

PAPER

Accelerating the estimation of collisionless energetic particle confinement statistics in stellarators using multifidelity Monte Carlo

To cite this article: Frederick Law *et al* 2022 *Nucl. Fusion* **62** 076019

View the [article online](#) for updates and enhancements.

You may also like

- [Physics analysis of results of stochastic and classic stellarator coil optimization](#)
Jim-Felix Lobsien, Michael Drevlak, Frank Jenko et al.
- [Improving fast-particle confinement in quasi-axisymmetric stellarator optimization](#)
S A Henneberg, M Drevlak and P Helander
- [Recent advances in stellarator optimization](#)
D.A. Gates, A.H. Boozer, T. Brown et al.

Accelerating the estimation of collisionless energetic particle confinement statistics in stellarators using multifidelity Monte Carlo

Frederick Law*, Antoine Cerfon[✉] and Benjamin Peherstorfer

Courant Institute of Mathematics, New York University, United States of America

E-mail: law@cims.nyu.edu

Received 18 August 2021, revised 16 December 2021

Accepted for publication 31 December 2021

Published 4 May 2022



Abstract

In the design of stellarators, energetic particle confinement is a critical point of concern which remains challenging to study from a numerical point of view. Standard Monte Carlo (MC) analyses are highly expensive because a large number of particle trajectories need to be integrated over long time scales, and small time steps must be taken to accurately capture the features of the wide variety of trajectories. Even when they are based on guiding center trajectories, as opposed to full-orbit trajectories, these standard MC studies are too expensive to be included in most stellarator optimization codes. We present the first multifidelity Monte Carlo (MFMC) scheme for accelerating the estimation of energetic particle confinement in stellarators. Our approach relies on a two-level hierarchy, in which a guiding center model serves as the high-fidelity model, and a data-driven linear interpolant is leveraged as the low-fidelity surrogate model. We apply MFMC to the study of energetic particle confinement in a four-period quasi-helically symmetric stellarator, assessing various metrics of confinement. Stemming from the very high computational efficiency of our surrogate model as well as its sufficient correlation to the high-fidelity model, we obtain speedups of up to 10 with MFMC compared to standard MC.

Keywords: stellarator, energetic particle confinement, Monte Carlo, variance reduction

(Some figures may appear in colour only in the online journal)

1. Introduction

In generic non-axisymmetric magnetic configurations, collisionless particle orbits are not guaranteed to be confined [26]. Energetic particle confinement therefore is a key point of concern for the design of tokamaks which have significant toroidal ripple and of stellarators [4, 6, 9, 18, 22, 27, 34, 37]. In order to design configurations which maximize energetic particle confinement, one must identify metrics which accurately capture the level and quality of confinement, and are at the same time as easy and as inexpensive to include in design optimization codes as possible.

One can distinguish two distinct approaches to estimating energetic particle confinement. The first approach may be viewed as a direct approach, or Monte Carlo (MC) approach.

* Author to whom any correspondence should be addressed.

One launches an ensemble of energetic particles, follows their individual orbits—usually their drift orbits—and then directly calculates the fraction of particles which have been lost and the time of flight of particles before being lost [1, 3, 5, 22, 27, 31, 39, 49, 52, 58]. On the one hand, this method can be considered as physically accurate, since it directly captures the effects on confinement of the wide variety of orbits energetic particles follow in nonaxisymmetric magnetic configurations. On the other hand, the convergence of the MC estimator as a function of the number of particles launched is slow, requiring a large ensemble of particle trajectories for accurate estimations. This means that for guiding center calculations this first method is more time-consuming than most other calculations involved in the design optimization of nonaxisymmetric equilibria, and prohibitively expensive for full-orbit trajectories. Furthermore, this method does not lend itself easily to the computation of

derivatives of the metric with respect to magnetic field variations, which implies that it cannot be easily implemented in gradient-based optimization schemes.

The second approach used for estimating the quality of energetic particle confinement is deterministic and relies on proxy functions with closed analytic forms, which can be directly computed from the output of a nonaxisymmetric equilibrium solver, and which have empirically been found to often correlate well with energetic particle confinement. One may for example make the approximation that good energetic particle confinement will be a direct consequence of good low collisionality neoclassical transport properties, and thus characterize the confinement performance through the evaluation of the effective ripple ϵ_{eff} first introduced by Nemov *et al* [40, 59]. In order to target energetic particle confinement more directly, one may instead rely on another metric recently proposed by Nemov, called the Γ_c metric [4, 41]. Γ_c is better suited for this purpose because it applies to fully collisionless orbits and unlike ϵ_{eff} it is also strongly influenced by particles close to the trapped/passing boundary, which often are the dominant contribution to energetic particle losses [4, 5]. The benefits of these proxy quantities for energetic particle confinement are three folds. First, they are significantly less expensive computationally than MC based solely on expensive, high-fidelity simulations. Second, their closed analytic forms, which were derived from physics principles, enable easier interpretation of the numerical results and identification of trends. Third, it may be easier to compute with high accuracy their derivatives with respect to optimization parameters, as was recently demonstrated [45].

Despite the advantages of the deterministic approach, its application may remain limited to preliminary optimization studies, because the currently used deterministic functions have known limitations in their predictive capabilities [1]. In the study of a vacuum quasi-helically symmetric configuration, it was for example recently found that energetic particle confinement could be anti-correlated with ϵ_{eff} [4]: energetic particle confinement was improved while at the same time the ϵ_{eff} metric increased. The Γ_c metric was found to be reliable in that same study [4], and a strong correlation between alpha particle energy loss and Γ_c in the outer half of the plasma was recently highlighted for a wide variety of stellarator configurations [3]. However, that correlation is not perfect. Stellarators with significantly different Γ_c values can have comparable levels of energy loss, and stellarators with comparable Γ_c values can have significantly different levels of energy loss. It therefore appears that for the goal of implementing reliable and efficient energetic particle confinement calculations in multi-objective optimization codes, it is necessary to include high-fidelity simulations with lower costs than a direct MC approach. This is precisely the purpose of this article: we apply the multifidelity Monte Carlo (MFMC) method [42, 47, 48] to reduce the variance of the MC estimator for energetic particle confinement, and thus reduce the computational cost of such an estimator.

As we will explain in more detail in the main text, the MFMC method is a strategy for variance reduction of MC estimators based on control variates; see, e.g. [23, 38]. In

MFMC, the control variates are given by low-fidelity models, i.e. approximations of the high-fidelity physics model one is interested in, which are carefully designed to satisfy two key criteria: they have a high correlation with the high-fidelity model, and they are significantly less expensive to evaluate than the high-fidelity model. These criteria are required to obtain variance reduction as shown in [42, 47, 48]. The MFMC approach is well suited to situations in which scientists rely on the MC method to propagate the uncertainty associated with the initial conditions and with parameters defining the model, as is precisely the case for the confinement of energetic particles in nonaxisymmetric magnetic configurations. The method and some of its close variants have recently been applied for the first time to kinetic plasma models [15, 16, 29]. In [15, 16], the authors demonstrate that physics based reduced models can be effectively used for variance reduction of the MC estimator for the solution of the Boltzmann equation subject to uncertainty, in highly collisional plasmas not subject to a far-field force field. In contrast, in [29], the authors rely on data-driven low-fidelity models to reduce the variance of their MC estimators for quantifying the uncertainty in kinetic microturbulence. To the best of our knowledge, our article represents the first time the MFMC method is applied to the study of energetic particle confinement in stellarators, and stellarator performance in general. As in [29], we rely on a data-driven low-fidelity model for MFMC based variance reduction, since traditional sources of low-fidelity models are inappropriate for our high-fidelity model. For example, our low-fidelity model cannot just utilize simplified physics because our high-fidelity model already is a reduced physics model, i.e. collisionless guiding center trajectories, and we are not aware of simpler physics based models which maintain a reasonably high correlation with the high-fidelity model during the entire energetic particle trajectory. We report that despite the large variety of particle orbits depending on the initial conditions, and despite the chaotic nature of the dynamics, our relatively simple data-driven model—which can be trained with standard tools available in common scientific software packages in Python and Matlab—can lead to significant variance reduction in the MFMC framework, and corresponding speedup.

The structure of the article is as follows. In section 2, we present our guiding center model for collisionless energetic particle trajectories, and briefly discuss why numerical simulations based on this model are intrinsically expensive. In section 3 we give a brief review of the standard MC approach for estimating energetic particle confinement, which we contrast with our new MFMC approach in section 4. We describe the set up for our numerical tests and demonstrate the speed-up obtained with our MFMC method in these tests in section 5. We summarize our work in section 6, where we also identify some of its limitations, and suggest corresponding directions for future work.

2. Energetic particle dynamics

The main purpose of our article is to demonstrate in a realistic setting the potential of the MFMC method to reduce the variance of the commonly used MC estimators for energetic

particle confinement, and thus to speed up the evaluation of these estimators for optimization applications. For simplicity, we will consider MC estimators based on collisionless orbits, as is still often done in stellarator optimization studies [4, 5, 14, 17, 27]. Our MFMC approach will be directly applicable to MC estimators based on single-particle orbits including collisions as well [27, 31, 49]. In this section, we begin by detailing the equations of motion we solve to compute the trajectories of energetic particles. We then highlight certain features of the model which underscore the expensive nature behind simulating particle trajectories.

2.1. Guiding center trajectories

We consider the collisionless dynamics of 3.5 MeV alpha particles as the byproduct of deuterium–tritium fusion in a stellarator magnetic field. We assume the magnetic configuration has nested flux surfaces which enables us to work in flux coordinates: (s, θ, ζ) are the coordinates of our flux coordinate system, where $s \in [0, 1]$ is the normalized flux surface label, θ is a poloidal angle, and ζ is a toroidal angle. In this coordinate system, we identify $s = 0$ with the magnetic axis and $s = 1$ with the last closed flux surface.

We model the dynamics of energetic particles using the *guiding center equations* [7, 11, 33] which arise from the asymptotic decoupling of the fast gyromotion from the full-orbit equations given by the Lorentz force. Our multi-fidelity MC approach is in principle also applicable to MC estimators based on full orbits of energetic particles. However, simulating such orbits on the desired long α -particle slowing down time [34, 57] remains prohibitively expensive for optimization studies on existing computing architectures. To highlight the immediate relevance of our work for tokamak and stellarator optimization studies, we thus chose to focus on drift orbits. The guiding center equations can be expressed in terms of a four-dimensional system of coupled ordinary differential equations (ODEs) for $(\mathbf{r}, v_{\parallel})$, where $\mathbf{r}$ is the position of the guiding center and $v_{\parallel}$ is the parallel velocity of the energetic particle. In flux coordinates (s, θ, ζ) , the guiding center equations take the form

$$\frac{ds}{dt} = \frac{1}{\Omega\sqrt{g}} \left[-\mu \frac{B_{\zeta}}{B} \frac{\partial B}{\partial \theta} + \mu \frac{B_{\theta}}{B} \frac{\partial B}{\partial \zeta} + \frac{v_{\parallel}^2}{B} \frac{\partial B_{\zeta}}{\partial \theta} - \frac{v_{\parallel}^2 B_{\zeta}}{B^2} \frac{\partial B}{\partial \theta} - \frac{v_{\parallel}^2}{B} \frac{\partial B_{\theta}}{\partial \zeta} + \frac{v_{\parallel}^2 B_{\theta}}{B^2} \frac{\partial B}{\partial \zeta} \right] \quad (1)$$

$$\frac{d\theta}{dt} = \frac{v_{\parallel} B^{\theta}}{B} + \frac{1}{\Omega\sqrt{g}} \left[\mu \frac{B_{\zeta}}{B} \frac{\partial B}{\partial s} - \mu \frac{B_s}{B} \frac{\partial B}{\partial \zeta} - \frac{v_{\parallel}^2}{B} \frac{\partial B_{\zeta}}{\partial s} + \frac{v_{\parallel}^2 B_{\zeta}}{B^2} \frac{\partial B}{\partial s} + \frac{v_{\parallel}^2}{B} \frac{\partial B_s}{\partial \zeta} - \frac{v_{\parallel}^2 B_s}{B^2} \frac{\partial B}{\partial \zeta} \right] \quad (2)$$

$$\frac{d\zeta}{dt} = \frac{v_{\parallel} B^{\zeta}}{B} + \frac{1}{\Omega\sqrt{g}} \left[-\mu \frac{B_{\theta}}{B} \frac{\partial B}{\partial s} + \mu \frac{B_s}{B} \frac{\partial B}{\partial \theta} + \frac{v_{\parallel}^2}{B} \frac{\partial B_{\theta}}{\partial s} - \frac{v_{\parallel}^2 B_{\theta}}{B^2} \frac{\partial B}{\partial s} - \frac{v_{\parallel}^2}{B} \frac{\partial B_s}{\partial \theta} + \frac{v_{\parallel}^2 B_s}{B^2} \frac{\partial B}{\partial \theta} \right] \quad (3)$$

$$\begin{aligned} \frac{dv_{\parallel}}{dt} = & -\frac{\mu v_{\parallel}}{B\Omega\sqrt{g}} \left[\frac{\partial B}{\partial s} \frac{\partial B_{\zeta}}{\partial \theta} - \frac{\partial B}{\partial s} \frac{\partial B_{\theta}}{\partial \zeta} - \frac{\partial B}{\partial \theta} \frac{\partial B_{\zeta}}{\partial s} \right. \\ & \left. + \frac{\partial B}{\partial \theta} \frac{\partial B_s}{\partial \zeta} + \frac{\partial B}{\partial \zeta} \frac{\partial B_{\theta}}{\partial s} - \frac{\partial B}{\partial \zeta} \frac{\partial B_s}{\partial \theta} \right] \\ & - \frac{\mu}{B} \left(B^{\theta} \frac{\partial B}{\partial \theta} + B^{\zeta} \frac{\partial B}{\partial \zeta} \right) \end{aligned} \quad (4)$$

in the domain $U \times \mathbb{R}$ where $U \subset \mathbb{R}^3$ is the stellarator volume. In flux coordinates, U is represented by $[0, 1] \times [0, 2\pi] \times [0, 2\pi]$, where $\{s = 0\} \times [0, 2\pi] \times [0, 2\pi]$ is the simple closed curve corresponding to the magnetic axis. The terms $(B_s, B_{\theta}, B_{\zeta})$ and (B^{θ}, B^{ζ}) are the covariant and contravariant components of the magnetic field $\mathbf{B}(\mathbf{r})$, $B = |\mathbf{B}(\mathbf{r})|$, and $\sqrt{g} = \det \left(\frac{\partial \mathbf{r}}{\partial (s, \theta, \zeta)} \right)$ is the Jacobian of the transformation from flux coordinates to Cartesian coordinates. Additionally, $\mu = v_{\perp}^2 / 2B$ is the magnetic moment, which is conserved for individual particles. For each particle it is chosen so that the total initial kinetic energy matches that of a 3.5 MeV particle. Moreover, $\Omega = 2eB_0/m$ is a characteristic gyrofrequency for the alpha particles within the stellarator. Here B_0 represents a characteristic field strength, which we take to be the field strength at $s = 0, \theta = 0, \zeta = 0$. Given initial conditions, we solve equations (1)–(4) up to T_{final} , and consider a particle confined if $s(t) < 1$ for $t < T_{\text{final}}$ and lost otherwise.

Note that for stellarators, the functions $B_s, B_{\theta}, B_{\zeta}, B^{\theta}, B^{\zeta}, B$, and $\sqrt{g}$ appearing on the right-hand side of (1)–(4) are often only accessible as numerical outputs of a 3D field solver (vacuum or finite pressure magnetohydrodynamics (MHD) solve) [28, 35, 53]. They are typically represented by a double Fourier representation in (θ, ζ) with an interpolant or spline in s for the coefficients, and are accessed in the form:

$$\begin{aligned} f(s, \theta, \zeta) &= \sum_{m,n} f_{mn}(s) \cos(m\theta - n\zeta) \quad \text{or} \\ f(s, \theta, \zeta) &= \sum_{m,n} f_{mn}(s) \sin(m\theta - n\zeta). \end{aligned} \quad (5)$$

2.2. High expense of simulating energetic particle dynamics

Considering equations (1)–(4), one may at first wonder why MC estimators need to be improved. Indeed, equations (1)–(4) ‘only’ is a four-dimensional system of coupled ODEs, and since the equations are not modified by the presence of the other energetic particles considered in the MC analysis, the calculations are embarrassingly parallelizable. It would be natural to conclude that the slow convergence of MC estimators can be easily offset by the relative ease of brute-force computations. We briefly review here why this is not the case, and why direct computations of energetic particle losses remain too expensive for many design optimization applications. The reasons are intrinsically tied to the nature of the motion of energetic particles in non-axisymmetric stellarator magnetic fields, and to the vastly different time scales of interest.

As we discussed previously, energetic particle confinement analyses require the simulation of their dynamics for a time at

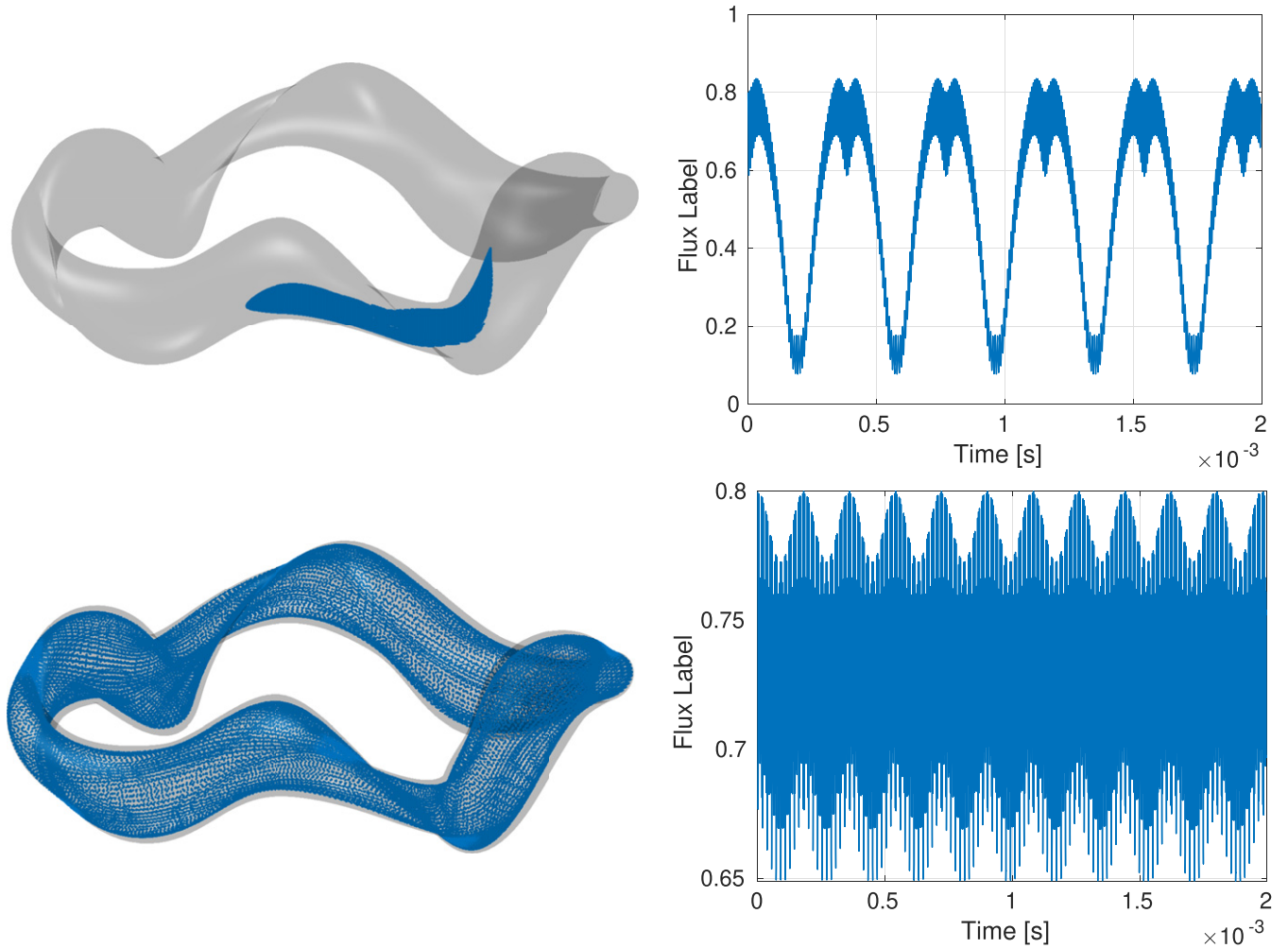


Figure 1. 3D trajectories and flux label vs time for trapped (top) and passing (bottom) particles launched from the same location. The flux label for trapped and passing particles exhibit significantly different temporal and spatial scales.

least as long as their thermalization time [4, 32], after which their orbits become comparable to those of the background thermal ions. We note that thermalization is due to collisions, which are absent in collisionless guiding center models for the study of alpha particle confinement in stellarators, such as the one we consider here. However, even for collisionless models, thermalization time is still often used as a characteristic time scale of interest to evaluate the confinement properties of a given magnetic configuration. This energetic particle thermalization time, corresponding to tens of thousands of toroidal transits for passing particles [36], and thousands of bounces for trapped particles, is significantly longer than the characteristic time scales of their single-particle dynamics. Because of the fast cross-field drift of the energetic particles and the high sensitivity of the orbits to local variations of the magnetic field, as illustrated in figure 1, the motion on these short time scales needs to be resolved with high accuracy. One is therefore faced with an intrinsically multi-scale problem, which is made even more difficult by the chaotic nature of some of the particle trajectories [1]. Numerically, this means that short time steps must be chosen when integrating equations (1)–(4) to accurately resolve the dynamics on the short time scales. At

the same time, high order integration schemes must be implemented, in order to prevent the numerical error from accumulating on the long thermalization time scale. The design of efficient and inexpensive numerical schemes satisfying this stringent criteria remains an open problem in computational science. Progress has been made for certain classes of multi-scale problems [19, 20, 61], but the potential of these methods to speed-up guiding center calculations for energetic particles has not been proven. To this day, integrating equations (1)–(4) with high accuracy on the long thermalization time scale for a single energetic particle remains an intrinsically expensive problem.

3. Estimating energetic particle losses with standard Monte Carlo

In this section, we first explain how we mathematically treat the uncertainty concerning the initial conditions required to compute the single particle motion according to equations (1)–(4), corresponding to the uncertainty in the location and velocity of the energetic particles at birth. We then present three different quantities we consider in this article to characterize energetic particle confinement. The first quan-

tivity is the energetic particle loss fraction, which is the standard figure of merit in optimization studies. The second and third quantities are the mean loss time and the mean flux surface position, which we motivate physically below. Finally, we briefly review the direct approach based on the standard MC method for estimating these quantities.

3.1. Uncertain initial conditions for guiding center dynamics

The energetic particles we consider are alpha particles, born as byproducts of deuterium–tritium fusion reactions. There is intrinsic uncertainty in where and in what direction the particles are emitted, as well as in their kinetic energy at birth. This manifests as uncertainty in the initial conditions needed to solve the guiding center equations (1)–(4).

In this work, we follow the standard practice of assuming that all alpha particles are born with the same kinetic energy of 3.5 MeV [1, 4, 27]. This is a simplification of the actual alpha particle birth process. For accurate alpha particle confinement results, one should in principle account for the energy spread in the alpha birth energy, due to the kinematics of the collision process [2, 8, 25, 62]. As we make this simplifying assumption, only two sources of uncertainty remain: the location of birth of alpha particles, and the velocity direction at birth. We model this uncertainty by imposing probability distributions on the initial position $\mathbf{R}_0$ and initial parallel velocity V_0 . Specifically, we model alpha particles as being born uniformly in the stellarator volume U :

$$\mathbf{R}_0 \sim \text{Uniform}(U). \quad (6)$$

We choose this model for its simplicity. It is physically relevant for flat density and temperature profiles, with a steep pedestal at the edge. For standard density and temperature profiles peaked on axis, our model overestimates the production of alpha particles at large radii. Consequently, the fraction of lost alpha particles in our setup is likely higher than those sought in practice. The method based on MFMC that we present below can be adapted to any choice of probability distribution for $\mathbf{R}_0$, and we will discuss later in this article the extent to which the numerical results we present could be affected by the choice of distribution. For the emission direction, we use the convention that particles are born isotropically in velocity space [1, 4]:

$$V_0 \sim \text{Uniform}(-V_{\max}, V_{\max}), \quad (7)$$

where $V_{\max}$ is the velocity of a particle with a kinetic energy of 3.5 MeV.

3.2. Metrics of energetic particle confinement

In order to estimate energetic particle losses, we must propagate the initial uncertainty through the guiding center dynamics to classify which initial conditions lead to lost particles. Let $\mathbb{E}$ denote the expectation under $\mathbf{R}_0$ and V_0 . Particle loss is thus measured by $\mathbb{E}[F^{\text{lost}}(\mathbf{R}_0, V_0)]$ where

$$F^{\text{lost}}(\mathbf{R}_0, V_0) := \begin{cases} 1, & \inf\{t : s(t; \mathbf{R}_0, V_0) = 1\} < T_{\text{final}} \\ 0, & \text{otherwise} \end{cases} \quad (8)$$

and $s(t; \mathbf{R}_0, V_0)$ comes from solving equations (1)–(4) using the initial conditions $(\mathbf{R}_0, V_0)$. Details on how $\mathbf{R}_0$ is converted to flux coordinates are discussed later in section 5. Since the dynamics are deterministic, we can interpret $F^{\text{lost}}(\mathbf{R}_0, V_0)$ as a Bernoulli random variable which classifies whether or not a particle crosses the last closed flux surface before $t = T_{\text{final}}$, based on where and in what direction it is born.

F^{lost} often is the key figure of merit considered in alpha particle confinement studies [1, 4, 27, 31, 34, 36, 37]. It corresponds to a discrete random variable. To demonstrate the applicability of our method to figures of merit corresponding to continuous random variables, we consider two other metrics which may be used for design optimization, and provide insights on a stellarator's efficacy in confining energetic particles. The first quantity of interest is the average time of loss of energetic particles. Its value for design optimization and performance analysis can be explained as follows. First, prompt losses, corresponding to alpha particles lost after a few transits or bounces, before they have significantly slowed down, lead to more damage to the first wall than slower losses. Second, alpha particles need to be confined for long enough that they heat the plasma as they slow down. Given two magnetic configurations with the same loss fraction after an alpha particle slowing down time, the more desirable one is the one with the longer mean time of loss. This figure of merit is also relevant for assessing the validity of the assumptions and models used for the analysis. If the order of magnitude of the mean time of loss is comparable or only slightly smaller than the alpha particle slowing down time, a large number of alpha particles remain confined long enough to be thermalized. Results obtained by simulating mono-energetic collisionless guiding center trajectories, as often done [1, 4, 27] and as we do for illustrative purposes in this work, should then be confirmed with codes including more detailed physics, and collisions in particular [18, 49]. As we consider this figure of merit, we however have to deal with a mathematical difficulty, associated with the fact that for fully confined alpha particles, the time of loss is infinity. We circumvent this difficulty by defining a modified loss time, such that all particles are considered 'lost' by T_{final} :

$$F^{\text{time}}(\mathbf{R}_0, V_0) := \min(\inf\{t : s(t; \mathbf{R}_0, V_0) = 1\}, T_{\text{final}}). \quad (9)$$

Our second physical quantity of interest associated with a continuous random variable is the mean flux surface location of energetic particles. This quantity is not directly related to alpha particle loss, but is a complementary measure of the quality of energetic particle confinement. It provides physical insights on why certain configurations have lower losses, as well as information on the location of alpha power deposition. We will thus consider the quantity:

$$F^{\text{flux}}(t, \mathbf{R}_0, V_0) := s(t; \mathbf{R}_0, V_0). \quad (10)$$

Therefore the *mean modified loss time* $\mathbb{E}[F^{\text{time}}(\mathbf{R}_0, V_0)]$ and *mean normalized flux surface label* $\mathbb{E}[F^{\text{flux}}(t, \mathbf{R}_0, V_0)]$ serve as alternative statistical metrics for a stellarator's effectiveness in confining energetic particles in our study. For the remainder of this article, we will take F to represent either F^{lost} , F^{time} , or F^{flux} at some fixed time t .

3.3. Uncertainty propagation using Monte Carlo

Now that we have identified energetic particle loss, along with two other metrics for energetic particle confinement, as an expectation of a function $F(\mathbf{R}_0, V_0)$, it remains to accurately estimate that expectation. A direct method for this is the MC estimator, whereby we draw p i.i.d. samples $\{(\mathbf{R}_0^{(i)}, V_0^{(i)})\}_{i=1}^p$ from (6) and (7) and then approximate $\mathbb{E}[F]$ by the unbiased estimator

$$\hat{F}_{\text{MC},p} := \frac{1}{p} \sum_{i=1}^p F(\mathbf{R}_0^{(i)}, V_0^{(i)}). \quad (11)$$

This standard MC approach is summarized in method 3.1. Recall that the variance of a random variable F is defined by $\text{Var}(F) = \mathbb{E}[(F - \mathbb{E}[F])^2]$. Since the MC estimator has variance $\text{Var}(F)/p$, it converges at the slow rate of $1/\sqrt{p}$ with respect to the root mean square error (RMSE). In order to compute the MC estimator, we need to solve the guiding center equations (1)–(4) a total of p times and then evaluate F using those trajectories. Due to the high computational cost of simulating guiding center trajectories of energetic particles over the appropriate time scale, as explained in section 2.2, using the MC estimator can rapidly become computationally intractable for large p .

The large cost required to produce an accurate MC estimator motivates us to aim for an estimator with lower variance via variance reduction. Such a procedure would lower the constant in front of $1/\sqrt{p}$, thereby yielding a more accurate estimator for the same computational effort.

Method 3.1. Monte Carlo estimator

- Draw p i.i.d. samples $(\mathbf{R}_0^{(1)}, V_0^{(1)}), (\mathbf{R}_0^{(2)}, V_0^{(2)}), \dots, (\mathbf{R}_0^{(p)}, V_0^{(p)})$ from (6) and (7).
- For each $i = 1, \dots, p$, solve the guiding center equations (1)–(4) using expensive numerical integration. Use the computed trajectories to evaluate $F(\mathbf{R}_0^{(i)}, V_0^{(i)})$ for $i = 1, \dots, p$.
- Estimate $\mathbb{E}[F]$ by

$$\hat{F}_{\text{MC},p} = \frac{1}{p} \sum_{i=1}^p F(\mathbf{R}_0^{(i)}, V_0^{(i)}).$$

4. Estimating energetic particle confinement with multifidelity Monte Carlo

Here, we first describe how one can leverage a surrogate model to construct a control variate based multifidelity estimator with smaller variance than the standard MC estimator. We then discuss an approach to construct highly correlated surrogate models for the analysis of energetic particle confinement.

4.1. Variance reduction with multifidelity Monte Carlo

Alongside the high-fidelity model F , which requires expensive numerical integration to evaluate, suppose we also have a low-fidelity, or *surrogate*, model G . In order to leverage G for variance reduction we utilize a control variate based MFMC method [48]. The MFMC estimator with computational budget p is the unbiased estimator

$$\hat{F}_{\text{MFMC},p} := \hat{F}_{\text{MC},n} + \alpha (\hat{G}_{\text{MC},m} - \hat{G}_{\text{MC},n}), \quad (12)$$

where α is a constant, n is the number of high-fidelity samples, and m is the number of low-fidelity samples. To minimize the variance of the MFMC estimator, α, n, m can be optimally chosen, and those choices satisfy

$$\begin{aligned} \alpha &= \rho(F, G) \sqrt{\frac{\text{Var}(F)}{\text{Var}(G)}}, \\ p &= n + \frac{m}{w}, \\ m &= n \sqrt{\frac{w \rho(F, G)^2}{1 - \rho(F, G)^2}} \end{aligned} \quad (13)$$

where $\rho(F, G)$ is the Pearson's correlation coefficient between the high-fidelity F and the surrogate G , which is defined by

$$\begin{aligned} \rho(F, G) &:= \frac{\text{Cov}(F, G)}{\sqrt{\text{Var}(F)} \sqrt{\text{Var}(G)}} \\ &= \frac{\mathbb{E}[(F - \mathbb{E}[F])(G - \mathbb{E}[G])]}{\sqrt{\text{Var}(F)} \sqrt{\text{Var}(G)}} \end{aligned} \quad (14)$$

and w is the ratio of cost of evaluating F to the cost of evaluating G . A practical implementation of this method is summarized in method 4.1. The variance of the resulting optimal MFMC estimator is given by

$$\text{Var}(\hat{F}_{\text{MFMC},p}) = \left[\left(\sqrt{1 - \rho(F, G)^2} + \sqrt{\frac{\rho(F, G)^2}{w}} \right)^2 \right] \times \text{Var}(\hat{F}_{\text{MC},p}). \quad (15)$$

In order to maximize the amount of variance reduction MFMC provides compared to MC, the bracketed term in (15) should be as small as possible. This means that we want the surrogate G to be both highly correlated to F , as well as much cheaper to evaluate than F . Although the MFMC estimator still converges at the same slow rate of $1/\sqrt{p}$ as the standard MC estimator, it can be more accurate with the same computational budget provided a sufficiently correlated and inexpensive surrogate.

Note that once one is given a surrogate G , the amount of speedup provided by MFMC is constant. That is, the construction of G is done offline, separate from the online phase of estimation. No matter how expensive the construction of G is (which could require many samples of F for instance), once we have that surrogate, the speedup from MFMC is guaranteed no matter how large p is taken to be; see [46] for an extension of MFMC that takes offline costs into account.

In the MFMC literature, the cost of sampling the initial uncertainty is not taken into account, as it is typically assumed to be orders of magnitude cheaper than everything else being considered. In practice, if sampling the input uncertainty requires non-trivial cost, then a better definition of the cost ratio w would be required as the current definition only accounts for the cost of propagating the uncertainty forward. While we do not account for this cost of input uncertainty to make a fair budget comparison in our MFMC experiments, we mention here the limits on implementation imposed by how the input uncertainty is sampled.

Method 4.1. Multifidelity Monte Carlo estimator

- (a) Given p , compute n and m as in (13)¹. Draw m i.i.d. samples $(\mathbf{R}_0^{(1)}, V_0^{(1)}), (\mathbf{R}_0^{(2)}, V_0^{(2)}), \dots, (\mathbf{R}_0^{(m)}, V_0^{(m)})$ from (6) and (7).
- (b) For each $i = 1, \dots, n$, solve the guiding center equations (1)–(4) using expensive numerical integration. Use the computed trajectories to evaluate $F(\mathbf{R}_0^{(i)}, V_0^{(i)})$ for $i = 1, \dots, n$. Evaluate $G(\mathbf{R}_0^{(i)}, V_0^{(i)})$ for $i = 1, \dots, m$.
- (c) Set α as in (13), replacing $\text{Var}(F)$ and $\text{Var}(G)$ with sample variances computed using the evaluations from (b).
- (d) Estimate $\mathbb{E}[F]$ by

$$\hat{F}_{\text{MFMC},p} = \hat{F}_{\text{MC},n} + \alpha (\hat{G}_{\text{MC},m} - \hat{G}_{\text{MC},n}).$$

Use an estimate $\hat{\rho}$ of $\rho(F, G)$. If an estimate is not readily available, one can be found easily using samples.

4.2. Choice of surrogate model

We now detail our choice of a data-fit surrogate model G designed to maximize correlation with F for our estimation of energetic particle confinement. We rely on data-fit surrogates because traditional sources of surrogates are not appropriate for energetic particle dynamics. For example, there is no natural hierarchy of simplified physics models to leverage [47], since the guiding center equations are already a simplification of the computationally intractable full-orbit dynamics, and a ‘beads on the wire’ model for the particle trajectories, which would omit all finite-gyroradius corrections, as in the

kinetic MHD model [10, 12, 21, 24, 30, 50, 51, 55], would not lead to good correlation since it neglects the large cross-field drifts of energetic particles. Similarly, we have empirically found that another usually powerful source of surrogates, arising from successively coarsening the grid [13], is not satisfactory either for our application. That is because the time step of the numerical ODE integrator is tightly constrained by the requirement to resolve the shortest characteristic time scales of the motion of both trapped and passing particles. When this time step is significantly increased, accuracy is significantly reduced, and correlation rapidly decreases. Lastly,

since the problem is transport dominated, popular surrogates from traditional model reduction, such as proper orthogonal decomposition or reduced basis methods, are a poor choice [44].

For a data-fit surrogate, we consider a linear basis function model of the form:

$$G^n(\mathbf{R}_0, V_0; \mathbf{P}^n) = \sum_{k=0}^n P_k^n \phi_k(\mathbf{R}_0, V_0), \quad (16)$$

where $\mathbf{P}^n \in \mathbb{R}^{n+1}$ and $\{\phi_k\}_{k=0}^n$ is any set of basis functions. Using this family of surrogates, we want to find the set of parameters $\mathbf{P}^n$ which maximizes the correlation with our high-fidelity model. This amounts to solving

$$\underset{\mathbf{P}^n}{\text{maximize}} \rho(F, G^n(\cdot; \mathbf{P}^n)). \quad (17)$$

However, a caveat of this formulation is that the optimization problem (17) has infinitely many maxima through scaling, since our model is linear, $\rho(F, G^n(\cdot; c\mathbf{P}^n)) = \rho(F, cG^n(\cdot; \mathbf{P}^n)) = \rho(F, G^n(\cdot; \mathbf{P}^n))$ for any $c > 0$.

It thus is better to consider $\mathbf{P}_*^n$, which solves the following optimization problem (18):

$$\underset{\mathbf{P}^n}{\text{minimize}} \mathbb{E}[(F - G^n(\cdot; \mathbf{P}^n))^2]. \quad (18)$$

Due to the linear nature of our family of surrogate models, the unique solution $\mathbf{P}_*^n$ of (18) is also one of the non-unique solutions for (17).

In practice, evaluating the objective function in (18) is as challenging as the original problem motivating our work, since it involves evaluating expectations of F . Finding the exact minimum $\mathbf{P}_*^n$ is therefore a computationally intractable problem, regardless of the choice of basis functions $\{\phi_k\}_{k=0}^n$. We now explain how we address this difficulty and construct a satisfactory approximation of $\mathbf{P}_*^n$ for the particular choice of basis functions $\{\phi_k\}_{k=0}^n$ we make for this study, namely piece-wise linear functions. The idea is to temporarily change the interpretation of F , and view it as a deterministic map from the four-dimensional space $(s, \theta, \zeta, v_{\parallel})$ to $\mathbb{R}$. First, in $(s, \theta, \zeta, v_{\parallel})$ coordinates, we construct a regular mesh on $[0, 1] \times [0, 2\pi] \times [0, \pi/2] \times [-V_{\max}, V_{\max}]$, and evaluate F on that mesh. Then we specifically choose our basis functions $\{\phi_k\}$ to be the hat functions on the mesh, so that our family of linear basis functions G^n is the family of piece-wise linear functions on the mesh. If F is a continuous map, and we determine $\tilde{\mathbf{P}}^n$ so that $G^n(\cdot; \tilde{\mathbf{P}}^n)$ is the corresponding interpolant of F on the mesh, then $\tilde{\mathbf{P}}^n$ will be a close approximation to $\mathbf{P}_*^n$. The reasoning behind this is that if F is continuous, as we refine the mesh then $G^n(\cdot; \tilde{\mathbf{P}}^n)$, where $\tilde{\mathbf{P}}^n$ solves (18) on successively finer meshes, will converge to F . Since $G^n(\cdot; \tilde{\mathbf{P}}^n)$ also converges to F as the mesh is refined, we expect $\tilde{\mathbf{P}}^n$ to be a strong approximation to $\mathbf{P}_*^n$, converging as the mesh refines. Although we cannot measure how close $\tilde{\mathbf{P}}^n$ is to $\mathbf{P}_*^n$, in practice we have observed that $G^n(\cdot; \tilde{\mathbf{P}}^n)$ is favorably correlated with F , even on a relatively coarse mesh.

5. Numerical results

We now conduct numerical tests to determine the level of variance reduction and the speedup one can obtain in practice with our MFMC approach in a realistic stellarator configuration. We first provide details for the setup of our numerical experiments. Then we demonstrate the effectiveness of MFMC for confinement studies done at high resolution, with a time step equal to the characteristic gyroperiod of alpha particles in our magnetic equilibrium. Since the time step is small, we cannot track particle orbits beyond $T_{\text{final}} = 0.1$ ms with the computing resources available to us. This is well below the typical alpha particle slowing down time, which is approximately 200 ms. In terms of confinement, the results of this analysis are mostly dominated by prompt particle losses. In the final part of this section, we decrease the time resolution of our high-fidelity numerical integrator, in order to compute particle trajectories up to $T_{\text{final}} = 20$ ms. While this remains below the typical alpha particle slowing down time, it is long enough to include more varied particle loss channels, and thus provide realistic estimates of the effectiveness of MFMC for alpha particle confinement studies.

5.1. Numerical setup

For our high-fidelity solver, we use a Runge–Kutta 4 (RK4) time integrator for the guiding center equations. The terms $B_s, B_{\theta}, B_{\zeta}, B^{\theta}, B^{\zeta}, \sqrt{g}$ are provided as numerical outputs of a VMEC solve [28]. These functions are given by the double Fourier representation (5) using 128 pairs (m, n) in the (θ, ζ) variables, where only cosine or sine series arise due to stellarator symmetry. For each (m, n) pair, $f_{mn}(s)$ is represented using a smoothing spline over 101 equispaced points in $[0, 1]$. We perform all tests on a Wistell-A vacuum magnetic configuration. Wistell-A is a quasihelically symmetric stellarator with four-fold stellarator symmetry [5]. For the time step size in our RK4 integrator, we choose $\Delta\tau = 2\pi/\Omega$ which is the characteristic gyroperiod of alpha particles in the Wistell-A magnetic field. This step size allows us to accurately resolve the guiding center dynamics, but limits us in how large T_{final} can be. We note that the gyroperiod is in principle not a characteristic time scale of guiding center motion. It may thus seem more natural to consider the transit time as the characteristic time scale, i.e. the time for an alpha particle to transit a field period. However, as our results below will show, for the Wistell-A configuration we study in this work, accurate results require the time step to be of the order of the gyroperiod.

We shall refer to our low-fidelity models as $G^{\text{lost}}, G^{\text{time}}$, and G^{flux} for the quantities of interest $F^{\text{lost}}, F^{\text{time}}$, and F^{flux} , respectively. To track the flux label in time, we specify N_{flux} time points $t_k = (\text{time step size}) \times N_{\text{skip}} k$ for $k = 1, \dots, N_{\text{flux}}$. Note that N_{flux} and N_{skip} change depending on the time step size, and are adjusted in order to get to the desired T_{final} . The reasoning behind recording the flux label every N_{skip} steps is that for any practical value of T_{final} , tracking the trajectory at every step requires an unreasonably large number of specified time points. For different time step sizes, we choose N_{skip} in an attempt to make the specified time points similar to one another. For example, if the time step size is halved, then N_{skip}

Table 1. Costs for the high-fidelity model in seconds, cost ratios, and N_{flux} and N_{skip} used for various time step sizes and integration lengths. We see that for $T_{\text{final}} = 20$ ms our low-fidelity models are several millions of times faster than the high-fidelity evaluation. Only the values for $(\Delta\tau, 0.1)$ were measured, all other columns result from scaling.

(Time step size, T_{final} (ms))	$(\Delta\tau, 0.1)$	$(3\Delta\tau, 20)$	$(5\Delta\tau, 20)$	$(7\Delta\tau, 20)$	$(10\Delta\tau, 20)$
High-fidelity cost (s)	20	1333	800	571	400
Cost ratio, G^{lost} and G^{time}	2×10^6	1.33×10^8	8×10^7	5.71×10^7	4×10^7
Cost ratio, G^{flux}	6×10^4	4×10^6	3.6×10^6	1.7×10^6	1.2×10^6
N_{flux}	684	760	684	978	684
N_{skip}	100	300	200	100	100

Table 2. Correlations between the high-fidelity and low-fidelity models for each quantity of interest using different time step sizes and T_{final} . For the normalized flux label we report both the mean correlation over time points as well as the correlation at T_{final} . Correlations are measured using 10^4 samples.

(Time step size, T_{final} (ms))	$(\Delta\tau, 0.1)$	$(3\Delta\tau, 20)$	$(5\Delta\tau, 20)$	$(7\Delta\tau, 20)$	$(10\Delta\tau, 20)$
Classification of lost particles	0.8776	0.9076	0.8996	0.8762	0.8460
Modified loss time	0.9444	0.9299	0.9426	0.9436	0.9247
Normalized flux label (average)	0.9846	0.9538	0.9517	0.9403	0.9111
Normalized flux label (final time)	0.9761	0.9486	0.9452	0.9270	0.8833

would be doubled. If the time step size is divided by three, then N_{skip} would be tripled, and N_{flux} adjusted as well. Thus, for an input $(\mathbf{R}_0, V_0)$, our high-fidelity flux label model outputs $\{F^{\text{flux}}(t_k, \mathbf{R}_0, V_0)\}_{k=1}^{N_{\text{flux}}}$. Our low-fidelity flux label model is then $\{G_k^{\text{flux}}(\mathbf{R}_0, V_0)\}_{k=1}^{N_{\text{flux}}}$ where G_k^{flux} is the piece-wise linear interpolant to $F^{\text{flux}}(t_k, \cdot)$ over a regular mesh in $(s, \theta, \zeta, v_{\parallel})$ as detailed in section 4.2. Similarly G^{time} is the piece-wise linear interpolant to F^{time} over the same mesh. For all interpolants we use 25 points in each dimension. For G^{lost} we specify an optimal threshold using G^{time} . Namely, we set

$$G^{\text{lost}}(\mathbf{R}_0, V_0) := \begin{cases} 1, & G^{\text{time}}(\mathbf{R}_0, V_0) \leq T_{\text{threshold}} \\ 0, & \text{otherwise,} \end{cases} \quad (19)$$

where $T_{\text{threshold}}$ is chosen to minimize the expected misclassification $\mathbb{E}[(F^{\text{lost}} - G^{\text{lost}})^2]$. To compute $T_{\text{threshold}}$, we use a sample average approximation of the expected misclassification using 10^4 samples, and then choose $T_{\text{threshold}}$ to be the minimizer of that sample average approximation evaluated at 10^4 equispaced points between 0 and T_{final} . Note that in our numerical experiments, the time $T_{\text{threshold}}$ is found to be comparable to T_{final} .

To sample $\mathbf{R}_0 \sim U$ in magnetic coordinates, we take advantage of the four-fold stellarator symmetry in Wistell-A and sample (s_0, θ_0, ζ_0) from the probability density proportional to $\sqrt{g}$ over $D = (0, 1) \times (0, 2\pi) \times (0, \pi/2)$. This is done by numerically integrating $\sqrt{g}$ over D to compute the normalization constant, and then performing rejection sampling with a uniform proposal. This step is relatively fast and requires on average only three rejections before a successful draw from the target distribution.

The costs of evaluating F^{lost} , F^{time} , and F^{flux} are all the same: the time required to numerically integrate to T_{final} using a given time step size. The cost per time step (which is roughly $4 \times$

the cost per right-hand side evaluation (1)–(4) in our implementation) is on the order of 10^{-3} s. For the values of T_{final} considered in this work, the total time is on the order of tens of minutes. For longer integration to $T_{\text{final}} = 200$ ms, this would be on the order of hours. The measured runtime is reported in table 1. In our experiments, constructing a surrogate G from given training data is on the order of milliseconds and thus very low because only a spline interpolant has to be built. If training data are available from previous simulations, then they can be directly used to construct a surrogate. If training data need to be generated first, then this can incur one-time high costs but more sophisticated surrogate models than the interpolants used here can help to reduce these one-time costs. We discuss this important point in more detail in section 6. The costs of evaluating G^{lost} and G^{time} are the same, but the cost of evaluating G^{flux} is slightly larger since it requires evaluating N_{flux} interpolants. In table 1 we report the costs of evaluating the high-fidelity model, the cost ratios for G^{lost} , G^{time} , and the cost ratios for G^{flux} across multiple time step sizes and values of T_{final} . We note that for $T_{\text{final}} = 20$ ms the low-fidelity models are all several millions of times faster than the high-fidelity models.

In table 2 we report the correlations between our high and low-fidelity models for each of the quantities of interest across multiple time step sizes and values of T_{final} . We measure the correlation coefficient using 10^4 samples.

5.2. High resolution high-fidelity model and short time integration

We numerically compare the MC and MFMC estimators for the loss fraction, mean modified loss time, and mean flux label using our low-fidelity models for a situation in which we have a high resolution high fidelity time integrator, with

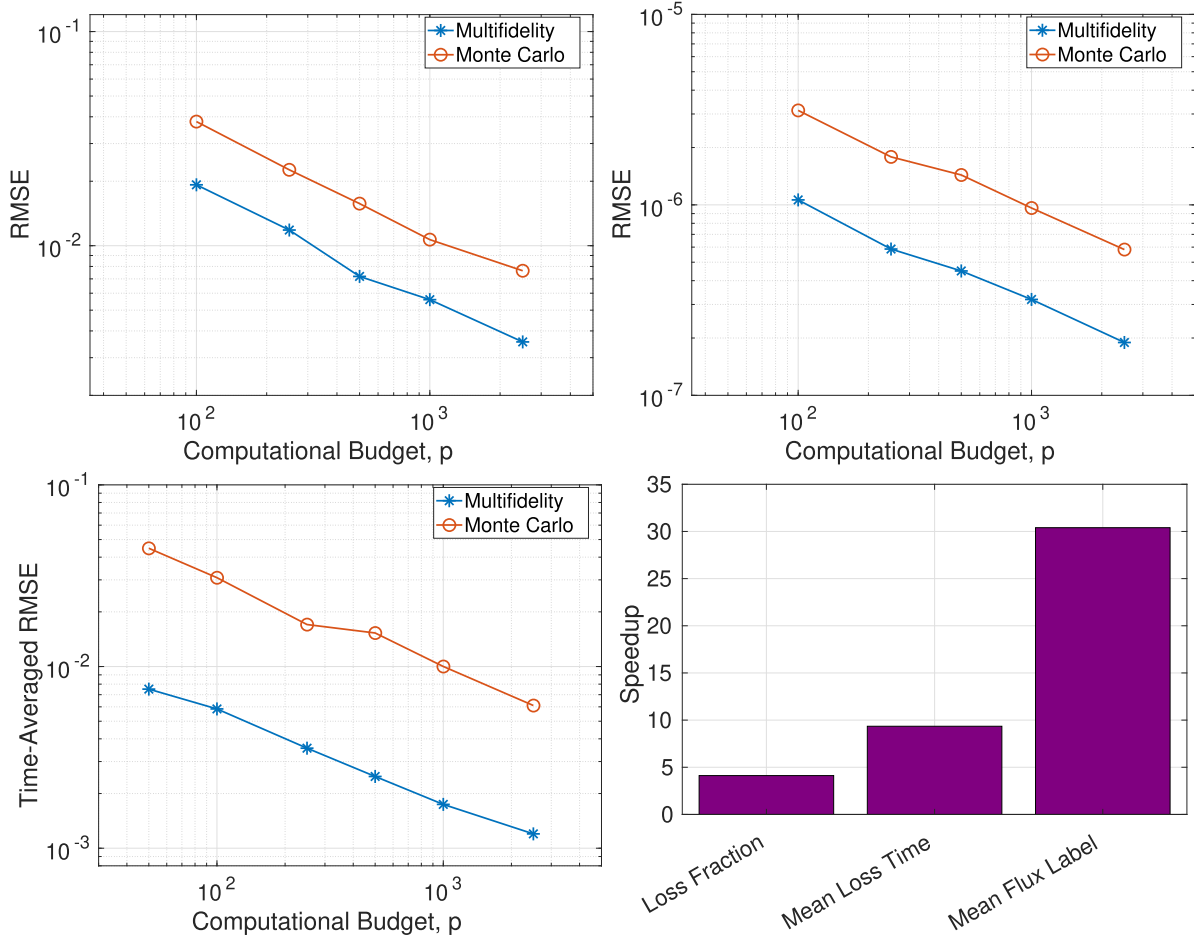


Figure 2. RMSE vs computational budget (seconds) for top left: loss fraction. Top right: mean modified loss time. Bottom left: mean normalized flux label. Bottom right: computational speedup of MFMC compared to MC for each of the quantities of interest. For the mean flux label, we report the median over speedups for each of the N_{flux} specified time points.

which we can only afford to compute trajectories over a relatively short time. Specifically, here we use time step size $\Delta\tau$ and choose T_{final} as 0.1 ms. For computational budgets $p = 100, 250, 500, 1000, 2500$, we measure the RMSE of MC and MFMC by using 100 replicates of each estimator and compare them to a reference solution generated using MC with 10^6 samples.

For a given computational budget p , we would ideally want to use the optimal MFMC estimator with n high-fidelity evaluations and m low-fidelity evaluations satisfying (13). However, in practice the resulting n and m are not necessarily integer. Thus we simply round down and choose

$$n = \left\lfloor \frac{p}{1 + \sqrt{\frac{\rho(F,G)^2}{w(1-\rho(F,G)^2)}}} \right\rfloor, \quad m = (p - n)w$$

and compare this MFMC estimator to the MC estimator with p high-fidelity model evaluations. In the case of the normalized flux label, we average the RMSE over the N_{flux} specified time points.

In figure 2 we see the clear variance reduction using MFMC with our low-fidelity models compared to standard MC. We see that for the same computational budget, for each quantity

of interest, MFMC is able to noticeably outperform MC. The most remarkable improvement is for the mean normalized flux label, where MFMC is nearly 30 times faster than MC. The improvement for the loss fraction is more modest, but MFMC is still four times faster than MC.

In the case of the normalized flux label, each of the N_{flux} specified time points has a different speedup, and in figure 2 we report the median over all such values. The different speedups at each specified time point t_k is due to the fact that the correlation between F^{flux} and G^{flux} decreases over time, which can be seen in figure 3. As a result the measured speedup for computing $\mathbb{E}[F^{\text{flux}}(t_1, \mathbf{R}_0, V_0)]$ is much larger than the measured speedup of, e.g. $\mathbb{E}[F^{\text{flux}}(T_{\text{final}}, \mathbf{R}_0, V_0)]$. Consequently, in utilizing the speedups for different time points we must be careful of outliers. This is the primary reasoning behind assessing the speedup for the whole trajectory by the median of the speedups at each time point, since the median is more robust to outliers than other measures of central tendency, such as the mean.

5.3. Lower resolution high-fidelity model and longer time integration

We now extend our analysis to a longer T_{final} of 20 ms. Since this is 200 times longer than the short time integration case,

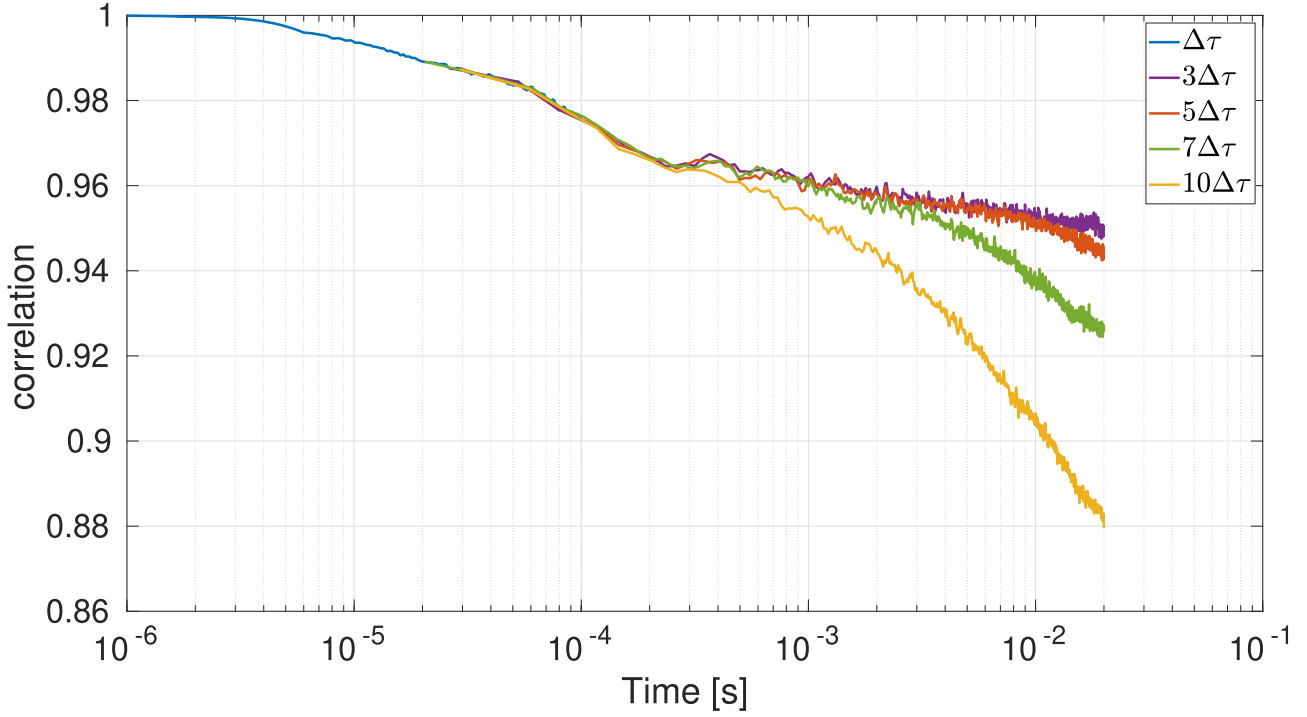


Figure 3. Correlation between F^{flux} and G^{flux} as a function of time, using different time step sizes for F^{flux} (which in turn leads to a different G^{flux} for each step size). We note that all the correlations stay practically the same until around 3×10^{-4} s. After that, the experiment with the lowest resolution high-fidelity model (i.e. $10\Delta\tau$) has a rapid drop off in correlation. The decrease with respect to different time step sizes is expected, with the highest resolution experiment maintaining the highest correlation.

we must increase the step size for the high-fidelity model by a factor of 3, 5, 7, and 10 for computational tractability. Note that changing the step size changes the high-fidelity models F^{lost} , F^{time} , and F^{flux} , which implicitly changes the corresponding low-fidelity models G^{lost} , G^{time} , and G^{flux} .

The results in table 2 indicate that accuracy of the high-fidelity model has a direct impact on the level of correlation between the high-fidelity and the low-fidelity models. The numerical experiments with $3\Delta\tau$ and $5\Delta\tau$ yield similar correlations by the final time, but we see that as the time step of the high-fidelity numerical integrator further increases, and the accuracy of the high-fidelity model thus further decreases, the correlation with the low fidelity model decays quickly. A partial exception to that conclusion applies to the modified loss time, for which the correlation remains steady as we increase the time step of the ODE integrator. For the moment, we do not have an explanation for this surprising observation.

The importance of relying on an accurate high-fidelity model can be seen in figure 3 as well, where we plot the correlation between F^{flux} and G^{flux} as a function of time, for different choices of step size. For short time all step sizes of the numerical ODE integrator provide the same correlation, but as time increases, we see that low resolution in the high-fidelity model leads to rapid loss of correlation. Moreover we notice the general pattern that, uniformly in time, the more resolution one has for the high-fidelity model, the more correlation one obtains with the corresponding low fidelity model.

Since the costs of our low-fidelity models remain the same, it is doubly beneficial for MFMC that correlation between the high-fidelity and low-fidelity models tends to increase as the

resolution, and thus cost, of the high-fidelity solver increases. Indeed, improving the resolution of the high-fidelity model not only yields better correlation to the resulting low-fidelity model, but also a larger cost ratio, which in turn means more variance reduction with MFMC.

Just by looking at tables 1 and 2 and figure 3, we might expect some of the speedup we obtained in our high resolution, short time integration analysis to extend to the current analysis with lower resolution and longer time integration. Although the correlation goes down for longer time, the cost ratio grows rapidly, which in the MFMC setting can offset the effects of decreasing correlation. Furthermore, we may expect that the models with smaller step sizes (e.g. $3\Delta\tau$) will yield better speedup than models with larger step sizes (e.g. $10\Delta\tau$). This is precisely what we now test numerically.

For computational budgets $p = 100, 250, 500, 1000$, we measure the RMSE of MC and MFMC by using 90 replicates for each estimator and compare them to a reference solution generated using MC with 10^5 samples. We rely on fewer replicates in estimating the RMSE and have less resolution in the reference solution because of the significantly increased computational cost for longer time integration of the particle orbits. We balance the costs in the same manner as in the short time integration case, and also assess the flux label speedup using the median, for the same reasons as before.

Our speedup results for the lower resolution, longer time integration experiments are displayed in figure 4. Broadly, we see that we lose some speedup for this larger T_{final} as compared to our short time integration experiment, but still achieve an order of magnitude speedup in estimating the mean flux

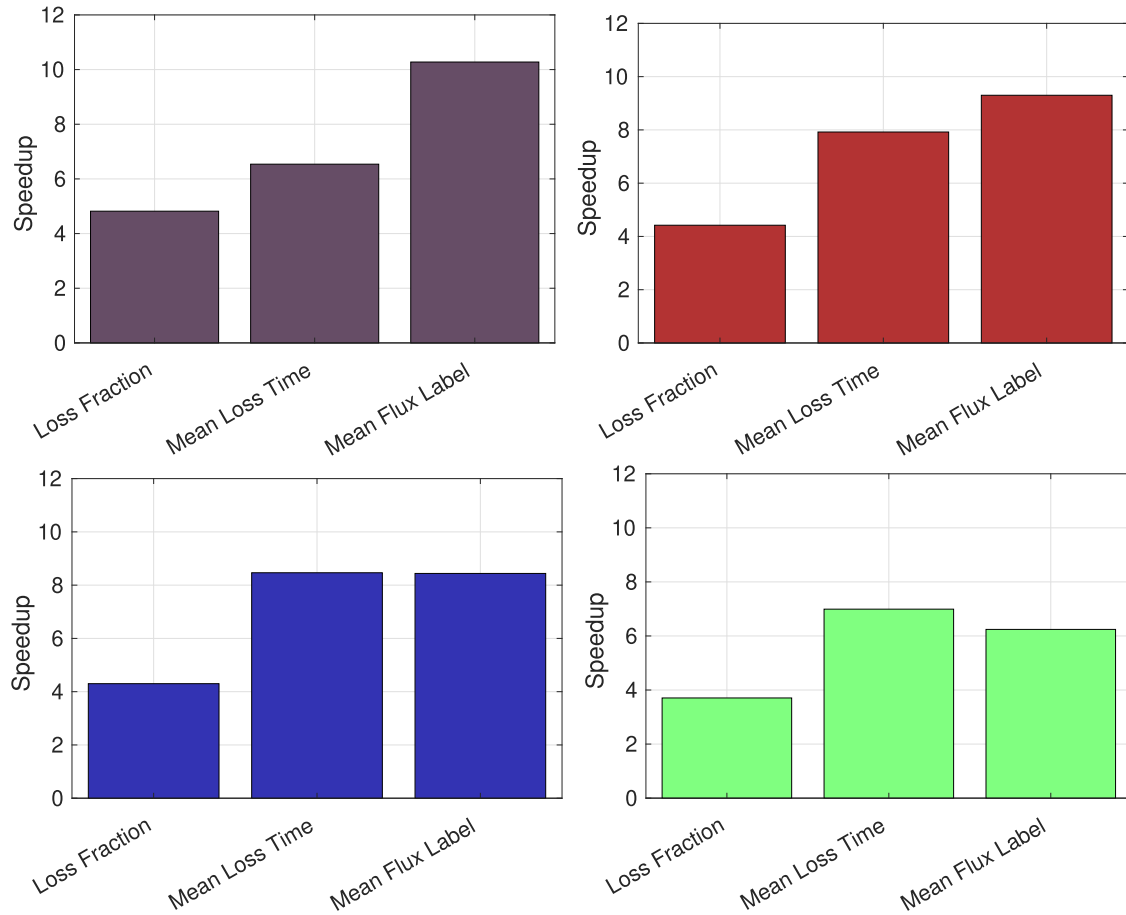


Figure 4. Measured speedup for $T_{\text{final}} = 0.02$ s for different time step sizes of the high-fidelity ODE integrator, using $\Delta\tau = 2\pi/\Omega$. Top left: $3\Delta\tau$. Top right: $5\Delta\tau$. Bottom left: $7\Delta\tau$. Bottom right: $10\Delta\tau$.

label when the step size $3\Delta\tau$ is used. For the other quantities of interest, the speedup is also quite large, the smallest being an almost four times speedup for estimating the loss fraction with the high fidelity integrator with time step $10\Delta\tau$. For all time step sizes, the speedup for estimating the loss fraction is about 4, as we had obtained for the experiment with a short time integration.

Comparing across experiments with different step sizes, we note that we typically get the most speedup for the most accurate, and most expensive, high-fidelity model, with step size $3\Delta\tau$. As the step size increases, we see that the speedup in estimating the loss fraction decreases the slowest, whereas the speedup for estimating the mean flux label decreases the fastest. This is consistent with the results from figure 3, and it means that when we increase the step size, the larger cost ratio does not seem to fully make up for the decreased correlation. The speedup for the mean modified loss time does not appear to depend on the step size in a clear manner, and is between around 6.5 and 8 depending on the step size. These results are also consistent with the correlations reported in table 2, where we did not observe any obvious trend with the step size for this quantity of interest. Note that for the largest step size $10\Delta\tau$, there is more speedup in estimating the mean modified loss time than in estimating the mean flux label.

6. Summary and discussion

We have developed and implemented the first MFMC scheme for estimating energetic particle confinement in stellarator configurations. For the Wistell-A magnetic equilibrium we considered, this has allowed us to gain over an order of magnitude in speedup compared to standard MC for certain cases and certain physical quantities. For other cases and physical quantities for which MFMC acceleration is less efficient, the speedup was still at least a factor of four. This speedup was gained by leveraging efficient data-driven surrogate models built from interpolants.

Although we found that MFMC provided the largest speedup for our highest resolution high fidelity model with particle trajectories which were integrated over a short time scale, we observed that the speedups for estimating the loss fraction and the modified mean loss time remained large and roughly constant as we significantly increased the final time for the integration of the particle trajectories and reduced the resolution of our high-fidelity models.

6.1. Benefits in early vs late design phases

As our results suggest, the acceleration obtained by replacing standard MC with MFMC will depend on the physical

quantity of interest, on the equilibrium magnetic configuration, and on the manner the uncertainty in the initial conditions is introduced. For example, one can mathematically show that speedup for the loss fraction deteriorates as the true loss fraction $\mathbb{E}[F^{\text{lost}}]$ goes to 0. If we had not assumed that alpha particles are born uniformly in the stellarator volume U , but instead had had all of them born on a flux surface close to the magnetic axis, as is sometimes done in confinement studies, the true loss fraction would have been smaller, and the MFMC speedup for this quantity of interest may have been smaller as well. On the other hand, the Wistell-A configuration is already well optimized for alpha particle confinement. If we had considered a magnetic configuration with significantly worse alpha particle confinement, we may have obtained a much higher speedup through our MFMC implementation. In the context of stellarator optimization, the approach we presented in this article may therefore be most useful in the early design optimization phase, when the loss fraction is high. In the later design phases, if the loss fraction has been optimized to become very low, the variance reduction provided by MFMC via control variates becomes ineffective for the estimation of the loss fraction (but not necessarily for other confinement quantities, as we showed in this work). For such situations, variance reduction methods based on, e.g. multi-fidelity importance sampling are more suitable [47].

6.2. Surrogate modeling

There are several open questions regarding the construction of surrogate models for estimating energetic particle confinement in stellarator configurations. In this work, we focused on interpolants because they can be readily constructed with off-the-shelf software packages that are widely used. If training data are available, then constructing an interpolant as surrogate model has negligible computational costs compared to, e.g. a high-fidelity model evaluation. However, if training data are unavailable, then they need to be obtained first, which can incur one-time high costs. One direction of future research is deriving surrogate models with adaptive methods that judiciously explore the parameter space and request new training samples only where necessary based on error indicators, rather than uniformly sampling the parameter space as in this work. For example, surrogate models based on sparse grids have been shown to require little data [29]. Another line of future work is to develop physics based reduced models that require no training [47].

The optimization loop for finding a stellarator design with good confinement properties poses additional challenges but also opportunities. In the early stages of optimization, when the magnetic configuration changes significantly from one optimization iteration to the next, surrogate models built for the previous iterations can become obsolete quickly. However, less accurate and quick to build surrogate models are sufficient because relatively crude estimators of the objective can provide the relevant information for finding descent directions. Physics based reduced models, which require no training, can play a critical role in these early stages of optimization. In later stages of the optimization, surrogate models may be re-used

for multiple iterations because they remain valid as the magnetic equilibrium changes only slightly from one optimization iteration to the next. Thus, it can be worthwhile to invest a larger budget in training a surrogate model to reduce the variance with multi-fidelity methods to obtain accurate estimators of the optimization objective. The development of multi-fidelity methods that trade-off the surrogate construction to exploit the properties of early vs later optimization stages is a challenging but promising research topic. Note in that regard that multi-fidelity methods have been successfully used for design in various engineering disciplines [42, 43, 54], but have not yet been applied to problems involving sensitive chaotic dynamics, as we have studied in this article.

6.3. Scope

We applied our approach to a relatively straightforward setup, but our MFMC framework for estimating energetic particle confinement lends itself to extensions to more detailed physics and design studies. For example, it would be straightforward to replace our numerical ODE integrator with existing high performance solvers for guiding center trajectories, which may also include collisions [18, 49] and wave-particle interactions [56]. Another immediate extension would be to consider more sophisticated models for the initial conditions of the alpha particles, accounting for energy spread at birth, and the fact that for peaked pressure profiles, most alpha particles are born in the core of the device. Our success with MFMC using guiding center orbits also suggests that estimating statistics based on full-orbit trajectories may become computationally tractable. While full-orbit simulations are extremely expensive to perform, if we can find a proper surrogate to leverage, MFMC would enable high accuracy estimation of full-orbit statistics while only requiring a very small amount of full-orbit simulations. These topics are all subjects of ongoing work. Finally, it is important to stress that for a reliable analysis of energetic particle confinement, single-particle codes and theories may not be sufficient. A more self-consistent drift kinetic theory has recently been proposed for tokamaks with magnetic field perturbations [60], and we do not yet know if MFMC is relevant for such theory, and could lead to speed-up.

Acknowledgments

We would like to thank Aaron Bader for providing the VMEC files for the Wistell-A magnetic equilibrium, and for many enlightening conversations on numerical methods for energetic particle confinement studies. We also would like to thank Alain Brizard and Matt Landreman for their comments on an early version of this manuscript, which helped us to improve it. Frederick Law was supported by the Department of Defense (DoD) through the National Defense Science & Engineering Graduate (NDSEG) Fellowship Program and was supported in part by the Research Training Group in Modeling and Simulation funded by the National Science Foundation via Grant RTG/DMS—1646339. Antoine Cerfon was partially supported for this work by the Simons Foundation

under Grant No. 560651, and by the United States National Science Foundation under Grant No. PHY-1820852. Benjamin Peherstorfer was partially supported by AFOSR under award number FA9550-21-1-0222 (Dr Fariba Fahroo) and by NSF under award number CMMI-1761068 and award number DMS-2012250.

ORCID iDs

Antoine Cerfon  <https://orcid.org/0000-0003-3326-4785>

References

- [1] Albert C.G., Kasilov S.V. and Kernbichler W. 2020 Accelerated methods for direct computation of fusion alpha particle losses within, stellarator optimization *J. Plasma Phys.* **86** 815860201
- [2] Appelbe B. and Chittenden J. 2011 The production spectrum in fusion plasmas *Plasma Phys. Control. Fusion* **53** 045002
- [3] Bader A. et al 2021 Modeling of energetic particle transport in optimized stellarators *Nucl. Fusion* **61** 116060
- [4] Bader A., Drevlak M., Anderson D.T., Faber B.J., Hegna C.C., Likin K.M., Schmitt J.C. and Talmadge J.N. 2019 Stellarator equilibria with reactor relevant energetic particle losses *J. Plasma Phys.* **85** 905850508
- [5] Bader A. et al 2020 Advancing the physics basis for quasi-helically symmetric stellarators *J. Plasma Phys.* **86** 905860506
- [6] Beidler C.D., Kolesnichenko Y.I., Marchenko V.S., Sidorenko I.N. and Wobig H. 2001 Stochastic diffusion of energetic ions in optimized stellarators *Phys. Plasmas* **8** 2731–8
- [7] Boozer A.H. 1980 Guiding center drift equations *Phys. Fluids* **23** 904–8
- [8] Brysk H. 1973 Fusion neutron energies and spectra *Plasma Phys.* **15** 611–7
- [9] Bunno M., Nakamura Y., Suzuki Y., Shinohara K., Matsunaga G. and Tani K. 2013 Fusion alpha-particle losses in a high-beta rippled tokamak *Phys. Plasmas* **20** 082511
- [10] Burby J.W. and Sengupta W. 2018 Hamiltonian structure of the guiding center plasma model *Phys. Plasmas* **25** 020703
- [11] Cary J.R. and Brizard A.J. 2009 Hamiltonian theory of guiding-center motion *Rev. Mod. Phys.* **81** 693–738
- [12] Cerfon A.J. and Freidberg J.P. 2011 Magnetohydrodynamic stability comparison theorems revisited *Phys. Plasmas* **18** 012505
- [13] Cliffe K.A., Giles M.B., Scheichl R. and Teckentrup A.L. 2011 Multilevel Monte Carlo methods and applications to elliptic PDEs with random coefficients *Comput. Vis. Sci.* **14** 3–15
- [14] Cole M.D.J., Hager R., Moritaka T., Lazerson S., Kleiber R., Ku S. and Chang C.S. 2019 Comparative collisionless alpha particle confinement in stellarator reactors with the XGC gyrokinetic code *Phys. Plasmas* **26** 032506
- [15] Dimarco G. and Pareschi L. 2019 Multi-scale control variate methods for uncertainty quantification in kinetic equations *J. Comput. Phys.* **388** 63–89
- [16] Dimarco G. and Pareschi L. 2020 Multiscale variance reduction methods based on multiple control variates for kinetic equations with uncertainties *Multiscale Model. Simul.* **18** 351–82
- [17] Drevlak M., Beidler C.D., Geiger J., Helander P. and Turkin Y. 2018 Optimisation of stellarator equilibria with ROSE *Nucl. Fusion* **59** 016010
- [18] Drevlak M., Geiger J., Helander P. and Turkin Y. 2014 Fast particle confinement with optimized coil currents in the W7-X stellarator *Nucl. Fusion* **54** 073002
- [19] Weinan E., Ren W. and Vanden-Eijnden E. 2009 A general strategy for designing seamless multiscale methods *J. Comput. Phys.* **228** 5437–53
- [20] Fatkullin I. and Vanden-Eijnden E. 2004 A computational strategy for multiscale systems with applications to Lorenz 96 model *J. Comput. Phys.* **200** 605–38
- [21] Grad H. 1966 Variational principle for a guiding-center plasma *Phys. Fluids* **9** 225–51
- [22] Grieger G. et al 1992 Physics optimization of stellarators *Phys. Fluids B* **4** 2081–91
- [23] Hammersley J.M. and Handscomb D.C. 1964 *Monte Carlo Methods* (London: Methuen)
- [24] Hazeltine R.D. and Waelbroeck F.L. 2018 *The Framework of Plasma Physics* (Boca Raton, FL: CRC Press)
- [25] Heidbrink W.W. and Sadler G.J. 1994 The behaviour of fast ions in tokamak experiments *Nucl. Fusion* **34** 535
- [26] Helander P. 2014 Theory of plasma confinement in non-axisymmetric magnetic fields *Rep. Prog. Phys.* **77** 087001
- [27] Henneberg S.A., Drevlak M., Nührenberg C., Beidler C.D., Turkin Y., Loizu J. and Helander P. 2019 Properties of a new quasi-axisymmetric configuration *Nucl. Fusion* **59** 026014
- [28] Hirshman S.P. and Whitson J.C. 1983 Steepest-descent moment method for three-dimensional magnetohydrodynamic equilibria *Phys. Fluids* **26** 3553–68
- [29] Konrad J., Farcas I.-G., Peherstorfer B., Di Siena A., Jenko F., Neckel T. and Bungartz H.-J. 2021 Data-driven low-fidelity models for multi-fidelity Monte Carlo sampling in plasma micro-turbulence analysis *J. Comput. Phys.* **451** 110898
- [30] Kruskal M.D. and Oberman C.R. 1958 On the stability of plasma in static equilibrium *Phys. Fluids* **1** 275–80
- [31] Ku L.P. and Garabedian P.R. 2006 New classes of quasi-axisymmetric stellarator configurations *Fusion Sci. Technol.* **50** 207–15
- [32] Ku L.P., Garabedian P.R., Lyon J., Turnbull A., Grossman A., Mau T.K. and Zarnstorff M. (ARIES Team) 2008 Physics design for ARIES-CS *Fusion Sci. Technol.* **54** 673–93
- [33] Littlejohn R.G. 1983 Variational principles of guiding centre motion *J. Plasma Phys.* **29** 111–25
- [34] Lotz W., Merkel P., Nührenberg J. and Strumberger E. 1992 Collisionless alpha-particle confinement in stellarators *Plasma Phys. Control. Fusion* **34** 1037–52
- [35] Merkel P. 1987 Solution of stellarator boundary value problems with external currents *Nucl. Fusion* **27** 867–71
- [36] Mynick H.E., Boozer A.H. and Ku L.P. 2006 Improving confinement in quasi-axisymmetric stellarators *Phys. Plasmas* **13** 064505
- [37] Najmabadi F. et al 2008 The ARIES-CS compact stellarator fusion power plant *Fusion Sci. Technol.* **54** 655–72
- [38] Nelson B.L. 1987 On control variate estimators *Comput. Oper. Res.* **14** 219–25
- [39] Nemov V.V., Kasilov S.V. and Kernbichler W. 2014 Collisionless high energy particle losses in optimized stellarators calculated in real-space coordinates *Phys. Plasmas* **21** 062501
- [40] Nemov V.V., Kasilov S.V., Kernbichler W. and Heyn M.F. 1999 Evaluation of $1/\nu$ neoclassical transport in stellarators *Phys. Plasmas* **6** 4622–32

- [41] Nemov V.V., Kasilov S.V., Kernbichler W. and Leitold G.O. 2008 Poloidal motion of trapped particle orbits in real-space coordinates *Phys. Plasmas* **15** 052501
- [42] Ng L.W.T. and Willcox K.E. 2014 Multifidelity approaches for optimization under uncertainty *Int. J. Numer. Methods Eng.* **100** 746–72
- [43] Ng L.W.T. and Willcox K.E. 2016 Monte Carlo information-reuse approach to aircraft conceptual design optimization under uncertainty *J. Aircr.* **53** 427–38
- [44] Ohlberger M. and Rave S. 2016 Reduced basis methods: success, limitations and future challenges *Proc. Conf. Algorithm* pp 1–12
- [45] Paul E.J., Antonsen T., Landreman M. and Cooper W.A. 2020 Adjoint approach to calculating shape gradients for three-dimensional magnetic confinement equilibria. Part 2. Applications *J. Plasma Phys.* **86** 905860103
- [46] Peherstorfer B. 2019 Multifidelity Monte Carlo estimation with adaptive low-fidelity models *SIAM/ASA J. Uncertain. Quantification* **7** 579–603
- [47] Peherstorfer B., Willcox K. and Gunzburger M. 2018 Survey of multifidelity methods in uncertainty propagation, inference, and optimization *SIAM Rev.* **60** 550–91
- [48] Peherstorfer B., Willcox K. and Gunzburger M. 2016 Optimal model management for multifidelity Monte Carlo estimation *SIAM J. Sci. Comput.* **38** A3163–94
- [49] Pfefferlé D., Cooper W.A., Fasoli A. and Graves J.P. 2016 Effects of magnetic ripple on 3D equilibrium and alpha particle confinement in the European DEMO *Nucl. Fusion* **56** 112002
- [50] Ramos J.J. 2015 Quasineutrality and parallel force balance in kinetic magnetohydrodynamics *J. Plasma Phys.* **81** 905810111
- [51] Ramos J.J. 2016 On stability criteria for kinetic magnetohydrodynamics *J. Plasma Phys.* **82** 905820607
- [52] Reiman A. *et al* 1999 Physics design of a high-beta quasi-axisymmetric stellarator *Plasma Phys. Control. Fusion* **41** B273–83
- [53] Reiman A. and Greenside H. 1986 Calculation of three-dimensional MHD equilibria with islands and stochastic regions *Comput. Phys. Commun.* **43** 157–67
- [54] Robinson T.D., Eldred M.S., Willcox K.E. and Haines R. 2008 Surrogate-based optimization using multifidelity models with variable parameterization and corrected space mapping *AIAA J.* **46** 2814–22
- [55] Rosenbluth M.N. and Rostoker N. 1959 Theoretical structure of plasma equations *Phys. Fluids* **2** 23–30
- [56] Sigmar D.J., Hsu C.T., White R. and Cheng C.Z. 1992 Alpha-particle losses from toroidicity-induced Alfvén eigenmodes. Part II: Monte Carlo simulations and anomalous alpha-loss processes *Phys. Fluids B* **4** 1506–16
- [57] Spitzer L. 1956 *Physics of Fully Ionised Gases* (New York: Interscience)
- [58] Spong D.A. 2011 Three-dimensional effects on energetic particle confinement and stability *Phys. Plasmas* **18** 056109
- [59] Spong D.A. 2015 3D toroidal physics: testing the boundaries of symmetry breaking *Phys. Plasmas* **22** 055602
- [60] Tolman E.A. and Catto P.J. 2021 Drift kinetic theory of alpha transport by tokamak perturbations *J. Plasma Phys.* **87** 855870201
- [61] Lee Y. and Engquist B. 2014 Variable step size multiscale methods for stiff and highly oscillatory dynamical systems *Discrete Contin. Dyn. Syst.* **34** 1079–97
- [62] Zweben S.J., Hammett G.W., Boivin R.L., Phillips C.K. and Wilson J.R. 1992 Loss of MeV ions during ^3He minority ion cyclotron resonance heating in TFTR *Nucl. Fusion* **32** 1823–33