Optimization of nonlinear turbulence in stellarators

P. Kim [1,2,†], **S. Buller** [3], **R. Conlin** [1], **W. Dorland** [2,3], **D.W. Dudt** [1], **R. Gaur** [1], **R. Jorge** [4,5], **E. Kolemen** [1,6], **M. Landreman** [3], **N.R. Mandell** [6] and **D. Panici** [1]

[1]Princeton University, Princeton, NJ 08544, USA

[2]Department of Physics, University of Maryland, College Park, MD 20742, USA

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

[4]Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico, Universidade de Lisboa, 1049-001 Lisboa, Portugal

[5]Department of Physics, University of Wisconsin-Madison, Madison, WI 53706, USA

[6]Princeton Plasma Physics Laboratory, PO Box 541, Princeton NJ 08543, USA

We present new stellarator equilibria that have been optimized for reduced turbulent transport using nonlinear gyrokinetic simulations within the optimization loop. The optimization routine involves coupling the pseudo-spectral GPU-native gyrokinetic code `GX` with the stellarator equilibrium and optimization code `DESC`. Since using `GX` allows for fast nonlinear simulations, we directly optimize for reduced nonlinear heat fluxes. To handle the noisy heat flux traces returned by these simulations, we employ the simultaneous perturbation stochastic approximation (SPSA) method that only uses two objective function evaluations for a simple estimate of the gradient. We show several examples that optimize for both reduced heat fluxes and good quasi-symmetry as a proxy for low neoclassical transport. Finally, we run full transport simulations using the `T3D` stellarator transport code to evaluate the changes in the macroscopic profiles.

Key words: fusion plasma, plasma confinement, plasma simulation

1. Introduction

Stellarators are one of the most promising designs for magnetic confinement fusion as they are inherently steady state and are less susceptible to current-driven instabilities (Helander 2014). However, stellarators have historically been plagued by large collisionally induced losses of particles and energy associated with cross-field drifts of particles caused by the inhomogeneity and curvature of the magnetic field. In principle, stellarators can be optimized to reduce these collisional losses. In practice, impressive improvements

† Email address for correspondence: pk2354@princeton.edu

in plasma confinement have been obtained. For example, experiments at the W7-X stellarator have shown greatly reduced neoclassical transport (Beidler *et al.* 2021). Furthermore, advances in stellarator optimization techniques have led to designs with precise quasi-symmetry (QS) (Landreman & Paul 2022; Landreman, Buller & Drevlak 2022) and quasi-isodynamicity (QI) (Goodman *et al.* 2022).

After neoclassical losses are minimized, confinement is limited by the anomalous transport of heat and particles by turbulence. This turbulence is driven by plasma microinstabilities on length scales of the gyroradius. For example, the ion-temperature gradient (ITG) instability is believed to be a major cause for ion-temperature clamping in W7-X, preventing heating of ions above 2 keV (Beurskens *et al.* 2021) (along with weak energy exchange between electrons and ions). While there have been recent attempts to optimize stellarators to reduce turbulence-induced transport, they have mainly relied on proxies based solely on the magnetic geometry (Mynick, Pomphrey & Xanthopoulos 2010; Proll *et al.* 2016; Roberg-Clark, Xanthopoulos & Plunk 2023) or linear simulations (Jorge *et al.* 2023). However, linear physics may not accurately predict nonlinear saturation mechanisms that ultimately determine the rate of heat and particle loss (McKinney *et al.* 2019). Unfortunately, using nonlinear analysis for optimization is usually very challenging. Gyrokinetics (Catto 1978; Antonsen & Lane 1980; Frieman & Chen 1982) is one of the most commonly used models to study turbulence in magnetic confinement fusion devices, and is also the model used for this paper. Typical nonlinear gyrokinetics simulations usually require hundreds to thousands of CPU hours, making them infeasible to use within an optimization loop.

In this work, we demonstrate the ability to reduce turbulent losses by optimizing stellarator configurations using nonlinear turbulence simulations directly rather than relying on proxies. To run nonlinear simulations inside the optimization loop, we use the new GPU-native gyrokinetic code GX (Mandell, Dorland & Landreman 2018; Mandell *et al.* 2023). GX uses pseudo-spectral methods in velocity space. GPU acceleration combined with flexible velocity resolution allows for nonlinear GX simulations that only take minutes to run. For this work, we focus on ITG turbulence, which contributes to major energy losses that limit plasma confinement and so is one of the most important microinstabilities to consider for reactor design (Kotschenreuther, Rewoldt & Tang 1995*b*; Horton 1999; Helander, Proll & Plunk 2013). To that end, we will be running electrostatic simulations with a Boltzmann (adiabatic) response assumed for the electrons.

The optimization is performed using the stellarator equilibrium and optimization code DESC (Dudt & Kolemen 2020; Conlin *et al.* 2023; Dudt *et al.* 2023; Panici *et al.* 2023). DESC also uses pseudo-spectral methods and directly solves the ideal MHD force-balance equation to compute the magnetic equilibrium. The quantities computed from the resulting magnetic fields are used as input for GX. Stochastic optimization methods are used to robustly handle the noisy landscape of turbulent heat fluxes.

The paper is organized as follows. The stochastic optimization method used in this work is described in § 2. Results are shown in § 3, with analysis on the effects of global magnetic shear for reduced turbulence explored in § 4. Transport simulations using the T3D stellarator transport code (Qian *et al.* 2022) are shown in § 5. Finally, the conclusions follow in § 6.

2. Optimization methods

In this optimization routine, we seek to minimize the nonlinear heat flux returned by GX. Specifically, GX returns as output the time trace of the heat flux (normalized to gyro-Bohm units). An example of some heat traces is shown in figure 5(*a*). At the beginning of the simulation, linear growth of the fastest-growing instability dominates. However, eventually

nonlinear effects cause the heat flux to decrease and saturate to a statistical steady state. We use the time average of this steady-state flux as our heat flux objective f_Q. To take the time average, we take the second half of the time trace and compute the weighted Birkoff average

$$f_Q = \frac{1}{I} \sum_i^N \exp(-\mathrm{i}/(N(1 - i/N)))q_i, \qquad (2.1)$$

where the sum is over each point in the trace (with N total points) and $I = \sum_i^N \exp(-\mathrm{i}/(N(1 - i/N)))$ is a normalization factor. This gives greater weight to values in the middle of the trace.

Since GX is a local flux-tube gyrokinetic code, each simulation is on a single field line on a single surface specified by the field line label α and the radial coordinate $\rho = \sqrt{\psi/\psi_b}$ (where $\boldsymbol{B} = \nabla\psi \times \nabla\alpha$ and ψ is the toroidal magnetic flux). Therefore, f_Q is also computed for only a single field line and surface. For all of the optimization examples, we only simulate on the $\rho = \sqrt{\psi/\psi_b} = \sqrt{0.5}$ surface and the $\alpha = 0$ (except for the multiple field line examples in § 3.3) field line. However, we will run post-processing simulations across different surfaces and field lines.

To run these simulations, GX requires a set of geometric quantities that can be computed from numerical equilibria. Given that the magnetic field is written in Clebsch form $\boldsymbol{B} = \nabla\psi \times \nabla\alpha$, the set of quantities needed are

$$\begin{aligned} \mathcal{G} = \{ &B, \boldsymbol{b} \cdot \nabla z, |\nabla\psi|^2, |\nabla\alpha|^2, \nabla\psi \cdot \nabla\alpha, \\ &\times (\boldsymbol{B} \times \nabla B) \cdot \nabla\psi, (\boldsymbol{B} \times \nabla B) \cdot \nabla\alpha, (\boldsymbol{b} \times \boldsymbol{\kappa}) \cdot \nabla\alpha \}, \end{aligned} \qquad (2.2)$$

where $\boldsymbol{b} = \boldsymbol{B}/B$, α is the straight field line label, $\boldsymbol{\kappa} = \boldsymbol{b} \cdot \nabla\boldsymbol{b}$ is the curvature and z is some coordinate representing the distance along the field line. For these simulations, the geometric toroidal angle ϕ is used for z. All of these quantities are easily computed using utility functions in DESC.

One issue with directly minimizing nonlinear heat fluxes is that their time traces are often very noisy. Even the resulting time averages are usually very noisy in parameter space. This can be seen in figure 1 showing the time-averaged nonlinear heat flux when scanning over the $Z_{m,n} = Z_{0,-1}$ boundary mode of the initial equilibrium used in this study. Optimizers designed for smooth objectives may easily get stuck in local minima and make very little progress. To help with this issue, we use the simultaneous perturbation stochastic approximation (SPSA) method (Spall 1987) to minimize the heat fluxes. In this algorithm, the ith component of the gradient at point x_n is approximated as

$$\hat{g}_n(x_n)_i = \frac{f(x_n + kc_n) - f(x_n - kc_n)}{2kc_{ni}}, \qquad (2.3)$$

where c_n is a random perturbation vector whose components are sampled from a Rademacher distribution, and k is a finite-difference step-size. Therefore, we are effectively using the finite-difference method in all directions at once. Finally, for this method, rather than specifying stopping tolerances, we instead specify a maximum number of iterations.

The main feature of the SPSA method is that it only requires two objective function measurements per iteration. This makes it suitable for high-dimensional optimization problems whose objective functions are expensive to compute. SPSA can also robustly handle noisy objective functions, and so has been used for simulation optimization, including Monte Carlo simulations (Chan, Doucet & Tadic 2003).
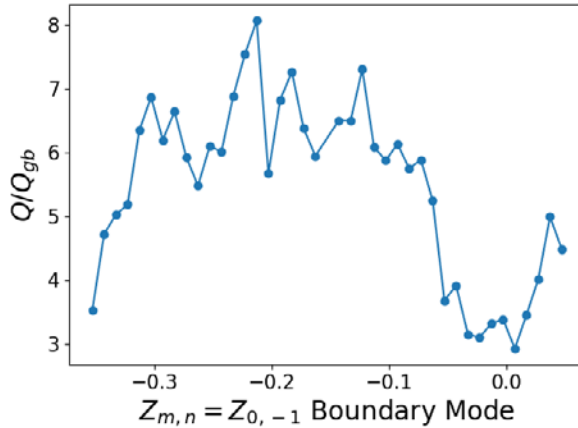
FIGURE 1. Time-averaged nonlinear heat flux computed by GX when scanning across the $Z_{0,-1}$ boundary mode.

To still use automatic differentiation and smooth optimization methods for other objectives like quasi-symmetry residuals, we split the optimization routine into two parts. We first minimize the turbulent heat flux using stochastic gradient descent. Next, we minimize the quasi-symmetry residuals using a least-squares method that uses automatic differentiation. This order is chosen arbitrarily, but the reverse ordering can also work. The two objectives are thus

$$f_1 = f_Q^2 + (A - A_{\text{target}})^2, \tag{2.4}$$

$$f_2 = f_{QS}^2 + (A - A_{\text{target}})^2, \tag{2.5}$$

where f_Q is the heat flux from GX and A is the aspect ratio. For these optimizations, we target an aspect ratio of $A_{\text{target}} = 8$. We use the two-term quasi-symmetry objective, where a magnetic field is quasi-symmetric if the quantity

$$C = \frac{(\boldsymbol{B} \times \boldsymbol{\nabla}\psi) \cdot \boldsymbol{\nabla}B}{\boldsymbol{B} \cdot \boldsymbol{\nabla}B} \tag{2.6}$$

is a flux function. Computationally, we evaluate the equivalent form

$$f_{QS} = (M - \iota N)(\boldsymbol{B} \times \boldsymbol{\nabla}\psi) \cdot \boldsymbol{\nabla}B - (MG + NI)\boldsymbol{B} \cdot \boldsymbol{\nabla}B \tag{2.7}$$

at several different flux surfaces and try to minimize the resulting values. For this study, we always choose flux surfaces at $\rho = 0.6,\ 0.8$ and 1. We target quasi-helical (QH) symmetry, so that the helicity is $(M, N) = (1, 4)$.

The GX simulation parameters used in the optimization loop and for post-processing, along with their justification, are in Appendix B.

Finally, the optimization routine is performed in stages. Each stage increments the maximum boundary Fourier mode being optimized over. For example, in the first stage, only boundary modes with modes m, n satisfying $|m| \leq 1$ and $|n| \leq 1$ are used as optimization variables. In the next stage, boundary modes with $|m| \leq 2$ and $|n| \leq 2$ are used. The $m = 0, n = 0$ mode is excluded to prevent the major radius from changing. This reduces the number of optimization variables at the beginning of the optimization and 'warm-starts' each successive stage. This type of method has been used with
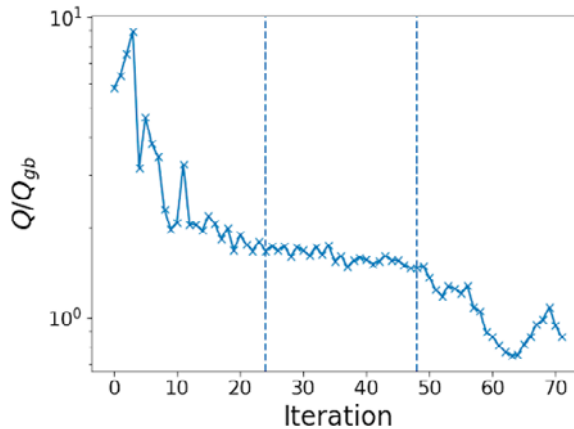
FIGURE 2. Normalized heat flux across each iteration. The dashed lines represent an increase in the maximum boundary mode number being optimized.

great success for previous stellarator optimization results, including for optimizing quasi-symmetry (Landreman & Paul 2022). After each stage, we start the next stage with an equilibrium chosen from the previous stage that had achieved both low heat flux and low quasi-symmetry error. No strict rule is used, and the choice depended on several factors like the value of the gradient estimate and the ι profile.

3. Results

3.1. *Turbulence optimization*

As an initial test of our stochastic optimization method, we begin by solely optimizing for turbulence. The optimization routine is the same as described in the previous section, except instead of optimizing for both quasi-symmetry and aspect ratio in the second half of each iteration, we only optimize for the target aspect ratio. Usually, combining the turbulence and aspect ratio objectives, as is done in the first part, is insufficient to reach our target aspect ratio. We have to force the optimizer to take more aggressive steps due to the simulation noise, which makes it more difficult to maintain the desired aspect ratio. For this optimization (and all of the examples in this work), the initial equilibrium is an approximately quasi-helically symmetric equilibrium with an aspect ratio of eight and four field-periods.

Figure 2 shows the normalized nonlinear heat flux across each iteration of the optimization process. Due to the stochastic nature of the optimization routine, the initial gradient estimates are very poor, leading to an increase in the heat flux in the first several iterations. However, as the optimization continued, there is eventually a rapid decrease in the heat flux as the gradient approximation begins to track the true gradient. The optimizer continues to steadily decrease the heat flux for most of the remaining iterations. Note that since we specify a maximum number of iterations rather than stopping tolerances, the optimization ends even when the heat flux had increased slightly at the end.

A scan of the nonlinear heat fluxes across ρ and optimized cross-sections of the flux surfaces are shown in figure 3. For this simulation (and all of the simulations in this section), the resolution is increased to the values in table 3. As seen in those plots, some surfaces see moderate to drastic improvements to the nonlinear heat flux. In particular, the $\rho = 0.4$ surfaces has a reduction of approximately an order of magnitude. However, other surfaces see much less improvement, such as at $\rho = 0.2$ and $\rho = 0.8$. It is not completely
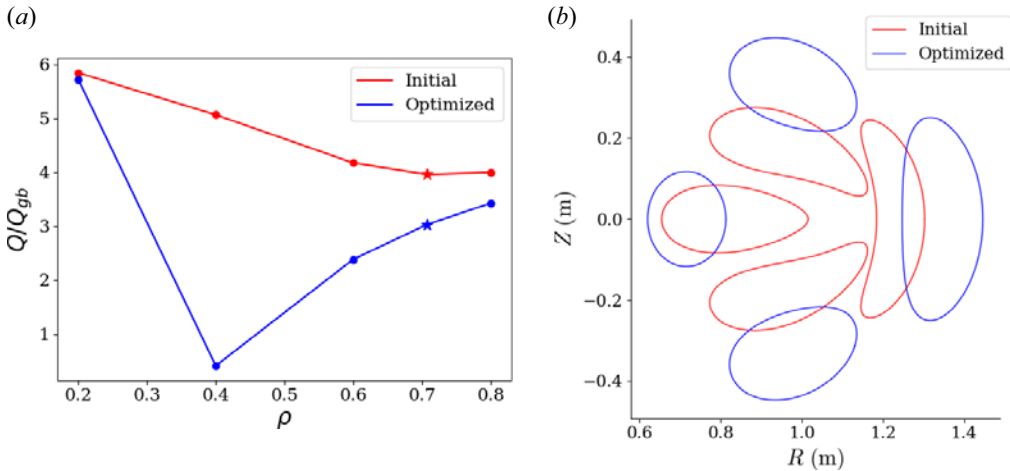
(*a*)

(*b*)

FIGURE 3. (*a*) Scan of the nonlinear heat flux for the initial (red) and optimized (blue) equilibria across different flux surfaces. (*b*) Cross-sections of the boundary magnetic flux surface of the initial (red) and optimized (blue) configurations. The star in panel (*a*) indicates the $\rho = \sqrt{0.5}$ surface that was chosen for the optimization loop.

well understood why that is, but it could be related to the fact that only the $\rho = \sqrt{0.5}$ surface is being optimized. This does reveal a potential weakness in the optimization strategy that will be addressed in future work. Nevertheless, the following sections show much more uniform improvement. Interestingly, based off of the surface boundary plots, the cross-sections of the optimized equilibria seem much less strongly shaped than in the initial equilibrium. Instead, the magnetic axis has significantly more torsion.

The contours of $|\boldsymbol{B}|$ in Boozer coordinates are plotted in figure 4. Surprisingly, the contours seem to resemble those from quasi-isodynamic equilibria despite not including a QI term in the objective functions. This may be related to the fact that it seems like the optimizer favoured a high mirror ratio. Additionally, it is well known that QI stellarators can be optimized to have the maximum-J property, which has numerous benefits including improved confinement of fast ions, neoclassical confinement of thermal ions (Sánchez *et al.* 2023; Velasco *et al.* 2023) and enhanced stabilization against trapped-electron modes (TEMs) (Proll *et al.* 2012; Helander *et al.* 2013; Helander 2014). It has been further theorized and shown in gyrokinetic simulations of W7-X that possessing the maximum-J property can also be beneficial for ITG turbulence (Proll *et al.* 2022). However, the TEM studies required modelling the full kinetic electron dynamics. In this study, we only use a Boltzmann response for the electrons.

3.2. *Combined turbulence–quasi-symmetry optimization*

Next, we include the two-term quasi-symmetry objective in the second part of the optimization loop. The final heat flux traces and optimized cross-sections of the flux surfaces are shown in figure 5. The cross-sections in figure 5(*b*) show relatively modest changes in the shape of the optimized stellarator. However, the time trace (simulated at the $\psi/\psi_b = 0.5$ surface) in figure 5(*a*) shows approximately a factor of three decrease in the nonlinear heat flux.

To ensure that the nonlinear heat flux was reduced throughout the plasma volume, we again ran simulations at radial locations of $\rho = 0.2$, 0.4, 0.6 and 0.8. The resulting heat fluxes as well as the maximum symmetry breaking residuals across ρ are plotted

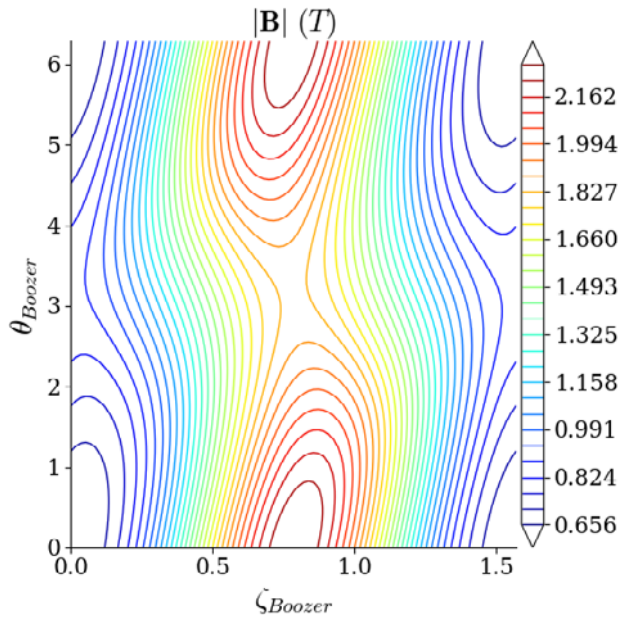FIGURE 4. The $|\boldsymbol{B}|$ contours in Boozer coordinates, which resemble contours of a QI equilibrium.
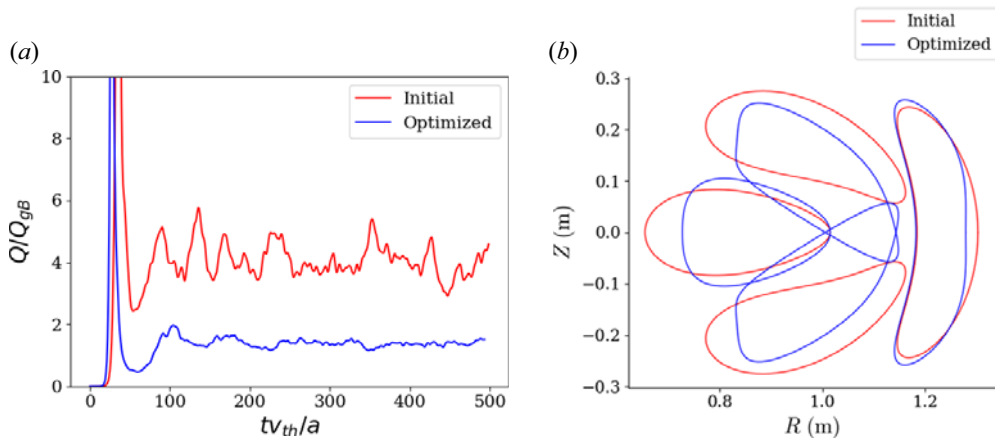

FIGURE 5. (*a*) Time traces of the nonlinear heat fluxes of the initial (red) and optimized (blue) configurations. (*b*) Cross-sections of the magnetic flux surfaces of the initial (red) and optimized (blue) configurations.

in figure 6. As seen in the plots, despite optimizing only for the $\rho = \sqrt{0.5}$ surface, the heat flux is reduced throughout the plasma volume. With regards to quasi-symmetry, for $\rho < 0.5$, the optimized equilibrium has smaller maximum symmetry-breaking modes than the initial equilibrium. However, this reverses in the outer surfaces. This is unexpected considering that the quasi-symmetry residuals were computed at $\rho = 0.6, 0.8$ and 1. Nevertheless, the degree of symmetry breaking is still comparable to WISTELL-A (Bader *et al.* 2020), another optimized QH stellarator. Indeed, the $|\boldsymbol{B}|$ plots in figure 7 show

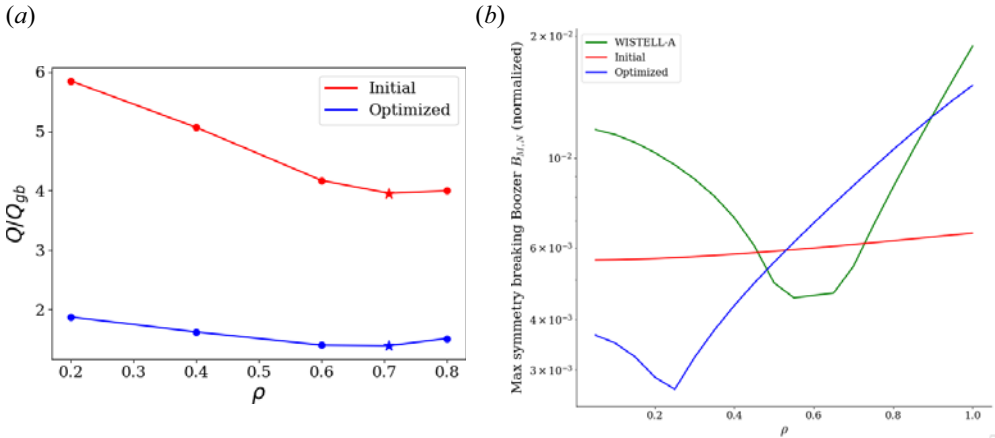(*a*)   (*b*)

FIGURE 6. Scans of the (*a*) nonlinear heat flux and (*b*) maximum symmetry breaking modes across different radial locations. The star in panel (*a*) indicates the $\rho = \sqrt{0.5}$ surface that was chosen for the optimization loop.
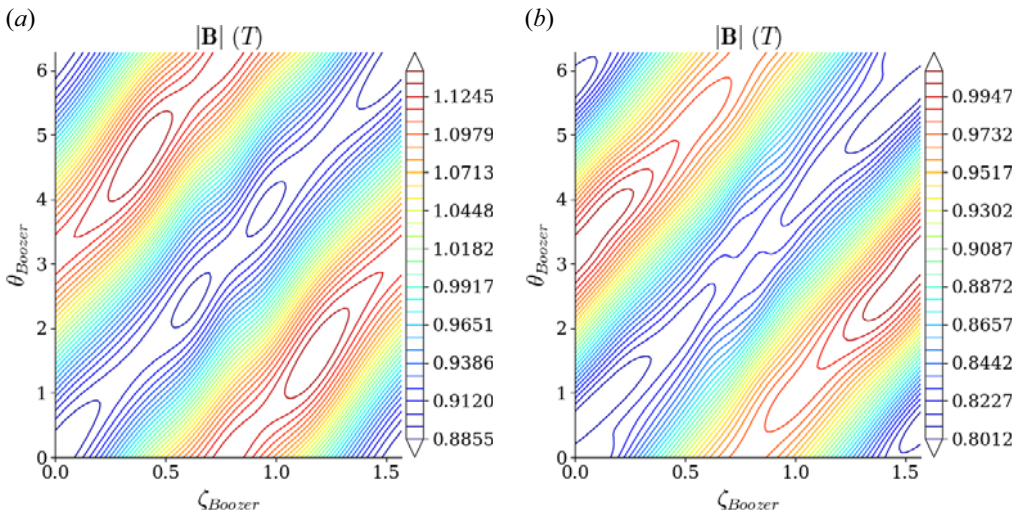
(*a*)   (*b*)

FIGURE 7. The |***B***| contours in Boozer coordinates for the (*a*) initial and
(*b*) optimized equilibria.

contours characteristic of a quasi-helically symmetric stellarator (plotted at the $\psi = 0.5$ surface).

Unlike in tokamaks, different field lines in stellarators experience different curvatures, magnetic fields, etc. and so may have very different fluxes (Dewar & Glasser 1983; Faber *et al.* 2015). Therefore, we also run simulations across α at the radial location $\rho = \sqrt{0.5}$. The resulting scan is shown in figure 8(*a*). At each α simulated, the optimized stellarator has a lower heat flux, indicating that the heat flux has been reduced across the entire flux surface.

It is interesting to see that for the initial geometry, there are large variations in the heat flux. This might indicate that some flux tubes are too short to sample enough area on the flux surface or that there is some coupling between the flux tubes. However, it should be
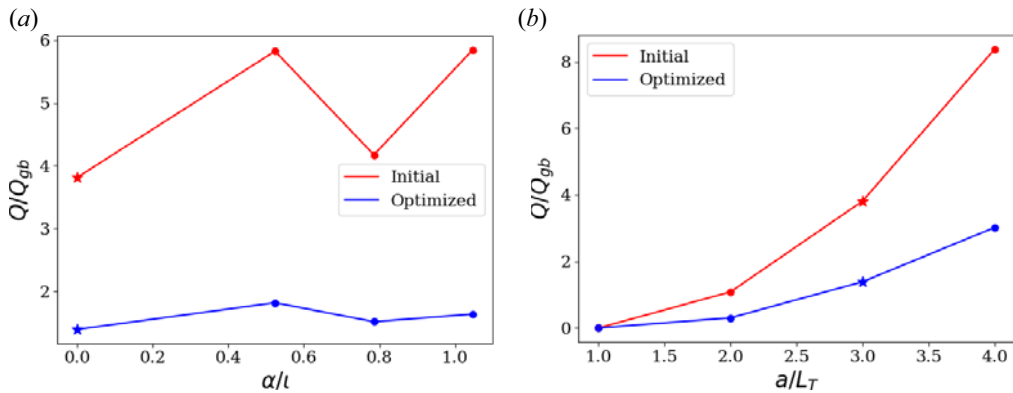
(a)

(b)

FIGURE 8. Heat fluxes across (a) different field lines α and (b) temperature gradients a/L_T for the initial (red) and optimized (blue) configurations. The star indicates the $\alpha = 0$ field line and the $a/L_T = 3$ temperature gradient that were chosen for the optimization loop.
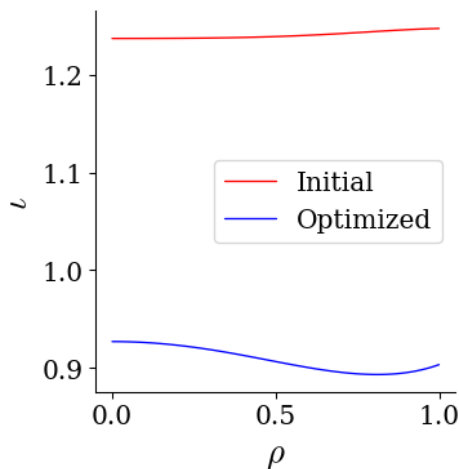

FIGURE 9. The ι profiles for the initial (red) and optimized (blue) equilibria.

noted that the ι of this equilibrium (plotted in figure 9) is close to 5/4. Recent work has shown a very large variation across α on low-order rational surfaces (Buller *et al.* 2023). In comparison, the optimized equilibrium has an ι that is slightly above 0.9. This could have negative impacts on the optimization routine as the optimizer may choose to approach equilibria with ι near low-order rationals. This could then lead to not only misleading heat fluxes, but also poor MHD stability. More work is needed to investigate the limitations of flux-tube codes, the effects of rational surfaces and the resulting consequences for optimization.

In a fusion reactor, since the transport is very stiff, the temperature profiles may evolve to approach the critical temperature gradient (the temperature gradient at which the heat flux is zero). Therefore, a decrease in the heat flux is less useful if the critical temperature gradient also decreases significantly. To verify that the critical temperature gradient did not decrease, we also ran several nonlinear simulations with different temperature gradients at $\rho = \sqrt{0.5}$. The results are shown in figure 8(b) which shows that the critical temperature gradient does not seem to change significantly. It would be more beneficial if the critical
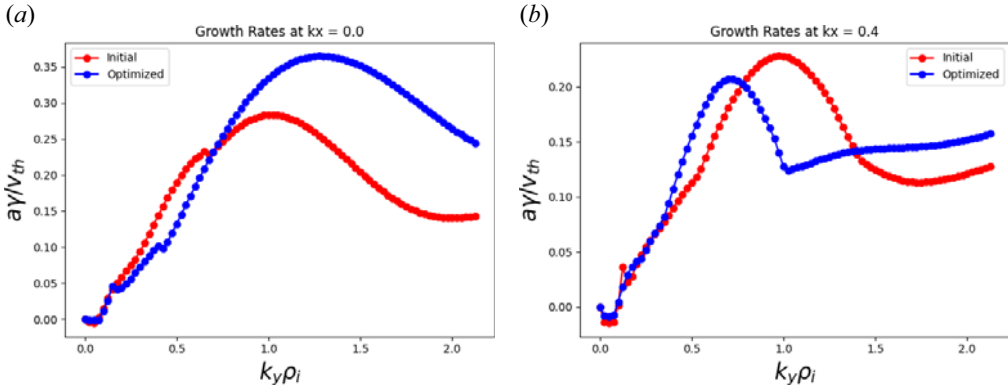
(*a*)

Growth Rates at kx = 0.0

(*b*)

Growth Rates at kx = 0.4

FIGURE 10. Linear growth rates across different k_y at (*a*) $k_x = 0$ and (*b*) $k_x = 0.4$.

temperature gradient had also increased. Specifically targeting the critical temperature gradient will be the focus of future work.

We also run linear simulations for the initial and optimized equilibria, with the growth rates shown in figure 10(*a*) at $k_x = 0$ and figure 10(*b*) at $k_x = 0.4$. Although the optimized equilibrium has lower heat fluxes, it has a significantly higher peak growth rate at $k_x = 0$. However, it has lower growth rates at lower k_y, and one would expect that the nonlinear heat flux scales like $\gamma / \langle k_\perp^2 \rangle$ (where the angle brackets indicate a flux-surface average) (Mariani *et al.* 2018). Therefore, this might result in the observed lower nonlinear heat fluxes. However, this trend reverses for the simulations at $k_x = 0.4$. Furthermore, recent work has shown that both peak growth rates and the quasilinear $\gamma / \langle k_\perp^2 \rangle$ estimate are both poor proxies for the nonlinear heat flux (Buller *et al.* 2023).

As a preliminary example, we replicate the plots from figure 10 but with a quasi-linear estimate of the form

$$f_{QL}(k_y) = \frac{\gamma(k_y)}{\langle k_\perp^2 \rangle}, \tag{3.1}$$

where $\langle k_\perp^2 \rangle$ is

$$\langle k_\perp^2 \rangle = \frac{\displaystyle\int \mathrm{d}z k_\perp^2 \phi^2 \sqrt{g}}{\displaystyle\int \mathrm{d}z \phi^2 \sqrt{g}}. \tag{3.2}$$

Here, ϕ is the electrostatic potential, $\sqrt{g}$ is the Jacobian and the integration is along the simulated field line. This expression is similar to that used by Mariani *et al.* (2018). The resulting plots are shown in figure 11. We have enforced that $\min(f_{QL}) = 0$, which assumes that only growing modes contribute to the flux. As seen in the plots, while the optimized configuration has a smaller f_{QL} at smaller k_y at $k_x = 0$, it has a much larger f_{QL} at $k_x = 0.4$. This makes it challenging to use such a quasilinear estimate of the heat flux as a proxy, necessitating our approach of directly computing the nonlinear heat flux.

Finally, we compare the optimized stellarator to another configuration solely optimized for quasi-symmetry. The plot of the heat fluxes across different surfaces is shown in figure 12. The heat fluxes for the precise QH equilibrium are higher than those from the approximately QH equilibrium used as the initial point. This shows that just optimizing for quasi-symmetry can be detrimental for turbulence and stresses the need to optimize for both.
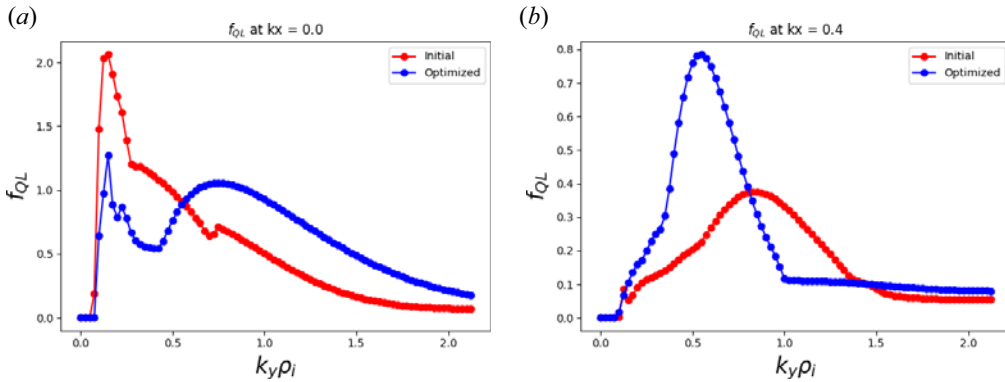
FIGURE 11. Quasilinear estimate for the initial (red) and optimized (blue) configurations at (*a*) $k_x = 0$ and (*b*) $k_x = 0.4$.
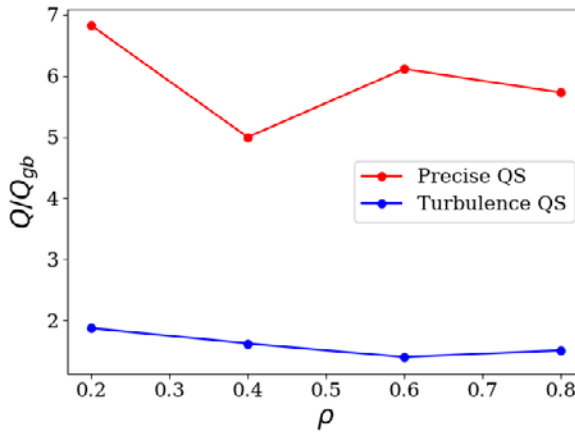

FIGURE 12. Heat fluxes across ρ for a precise QH equilibrium (red) and the turbulence optimized equilibrium (blue).

3.3. *Optimization with multiple field lines*

As described in the previous section, several physical and computational factors can lead to large variations in the heat fluxes at different α. To avoid this issue, we rerun the optimization loop but simulate on field lines of both $\alpha = 0$ and $\alpha = \pi\iota/4$. This makes the new objective function

$$f_Q = f_{Q,\alpha=0}^2 + f_{Q,\alpha=\pi\iota/4}^2 + (A - A_{\text{target}})^2. \tag{3.3}$$

This new objective function serves as a way of reducing the heat flux across multiple field lines while also testing our stochastic optimizer against additional objectives. Similar situations include trying to optimize across different flux surfaces, different temperature/density gradients, etc.

The plots in figure 13 show the cross-sections, maximum symmetry breaking modes, and heat fluxes across ρ and α for the initial equilibrium and the final optimized equilibria after optimizing at one and two field lines. While this new equilibrium has higher heat fluxes than when optimizing for just a single field line, it instead has lower maximum symmetry breaking modes and better quasisymmetry.
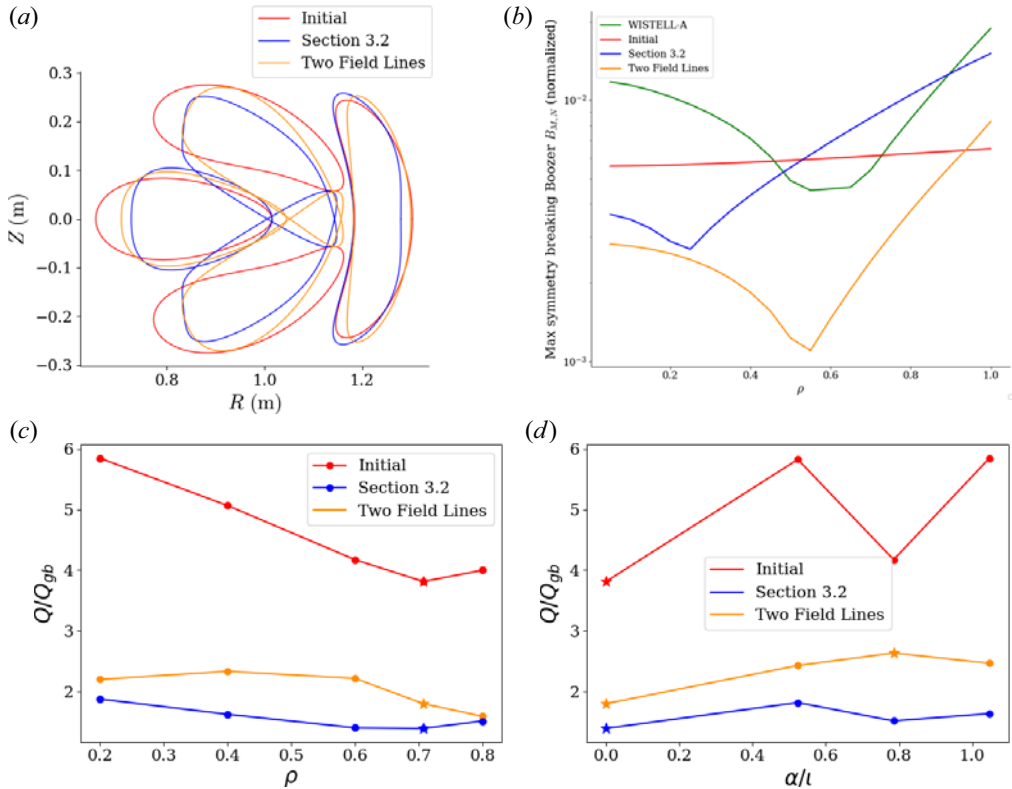
FIGURE 13. (*a*) Cross-sections for the initial equilibrium (red), the single field line optimized equilibrium (blue) from § 3.2 and the new two field line optimized equilibrium. (*b*) Maximum symmetry-breaking modes for all three equilibria and WISTELL-A. (*c*) Heat flux scans across ρ. (*d*) The heat flux scans across α. The stars in panels (*c,d*) indicate parameters that were chosen for optimization.

The scans across multiple α in figure 13(*d*) show approximately a 50 % variation in the heat flux compared with the $\alpha = 0$ point. Unfortunately, this is larger than the approximately 30 % variation in the single-field line case and comparable to the relative variation in the initial equilibrium. Nevertheless, the heat fluxes across each α are still 2–3 times smaller than in the initial equilibrium and this shows that the stochastic optimizer is still effective with additional terms in the objective function. More investigation is needed to more effectively and precisely optimize with multiple field lines.

3.4. *Fluid approximation*

In this final case, we change the simulation parameters within the optimization loop to approach the fluid limit in GX. This is based off of previous work that showed a fluid approximation with only a few velocity moments can accurately match the true heat fluxes at higher velocity resolution at moderate collision frequencies (Buck *et al.* 2022). That study was motivated by a recently developed three-field model for the density, temperature and momentum to approximate growth rates and nonlinear heat fluxes (Hegna, Terry & Faber 2018).

To approximate this model, we reduce the velocity resolution to four Hermite moments and two Laguerre moments, as in the gyrofluid model. While normally this requires
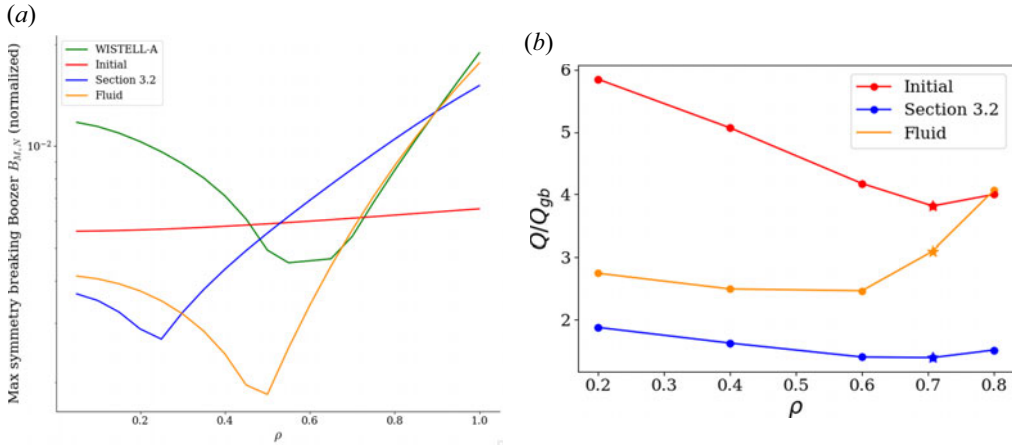
FIGURE 14. (*a*) Maximum symmetry-breaking modes and (*b*) heat flux scan across ρ for the initial equilibrium, the kinetic case from § 3.2 and the fluid case (as well as WISTELL-A in panel *a*). The star in panel (*b*) indicates the $\rho = \sqrt{0.5}$ surface that was chosen for the optimization loop.

closure relations like those of Beer (1995) and Mandell *et al.* (2018), we instead increase the collisionality to damp the high wavenumber modes and increase the temperature gradient to $a/L_T = 5$. We employ a Dougherty collision operator (Dougherty 1964).

The plots of the maximum-symmetry breaking modes and the heat flux across ρ for the kinetic and fluid cases are shown in figure 14. While the new equilibrium does not achieve lower heat fluxes than in the kinetic case, it still achieves approximately a factor of two reduction compared to the initial equilibrium except at $\rho = 0.8$ despite using very different physical parameters. The fluid case retains comparable levels of quasi-symmetry as well. These results indicate that the optimization routine is robust against varying levels of fidelity. This opens new possibilities of potentially varying the fidelity across iterations. This can either decrease the computational cost further or allow us to increase other resolution parameters.

4. Effects of magnetic shear on reduced turbulence

It is well known since the findings of Greene & Chance (1981) that sufficient positive global magnetic shear can be destabilizing for intermediate to high wavenumber instabilities like the ballooning mode. An analogous relationship for transport was investigated by Kotschenreuther *et al.* (1995*a*), where it was found that while increased shear can reduce thermal transport, it can also negatively impact the critical gradient.

To further investigate these effects, we optimized for several more precisely quasi-symmetric equilibria, but with different target shears. The nonlinear heat flux traces are shown in figure 15. The shear $= 0.01$ case is when there is no shear target. While increasing the shear slightly does reduce the heat flux, continuing to increase it eventually increases the heat flux again, consistent with the findings of Kotschenreuther *et al.* (1995*a*). It is somewhat surprising that the heat flux is this sensitive to the shear since the shear is very small. For comparison, the global magnetic shear for W7-X in these units is approximately 0.16. Future work will better address the relationship between shear and turbulence in stellarators. This relationship can be very important as very low global magnetic shear is characteristic of vacuum quasi-symmetric stellarators (Landreman &
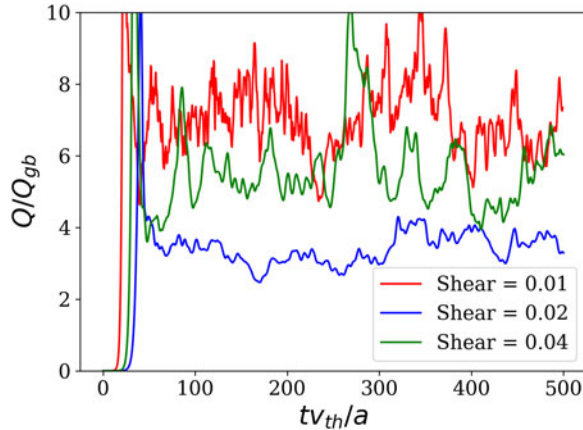
FIGURE 15. Nonlinear heat flux traces for quasi-symmetric equilibria with different shear targets.

Paul 2022). While the shear increases for finite β QS stellarators (Landreman *et al.* 2022), it is still small compared with shear in tokamaks.

5. Transport simulations

To study how the changes in nonlinear heat fluxes affected the macroscopic profiles, we used the T3D transport code (Qian *et al.* 2022). T3D uses the same algorithm as Trinity (Barnes *et al.* 2010), but is written in PYTHON and adapted to work with stellarator geometries. T3D takes advantage of the separation of scales in the local δf gyrokinetic model where the distribution function is written as $F = F_{0s} + \delta f_s$. It can be shown that F_{0s} is a Maxwellian and it is assumed that the perturbation δf scales like $\delta f \sim \epsilon F_{0s}$, where $\epsilon = \rho/a$. The Maxwellian F_{0s} evolves slower than the perturbation $\partial F_{0s}/\partial t \sim \epsilon^2 \partial \delta f_s/\partial t$. This allows for T3D to evolve the macroscopic density, pressure and temperature profiles using heat and particle fluxes computed by gyrokinetic codes like GX.

We run T3D simulations for the initial approximately QH equilibrium and optimized equilibrium from § 3.2. Both equilibria were scaled to the same major radius and on-axis magnetic field as W7-X. An adiabatic response is still assumed for the electrons, and neoclassical contributions are ignored to isolate the effects from turbulence. The plots of the final steady-state temperature profiles for both equilibria are shown in figure 16.

From the plot of the temperature profiles, there is a clear increase in the temperature for the optimized equilibrium. The temperature at the innermost simulated point increased by approximately 10 %. This seems lower than expected considering that the optimized equilibrium had lower heat fluxes by approximately a factor of three. The true heat flux scales like $Q \sim Q_{\text{sim}} n T v_{\text{th}} (\rho/a)^2$, where Q_{sim} is the heat flux from the simulation. Since $v_{\text{th}} \sim T^{1/2}$ and $\rho \sim v_{\text{th}}$, the heat flux scales like $Q \sim Q_{\text{sim}} T^{5/2}$. For equivalent sources, $T \sim Q_{\text{sim}}^{2/5}$. Therefore, we would expect the temperature to increase by a factor of $Q^{0.4} \sim 1.55$. However, the simulations from § 3.2 assumed a temperature gradient of $a/L_{Ti} = 3$. From the plot of the temperature gradients, most points are simulated with lower gradients, and the equilibria have the same critical temperature gradient (as seen in figure 8*b*). Nevertheless, the improvement is still very encouraging. Future work will investigate using simulations to optimize for a higher critical gradient or for specific target profiles.
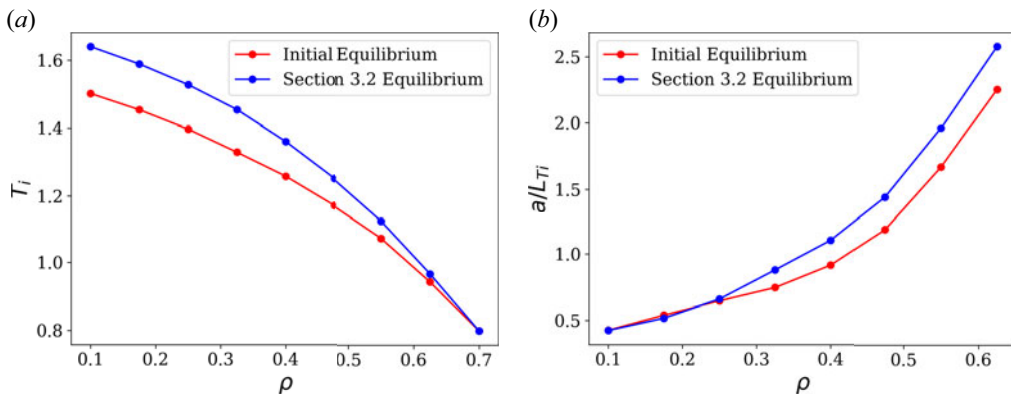
FIGURE 16. (*a*) Final temperature and (*b*) temperature gradient profiles for the initial and optimized equilibria.

6. Conclusions

In this work, we directly optimized stellarators for reduced nonlinear heat fluxes and good quasi-symmetry by coupling `GX` and `DESC`. By directly running nonlinear simulations, we include the nonlinear saturation mechanisms that determine the steady-state heat flux, and so avoid potential limitations of linear or quasilinear models. The SPSA method is used to handle the noisy heat flux objective and to cheaply estimate the gradient. The newly optimized equilibria show factors of 2–4 improvement in the nonlinear heat flux across several flux surfaces and multiple field lines while also having comparable or improved quasi-symmetry. We also investigated how global magnetic shear affects the nonlinear heat flux and found that slightly increasing it can significantly reduce the flux. However, increasing the shear too much led to increases in the heat flux. Our `T3D` simulations showed that by reducing the heat fluxes, the steady-state temperature profiles did increase slightly. Unfortunately, the transport is sufficiently stiff so that macroscopic profiles will approach the critical gradient. Since the initial and optimized equilibria have similar critical gradients, this limits the improvements in the temperature profile. Nevertheless, these results demonstrate that we can efficiently include nonlinear gyrokinetic simulations within the optimization loop. Future work will include adding an objective for MHD stability, optimizing for the nonlinear critical gradient and also directly optimize for desired macroscopic profiles. More improvements will also be made to the SPSA optimizer to make it more effective near minima. Currently, near the minima, the true gradient is small enough to be comparable to the simulation noise. This leads to poorer convergence near the minima and so requires more sophisticated optimization methods. Consequently, we will implement other optimization methods that have been effective for stochastic optimization, such as Bayesian optimization.

Editor Per Helander thanks the referees for their advice in evaluating this article.

Declaration of interests

The authors report no conflict of interest.

Data availability statement

The data that support the findings of this study are openly available in Zenodo at http://doi.org/10.5281/zenodo.10277921 (Kim 2024).

Appendix A. Codes

A.1. GX

GX employs the radially local δf approach to solve the gyrokinetic equation. In this approximation, the distribution function F_s is represented as $F_s = F_{0s} + \delta f_s = F_{Ms}(1 - Z_s\Phi/\tau_s) + h_s$, where F_{0s} is Maxwellian and δf_s is a perturbation consisting of a Boltzmann part proportional to the electrostatic potential Φ and a general perturbation h_s. The subscript s labels species. Assuming electrostatic fluctuations, the gyroaveraged perturbation $g_s = \langle f_s \rangle$ then obeys the electrostatic gyrokinetic equation

$$
\frac{\partial g_s}{\partial t} + v_\parallel \boldsymbol{b} \cdot \nabla z \left(\frac{\partial g_s}{\partial t} + \frac{q_s}{T_s} \frac{\partial \langle \phi \rangle}{\partial t} F_s \right) - \frac{\mu}{m_s} \boldsymbol{b} \cdot \nabla z \frac{\partial B}{\partial z} \frac{\partial g_s}{\partial v_\parallel}
$$

$$
+ \boldsymbol{v}_{Ms} \cdot \left(\nabla_\perp g_s + \frac{q_s}{T_s} \nabla_\perp \langle \phi \rangle F_s \right)
$$

$$
+ \langle \boldsymbol{v}_E \rangle \cdot \nabla_\perp g_s + \langle \boldsymbol{v}_E \rangle \cdot \nabla F_s = \langle C(\delta f_s) \rangle, \tag{A1}
$$

where $v_\parallel$ is the velocity parallel to the magnetic field, $\boldsymbol{b} = \boldsymbol{B}/B$, μ is the magnetic moment, z is some coordinate along the field line, κ is the field line curvature, $\boldsymbol{v}_{Ms}$ is the sum of the magnetic and curvature drift velocities, and $\boldsymbol{v}_E$ is the $\boldsymbol{E} \times \boldsymbol{B}$ velocity. For this work, electromagnetic effects are ignored and a Boltzmann response is used to model the perturbed electron distribution.

In GX, (A1) is projected onto the Hermite and Laguerre basis functions

$$\psi^l(\mu B) = (-1)^l \, e^{-\mu B} L_l(\mu B), \tag{A2}$$

$$\phi^m(v_\parallel) = \frac{\exp(-v_\parallel^2/2) He_m(v_\parallel)}{\sqrt{(2\pi)^3 m!}}, \tag{A3}$$

where $He_m(x)$ and $L_l(x)$ are the (probabilist's) Hermite and Laguerre polynomials, respectively. More details on the expansion and numerical algorithm can be found from Mandell *et al.* (2018) and Mandell *et al.* (2023).

By choosing a Hermite–Laguerre basis, the resulting equations for the spectral coefficients reduce to the gyrofluid equations at low resolution (Dorland & Hammett 1993; Beer 1995). Particle number, momentum and energy are also conserved at low resolution. Overall, any inaccuracies are due to the closure or dissipation model used. Therefore, GX allows for lower velocity resolution than other similar codes like GS2 (Kotschenreuther *et al.* 1995*b*; Dorland *et al.* 2000) and stella (Barnes, Parra & Landreman 2019) that use finite-difference methods in velocity space. Flexible velocity resolution combined with a GPU implementation allows GX to run nonlinear (electrostatic with adiabatic electrons) gyrokinetic simulations in minutes rather than hours or days.

A.2. DESC

DESC (Dudt & Kolemen 2020; Conlin *et al.* 2023; Dudt *et al.* 2023; Panici *et al.* 2023) is a new stellarator equilibrium and optimization code. Two of the main features of DESC are its pseudo-spectral representation of the magnetic geometry and the use of automatic differentiation to compute exact derivatives.

DESC directly solves the ideal MHD force balance equation

$$\boldsymbol{J} \times \boldsymbol{B} = \nabla p. \tag{A4}$$

DESC uses as its computational domain the coordinates (ρ, θ, ζ), defined as

$$\rho = \sqrt{\frac{\psi}{\psi_a}}, \tag{A5a}$$

$$\theta = \theta^* - \lambda(\rho, \theta, \zeta), \tag{A5b}$$

$$\zeta = \phi, \tag{A5c}$$

where ψ is the enclosed toroidal flux, ψ_a is the total enclosed toroidal flux in the plasma, θ^* is a straight field line poloidal angle, λ is a stream function and ϕ is the geometric toroidal angle. It then expands λ, and the cylindrical coordinates R and Z in terms of a Fourier–Zernike basis,

$$R(\rho, \theta, \zeta) = \sum_{m=-M, n=-N, l=0}^{M,N,L} R_{lmn} \mathcal{Z}_l^m(\rho, \theta) \mathcal{F}^n(\zeta), \tag{A6a}$$

$$\lambda(\rho, \theta, \zeta) = \sum_{m=-M, n=-N, l=0}^{M,N,L} \lambda_{lmn} \mathcal{Z}_l^m(\rho, \theta) \mathcal{F}^n(\zeta), \tag{A6b}$$

$$Z(\rho, \theta, \zeta) = \sum_{m=-M, n=-N, l=0}^{M,N,L} Z_{lmn} \mathcal{Z}_l^m(\rho, \theta) \mathcal{F}^n(\zeta), \tag{A6c}$$

Parameter	Value
Normalized toroidal flux (s)	0.5
Field line label (α)	0.0 (and $\pi\iota/4$) for multiple α
Number of poloidal turns (npol)	1
Parallel resolution (ntheta)	64
Radial resolution (nx)	64
Field line label resolution (ny)	64
Hermite resolution (nhermite)	8
Laguerre resolution (nlaguerre)	4
Normalized temperature gradient (tprim)	3.0
Normalized density gradient (fprim)	1.0

TABLE 1. Simulation parameters used for the GX simulations within the optimization loop for §§ 3.1–3.3.

where $\mathcal{R}_l^{|m|}$, $\mathcal{Z}_l^m$ and $\mathcal{F}$ are the shifted Jacobi polynomials, Zernike polynomials and Fourier series, respectively.

It can be shown that the force error $\boldsymbol{J} \times \boldsymbol{B} - \nabla p$ has only two independent components

$$F_\rho = \sqrt{g}(J^\zeta B^\theta - J^\theta B^\zeta) + p', \tag{A7}$$

$$F_\beta = \sqrt{g}J^\rho, \tag{A8}$$

where $\sqrt{g}$ is the Jacobian and the superscripts indicate the contravariant components. DESC then solves for the spectral coefficients R_{lmn}, Z_{lmn} and λ_{lmn} that minimize this force error.

DESC also uses automatic differentiation through JAX (Bradbury *et al.* 2018) in its optimization routines. Briefly, in automatic differentiation, the chain rule is applied through the code, allowing DESC to compute exact derivatives faster and more accurately than derivatives from finite differencing routines. However, at the time of this writing, this requires writing the objective function in native PYTHON and GX is written in C++/CUDA. Furthermore, it is unclear how effective it would be to differentiate through the time stepping of such a noisy system. Therefore, this would require extensive code developments for both DESC and GX, and so AD will not be used when optimizing for reduced turbulence.

Appendix B. Simulation parameters used in optimization

B.1. *Turbulence, turbulence-QS and multiple α optimization*

For the GX simulations within the optimization loop, we use the simulation parameters listed in table 1.

The number of poloidal turns, parallel resolution and number of simulated modes (proportional to n_x and n_y) are chosen to enable cheaper simulations. However, one poloidal turn has been used in previous nonlinear W7-X gyrokinetic benchmark studies (González-Jerez *et al.* 2022*a*). Furthermore, from a quasilinear estimate, the heat flux scales like $1/\langle k_\perp^2 \rangle$ (where the angle brackets indicate a flux-surface average) (Mariani *et al.* 2018). Consequently, higher wavenumber modes contribute less to the heat flux. The gradients are typical for experimental profiles for W7-X (Beurskens *et al.* 2021) and have been previously used for benchmark, transport and optimization studies (González-Jerez *et al.* 2022*b*; Roberg-Clark *et al.* 2023; Bañón Navarro *et al.* 2023). Finally, it has been

Parameter	Value
Hermite resolution (nhermite)	4
Laguerre resolution (nlaguerre)	2
Collision frequency (vnewk)	2.0
Normalized temperature gradient (tprim)	5.0
Normalized density gradient (fprim)	1.0

TABLE 2. Simulation parameters used for the GX simulations within the optimization loop for the fluid case in § 3.4.

Parameter	Value
Number of poloidal turns (npol)	2
Parallel resolution (ntheta)	128
Radial resolution (nx)	128
Field line label resolution (ny)	128
Hermite resolution (nhermite)	16
Laguerre resolution (nlaguerre)	8

TABLE 3. Simulation parameters used for the GX simulations for post-processing.

Parameter	Value
Spectral resolution (M, N, L)	(8, 8, 8)
Grid resolution (M_{grid}, L_{grid}, N_{grid})	(16, 16, 16)
Total toroidal flux (Psi)	0.03817902
Major radius (R0)	1.0
Aspect ratio (R0/a)	8.0

TABLE 4. Resolution parameters for the DESC equilibria in this study.

demonstrated that GX simulations can yield accurate heat flux traces using the specified number of Hermite and Laguerre polynomials (Mandell *et al.* 2023). To check our results, the post-processing simulations are run at higher resolution.

B.2. *Fluid approximation optimization*

When running the optimization at fluid resolution, the simulation parameters are identical except those listed in table 2.

The velocity resolution is decreased to that of the gyrofluid model and the collision frequency is increased. Since there is far greater dissipation in the model, we also increase the temperature gradient.

B.3. *Post-processing simulations*

To ensure our final results are well converged, we increase the resolution of the post-processing simulations to those listed in table 3.

The scans over ρ, α and a/L_T of course include additional values.

Parameter	Value
GX evaluation	3
GX optimization	10
Quasi-symmetry optimization	4

TABLE 5. Approximate timings for different parts of the optimization loop.

B.4. DESC *equilibria parameters*

The parameters used for the equilibria are shown in table 4. These are considered 'moderate' resolution and are typical values used for optimization. This is a vacuum case, so both the pressure and current are zero. When optimized for precise QH symmetry, the resulting equilibrium is similar to the QH equilibrum of Landreman & Paul (2022).

B.5. *Optimization timings*

Table 5 shows the approximate timings of different parts of the optimization loop using a single Tesla V100 GPU on the Princeton Traverse cluster. All values are in minutes. In total, each optimization completed in approximately 20 hours. It takes slightly more than 12 hours on a single NVIDIA A100 GPU on the NERSC Perlmutter cluster. The 'optimization' timings are for a single iteration.

Appendix C. DESC input files

C.1. *Initial equilibrium*

```
# DESC input file generated from the output file:
# test_equilibria/traverse/input.nfp4_QH_desc_output.h5

# global parameters
sym = 1
NFP = 4
Psi = 3.81790200E-02

# spectral resolution
L_rad = 8
M_pol = 8
N_tor = 8
L_grid = 16
M_grid = 16
N_grid = 16

# solver tolerances
ftol = 1.000E-02
xtol = 1.000E-06
gtol = 1.000E-06
maxiter = 100

# solver methods
optimizer = lsq-exact
objective = force
spectral_indexing = ansi
```

```
# pressure and rotational transform/current profiles
l: 0 p = 0.00000000E+00 c = 6.27347854E-14
l: 1 p = 0.00000000E+00 c = 1.71917298E-10
l: 2 p = 0.00000000E+00 c = -5.44257894E-09
l: 3 p = 0.00000000E+00 c = 5.92952658E-08
l: 4 p = 0.00000000E+00 c = -2.95964833E-07
l: 5 p = 0.00000000E+00 c = 7.63421786E-07
l: 6 p = 0.00000000E+00 c = -1.05372308E-06
l: 7 p = 0.00000000E+00 c = 7.38152073E-07

# fixed-boundary surface shape
l: 0 m: -3 n: -3 R1 =  1.06009580E-04 Z1 =  0.00000000E+00
l: 0 m: -2 n: -3 R1 =  1.94431471E-03 Z1 =  0.00000000E+00
l: 0 m: -1 n: -3 R1 = -4.45641661E-03 Z1 =  0.00000000E+00
l: 0 m: -3 n: -2 R1 = -1.41499877E-03 Z1 =  0.00000000E+00
l: 0 m: -2 n: -2 R1 =  3.96888996E-03 Z1 =  0.00000000E+00
l: 0 m: -1 n: -2 R1 = -3.13235773E-02 Z1 =  0.00000000E+00
l: 0 m: -3 n: -1 R1 = -4.34044222E-04 Z1 =  0.00000000E+00
l: 0 m: -2 n: -1 R1 =  2.23464111E-02 Z1 =  0.00000000E+00
l: 0 m: -1 n: -1 R1 = -1.16349515E-01 Z1 =  0.00000000E+00
l: 0 m:  0 n:  0 R1 =  1.00000000E+00 Z1 =  0.00000000E+00
l: 0 m:  1 n:  0 R1 =  1.38039047E-01 Z1 =  0.00000000E+00
l: 0 m:  2 n:  0 R1 =  1.53605591E-02 Z1 =  0.00000000E+00
l: 0 m:  3 n:  0 R1 =  1.06941717E-04 Z1 =  0.00000000E+00
l: 0 m:  0 n:  1 R1 =  1.79819788E-01 Z1 =  0.00000000E+00
l: 0 m:  1 n:  1 R1 = -5.82857335E-02 Z1 =  0.00000000E+00
l: 0 m:  2 n:  1 R1 =  1.99922621E-02 Z1 =  0.00000000E+00
l: 0 m:  3 n:  1 R1 =  5.62751826E-04 Z1 =  0.00000000E+00
l: 0 m:  0 n:  2 R1 =  1.73495115E-02 Z1 =  0.00000000E+00
l: 0 m:  1 n:  2 R1 = -1.75730831E-02 Z1 =  0.00000000E+00
l: 0 m:  2 n:  2 R1 =  6.57072137E-03 Z1 =  0.00000000E+00
l: 0 m:  3 n:  2 R1 = -7.49945379E-04 Z1 =  0.00000000E+00
l: 0 m:  0 n:  3 R1 =  2.57716016E-03 Z1 =  0.00000000E+00
l: 0 m:  1 n:  3 R1 = -2.82630336E-03 Z1 =  0.00000000E+00
l: 0 m:  2 n:  3 R1 =  1.99926226E-03 Z1 =  0.00000000E+00
l: 0 m:  3 n:  3 R1 = -6.16014311E-05 Z1 =  0.00000000E+00
l: 0 m:  0 n: -3 R1 =  0.00000000E+00 Z1 = -1.19971129E-03
l: 0 m:  1 n: -3 R1 =  0.00000000E+00 Z1 =  2.92305608E-03
l: 0 m:  2 n: -3 R1 =  0.00000000E+00 Z1 = -3.11190235E-03
l: 0 m:  3 n: -3 R1 =  0.00000000E+00 Z1 =  6.31026425E-04
l: 0 m:  0 n: -2 R1 =  0.00000000E+00 Z1 = -1.86633423E-02
l: 0 m:  1 n: -2 R1 =  0.00000000E+00 Z1 =  1.74756774E-02
l: 0 m:  2 n: -2 R1 =  0.00000000E+00 Z1 = -8.13554668E-03
l: 0 m:  3 n: -2 R1 =  0.00000000E+00 Z1 =  4.05079520E-04
l: 0 m:  0 n: -1 R1 =  0.00000000E+00 Z1 = -1.52687429E-01
l: 0 m:  1 n: -1 R1 =  0.00000000E+00 Z1 =  1.24701868E-02
l: 0 m:  2 n: -1 R1 =  0.00000000E+00 Z1 =  1.90954690E-02
l: 0 m:  3 n: -1 R1 =  0.00000000E+00 Z1 =  4.56955309E-04
l: 0 m: -3 n:  0 R1 =  0.00000000E+00 Z1 = -3.80285523E-04
l: 0 m: -2 n:  0 R1 =  0.00000000E+00 Z1 = -9.45114014E-03
```

```
l: 0 m: -1 n:  0 R1 =  0.00000000E+00 Z1 = -1.31096801E-01
l: 0 m: -3 n:  1 R1 =  0.00000000E+00 Z1 = -1.82574602E-03
l: 0 m: -2 n:  1 R1 =  0.00000000E+00 Z1 = -1.82165803E-02
l: 0 m: -1 n:  1 R1 =  0.00000000E+00 Z1 = -7.78740189E-02
l: 0 m: -3 n:  2 R1 =  0.00000000E+00 Z1 = -8.84087119E-04
l: 0 m: -2 n:  2 R1 =  0.00000000E+00 Z1 =  6.09077798E-03
l: 0 m: -1 n:  2 R1 =  0.00000000E+00 Z1 = -3.10828300E-02
l: 0 m: -3 n:  3 R1 =  0.00000000E+00 Z1 = -6.25309212E-04
l: 0 m: -2 n:  3 R1 =  0.00000000E+00 Z1 =  3.23790883E-03
l: 0 m: -1 n:  3 R1 =  0.00000000E+00 Z1 = -4.64192523E-03
```

C.2. *Optimized equilibrium from § 3.2*

```
# DESC input file generated from the output file:
# test_equilibria/traverse/swap_gx_end_case23_n_3_i_14.h5

# global parameters
sym = 1
NFP = 4
Psi = 3.81790200E-02

# spectral resolution
L_rad = 8
M_pol = 8
N_tor = 8
L_grid = 16
M_grid = 16
N_grid = 16

# solver tolerances
ftol = 1.000E-02
xtol = 1.000E-06
gtol = 1.000E-06
maxiter = 100

# solver methods
optimizer = lsq-exact
objective = force
spectral_indexing = ansi

# pressure and rotational transform/current profiles
l: 0 p = 0.00000000E+00 c = 1.41645009E-12
l: 1 p = 0.00000000E+00 c = -9.25103059E-10
l: 2 p = 0.00000000E+00 c = 2.19561377E-08
l: 3 p = 0.00000000E+00 c = -1.86967349E-07
l: 4 p = 0.00000000E+00 c = 7.72543397E-07
l: 5 p = 0.00000000E+00 c = -1.72630498E-06
l: 6 p = 0.00000000E+00 c = 2.13214903E-06
l: 7 p = 0.00000000E+00 c = -1.36872255E-06
```

```
# fixed-boundary surface shape
l: 0 m: -3 n: -3 R1 = -9.31357764E-03 Z1 =  0.00000000E+00
l: 0 m: -2 n: -3 R1 =  5.71628208E-03 Z1 =  0.00000000E+00
l: 0 m: -1 n: -3 R1 =  3.82681062E-03 Z1 =  0.00000000E+00
l: 0 m: -3 n: -2 R1 = -2.29388344E-03 Z1 =  0.00000000E+00
l: 0 m: -2 n: -2 R1 =  1.66286469E-02 Z1 =  0.00000000E+00
l: 0 m: -1 n: -2 R1 = -2.90057270E-02 Z1 =  0.00000000E+00
l: 0 m: -3 n: -1 R1 =  5.54126280E-03 Z1 =  0.00000000E+00
l: 0 m: -2 n: -1 R1 = -5.52415883E-03 Z1 =  0.00000000E+00
l: 0 m: -1 n: -1 R1 = -9.11377065E-02 Z1 =  0.00000000E+00
l: 0 m:  0 n:  0 R1 =  1.00000000E+00 Z1 =  0.00000000E+00
l: 0 m:  1 n:  0 R1 =  1.22407301E-01 Z1 =  0.00000000E+00
l: 0 m:  2 n:  0 R1 =  1.27821030E-02 Z1 =  0.00000000E+00
l: 0 m:  3 n:  0 R1 = -4.34547934E-03 Z1 =  0.00000000E+00
l: 0 m:  0 n:  1 R1 =  1.87116351E-01 Z1 =  0.00000000E+00
l: 0 m:  1 n:  1 R1 = -3.72058062E-02 Z1 =  0.00000000E+00
l: 0 m:  2 n:  1 R1 = -1.28924376E-02 Z1 =  0.00000000E+00
l: 0 m:  3 n:  1 R1 = -3.16332008E-03 Z1 =  0.00000000E+00
l: 0 m:  0 n:  2 R1 =  1.98088376E-02 Z1 =  0.00000000E+00
l: 0 m:  1 n:  2 R1 = -1.77228005E-02 Z1 =  0.00000000E+00
l: 0 m:  2 n:  2 R1 =  1.84202872E-02 Z1 =  0.00000000E+00
l: 0 m:  3 n:  2 R1 = -4.06745395E-03 Z1 =  0.00000000E+00
l: 0 m:  0 n:  3 R1 = -2.60990147E-04 Z1 =  0.00000000E+00
l: 0 m:  1 n:  3 R1 =  6.53196582E-04 Z1 =  0.00000000E+00
l: 0 m:  2 n:  3 R1 =  5.67484731E-03 Z1 =  0.00000000E+00
l: 0 m:  3 n:  3 R1 = -8.14020474E-03 Z1 =  0.00000000E+00
l: 0 m:  0 n: -3 R1 =  0.00000000E+00 Z1 = -2.70012676E-03
l: 0 m:  1 n: -3 R1 =  0.00000000E+00 Z1 =  2.98740559E-03
l: 0 m:  2 n: -3 R1 =  0.00000000E+00 Z1 = -6.43349941E-03
l: 0 m:  3 n: -3 R1 =  0.00000000E+00 Z1 =  4.61608984E-03
l: 0 m:  0 n: -2 R1 =  0.00000000E+00 Z1 = -1.16055055E-02
l: 0 m:  1 n: -2 R1 =  0.00000000E+00 Z1 =  4.49890508E-03
l: 0 m:  2 n: -2 R1 =  0.00000000E+00 Z1 = -1.21016790E-02
l: 0 m:  3 n: -2 R1 =  0.00000000E+00 Z1 =  1.27557793E-03
l: 0 m:  0 n: -1 R1 =  0.00000000E+00 Z1 = -1.07649166E-01
l: 0 m:  1 n: -1 R1 =  0.00000000E+00 Z1 =  2.76464573E-02
l: 0 m:  2 n: -1 R1 =  0.00000000E+00 Z1 = -9.28264520E-03
l: 0 m:  3 n: -1 R1 =  0.00000000E+00 Z1 =  1.51173953E-03
l: 0 m: -3 n:  0 R1 =  0.00000000E+00 Z1 =  7.72083960E-03
l: 0 m: -2 n:  0 R1 =  0.00000000E+00 Z1 = -2.60627865E-03
l: 0 m: -1 n:  0 R1 =  0.00000000E+00 Z1 = -1.56329762E-01
l: 0 m: -3 n:  1 R1 =  0.00000000E+00 Z1 =  6.64991771E-03
l: 0 m: -2 n:  1 R1 =  0.00000000E+00 Z1 =  1.41663507E-02
l: 0 m: -1 n:  1 R1 =  0.00000000E+00 Z1 = -7.37699026E-02
l: 0 m: -3 n:  2 R1 =  0.00000000E+00 Z1 = -1.91298905E-03
l: 0 m: -2 n:  2 R1 =  0.00000000E+00 Z1 =  1.48107607E-02
l: 0 m: -1 n:  2 R1 =  0.00000000E+00 Z1 = -1.58727989E-02
l: 0 m: -3 n:  3 R1 =  0.00000000E+00 Z1 = -5.73744002E-03
l: 0 m: -2 n:  3 R1 =  0.00000000E+00 Z1 =  5.43576784E-03
l: 0 m: -1 n:  3 R1 =  0.00000000E+00 Z1 =  8.48615937E-04
```

REFERENCES

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

BADER, A., FABER, B.J., SCHMITT, J.C., ANDERSON, D.T., DREVLAK, M., DUFF, J.M., FRERICHS, H., HEGNA, C.C., KRUGER, T.G., LANDREMAN, M., *et al.* 2020 Advancing the physics basis for quasi-helically symmetric stellarators. *J. Plasma Phys.* **86** (5), 905860506.

BAÑÓN NAVARRO, A., DI SIENA, A., VELASCO, J.L., WILMS, F., MERLO, G., WINDISCH, T., LODESTRO, L.L., PARKER, J.B. & JENKO, F. 2023 First-principles based plasma profile predictions for optimized stellarators. *Nucl. Fusion* **63** (5), 054003.

BARNES, M., ABEL, I.G., DORLAND, W., GÖRLER, T., HAMMETT, G.W. & JENKO, F. 2010 Direct multiscale coupling of a transport code to gyrokinetic turbulence codes. *Phys. Plasmas* **17** (5), 056109.

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

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

BEIDLER, C.D., SMITH, H.M., ALONSO, A., ANDREEVA, T., BALDZUHN, J., BEURSKENS, M.N., BORCHARDT, M., BOZHENKOV, S.A., BRUNNER, K.J., DAMM, H., *et al.* 2021 Demonstration of reduced neoclassical energy transport in Wendelstein 7-X. *Nature* **596**, 221–226.

BEURSKENS, M.N.A., BOZHENKOV, S.A., FORD, O., XANTHOPOULOS, P., ZOCCO, A., TURKIN, Y., ALONSO, A., BEIDLER, C., CALVO, I., CARRALERO, D., *et al.* 2021 Ion temperature clamping in Wendelstein 7-X electron cyclotron heated plasmas. *Nucl. Fusion* **61** (11), 116072.

BRADBURY, J., FROSTIG, R., HAWKINS, P., JOHNSON, M.J., LEARY, C., MACLAURIN, D., NECULA, G., PASZKE, A., VANDERPLAS, J., WANDERMAN-MILNE, S., *et al.* 2018 JAX: composable transformations of Python+NumPy programs.

BUCK, B., DORLAND, W., MANDELL, N., KIM, P., FISCHER, S. & QIAN, T. 2022 A comparison of numerical and analytic ITG turbulence models in the gyrokinetic code GX.

BULLER, S., MANDELL, N., PARISI, J., KIM, P., ADKINS, T., GAUR, R., DORLAND, W. & LANDREMAN, M. 2023 Linear physics proxies for the nonlinear heat flux in stellarators. In *Simons-CIEMAT Joint Meeting on Stellarator Turbulence Optimization*.

CATTO, P.J. 1978 Linearized gyro-kinetics. *Plasma Phys.* **20** (7), 719–722.

CHAN, B.L., DOUCET, A. & TADIC, V.B. 2003 Optimisation of particle filters using simultaneous perturbation stochastic approximation. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings*, vol. 6, pp. VI–681.

CONLIN, R., DUDT, D.W., PANICI, D. & KOLEMEN, E. 2023 The DESC stellarator code suite. Part 2. Perturbation and continuation methods. *J. Plasma Phys.* **89** (3), 955890305.

DEWAR, R.L. & GLASSER, A.H. 1983 Ballooning mode spectrum in general toroidal systems. *Phys. Fluids* **26** (10), 3038–3052.

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

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

DOUGHERTY, J.P. 1964 Model Fokker-Planck equation for a plasma and its solution. *Phys. Fluids* **7** (11), 1788–1799.

DUDT, D.W., CONLIN, R., PANICI, D. & KOLEMEN, E. 2023 The DESC stellarator code suite. Part 3. Quasi-symmetry optimization. *J. Plasma Phys.* **89** (2), 955890201.

DUDT, D.W. & KOLEMEN, E. 2020 DESC: a stellarator equilibrium solver. *Phys. Plasmas* **27** (10), 102513.

FABER, B.J., PUESCHEL, M.J., PROLL, J.H.E., XANTHOPOULOS, P., TERRY, P.W., HEGNA, C.C., WEIR, G.M., LIKIN, K.M. & TALMADGE, J.N. 2015 Gyrokinetic studies of trapped electron mode turbulence in the Helically Symmetric eXperiment stellarator. *Phys. Plasmas* **22** (7), 072305.

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

GONZÁLEZ-JEREZ, A., XANTHOPOULOS, P., GARCÍA-REGAÑA, J.M., CALVO, I., ALCUSÓN, J., BAÑÓN NAVARRO, A., BARNES, M., PARRA, F.I. & GEIGER, J. 2022*a* Electrostatic gyrokinetic

simulations in Wendelstein 7-X geometry: benchmark between the codes stella and GENE. *J. Plasma Phys*. **88** (3), 905880310.

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

GOODMAN, A., MATA, K.C., HENNEBERG, S.A., JORGE, R., LANDREMAN, M., PLUNK, G., SMITH, H., MACKENBACH, R. & HELANDER, P. 2022 Constructing precisely quasi-isodynamic magnetic fields. *J. Plasma Phys*. **89** (5), 905890504.

GREENE, J.M. & CHANCE, M.S. 1981 The second region of stability against ballooning modes. *Nucl. Fusion* **21** (4), 453–464.

HEGNA, C.C., TERRY, P.W. & FABER, B.J. 2018 Theory of ITG turbulent saturation in stellarators: identifying mechanisms to reduce turbulent transport. *Phys. Plasmas* **25** (2), 022511.

HELANDER, P. 2014 Theory of plasma confinement in non-axisymmetric magnetic fields. *Rep. Prog. Phys*. **77** (8), 087001.

HELANDER, P., PROLL, J.H.E. & PLUNK, G.G. 2013 Collisionless microinstabilities in stellarators. I. Analytical theory of trapped-particle modes. *Phys. Plasmas* **20** (12), 122505.

HORTON, W. 1999 Drift waves and transport. *Rev. Mod. Phys*. **71** (3), 735–778.

JORGE, R., DORLAND, W., KIM, P., LANDREMAN, M., MANDELL, N.R., MERLO, G. & QIAN, T. 2023 Direct microstability optimization of stellarator devices.

KIM, P. 2024 Data and scripts for 'Optimization of nonlinear turbulence in stellarators'.

KOTSCHENREUTHER, M., DORLAND, W., BEER, M.A. & HAMMETT, G.W. 1995*a* Quantitative predictions of tokamak energy confinement from first-principles simulations with kinetic effects. *Phys. Plasmas* **2** (6), 2381–2389.

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

LANDREMAN, M., BULLER, S. & DREVLAK, M. 2022 Optimization of quasi-symmetric stellarators with self-consistent bootstrap current and energetic particle confinement. *Phys. Plasmas* **29** (8), 082501.

LANDREMAN, M. & PAUL, E. 2022 Magnetic fields with precise quasisymmetry for plasma confinement. *Phys. Rev. Lett*. **128** (3), 035001.

MANDELL, N.R., DORLAND, W., ABEL, I., GAUR, R., KIM, P., MARTIN, M. & QIAN, T. 2023 GX: a GPU-native gyrokinetic turbulence code for tokamak and stellarator design.

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

MARIANI, A., BRUNNER, S., DOMINSKI, J., MERLE, A., MERLO, G., SAUTER, O., GÖRLER, T., JENKO, F. & TOLD, D. 2018 Identifying microturbulence regimes in a TCV discharge making use of physical constraints on particle and heat fluxes. *Phys. Plasmas* **25** (1), 012313.

MCKINNEY, I.J., PUESCHEL, M.J., FABER, B.J., HEGNA, C.C., TALMADGE, J.N., ANDERSON, D.T., MYNICK, H.E. & XANTHOPOULOS, P. 2019 A comparison of turbulent transport in a quasi-helical and a quasi-axisymmetric stellarator. *J. Plasma Phys*. **85** (5), 905850503.

MYNICK, H.E., POMPHREY, N. & XANTHOPOULOS, P. 2010 Optimizing stellarators for turbulent transport. *Phys. Rev. Lett*. **105** (9), 095004.

PANICI, D., CONLIN, R., DUDT, D.W., UNALMIS, K. & KOLEMEN, E. 2023 The DESC stellarator code suite. Part 1. Quick and accurate equilibria computations. *J. Plasma Phys*. **89** (3), 955890303.

PROLL, J.H.E., HELANDER, P., CONNOR, J.W. & PLUNK, G.G. 2012 Resilience of quasi-isodynamic stellarators against trapped-particle instabilities. *Phys. Rev. Lett*. **108** (24), 245002.

PROLL, J.H.E., MYNICK, H.E., XANTHOPOULOS, P., LAZERSON, S.A. & FABER, B.J. 2016 TEM turbulence optimisation in stellarators. *Plasma Phys. Control. Fusion* **58** (1), 014006.

PROLL, J.H.E., PLUNK, G.G., FABER, B.J., GÖRLER, T., HELANDER, P., MCKINNEY, I.J., PUESCHEL, M.J., SMITH, H.M. & XANTHOPOULOS, P. 2022 Turbulence mitigation in maximum-J stellarators with electron-density gradient. *J. Plasma Phys*. **88** (1), 905880112.

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

ROBERG-CLARK, G.T., PLUNK, G.G., XANTHOPOULOS P., NÜHRENBERG C., HENNEBERG S.A. & SMITH H.M. 2023 Critical gradient turbulence optimization toward a compact stellarator reactor concept. *Phys. Rev. Res.* **5** (3), L032030.

SÁNCHEZ, E., VELASCO, J.L., CALVO, I. & MULAS, S. 2023 A quasi-isodynamic configuration with good confinement of fast ions at low plasma β. *Nucl. Fusion* **63** (6), 066037.

SPALL, J.C. 1987 A stochastic approximation technique for generating maximum likelihood parameter estimates. In *1987 American Control Conference*, pp. 1161–1167.

VELASCO, J.L., CALVO, I., SÁNCHEZ, E. & PARRA, F.I. 2023 Robust stellarator optimization via flat mirror magnetic fields. *Nucl. Fusion* **63** (12), 126038.