cufinufft: a load-balanced gpu library for general-purpose nonuniform ffts*, arXiv: 2102.08463.
- SPITZER JR, LYMAN 1958 *The stellarator concept*. *The Physics of Fluids* **1**, 253–264.
- SPONG, D. A., HIRSHMAN, S. P., WHITSON, J. C., BATCHELOR, D. B., CARRERAS, B. A., LYNCH, V. E. & ROME, J. A. 1998 *J^* optimization of small aspect ratio stellarator/tokamak hybrid devices*. *Physics of Plasmas* **5** (5), 1752–1758.
- SÜLI, ENDRE & MAYERS, DAVID F. 2003 *An Introduction to Numerical Analysis*. Cambridge University Press.
- VELASCO, JL, CALVO, I, MULAS, S, SÁNCHEZ, E, PARRA, FI, CAPPA, A & OTHERS 2021 *A model for the fast evaluation of prompt losses of energetic ions in stellarators*. *Nuclear Fusion* **61**, 116059.
- VELASCO, J.L., CALVO, I., PARRA, F.I. & GARCÍA-REGAÑA, J.M. 2020 *Knosos: A fast orbit-averaging neoclassical code for stellarator geometry*. *Journal of Computational Physics* **418**, 109512.