MSDE

Molecular Systems Design & Engineering

Accepted Manuscript

MSDE

Molecular Systems Design & Engineering

rsc.li/molecular-engineering

ISSN 2058-9689

ROYAL SOCIETY
OF CHEMISTRY

IChemE

ROYAL SOCIETY
OF CHEMISTRY

rsc.li/molecular-engineering

Molecular Systems Design & Engineering Accepted Manuscript

Design, System, Application statement

Field-theoretic simulations (FTS) and self-consistent field theory (SCFT) are powerful tools for probing the phase behavior of multi-component polymer systems, but their efficiency and stability is highly sensitive to the choice of numerical relaxation coefficients used to evolve auxiliary fields. In this work, we introduce a Bayesian optimization framework to systematically identify optimal relaxation coefficients that accelerate convergence and improve numerical stability. By combining field-based polymer models with surrogate modeling and adaptive parameter tuning, our approach dramatically reduces simulation cost, even in high-dimensional systems with ten or more distinct components. We demonstrate broad utility across both SCFT and FTS, including systems with explicit solvents and complex chain architectures. Our workflow is broadly applicable to polymer physics, biomolecular design, and coarse-grained modeling efforts where field-based methods are used. Looking forward, this approach enables high-throughput exploration of chemically specific systems and introduces a general framework for intelligent parameter tuning in field-based simulations of systems that are otherwise challenging or impossible to simulate.

Journal Name

ARTICLE TYPE

Accelerating multi-species field-theoretic simulations using Bayesian optimization[†]

Ritvind Suketana,[a] Andrew Golembeski,[a] and Joshua Lequieu[a‡]

Field-based simulations can be challenging in multi-component polymer systems and are highly sensitive to the choice of relaxation coefficients (λ) used in the field update algorithms. Judiciously chosen relaxation coefficients are critical for both the stability and convergence of field-based simulations, yet their selection is challenging when the number of unique chemical species in the system is large. In this work, we develop a new method to automatically and efficiently locate optimal relaxation coefficients in systems with large numbers of species. We begin by analyzing the effects of relaxation coefficients in two- and three-species systems and demonstrate that regions of high-performance are both narrow and system-specific. Based on these findings, we next develop a method based on Bayesian optimization that automatically locates relaxation coefficients that are stable and exhibit good performance. We demonstrate that our method is considerably faster than naive search methods and becomes particularly efficient as the system complexity increases. This work demonstrates that Bayesian optimization can be used to stabilize and accelerate field-based simulations that contain many different chemical species.

1 Introduction

Simulations based on polymer field theory are widely used to examine the properties of polymeric materials. In these methods, a particle-based model is first converted into a statistical field theory and is subsequently analyzed using numerical simulation[1–3]. The mathematical structure of statistical field theories are particularly well-suited for the simulation of polymers and so field-based simulations tend to become more efficient as the polymers become long and as the system density increases[1]. The mathematical structure of these methods also enable direct access to the free energy[4] and can efficiently handle long-ranged Coulombic interactions without requiring Ewald-based methods[5,6]. As a consequence, field-based simulations can be many orders of magnitude faster than particle-based simulations, despite giving identical results[7,8]. The two most common variants of these methods are self-consistent field theory (SCFT), which invokes a mean-field approximation, and field-theoretic simulations (FTS), which directly sample the statistical field theory without any simplifying approximations.

Most prior studies using field-theoretic simulations have employed models that contain only two species or bead types, typically denoted by A and B. In the majority of these past models,

these two species interact through a Flory-Huggins parameter χ and require two auxiliary fields to decouple the pairwise interactions: a pressure-like field w_+ and an exchange field w_-. Field-theoretic simulations of these models have been extensively used to examine the self-assembly of homopolymers, block polymers and blends thereof[5,9–17]. Another commonly employed model instead permits these two monomers to carry charges and to interact through Coulombic interactions[6,18–20]. Field-theoretic simulations of this model have been used to examine a variety of polyelectrolytes and polyampholytes and also require two fields: an auxiliary field w to decouple the excluded volume interactions and an electrostatic field φ to decouple the electrostatic interactions.

In recent years, there has been a considerable effort to extend field-theoretic simulations to systems that contain more than two species. One of the primary motivations for these efforts is that multi-species FTS can incorporate more chemical specificity than traditional two-species FTS. For example, whereas each species in traditional FTS typically represents a monomer or collection of monomers, each species in a multi-species FTS can now represent a collection of atoms or distinct chemical groups. This allows multi-species FTS to incorporate more chemical information and can result in models that are increasingly predictive. Fredrickson, Shell and co-workers have recently developed a powerful approach to parameterize molecularly informed field theories using atomistic particle-based simulations[21,22]. Multi-species field-based simulations parameterized using this approach have been used to examine surfactant phase behavior[23], the critical micelle

[a] Department of Chemical and Biological Engineering, Drexel University, Philadelphia, Pennsylvania, 19104, United States.

Molecular Systems Design & Engineering Accepted Manuscript

concentration of intrinsically disordered proteins[24], cellulose acetate solubility[25], and block copolymer solution self-assembly[26]. Independent work by Pert *et al.* have also used multi-species field-based simulations to examine the morphologies formed by mRNA encapsulating nanoparticles[27].

One of the major challenges with multi-species field-based simulations is that they can be difficult to numerically stabilize. Most existing multi-species field-based simulations rely on the framework developed by Düchs, Delaney and Fredrickson[28] where the pairwise interaction matrix is diagonalized in order to determine the fields that will decouple the interactions within the model. A challenge with this approach is that most polymeric models result in a spectrum of eigenvalues that is very broad, which can lead to numerical challenges in the resulting simulations. The most common solution to this problem is to empirically tune the rate at which different fields are updated so that the fields associated with different eigenvalues are updated more or less quickly[28,29]. While this empirical tuning can be tolerated if a small number of species are present, it becomes increasingly burdensome as the number of species within the system grows larger. Other strategies for handling these numerical difficulties are to ignore the effects of fluctuations by only focusing on SCFT[23-26], or to avoid numerical simulations altogether and to analyze multi-species field theories analytically[21,22]. As a consequence of these numerical challenges, past work on fully-fluctuating FTS with many different species is still relatively rare[8,27].

In this paper, our first objective is to explore why multi-species field-based simulations are difficult to stabilize numerically. To explore this question, we build on past work[28,29] that focused on tuning the relaxation coefficients of the different auxiliary fields in order to enhance simulation performance. We first comprehensively examine a two-species system to examine the effects of polymer connectivity (e.g. diblocks vs homopolymer blends), morphology (e.g. lamellar vs gyroid phases), and fluctuations (e.g. SCFT vs FTS). Our analysis shows that the qualitative effects of relaxation coefficients are generally conserved across these different systems. We next extend our analysis to multi-species systems and show that the effect of relaxation coefficients depends strongly on the interactions between the different species. Notably, we show that even if the eigenvalues of the interaction matrix are identical, the underlying effect of relaxation coefficients can still be quite different.

From these findings, we then turn to our second objective where we seek an automated strategy that can optimize the stability and performance of multi-species SCFT and FTS. We demonstrate that Bayesian optimization (BayesOpt) is particularly well-suited for this task and can be used to efficiently locate both stable and high-performing relaxation coefficients for SCFT and FTS. A key finding from our analysis is that tuning our BayesOpt implementation is critical for achieving good performance: the choices of surrogate model initialization, kernel, acquisition function and objective function all have a significant impact on overall performance. Nonetheless if these subtleties are accounted for, BayesOpt can lead to orders of magnitude improvements in multi-species SCFT and FTS performance. In summary, the strategy that we have presented here can both stabilize and accelerate

multi-species SCFT and FTS and is envisioned to be useful as field-based methods are extended to systems with increased chemical specificity.

2 Methods

2.1 Multi-species exchange model

To perform field-based simulations, we use a slightly modified version of the multi-species exchange model of Düchs, Delaney and Fredrickson[28] that includes both Gaussian regularization[30,31] and discrete polymer chains. Our models involves n total polymer molecules in a volume V at temperature T. Each polymer molecule of type m has n_m indistinguishable copies and each polymer molecule is composed of N_m covalently bonded beads. Each bead is chosen from a total of S species types (or bead types) within the system. In addition to bonded interactions, beads of type i and j interact through a Flory-Huggins interaction parameter χ_{ij}, and a Helfand compressibility parameter ζ.

One key difference between our model and the multi-species exchange model of Düchs, Delaney and Fredrickson[28] is that we regularize our model using a Gaussian smearing function[30,31]. Formally, this approach involves a convolution of the density of each species K with a Gaussian distribution function, $\Gamma_K(\boldsymbol{r}) = (2\pi a_K^2)^{-3/2} \exp[-r^2/(2a_K^2)]$, in order to convert the microscopic species density $\hat{\rho}_K(\boldsymbol{r})$ into a smeared microscopic species density $\check{\rho}_K(\boldsymbol{r}) = \int d\boldsymbol{r}' \Gamma_K(\boldsymbol{r}-\boldsymbol{r}')\hat{\rho}_K(\boldsymbol{r}') = \Gamma_K \star \hat{\rho}_K$, where $K = \{1,...,S\}$ and $\star$ denotes a spatial convolution. Once these smeared densities have been defined, the non-bonded energy within the system becomes

$$\beta U_{nb} = \frac{1}{2\rho_0} \sum_{i,j=1}^{S} \int_V d^3r\, \check{\rho}_i(\boldsymbol{r})\chi_{ij}\check{\rho}_j(\boldsymbol{r})$$

$$+ \frac{\zeta}{2\rho_0} \int_V d^3r \left(\sum_{j=1}^{S} \check{\rho}_j(\boldsymbol{r}) - \rho_0 \right)^2 \qquad (1)$$

where $\beta = 1/k_b T$, k_b is the Boltzmann constant and $\rho_0 = \sum_m^M n_m N_m / V$ is the total density of the system. Bonded interactions are represented using harmonic bonds so that the total bonded energy of the system is

$$\beta U_b = \sum_{m=1}^{M} \sum_{i=1}^{n_m} \sum_{j=1}^{N_m-1} \frac{3}{2b} \left(r_{j,j+1}^{(m,i)} \right)^2 \qquad (2)$$

where $r_{j,j+1}^{(m,i)}$ is the spatial separation between the j and $j+1$ bead of the ith copy of molecule type m and b is the statistical segment length corresponding to that bond.

By following the approach described in Ref.[28], this model can exactly be converted into a field theory,

$$\mathcal{Z} = \mathcal{Z}_0 \int \mathcal{D}\boldsymbol{\mu}\, exp\left(-H[\boldsymbol{\mu}]\right) \qquad (3)$$

where $\boldsymbol{\mu} = \{\mu_1, \mu_2, ..., \mu_S\}$ contains S newly introduced auxiliary fields and the prefactor $\mathcal{Z}_0$ is slightly modified from the expression in Ref[28] to account for our use of discrete Gaussian chains[3,8].

$H[\boldsymbol{\mu}]$ is a field-theoretic Hamiltonian given by

$$H[\boldsymbol{\mu}] = C \left[-\sum_{i=1}^{S} \frac{(\zeta N)^{-\frac{1}{2}}}{2\Lambda_i} \int_{\tilde{V}} d^3 r\, \mu_i^2(\mathbf{r}) - \sum_{i,j=1}^{S} \frac{O_{ji}}{\Lambda_i(\zeta N)^{-\frac{1}{2}}} \right.$$
$$\left. \times \int_{\tilde{V}} d^3 r\, \mu_i(\mathbf{r}) - \tilde{V} \sum_{m=1}^{M} \frac{\bar{\phi}_m}{\alpha_m} \ln Q_m[\boldsymbol{\Omega}] \right] \quad (4)$$

where $C = \rho_0 R_g^3/N$ is a reduced chain density, $\tilde{V} = V/R_g^3$ is a reduced volume, $\bar{\phi}_m = n_m N_m V/\rho_0$ is the volume fraction of molecule m, $\alpha_m = N_m/N$ is a normalized chain length, and $R_g = b((N-1)/6)^{1/2}$ is the unperturbed radius of gyration of a chain consisting of N beads and statistical segment length b. In this equation, Λ_i are eigenvalues of the $S \times S$ matrix $\mathbf{X} = \left(\boldsymbol{\chi} N(\zeta N)^{-\frac{1}{2}} + (\zeta N)^{\frac{1}{2}} \mathbf{1} \right)$ where $\mathbf{1}$ is a $S \times S$ matrix with all entries equal to 1. The eigenvalues of $\mathbf{X}$ are contained in the columns of the $S \times S$ matrix $\mathbf{O}$ with elements O_{ij} and correspond to the linear transformation from the exchange fields $\boldsymbol{\mu}$ to the species fields $\boldsymbol{\Omega}$. In this work, we only consider matricies $\mathbf{X}$ where all eigenvalues are non-zero and non-degenerate. For zero-valued or degenerate eigenvalues, extra considerations are required[3,8,28].

The final term to be specified in Eq. 4 is Q_m, the single-chain partition function of molecule m. Q_m is a functional of the species fields $\boldsymbol{\Omega} = \{\Omega_1, \Omega_2, ..., \Omega_S\}$

$$\Omega_K(r) = \Gamma_K \star \sum_{j=1}^{S} O_{Kj} \mu_j(r) \quad (5)$$

for $K = 1, ..., S$ and is defined as

$$Q_m[\boldsymbol{\Omega}] = \frac{1}{\tilde{V}} \int_{\tilde{V}} d^3 r\, q_{m,N_m}(\boldsymbol{r}; [\boldsymbol{\Omega}]). \quad (6)$$

In this expression, $q_{m,j}(\boldsymbol{r})$ is the propagator corresponding to the statistical weight for a molecule m at bead index j at position $\boldsymbol{r}$ for fields $\boldsymbol{\Omega}$ and is calculated from a Chapman-Kolmogorov equation as described elsewhere[1,3].

To sample the field-based partition function in FTS, field configurations for each exchange field $\mu_i(r)$ are evolved in fictitious time t using complex Langevin[32,33] (CL) dynamics,

$$\frac{\partial \mu_i(r,t)}{\partial t} = -\lambda_i \gamma_i^2 \frac{\delta H[\boldsymbol{\mu}]}{\delta \mu_i(r,t)} + \gamma_i \eta_i(r,t) \quad (7)$$

where

$$\gamma_i = \begin{cases} 1, & \Lambda_i < 0 \\ \sqrt{-1}, & \Lambda_i > 0 \end{cases}$$

and $\eta_i(r,t)$ is Gaussian white noise with moments $\langle \eta_i(\boldsymbol{r},t) \rangle = 0$ and $\langle \eta_i(\boldsymbol{r},t)\eta_i(\boldsymbol{r}',t') \rangle = 2\lambda_i \delta(\boldsymbol{r}-\boldsymbol{r}')\delta(t-t')$. Eq. 7 can also be used to locate mean-field configurations (i.e. SCFT) by setting $\eta_i = 0$. The parameter λ_i in Eq. 7 corresponds to the real relaxation coefficient of the ith field.

An important aspect of Eq. 7 to emphasize is that the CL dynamics are not physical and simply correspond to a numeric scheme to sample field configurations $\boldsymbol{\mu}$ according to the weights $\exp(-H[\boldsymbol{\mu}])$. As a consequence, the relaxation coefficients $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, ..., \lambda_S\}$ are not constrained to any specific physical values and can instead be tuned to obtain optimal sampling per-

formance. One of the central objectives of this work is to develop tools that can automatically tune these relaxation coefficients for systems with many components. This approach will be discussed extensively in Section 3.

In order to numerically integrate equation 7 we use a Euler-Maruyama Predictor-Corrector (EMPEC) algorithm which has second order accuracy in fictitious time. The EMPEC algorithm first conducts a predictor step which calculates a preliminary guess for the update field at time $\tilde{t}$. This is followed by a corrector step which refines the field update by averaging the forces evaluated at the current step the predicted step. The same noise $\eta_i(t)$ is used in both steps[28,29]

$$\mu_i(\mathbf{r},\tilde{t}) = \mu_i(\mathbf{r},t) - \lambda_i \Delta t \gamma_i^2 \frac{\delta H[\boldsymbol{\mu}]}{\delta \mu_i(\mathbf{r},t)} + \gamma_i \eta_i(\mathbf{r},t) \quad (8)$$

$$\mu_i(\mathbf{r},t+\Delta t) = \mu_i(\mathbf{r},t) - \frac{\lambda_i \Delta t \gamma_i^2}{2} \left(\frac{\delta H[\boldsymbol{\mu}]}{\delta \mu_i(\mathbf{r},t)} + \frac{\delta H[\boldsymbol{\mu}]}{\delta \mu_i(\mathbf{r},\tilde{t})} \right) + \gamma_i \eta_i(\mathbf{r},t) \quad (9)$$

where Δt is the timestep. In this expression, the product $\lambda_i \Delta t$ governs the rate at which exchange field μ_i is updated. In our convention, we have chosen a constant timestep Δt for all fields, and have varied the λ_i for each field individually.

For all simulation in this work we use $\zeta N = 100$, smearing length $a = 0.15 R_g$, reference chain length $N = 100$, chain number density $C = 10$, and statistical segment length $b = 1$. The timestep Δt is held fixed at 1.0 for SCFT and at 0.02 for FTS. All SCFT and FTS were performed using an in-house software called OpenFTS.

2.2 Quantification of SCFT/FTS Performance

Throughout this work, one of our central objectives is to quantify the performance of SCFT and FTS and how this performance can be optimized by tuning the relaxation coefficients $\boldsymbol{\lambda}$. In SCFT, performance is quantified by the number of timesteps required to converge the field configuration to a saddle point within an error of $\delta H[\boldsymbol{\mu}]/\delta \mu_i < 10^{-5}$ for $i = \{1, ..., S\}$. Since the dynamic evolution of the fields in SCFT is typically unimportant, optimal performance in SCFT is achieved by locating a saddle point in as few timesteps as possible.

In FTS, the objective of the simulation is to instead stochastically sample field configurations so that time averages of field-based operators converge to their equilibrium values. Since field-based operators at subsequent timesteps will typically be correlated, optimal performance in FTS is achieved by minimizing the number of timesteps required to obtain decorrelated samples of different field-based operators. Throughout this work, we quantify the performance of FTS by measuring the autocorrelation time of the excess chemical potential operator (Fig. S2). Shorter values of the correlation time in FTS correspond to better simulation performance. In the discussion of our results below, we denote both the performance of SCFT (i.e. timesteps to converge to a saddle point) and the performance of FTS (i.e. autocorrelation time) as $P(\boldsymbol{\lambda})$, where the dependence of simulation performance on the relaxation coefficients $\boldsymbol{\lambda}$ is explicit.

In both SCFT and FTS, the performance will depend on the field

configurations used to initialize the simulation. For SCFT, we use initial field configurations from a converged simulation at slightly different interaction parameters χ. For consistency throughout the many systems considered in this work, we evaluate SCFT performance using interactions parameters χN that are multiples of five using initial field configurations obtained using interaction parameters that are multiples of four. For example, when evaluating the performance of a simulation at $\chi N = \{5, 15, 20\}$ we initialize these simulation using converged fields for $\chi N = \{4, 12, 16\}$. When initializing FTS, we first converge fields using SCFT for the given interaction parameters and then perform FTS on those fields until all operators reach their equilibrium values. Once equilibrium is reached, we then calculate the correlation time.

All SCFT calculations, excluding those for the double gyroid, are performed in a one-dimensional (1D) box with a length of 6 R_g and 64 plane waves (M). The double gyroid simulations are run in a three-dimensional (3D) cubic box with length 9.6 R_g and M=64^3. SCFT is run until the summation of field forces over all fields is below 10^{-5} or for a maximum of 1,500,000 time steps — 1.5 times the steps required for the slowest relaxation coefficients to converge. FTS is run in a 3D cubic box with a length of 6 R_g and $M = 48^3$. Only macrophase separating homopolymer blend systems are explored using FTS. When evaluating different relaxation coefficients, each λ_i is selected from 30 logarithmically spaced points ranging from 0.001 to 100. Further details are provided in the Supporting Information.

2.3 Bayesian Optimization

We use Bayesian optimization (BayesOpt) to automatically tune the relaxation coefficients λ in order to achieve optimal SCFT and FTS performance. In our BayesOpt implementation, we use a Gaussian process regression surrogate model and an expected improvement acquisition function[34,35]. Our objective function is defined as $(P(\lambda))^{-1}$ and is normalized to the range of zero to unity using the minimum and maximum measured values of $(P(\lambda))^{-1}$. Since optimal performance of SCFT/FTS corresponds to small values of $P(\lambda)$, we seek to maximize this objective function.

While BayesOpt is a widely applied method, our specific application of BayesOpt led to some challenges that warrant additional discussion. The first challenge is that many values of the relaxation coefficients, λ, result in simulations that are divergent or that do not converge (see Fig. 1 and Fig. 2). Since a nonconvergent simulation corresponds to performance $P(\lambda) \approx \infty$, the objective function is rugged and contains many sharp features. We find that it can be challenging to approximate these sharp features with a Gaussian process surrogate model. A related challenge is that the optimal relaxation coefficients λ^* are often in the immediate vicinity of these non-convergent regions and so the incorporation of these sharp features into the surrogate model are critical for good BayesOpt performance. By empirically varying many aspects of our BayesOpt implementation, we find that both these challenges could be addressed by using a suitably tuned Matérn kernel[36], an acquisition function that favored exploitation[37], and a dynamically adjusted penalty associated with non-

convergent trajectories (see Supporting Information).

We also faced challenges with our BayesOpt implementation when we extended it to systems containing large numbers of species. In our BayesOpt implementation, the memory required to evaluate the acquision function scales exponentially with the number of species in the system. As the number of species becomes large, this scaling can result in memory requirements exceeding 100GB that preclude the use of our typical computational hardware. In order to overcome the memory requirements of these systems, we use an adaptive BayesOpt scheme where the resolution and range of the surrogate model is dynamically updated as the optimization proceeds (see Supporting Information). We find that our adaptive BayesOpt performs similarly to our original BayesOpt implementation despite using a fraction of the required memory (Fig. S7). Throughout this work, we use our adaptive BayesOpt for systems containing more than five species.

We also find that the relaxation coefficients λ used to initialize the surrogate model are extremely important for BayesOpt performance. Notably, the performance of BayesOpt suffered when it was initialized with conventional schemes such as randomly selected or space-filling relaxation coefficients. After trying numerous initialization schemes, we found that choosing relaxation coefficients that were proportional to the square root of the absolute value of that field's eigenvalue tended to give the best BayesOpt performance. Specifically, the initial relaxation coefficients are $\lambda_i = \alpha |\Lambda_i|^{1/2}$, where Λ_i is the eigenvalue defined in Eq. 4 and α is a randomly selected proportionality constant that is the same for all fields i. One of the primary advantages of this approach is that the initial relaxation coefficients are chosen along a one-dimensional line, even if the overall dimension of the search space for λ is much larger. By initializing our surrogate model in this way, we are able to efficiently locate convergent relaxation coefficients that can be subsequently refined using BayesOpt. Additional details of our BayesOpt implementation are provided in the Supporting Information.

3 Results and Discussion

3.1 Effect of relaxation coefficients on SCFT/FTS performance

To demonstrate the importance of the relaxation coefficients λ on simulation performance, we first examine several simple two-species systems consisting of either neat diblock copolymers or homopolymer blends. We begin with these simple systems because the simulations are very efficient and enable us to exhaustively examine the role of relaxation coefficients λ on SCFT/FTS performance. We will consider systems of higher complexity in subsequent sections.

We begin by examining the number of timesteps required by SCFT to converge a diblock copolymer double gyroid phase for different relaxation coefficients $\lambda = \{\lambda_1, \lambda_2\}$ (Fig. 1A). In these calculations, the A block fraction $f_A = 0.34$, the segregation strength $\chi N = 30$, number of plane waves $M = 64^3$ and the initial fields are obtained from a converged unit cell at $\chi N = 20$ and $f_A = 0.37$. From these calculations, we observe a wide variation in timesteps to convergence with respect to the
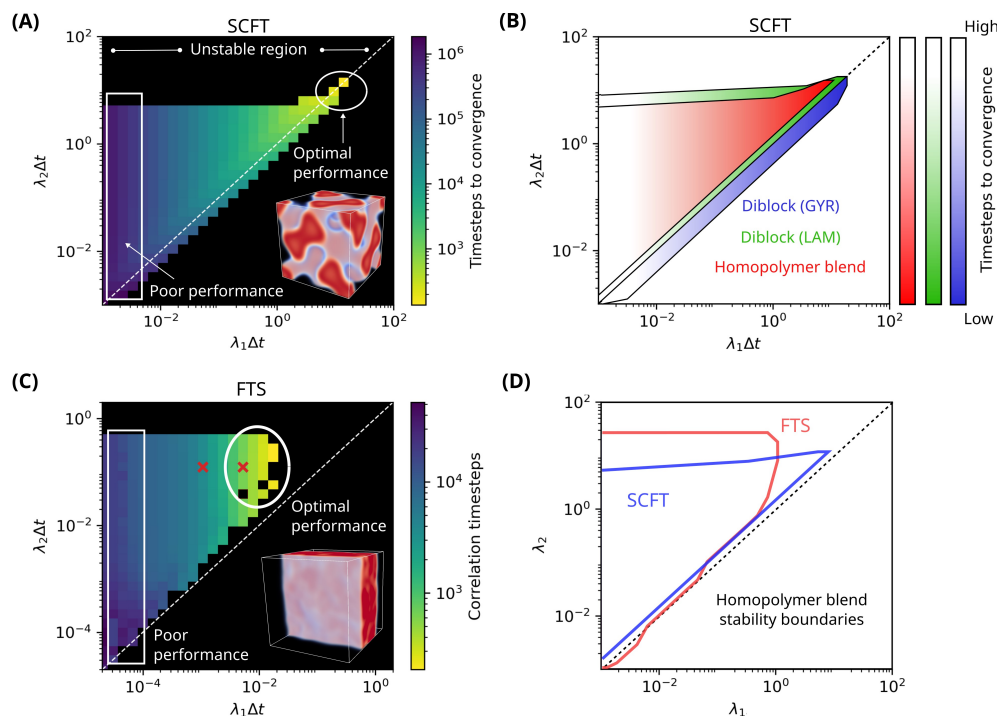
Fig. 1 Effect of relaxation coefficients $\lambda = \{\lambda_1, \lambda_2\}$ on SCFT and FTS performance for various two-species systems. (A) SCFT performance for a gyroid-forming diblock copolymer melt. (B) Comparison of SCFT performance for different polymer architectures (i.e. diblock vs. homopolymer blends) and microphases (i.e. lamellar vs. gyroid). (C) FTS performance for a homopolymer blend. Detailed trajectories for two values of λ (red X markers) are given in Fig. S2. (D) Stability boundaries of SCFT and FTS for a homopolymer blend. Dotted lines correspond to $\lambda_1 = \lambda_2$.

relaxation coefficients λ (Fig. 1A). The performance of SCFT ranges from poorly performing relaxation coefficients that require greater than 10^6 timesteps to the best performing relaxation coefficients $\lambda^* = \{14.2, 14.2\}$ that converge in 140 timesteps. Moreover, many relaxation coefficients never yield a converged SCFT solution, either due to numeric instabilities (i.e. divergent trajectories) or the inability to reduce field errors beneath the prescribed threshold of 10^{-5}. It is also noteworthy that the optimal relaxation coefficients λ^* that give the best SCFT performance are immediately adjacent to these non-convergent relaxation coefficients.

In order to understand whether our findings in Fig. 1A generalize to other systems and conditions, we also examine the influence of relaxation coefficients on SCFT performance for a lamellar-forming diblock copolymer melt ($f_A = 0.5$, $\chi N = 30$) and a symmetric two component homopolymer blend with $\bar{\phi} = 0.5$ and $\chi N = 30$ (Fig. 1B). These calculations are initialized from converged SCFT simulations at $\chi N = 20$ and $f_A = 0.5$ for the diblock, and at $\chi N = 20$ and $\bar{\phi}_A = 0.5$ for the homopolymer blend.

In general, our observations for the gyroid phase are very similar to our observations for the lamellar phase and the homopolymer blend: the optimal relaxation coefficients λ^* are many orders of magnitude faster than the poorly performing relaxation coefficients and the optimal relaxation coefficients λ^* are adjacent to non-convergent relaxation coefficients. Furthermore, the general shape of the convergent relaxation coefficients are essentially unchanged across these three different systems (Fig. 1B). It is also noteworthy that the general shape of the convergent relaxation

coefficients are minimally affected by the convergence tolerance used in SCFT (Fig. S9).

We are also interested in how the choice of relaxation coefficients affect the performance of FTS (Fig. 1C). In our analysis, we consider a homopolymer blend with $\bar{\phi}_A = 0.5$, $\chi N = 30$ and quantify the performance of FTS using the correlation time of the chemical potential operator (see Methods). In order to obtain accurate correlation time measurements when the relaxation coefficients λ_i are small, all simulations are run for 320,000 timesteps. Additionally, to account for the stochastic nature of FTS, three independent replicas are performed for every combination of relaxation coefficients and the mean correlation time is reported. If one or more replicas diverge before completing the total number of timesteps, then the corresponding relaxation coefficients are classified as non-convergent.

The performance of FTS is qualitatively similar to our results for SCFT in Fig. 1A-B. A notable difference between SCFT and FTS is that FTS requires values of $\lambda \Delta t$ that are approximately 50 times smaller than for SCFT ($\Delta t = 0.02$ for FTS, 1.0 for SCFT). Another difference is the less pronounced optimal region for FTS whereas SCFT has a sharply peaked region of optimal performance. Both of these differences associated with FTS are not surprising given the numeric challenges associated with integrating the stochastic CL dynamics in FTS, which require smaller timesteps and lead to broader optimal parameter regions due to the increased sensitivity to stochastic fluctuations as the relaxation coefficients increase. Nonetheless, if the different timesteps required by SCFT and FTS are accounted for, then the effects of λ on the stability

of SCFT and FTS are largely in agreement (Fig. 1D).

A more subtle difference between SCFT and FTS is that equal-valued relaxation coefficients $\lambda_1 = \lambda_2$ are generally convergent in SCFT (Fig. 1A) but are unstable in FTS (Fig. 1C). It is also noteworthy that the optimal relaxation coefficients generally occur when $\lambda_1 \approx \lambda_2$ in SCFT, but for $\lambda_1 \neq \lambda_2$ in FTS. This observation underscores the difficulty of running efficient FTS compared to SCFT and is reasonable given the presence of fluctuations in FTS.

Taken together, our results in Fig 1 demonstrate that tuning the relaxation coefficients λ is critical for both SCFT and FTS performance and that the choice of optimal relaxation coefficients λ^* can lead to simulations that are orders of magnitude faster. We also demonstrate that the general effects of relaxation coefficients on simulation performance are qualitatively similar across different systems and simulation techniques. This overarching similarity suggests that an approach to optimize the relaxation coefficients in one context will likely be applicable to other contexts as well. In particular, it motivates a general-purpose strategy for optimizing the relaxation coefficients in SCFT/FTS. The goal of this paper is to describe such a strategy that can efficiently locate these optimal relaxation coefficients λ^*. However, before we do this, it is useful to consider how the optimal relaxation coefficients λ^* change as the number of species in the system increase beyond two.

3.2 Challenges in systems with many species

Now that we have established the importance of the relaxation coefficients λ in two-species systems, we turn to more complex systems that contain additional species. As we will demonstrate, for systems that contain three or more species, there are several additional complexities that make it more challenging to locate the optimal relaxation coefficients λ^*. In the analysis and discussion that follows, we will focus on systems that contain three species but our findings are anticipated to be generally applicable to any systems whose total number of species exceed two.

One of the most obvious complexities that emerges from a three-species system is that three distinct exchange fields are now necessary to decouple the non-bonded interactions in the model. This now results in three relaxation coefficients $\lambda = \{\lambda_1, \lambda_2, \lambda_3\}$ and thus a higher dimensional space to search over in order to locate the optimal relaxation coefficients λ^*. A more subtle, but nonetheless important, consequence of moving to three-species systems is that the mapping from the exchange fields μ to species fields Ω now depends on the interaction matrix χ. In order to illustrate this dependence, is is helpful to first consider a two species system where $\chi = \chi_{12}$. For this two species system, the $\mathbf{X}$ matrix becomes

$$\mathbf{X} = \begin{pmatrix} (\zeta N)^{1/2} & (\zeta N)^{1/2} + \chi N (\zeta N)^{-1/2} \\ (\zeta N)^{1/2} + \chi N (\zeta N)^{-1/2} & (\zeta N)^{1/2} \end{pmatrix} \quad (10)$$

whose eigenvectors are the columns of $\mathbf{O}$ where

$$\mathbf{O} = \frac{1}{\sqrt{2}} \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}. \quad (11)$$

A noteworthy feature of this two species system is that the eigenvectors contained in $\mathbf{O}$ are independent of interaction parameters χ. Since $\mathbf{O}$ corresponds to the linear transform between the exchange fields μ and the species fields Ω, this independence means that the definition of the species fields in a two species system are always the same, independent of how the different species interact. This is one reason why the effects of the relaxation coefficients λ on system performance for the two species system are largely independent of interaction parameters (Fig. 1).

For systems with three or more species, this independence is lost and the mapping between the exchange fields μ and the species fields Ω will now depend on the interaction parameters χ. As a direct consequence of this interdependence, the relaxation coefficients λ will now depend on the interaction parameters χ. This means that it is generally more challenging to identify stable relaxation coefficients λ in systems with three or more species than in systems that contain only two.

In order to illustrate these challenges, we first examine a three-component homopolymer blend consisting of polymers A, B and C, with volume fractions $\bar{\phi}_A = \bar{\phi}_B = 0.33$, $\bar{\phi}_C = 0.34$, and interaction parameters $\chi_{AB}N = 15$, $\chi_{BC}N = 20$ and $\chi_{AC}N = 15$. To keep calculations computationally tractable, our analysis will focus on SCFT, however, we expect qualitatively similar results if the analysis were instead performed using FTS (c.f. Fig. 1D). In order to visualize the impact of relaxation coefficients λ for this system, we plot two different orientations of the complete three-dimensional data (Fig. 2A) and several two-dimensional slices at fixed values of $\lambda_2 \Delta t$ (Fig. 2B). At least qualitatively, the effect of relaxation coefficients λ for this three species system is similar to our results for two species systems discussed above (Fig. 1). As with the two species system, the performance of SCFT varies widely with respect to λ, with poor performing λ requiring close to 10^6 timesteps for convergence, while the optimal $\lambda^* = \{9.2, 13.7, 32.3\}$ converges in just 80 timesteps. The stable λ space also narrows as λ_i values increase, resulting in a small region of optimal performance with the optimal relaxation coefficients λ^* located immediately adjacent to non-convergent regions. It is also noteworthy that most λ values fail to yield a converged SCFT solution.

Yet beyond these qualitative similarities to the two species systems, this three species system also contains additional complexities that make it more challenging to identify efficient relaxation coefficients λ. For example, in contrast to the two species system discussed in Fig 1, this three species system does not converge if all relaxation coefficients are equal $\lambda_1 = \lambda_2 = \lambda_3$ (Fig. 2A, dotted black line). This suggests that the tuning of relaxation coefficients for different exchange fields is essential for achieving stable SCFT/FTS of systems with more than two species. Another complexity is that the region of optimal relaxation coefficients is much narrower (Fig 2B) and only exists for a small range of relaxation coefficients λ. This narrow region of optimal performance can result in over an order of magnitude increase in performance (e.g. Fig 2B, $\lambda_2 \Delta t = 0.08$ versus $\lambda_2 \Delta t = 13.7$) and so there is a considerable computational benefit if this narrow region can be found.

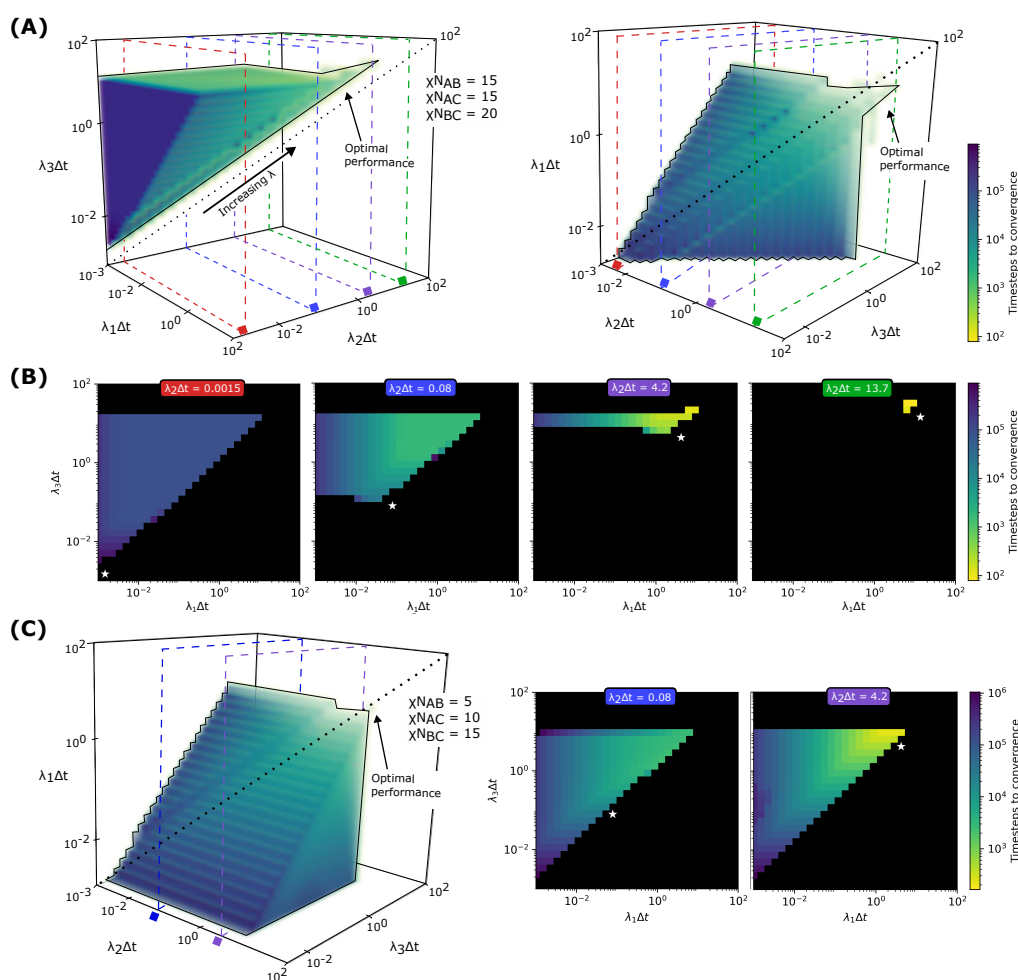Lastly, and perhaps most significantly, the performance of dif-
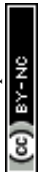
Fig. 2 Effect of relaxation coefficients $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2, \lambda_3\}$ on SCFT performance for a three-species homopolymer blend. (A) Complete three-dimensional representation of SCFT performance for interaction parameters $\chi_{AB}N = 15$, $\chi_{BC}N = 20$ and $\chi_{AC}N = 15$. Two different orientations of the performance space are presented. The dotted black line denotes points where $\lambda_1 = \lambda_2 = \lambda_3$. (B) Representative two-dimensional slices of SCFT performance for fixed values of $\lambda_2\Delta t$. Note that the colors of the cross-sections correspond to the marker colors in the complete representation. The white star denotes points where $\lambda_1 = \lambda_2 = \lambda_3$. (C) For different interaction parameters $\chi_{AB}N = 5$, $\chi_{AC}N = 10$ and $\chi_{BC}N = 15$, the three-dimensional representation of SCFT performance (left) and corresponding two-dimensional slices of SCFT performance for fixed values of $\lambda_2\Delta t$. Changes to the interaction parameters affect the performance of relaxation coefficients $\boldsymbol{\lambda}$ and therefore the location of their optimal value $\boldsymbol{\lambda}^*$.
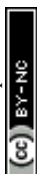
ferent relaxation coefficients in this three species system now depends on the interaction parameters $\boldsymbol{\chi}$. To illustrate this dependence, we have recomputed the performance of SCFT for this same three species system with slightly different interaction parameters $\chi_{AB}N = 5$, $\chi_{AC}N = 10$ and $\chi_{BC}N = 15$ (Fig. 2C). This change in interaction parameters has a significant effect on the role of different relaxation coefficients and where the optimal relaxation coefficients might be found.

These results and the others presented in Fig. 2 collectively demonstrate the challenges of locating stable and high-performing relaxation coefficients in systems that contain three (or more) unique chemical species. In our treatment so far, we were able to locate these optimal relaxation coefficients through brute force sampling: we discretized the range of relaxation coefficients, explicitly performed simulations at each of these different relaxation coefficient values and then quantified the simulation performance. Since the overall search space increases exponentially with the number of species, this brute force approach quickly becomes intractable as the number of species increases beyond three. It is therefore of great interest to develop methods that can locate these optimal relaxation coefficients without brute force sampling.

One particularly attractive strategy would be to use a closed-form expression to determine the optimal relaxation coefficients from the interaction parameters $\boldsymbol{\chi}$. For example, Düchs, Delaney and Fredrickson[28] previously suggested that the eigenvalue magnitude $|\Lambda_i|$ might be related to the optimal relaxation coefficient λ_i^* of that field. Yet, Düchs, Delaney and Fredrickson[28] also demonstrated that non-linear couplings between different fields through the $\ln Q_m$ term in Eq. 4 makes this scheme too simplistic to yield optimal relaxation coefficients in practice. We also examined the possibility of closed-form predictions of the optimal relaxation coefficients but were unable to identify any general purpose approach that reliably worked for multi-species systems. In our analysis, we found that the closed-form prediction of relaxation coefficients is complicated by the fact that the optimal relaxation coefficients can vary for systems with different interaction parameters $\boldsymbol{\chi}$ but equivalent eigenvalues $\boldsymbol{\Lambda}$ (Fig. S5), and for systems with equivalent interaction parameters $\boldsymbol{\chi}$, but different molecular architectures (Fig. 1B). Our conclusion is thus consistent with Düchs, Delaney and Fredrickson[28] that numeric schemes are needed to determine the optimal relaxation coefficients $\boldsymbol{\lambda}^*$. In the following section, we demonstrate that Bayesian optimization is particularly well-suited for locating optimal relaxation coefficients in complex polymeric systems containing many chemical species.

3.3 Bayesian optimization for relaxation coefficient selection

Bayesian optimization (BayesOpt) is a method for maximizing objective functions that are both expensive to evaluate and whose derivatives are not known[34]. The two key components of BayesOpt are (1) a probabilistic surrogate model that is used to approximate the objective function and (2) an acquisition function that leverages this surrogate model to decide the next point

to sample. A particularly advantageous feature of BayesOpt is that the method can be used to maximize virtually any objective function, even if the underlying functional form is not known or does not exist. BayesOpt is also well-suited for objective functions that are expensive to evaluate since the objective function is only evaluated for points that the surrogate model predicts to be most useful.

In order to explore whether BayesOpt could be used to optimize the performance of SCFT and FTS, we now return to the simple two species homopolymer blend previously considered in Sec. 3.1. In particular, we are interested in whether BayesOpt can automatically locate the optimal relaxation coefficients $\boldsymbol{\lambda}^*$ that we previously located using brute force sampling. Fig. 3A shows a representative evolution of a BayesOpt-driven search. We find that BayesOpt quickly locates the region of optimal performance in approximately 10 iterations and then samples this region until the globally optimal relaxation coefficients $\boldsymbol{\lambda}^*$ are found. In order to quantify the performance of BayesOpt, we compare the best performing $\boldsymbol{\lambda}$ found by BayesOpt to those obtained using random sampling (Fig. 3B). We see that BayesOpt reliably locates high-performing relaxation coefficients $\boldsymbol{\lambda}^*$ in 20 evaluations and results in SCFT simulations that are 30 times faster than $\boldsymbol{\lambda}$ chosen at random.

While the performance achieved using BayesOpt is quite impressive, it is important to emphasize that the choice of hyperparameters can have a significant effect on BayesOpt performance. In general, we found that the performance of naive BayesOpt implementations were quite poor and that good performance required the careful tuning of the surrogate model kernel, the acquisition function and the objective function. The details of our specific BayesOpt implementation are described in Sec. 2.3 and in the Supporting Information. It is also important to note that there are many strategies that could be used to generate relaxation coefficients in addition to BayesOpt and random sampling. We have chosen to compare these two methods in order to give a baseline measure of performance and not an exhaustive comparison of different methods.

3.4 Bayesian optimization in systems with many species

Based on these promising initial results for BayesOpt in a two-species system, we next analyze whether BayesOpt can be used to optimize the relaxation coefficients in systems with many species. Specifically, we consider homopolymer blends containing two, three, four and five unique chemical species and examine how BayesOpt selects their relaxation coefficients in SCFT (Fig. 4). Even though this is a relatively simple system only capable of macrophase separation, we expect that our results will generalize to other systems of higher complexity where microphase separation is present (c.f. Fig. 1B). The specific interaction parameters $\boldsymbol{\chi}$ for these systems are given in the Supporting Information.

In order to quantify the performance of BayesOpt for these different systems, we once again examine the best performing $\boldsymbol{\lambda}$ found by BayesOpt relative to random at each SCFT evaluation (Fig. 4A,B,C,D). For each of these systems, the relaxation coefficients optimized by BayesOpt converge considerably faster than
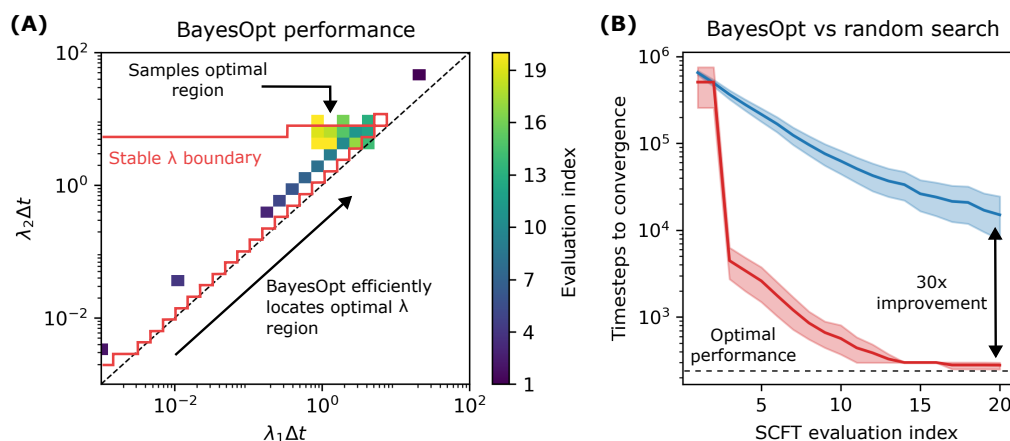
Fig. 3 Selection of relaxation coefficients $\boldsymbol{\lambda} = \{\lambda_1, \lambda_2\}$ using Bayesian optimization (BayesOpt) for SCFT of a two-species homopolymer blend. (A) BayesOpt rapidly identifies optimal regions of relaxation coefficients $\boldsymbol{\lambda}$ in fewer than 10 iterations and subsequently focuses its sampling in this region to locate the best performing $\boldsymbol{\lambda}$. (B) Comparison of BayesOpt and random sampling. BayesOpt locates relaxation coefficients that are considerably faster than those found with random sampling. The shaded regions are 95% confidence intervals.

those obtained with random sampling. In general, we find that the advantages of BayesOpt tend to increase with the number of species, with a 190-fold performance improvement of BayesOpt relative to random for the five-species system. This result indicates that BayesOpt can efficiently navigate the complexities of systems that contain many species and that it can continue to locate relaxation coefficients with good SCFT performance.

Another metric that we use to quantify BayesOpt's performance is the fraction of total SCFT simulations that are non-convergent. As demonstrated in Fig. 1 and Fig. 2, it is common for the vast majority of relaxation coefficients $\boldsymbol{\lambda}$ to not yield a convergent SCFT solution. Thus in addition to finding $\boldsymbol{\lambda}$ that have good SCFT performance (Fig. 4A,B,C,D), BayesOpt quickly finds $\boldsymbol{\lambda}$ that are able to converge. While random sampling tends to find $\boldsymbol{\lambda}$ of which $> 75\%$ are non-convergent, BayesOpt is able to focus its sampling on $\boldsymbol{\lambda}$ of which $< 30\%$ are non-convergent on average (Fig. 4E,F,G,H). It is also noteworthy that the fraction of non-convergent simulations tends to decrease throughout the course of a BayesOpt-driven search, resulting in simulations where only $\approx 10\%$ are non-convergent for a five-species homopolymer blend. Thus, in addition to finding relaxation coefficients with good performance, BayesOpt is also able to focus sampling on those relaxation coefficients that yield convergent SCFT solutions.

We next examine whether BayesOpt can be used to optimize the relaxation coefficients $\boldsymbol{\lambda}$ in FTS for systems with many species (Fig. 5). As we observed for SCFT, BayesOpt performs well for two- to five-species systems and results in 7- to 21-fold performance improvements relative to random sampling (Fig. 5A,B,C,D). BayesOpt also similarly locates convergent relaxation coefficients much more efficiently than random, with non-convergent fractions of $\approx 40\%$ in BayesOpt on average versus $> 85\%$ for random (Fig. 5E,F,G,H).

It is also illustrative to compare the performance of BayesOpt between SCFT (Fig. 4) and FTS (Fig. 5). In general, BayesOpt leads to more significant performance gains relative to random in SCFT than in FTS. One reason for this difference is a super-

ficial artifact from how we have defined the correlation time of non-convergent trajectories in FTS. Since these three-dimensional FTS are considerably more expensive than these one-dimensional SCFT, we set the maximum number of timesteps to be considerably lower in FTS than in SCFT (5×10^4 versus 1.5×10^6). One direct consequence of this choice is that the maximum stable correlation time possible in our FTS is 1250 timesteps and non-convergent FTS trajectories are therefore assigned a correlation time of 1250 timesteps (see Sec. 2 and Supporting Information). Furthermore, for computational efficiency, relaxation coefficients with correlation times exceeding this threshold are labeled as non-convergent. This choice limits the amount of information that BayesOpt can gain from each evaluation of FTS, and also sets a ceiling for the maximum correlation time which leads to more modest improvements in FTS (e.g. 7-21x) than in SCFT (e.g. 30-190x) relative to random. While we could increase the maximum number of timesteps in FTS, we feel that our choice is computationally prudent because it avoids wasting resources on relaxation coefficients with correlation times are exceedingly large. In summary, our choice to truncate inefficient relaxation coefficients improves the absolute performance (i.e. walltime) of BayesOpt FTS at the expense of penalizing the relative performance of BayesOpt FTS compared to random.

Another important difference between BayesOpt in SCFT and FTS is the role of stochasticity in the evaluation of simulation performance. In SCFT, simulation performance is quantified by the number of timesteps to convergence, which is a deterministic quantity. In contrast, simulation performance in FTS is quantified using correlation time which is a stochastic quantity that will depend on the realization of the random noise described in Eq. 7. We anticipate that accounting for this noise in our BayesOpt implementation could improve performance[34,35], but for simplicity we do not pursue these improvements here.

As an additional test for our BayesOpt implementation we consider two different systems. These systems represent the most complex systems we have examined thus far and serve to rig-
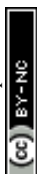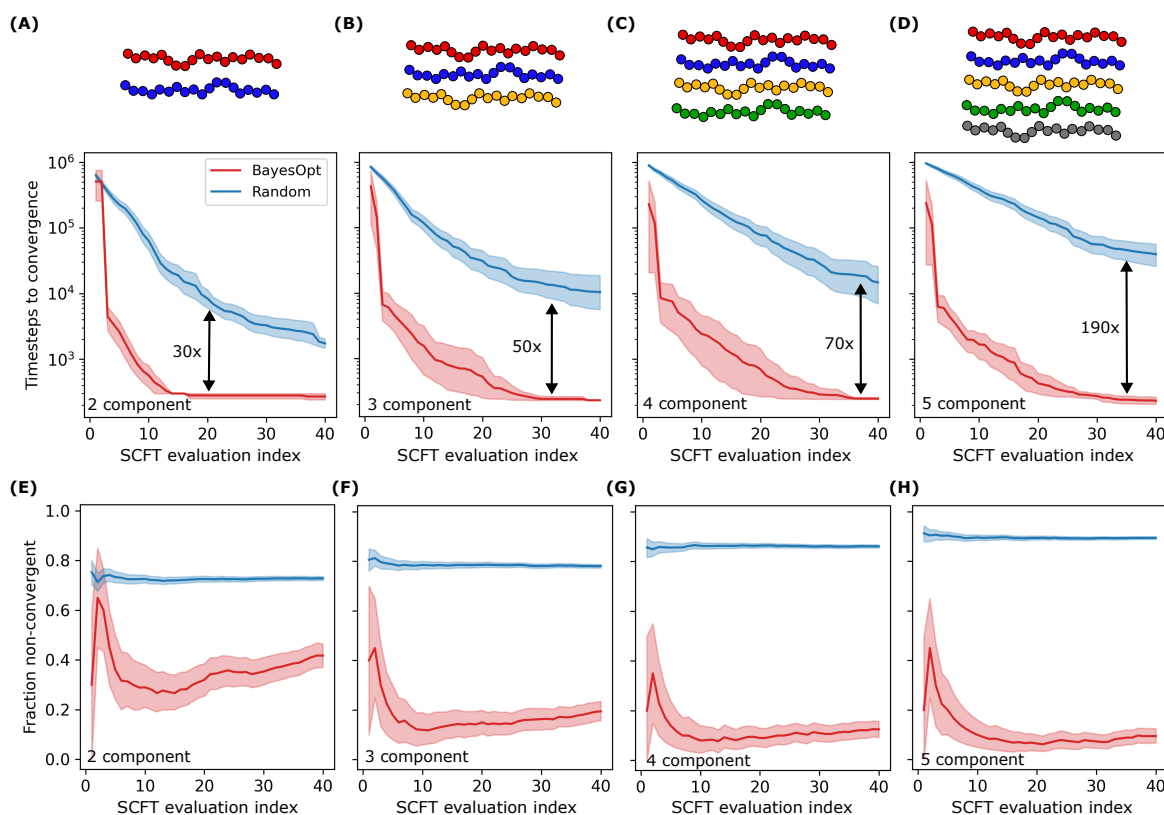
Fig. 4 BayesOpt enhances SCFT performance in multi-species homopolymer blends. (A) Comparison of BayesOpt and random sampling for two-species, (B) three-species, (C) four-species and (D) five-species homopolymer blends. BayesOpt substantially outperforms random sampling and achieves a 190-fold improvement for the five-species homopolymer blend. This improvement is calculated at the first evaluation where BayesOpt is within 10% of its best value. Cartoons above each panel depict the molecular composition of the corresponding systems. (E) Fraction of non-convergent relaxation coefficients λ as a function of SCFT evaluation for two-species (F) three-species, (G) four-species and (H) five-species homopolymer blends. BayesOpt reduces the sampling of non-convergent λ across all systems. In contrast, random sampling frequently samples non-convergent regions, especially as the number of species increase. The shaded regions are 95% confidence intervals.

orously test whether BayesOpt can optimize relaxation coefficients under increased system complexity. For the first system, we choose a five-component mixture containing homopolymer chains with 100 or 150 discrete beads in an explicit solvent (details in Supporting Information). The asymmetric chain lengths coupled with the explicit solvent make it very difficult to obtain convergent FTS relaxation coefficients.

As before, we compare BayesOpt to random sampling over many independent replicas (Fig. 6A). Whereas the performance of randomly sampled relaxation coefficients is relatively constant near the maximum correlation time of 1250 timesteps, BayesOpt identifies relaxation coefficients with correlation times < 100 timesteps after only 10 FTS evaluations. On average, this corresponds to a 33-fold performance improvement relative to random. It is also noteworthy that convergent relaxation coefficients are very difficult to locate for this system as > 95% of relaxation coefficients sampled randomly are non-convergent (Fig. 6C). In contrast, we find that BayesOpt is able to focus its sampling towards more stable relaxation coefficients, typically achieving non-convergent rates of $\approx 50\%$ after its first 10 evaluations. Thus, whereas randomly sampled relaxation coefficients are completely ineffective at generating stable FTS for this system, BayesOpt is

able to robustly locate stable relaxation coefficients with good performance.

For the second system, we examine a ten-component SCFT homopolymer blend in which each species is present at equal volume fraction and each homopolymer consists of 100 discrete beads. This system represents the highest dimensional search space considered in this study and poses a significant challenge for our optimization method due to the rapid increase in possible λ as the number of species increases. As the number of species increases beyond five, the computational and memory demands of BayesOpt become prohibitive due our Gaussian process regression surrogate model. This is because the number of potential relaxation coefficients λ that BayesOpt needs to estimate grows exponentially with the number of species. Without a more efficient strategy for navigating the increasingly large search space, BayesOpt is computationally intractable for this system due to both time and memory constraints. To address these constraints, we develop a modified BayesOpt workflow, referred to as Adaptive BayesOpt, where the resolution is lowered but the range of the search space is dynamically updated as BayesOpt samples λ (see Sec. 2 and Supporting Information).

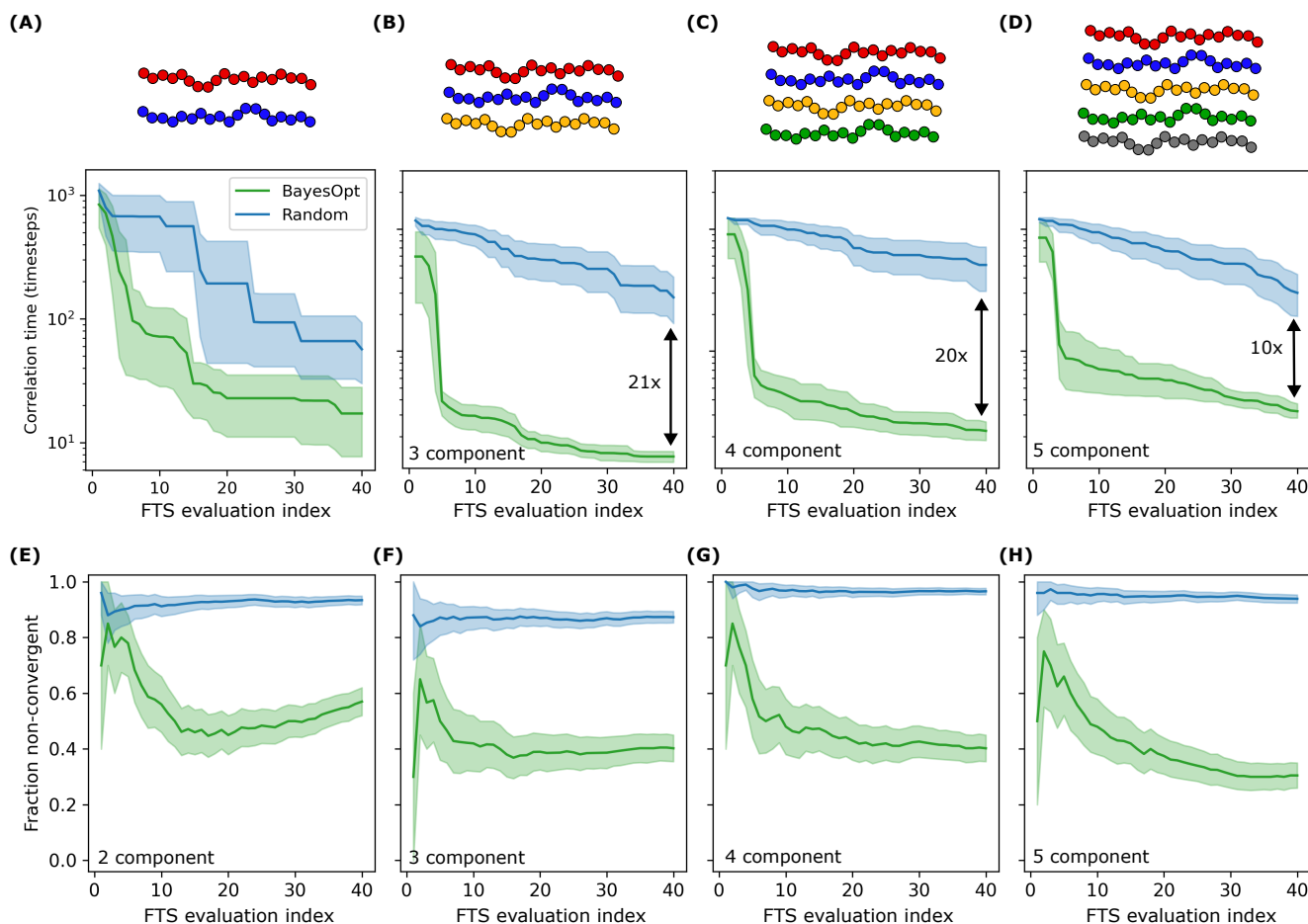We apply Adaptive BayesOpt to this system and compare its

Fig. 5 BayesOpt enhances FTS performance in multi-species homopolymer blends. (A) Comparison of BayesOpt and random sampling for two-species, (B) three-species, (C) four-species and (D) five-species homopolymer blends. BayesOpt significantly reduces correlation times compared to random sampling, despite the inherent stochastic noise present in FTS. Cartoons above each panel depict the molecular composition of the systems. (E-F) Fraction of non-convergent relaxation coefficients λ as a function of FTS evaluation for two-species (F) three-species, (G) four-species and (H) five-species homopolymer blends. BayesOpt decreases the likelihood of evaluating non-convergent λ, while random sampling frequently evaluates unstable regions, especially as number of species increase. The shaded regions are 95% confidence intervals.

performance to random sampling (Fig. 6B). As in the previous examples, we track the best performing λ found after each evaluation. For random sampling, performance remains relatively poor as new λ are evaluated, with the best performing λ yielding only modest improvements in SCFT performance. In contrast, Adaptive BayesOpt quickly identifies λ with substantially better performance. Adaptive BayesOpt then refines and improves performance over subsequent evaluations and ultimately locates λ with a 125-fold improvement in simulation performance relative to random. As with the five-component system, this improvement is not only due to BayesOpt finding better performing relaxation coefficients, but also due to a reduction in the fraction of non-convergent evaluations. As shown in Fig. 6D, random sampling continues to produce a high proportion of non-convergent λ ($> 80\%$), whereas Adaptive BayesOpt instead prioritizes sampling in convergent regions of the search space ($< 10\%$). Taken together, these results demonstrate that even in multi-species settings where the search space becomes prohibitively large, BayesOpt can still efficiently identify high-performing and

stable relaxation coefficients.

3.5 Transferability of relaxation coefficients

Another useful test for the relaxation coefficients obtained using BayesOpt is to examine how these relaxation coefficients perform for different systems and conditions. For example, when performing SCFT or FTS, it is typically of interest to perform simulations for many different interactions parameters χ not just a single value of χ as we have examined so far in this work. Moreover, the benefits of our BayesOpt approach would be significantly diminished if a new BayesOpt calculation was required for every new set of interaction parameters or system conditions. Thus it is useful to examine how the relaxation coefficients obtained using BayesOpt for a single χ perform for interaction parameters that are different.

To investigate the transferability of our BayesOpt relaxation coefficients, we focus on the five-species homopolymer blend examined previously in Fig. 4D,H and Fig. 5D,H. The original interaction parameters for this system used with BayesOpt are denoted
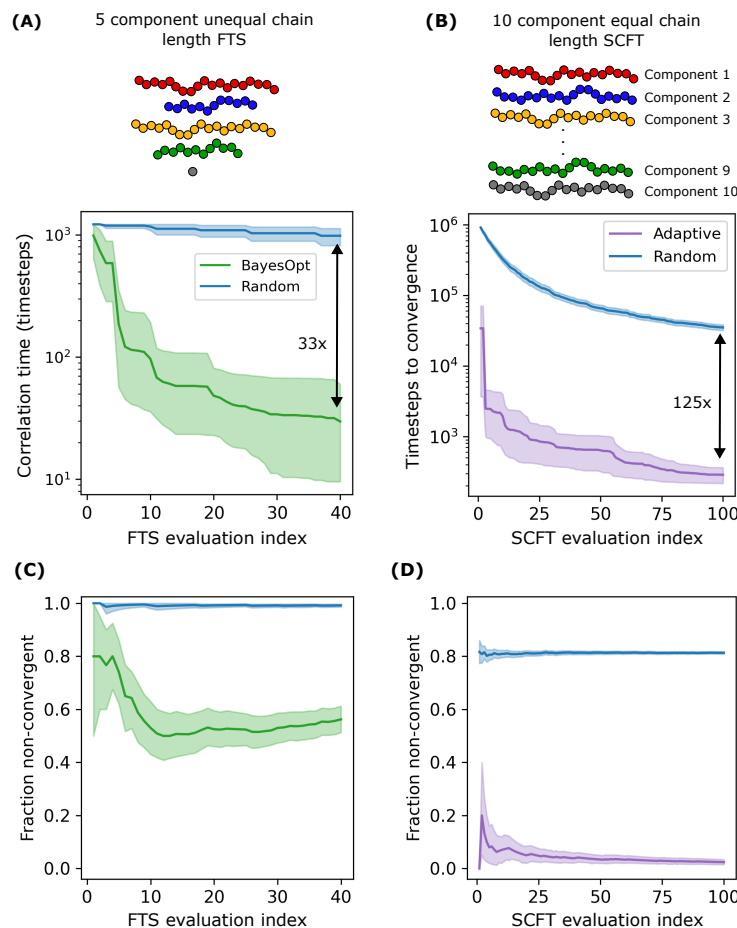
Fig. 6 BayesOpt performance in systems of increased complexity. Comparison of BayesOpt and random sampling of (A) FTS for a five-component system with homopolymer chains of varying length in explicit solvent and (B) SCFT for a ten-component homopolymer blend. Cartoons above each panel depict the molecular composition of the systems. Even for these complex multi-species systems, BayesOpt effectively identifies high-performing relaxation coefficients $\boldsymbol{\lambda}$. (C) Fraction of non-convergent relaxation coefficients evaluated with FTS of the five-component system. BayesOpt reduces the likelihood of sampling non-convergent relaxation parameters compared to random sampling, which maintains a consistently high failure rate close to 100% across all evaluations. (D) Fraction of non-convergent relaxation coefficients evaluated with SCFT for the ten-component homopolymer blend. Adaptive BayesOpt reaches a negligible non-convergent fraction while random sampling continues to frequently evaluate unstable relaxation coefficients throughout the optimization process.

by $\boldsymbol{\chi}_0 N$ and are given in Table S5. In order to calculate the transferability of the optimal $\boldsymbol{\lambda}^*$ found by BayesOpt, we first define a perturbation magnitude ε that parameterizes the deviation of the new interaction parameters $\tilde{\boldsymbol{\chi}}N$, relative to their original values $\boldsymbol{\chi}_0 N$. Specifically, $\tilde{\boldsymbol{\chi}}N$ is obtained by multiplying each element of $\boldsymbol{\chi}_0 N$ by a random factor between $1 - \varepsilon$ and $1 + \varepsilon$. When examining a range of different ε, we select values of $\tilde{\boldsymbol{\chi}}N$ that are unique to a specific value of ε and that cannot be generated using other values. In order to obtain statistics for different values of ε, we choose five different values of $\boldsymbol{\lambda}^*$ obtained from independent BayesOpt calculations and generate 40 random realizations of $\tilde{\boldsymbol{\chi}}N$.

We first examine whether the optimal $\boldsymbol{\lambda}^*$ yield convergent SCFT simulations for different values of ε (Fig. 7A). The fraction of non-convergent trajectories is observed to increase with ε and suggests a rather poor transferability of $\boldsymbol{\lambda}^*$, especially when ε is large. We hypothesize that this low transferability emerges because the optimal relaxation coefficients located by BayesOpt

are deep inside the narrow optimal performance region discussed previously for two- and three-species systems in Fig. 2A and Fig. 3A, respectively. Though this optimal performance region results in exceptional performance for $\boldsymbol{\chi}_0 N$ (Fig. 4D), it leads to poor transferability to the new values of $\tilde{\boldsymbol{\chi}}N$.

One natural solution to this problem is to instead modify $\boldsymbol{\lambda}^*$ to slightly decrease performance in order to improve transferability. A simple approach to achieve this goal is to project $\boldsymbol{\lambda}^*$ onto the vector given by the square root of eigenvalue magnitudes $|\Lambda_i|^{1/2}$ for the unperturbed $\boldsymbol{\chi}_0 N$, compute the norm of this projection, and multiply the unit vector corresponding to $|\Lambda_i|^{1/2}$ for the perturbed $\tilde{\boldsymbol{\chi}}N$ by half this norm. This procedure to obtain a projected $\boldsymbol{\lambda}^*$ effectively uses the original $\boldsymbol{\lambda}^*$ to obtain a proportionality constant that is then used to set the relaxation coefficients for the new $\tilde{\boldsymbol{\chi}}N$. We observe that these projected $\boldsymbol{\lambda}^*$ exhibit excellent transferability and result in stable SCFT simulations even for $\varepsilon = 1$. The performance of the projected $\boldsymbol{\lambda}^*$ are also comparable to the performance obtained if the original $\boldsymbol{\lambda}^*$ are used (Fig. 7B). These
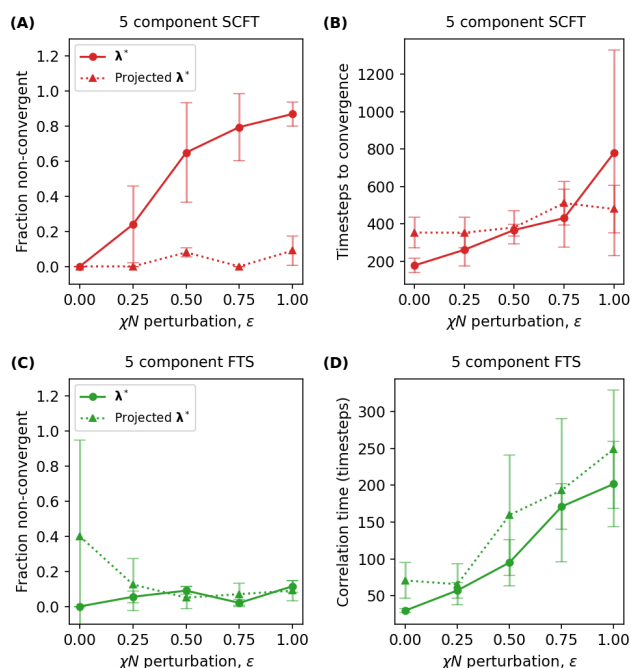
Fig. 7 Transferability of optimal BayesOpt relaxation coefficients $\boldsymbol{\lambda}^*$ to different interaction parameters. (A) Fraction of non-convergent SCFT simulations for different magnitudes of $\boldsymbol{\chi} N$ perturbation and (B) corresponding SCFT simulation performance. (C) Fraction of non-convergent FTS for different magnitudes of $\boldsymbol{\chi} N$ perturbation and (D) corresponding FTS performance. The definition of ε and the procedure use to obtain the projected $\boldsymbol{\lambda}^*$ are described in the text.

results indicate that our simple projection strategy is very effective at producing relaxation coefficients for SCFT that are both high-performing and convergent for a wide range of different interaction parameters.

We next examine how the FTS performance of the $\boldsymbol{\lambda}^*$ obtained using BayesOpt transfer to different values of $\tilde{\boldsymbol{\chi}} N$. In contrast to SCFT, we observe that the optimal $\boldsymbol{\lambda}^*$ result in a low fraction of non-convergent FTS simulations across all values of ε (Fig. 7C). Moreover, the correlation time in FTS only increases slightly with respect to ε and still yields relatively low correlation times even when ε is large (Fig. 7D). We attribute the excellent transferability of $\boldsymbol{\lambda}^*$ in FTS to the relatively broad region of optimal performance in FTS, especially relative to SCFT (c.f. Fig. 1A,C). Thus, even though BayesOpt leads to more modest performance gains in FTS relative to SCFT (Fig. 4, Fig. 5), the transferability of $\boldsymbol{\lambda}^*$ to different interaction parameters is quite good.

Taken together, the results presented in Fig. 7 demonstrate that the relaxation coefficients obtained using BayesOpt can result in convergent and high-performing SCFT and FTS calculations at different values of interaction parameters. We have also examined the transferability of relaxation coefficients to different polymer volume fractions and observe similar results (Fig. S8). These results demonstrate that a single BayesOpt calculation can be used to obtain relaxation coefficients that are broadly useful across a range of different systems and conditions.

Conclusions

This work introduces a general framework for optimizing relaxation coefficients ($\boldsymbol{\lambda}$) in SCFT and FTS of multi-species polymer systems. We show that simulation performance is highly sensitive to relaxation coefficients, with optimal choices yielding orders of magnitude improvements in performance. We also observe that the effects of relaxation coefficients can be system-specific and that they are difficult to anticipate or predict, especially as the number of species increases.

To address these challenges, we develop a Bayesian optimization (BayesOpt) workflow tailored to optimize SCFT and FTS. Our results demonstrate that BayesOpt consistently outperforms random sampling and rapidly identifies high-performing relaxation coefficients while avoiding non-convergent regions of the search space. We show that our approach can optimize over high-dimensional systems with ten or more species that would otherwise be intractable. We also show that the relaxation coefficients obtained using BayesOpt transfer reasonably well to different systems and conditions. Furthermore, our approach is easily extended to other field update algorithms[29], alternative theoretical frameworks for performing multi-species FTS[38] and is compatible with other recent numerical strategies for stabilizing FTS[39].

Though our study has focused on BayesOpt, there are numerous other optimization algorithms that could also be used to identify relaxation coefficients. While we have not examined alternatives to BayesOpt in this work, there are several general criteria that should be satisfied in order for an algorithm to be suitable for relaxation coefficient optimization. The first and most important criteria is for the algorithm to minimize the number of required evaluations, especially for FTS where each evaluation can be very computationally expensive. Another criteria is for the algorithm to be able to locate the optimal relaxation coefficients despite their proximity to large non-convergent regions. These two criteria suggest that any successful algorithm should exhibit a multi-scale character: it should first make large changes to the relaxation coefficients in order to locate the optimal region and then it should focus on sampling this region despite the occurrence of non-convergent simulations. Our BayesOpt implementation is able to achieve this (c.f. Fig. 3A) and we expect that other successful algorithms for relaxation coefficient optimization will do the same.

It is also important to note that there are numerous methods to locate the SCFT field configurations other than solving the fictitious dynamics (i.e. Eq. 7 with $\eta_i = 0$) used throughout this work. One prominent and powerful example is Anderson mixing[40-42] which is implemented in the open-source PSCF code[43,44]. Vigil, Delaney and Fredrickson recently compared the performance of SCFT using either fictitious dynamics or Anderson mixing and observed that both methods had similar performance so long as each method was properly tuned[29]. Since this past work only considered two-species systems, it would be very interesting to compare how the performance of our optimal relaxation coefficients compare to Anderson mixing when the number of species become large.

Another interesting extension of our work would be to perform

BayesOpt using a relatively loose convergence tolerance in SCFT. Since our initial results suggest that the relative performance of different relaxation coefficients is largely independent of convergence tolerance (Fig. S9), it could be possible to identify λ^* more quickly by first performing BayesOpt using a loose tolerance and then performing production SCFT simulations with a tolerance that is tighter. This could dramatically accelerate the 10–100 SCFT evaluations typically required for BayesOpt to find the optimal relaxation coefficients, especially if three-dimensional SCFT simulations are required.

Taken together, our work represents an advance in the automation and acceleration of field-based simulations, especially when many species are present. By reducing the need for manual tuning, our BayesOpt method broadens the practical scope of SCFT and FTS and paves the way for high-throughput and chemically specific simulations in soft matter and biomolecular physics.

Conflicts of interest

There are no conflicts to declare.

Data availability

Data for this article, including raw data used to generate all figures, are available at Zenodo at [URL to be determined]

Acknowledgements

Notes and references

1 G. Fredrickson, *The Equilibrium Theory of Inhomogeneous Polymers*, Oxford University Press, Oxford, 2006.

2 M. W. Matsen, *Journal of Chemical Physics*, 2020, **152**, 110901.

3 G. H. Fredrickson and K. T. Delaney, *Field-Theoretic Simulations in Soft Matter and Quantum Fluids*, Oxford University Press, 1st edn, 2023.

4 G. H. Fredrickson and K. T. Delaney, *Proceedings of the National Academy of Sciences*, 2022, **119**, e2201804119.

5 G. H. Fredrickson, V. Ganesan and F. Drolet, *Macromolecules*, 2002, **35**, 16–39.

6 Y. O. Popov, J. Lee and G. H. Fredrickson, *J. Polym. Sci. B Polym. Phys.*, 2007, **45**, 3223–3230.

7 J. Lequieu, *J. Chem. Phys.*, 2023, **158**, 244902.

8 J. Lequieu, *Macromolecules*, 2024, **57**, 10870–10884.

9 V. Ganesan and G. H. Fredrickson, *Europhysics Letters*, 2001, **55**, 814–820.

10 D. Düchs, V. Ganesan, G. H. Fredrickson and F. Schmid, *Macromolecules*, 2003, **36**, 9237–9248.

11 K. T. Delaney and G. H. Fredrickson, *J. Phys. Chem. B*, 2016, **120**, 7615–7634.

12 Y. X. Liu, K. T. Delaney and G. H. Fredrickson, *Macromolecules*, 2017, **50**, 6263–6272.

13 R. K. Spencer and M. W. Matsen, *Journal of Chemical Physics*, 2018, **148**, 204907.

14 M. W. Bates, J. Lequieu, S. M. Barbon, R. M. Lewis, K. T. Delaney, A. Anastasaki, C. J. Hawker, G. H. Fredrickson and C. M. Bates, *Proc. Nat. Acad. Sci. USA.*, 2019, **116**, 13194–13199.

15 M. W. Matsen, T. M. Beardsley and J. D. Willis, *Phys. Rev. Lett.*, 2023, **130**, 248101.

16 J. D. Willis and M. W. Matsen, *The Journal of Chemical Physics*, 2024, **160**, 024906.

17 M. W. Matsen, J. D. Willis and T. M. Beardsley, *Macromolecules*, 2024, **57**, 4312–4322.

18 J. Lee, Y. O. Popov and G. H. Fredrickson, *J. Chem. Phys.*, 2008, **128**, 224908.

19 R. A. Riggleman, R. Kumar and G. H. Fredrickson, *Journal of Chemical Physics*, 2012, **136**, 024903.

20 K. T. Delaney and G. H. Fredrickson, *J. Chem. Phys.*, 2017, **146**, 224902.

21 N. Sherck, K. Shen, M. Nguyen, B. Yoo, S. Köhler, J. C. Speros, K. T. Delaney, M. S. Shell and G. H. Fredrickson, *ACS Macro Letters*, 2021, **10**, 576–583.

22 M. Nguyen, N. Sherck, K. Shen, C. E. Edwards, B. Yoo, S. Köhler, J. C. Speros, M. E. Helgeson, K. T. Delaney, M. S. Shell and G. H. Fredrickson, *Macromolecules*, 2022, **55**, 9868–9879.

23 K. Shen, M. Nguyen, N. Sherck, B. Yoo, S. Köhler, J. Speros, K. T. Delaney, M. S. Shell and G. H. Fredrickson, *Journal of Colloid And Interface Science*, 2023, **638**, 84–98.

24 My. V. T. Nguyen, K. Dolph, K. T. Delaney, K. Shen, N. Sherck, S. Köhler, R. Gupta, M. B. Francis, M. S. Shell and G. H. Fredrickson, *The Journal of Chemical Physics*, 2023, **159**, 244904.

25 M. V. T. Nguyen, N. Sherck, S. Köhler, E. Schreiner, R. Gupta, G. H. Fredrickson and M. S. Shell, *Biomacromolecules*, 2024, **25**, 5809–5818.

26 C. Li, E. A. Murphy, S. J. Skala, K. T. Delaney, C. J. Hawker, M. S. Shell and G. H. Fredrickson, *J. Am. Chem. Soc.*, 2024, **146**, 29751–29758.

27 E. K. Pert, P. J. Hurst, R. M. Waymouth and G. M. Rotskoff, *The Journal of Chemical Physics*, 2025, **162**, 074902.

28 D. Düchs, K. T. Delaney and G. H. Fredrickson, *J. Chem. Phys.*, 2014, **141**, 174103.

29 D. L. Vigil, K. T. Delaney and G. H. Fredrickson, *Macromolecules*, 2021, **54**, 9804–9814.

30 Z. G. Wang, *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 2010, **81**, 1–12.

31 M. C. Villet and G. H. Fredrickson, *J. Chem. Phys.*, 2014, **141**, 224115.

32 G. Parisi, *Physics Letters B*, 1983, **131**, 393–395.

33 J. R. Klauder, *Physical Review A*, 1984, **29**, 2036–2047.

34 P. I. Frazier, *A Tutorial on Bayesian Optimization*, 2018.

35 X. Wang, Y. Jin, S. Schmitt and M. Olhofer, *ACM Comput. Surv.*, 2023, **55**, 1–36.

36 C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*, MIT Press, 2006.

37 E. Brochu, V. M. Cora and N. de Freitas, *A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning*, 2010.

38 D. Morse, D. Yong and K. Chen, *Macromolecules*, 2025, **58**, 816–825.

39 J. D. Willis and M. W. Matsen, *The Journal of Chemical Physics*, 2024, **161**, 244903.

40 R. B. Thompson, K. O. Rasmussen and T. Lookman, *The Journal of Chemical Physics*, 2004, **120**, 31–34.

41 M. W. Matsen, *Eur. Phys. J. E*, 2009, **30**, year.

42 P. Stasiak and M. W. Matsen, *Eur. Phys. J. E*, 2011, **34**, year.

43 A. Arora, J. Qin, D. C. Morse, K. T. Delaney, G. H. Fredrickson, F. S. Bates and K. D. Dorfman, *Macromolecules*, 2016, **49**, 4675–4690.

44 G. K. Cheong, A. Chawla, D. C. Morse and K. D. Dorfman, *Eur. Phys. J. E*, 2020, **43**, 15.

Data availability statement

Data for this article, including raw data used to generate all figures, are available at Zenodo at [URL to be determined]