University of California
Santa Barbara

Heuristic algorithms for agnostically identifying the globally stable and competitive metastable morphologies of block copolymer melts

A dissertation submitted in partial satisfaction
of the requirements for the degree

Doctor of Philosophy
in
Chemistry

by

Carol Leanne Tsai

Committee in charge:

Professor Glenn Fredrickson, Co-Chair
Professor Craig Hawker, Co-Chair
Professor Joan Shea
Professor Frank Brown
Professor Songi Han

December 2018

The Dissertation of Carol Leanne Tsai is approved.

Professor Joan Shea

Professor Frank Brown

Professor Songi Han

Professor Craig Hawker, Committee Co-Chair

Professor Glenn Fredrickson, Committee Co-Chair

November 2018

Heuristic algorithms for agnostically identifying the globally stable and competitive metastable morphologies of block copolymer melts

by

Carol Leanne Tsai

For my mother, who taught me the value and necessity of education.

Acknowledgements

First and foremost, I want to thank my advisor, Glenn Fredrickson, for being a truly wonderful and patient human being and guide. I truly believe he is the best advisor I could have possibly had. Glenn has always given me enormous freedom in pursuing my education and in my research, while also always being available to offer great advice and guidance as necessary. This kind of balance is, to my understanding, incredibly rare, and highly sought-after in the academic world, and I feel incredibly fortunate that I had the privilege of benefiting from it.

I think most people I have met in grad school have struggled in some way at one point or another. Some might have felt like they weren't good enough, while others may have lost steam at various times, out of frustration with their projects or perhaps a kind of ennui. Some people had partners, families who they were trying to build a life with, or care for in some capacity. Others still had the devastating experience of losing loved ones during their time in school, leaving them to wonder whether they had shared enough good time, used their moments wisely. Some people whose work and life otherwise seemed to be going well were met very suddenly with life-threatening illness. I guess what I am trying to say, is that the rest of life does not suddenly get put on hold just because one is in graduate school.

During my years here I have met some superposition of these experiences. I saw other people endure such adversity as well, and I was always so impressed with the grace, dignity, and strength with which people approached their respective challenges. I was also floored, every time something "went awry" in my life by the incredible support and encouragement Glenn and so many people at this institution offered. I am grateful, because every time I thought I couldn't go any further, Glenn and all these other people encouraged me to continue putting one foot in front of another.

I am so fortunate that the good people in my life are too numerous to name, and

from various facets of my world. For being so welcoming during my rotation with Glenn and beyond, and the camaraderie during the 15-21 hour days I was often pulling in my first year, I am deeply grateful to my office mates from the "6 pack," especially Dr.("Mustache") Michael Gaultois, Dr. Christopher Freeze, and Dr. Evgeny Mikhail. I also want to thank Professor Scott Shell, Professor Ram Seshadri, Professor Matthew Fisher, Professor Songi Han, Professor Joan Shea, and Professor Frank Brown, for their outstanding instruction and kind advice, especially in my first couple of years.

I cannot begin to say thank you enough to the various colleagues whose impact have been educational, stimulating, real, and indelible in so many ways. I learned so much from my rich interactions with the various members of the study groups I have participated in over the years: Dr. Sean Cray, Dr. Xingkun Man, Dr. Wei Li, Dr. Nabil Laachi, Dr. Tatsuhiro Iwama, Dr. Michael Carrilli, Yassine Dhane, Dr. Sean Paradiso, Dr. Mihir Khadilkar, Jonathan Martin, Dr. Jimmy Liu, Dr. Scott Danielson, Jan Garcia, Daniel Vigil, Dr. Mukul Tikekar, Dr. Douglas Tree, and Dr. Joshua Lequieu. I would like to acknowledge particularly the following group members who have shared laugher, frustration, and wonderful conversation with me over the years: Dr. Corinne Carpenter, Dr. Kate Barteau, and Dr. Douglas Grzetic.

I am also grateful for my time mentoring Sean Friedowitz, who needed very little guidance and often seemed to be teaching me things!

I feel incredibly indebted to Dr. Kris Delaney, who has the patience of a saint, a boundless enthusiasm for science that fed mine, the uncanny ability to repair any confusion I had (and there was plenty of it!), and just the right encouraging words at any given time. I would also like to thank Nathan "Fuzzy" Rogers, the Research Computing Administrator at the MRL, whose humor, patience, and computing wisdom I always deeply appreciated. Additionally, the fantastic staff at the MRL and the Chemistry Department made my life at UCSB flow so much more easily, especially

Sara Bard, Sylvia Vogel, and Ericka James.

I would also like to thank the many wonderful professors and colleagues I encountered at the UCSB music department.

In my life outside of the academic setting, I am grateful to Jon Nathan, for the opportunity to coach the many wonderful students of the musical theater departments at UCSB and Santa Barbara High School. I also am thankful to the churches I sang for and belonged to during my stay in Santa Barbara. I appreciate the wonderful people I got to make music with each week for so many years: especially the various members of the SB Mission Schola under the direction of Roy Spicer. The fabulous Nelson Huber and the kind Temmo Korishelli at All Saints By The Sea have also enhanced my musical and spiritual life tremendously.

I am also indebted to the professionals at Acacia Counseling and Wellness, CAPS, and CARE for their tremendous support. Their work is very important, and often unacknowledged because of the circumstances surrounding their efforts at assisting members of the university.

I deeply appreciate the many remarkable friends I have had the privilege of knowing here in Santa Barbara, particularly Dr. Nancy Eisenmenger (and her family), Kristen Moore (and her family), Nick Bedi (and his family), Dr. Caitlin McDowell, and Daniil Bochkov for their unconditional and unceasing care and support, especially in the final year of my program, and for all of the wonderful adventures we have shared over the years. I believe I would not have survived 2018 without them.

To Hui Jia, Hannah Wolfe, Dr. Maher Tahan, Emily Fujimoto, Dr. Alex Morriss-Andrews, Dr. Dominic Else, Neeraj Kumar, Andrew Rowberg, Dr. Alexandra Miller, Abe Pressman, Dr. Michael Swift, Dr. Lucas Brady, Nicholas Adamski, David McCarthy, Muna Saber, Katharine Dickson, Dr. Ryan Barnes, Dr. Tracy Chuong, Dr. Bi Nguyen, and Naomi Merer: thanks for all of the adventures, heart-to-hearts, meals,

and camaraderie.

I need to thank three amazing women in my life. First, my MM Vocal Performance professor, Dr. Isabel Bayrakdarian, for the immense, positive, and enduring impact she has had on my music, my worldview, and my life in general. I will be eternally grateful to her for this. Secondly, Dr. Claudia Mazzotti, who in many ways has been a maternal figure to me, and I imagine to many other students in the MRL who are involved with ScienceLine. Claudia's boundless, unconditional love and wisdom were exactly what I needed at very critical points in the past several years. I would also like to thank Chantal Evrard, whose life, kindness, and efforts at true peace have inspired me to be the best version of myself that I can be.

I would also like to thank people from my life before UCSB. I would never have even considered grad school if not for the fantastic mentoring I received from my undergraduate advisor, Professor Douglas Tobias and the members of his group. I am grateful for my conversations with my freshman physics instructors, Professor Roger McWilliams and Dr. Anne Kirkby, as well as my choir director Dr. Joseph Huszti, who inspired me to approach my work with zest and appreciation. I believe I received a very fine education from being a part of the Campuswide Honors Program at UC Irvine, and was privileged to receive an equally fine education from my K-12 years in the Tustin Unified School District. In particular, I thank my high school orchestra director, Jim Kollias, who told us every morning that we should find something we love and keep doing it for as long as we can, and my AP English teacher, Ms. Ryan, who believed in and encouraged me to no end. I also thank my middle school art teacher, Ms. Lori Spiak, and my many music instructors growing up, who taught me the importance of channeling the beauty and grit of life in forms which speak to other people. I am a firm believer in the importance of early education, and do not think I would have been able to even consider some of the academic goals I have without

the foundation which my early teachers supplied me.

With regard to my life pre-UCSB: I don't know where I would be without the consistent encouragement, love, and humor of Jillian Ma, Christopher Larsen, Yilin Hsu-Wendtland, Betty Lee, Nader Heidari, Pio Dumatol, Dr. Sarah Massatt (and the Massatt family), and Dr. Kevin Slagle (and the Slagle family). These are all people I have known for at least a decade (in some cases almost 20 years), and they have each had an extensive and unique influence on my perspective.

There are many other people I have not listed who have been part of the village it takes to survive in grad school and life in general. I appreciate all the ways they have affected and inspired me.

I thank my brother, for the years we shared growing up. He is one of the most insightful people I know; he can look at a situation and understand it in remarkable depth, and that has always amazed me.

Lastly, I want to thank my mother. Growing up in a single parent household was tough but she always made sure we had a roof over our heads and food on the table, and that we understood that education was so necessary, for our enlightenment and also as a way to a better life for ourselves. She taught me how to focus on the abundant good that life has to offer, to be resilient when difficult times come, and to make the most of whatever is in front of you, which is always plenty, and always enough to share with others. She is proud of how far she's come from two suitcases and $50 in her pocket back in 1976. I am ineffably proud of her as well, and hope that I've been able to do all of her years of hard work and sacrifice some justice.

Curriculum Vitæ

Carol Leanne Tsai

Education

2018 Ph.D. in Physical Chemistry (Expected), University of California, Santa Barbara.

2018 M.M. in Vocal Performance, University of California, Santa Barbara.

2011 B.S. in Chemistry, University of California, Irvine, with departmental Honors, *Cum Laude*

2011 B.S. in Biology, Minor in Earth System Science, University of California, Irvine, *Cum Laude*

Awards and Honors

Graduate

National Science Foundation Graduate Research Fellowship	2013
UCSB Special Regents Fellowship	2012
UCSB Chemistry Department Teaching Award	2016
UCSB Scienceline Outreach Awards	2013-2015, 2017, 2018

Undergraduate

Phi Beta Kappa (ΦBK) Society	2011
Phi Lambda Upsilon (ΦΛϒ) Chemistry Honors Society	2011
UCI Regents Scholarship	2007-2011
Campuswide Honors Program at UCI	2007-2011
Dean's List, UCI	2007-2011
Irvine Company Student Leadership Scholarship	2007
Literary Authors Guild Scholarship	2007

Publications

C.L. Tsai, K.T. Delaney, G.H. Fredrickson "Genetic Algorithm for Discovery of Globally Stable Phases in Block Copolymers." *Macromolecules.* **2016**, *49* (17), pp 6558–6567

C.L. Tsai, S. Friedowitz, A.B. Chang, B. Kim, R.J. Macfarlane, K.T. Delaney, R.H. Grubbs, G.H. Fredrickson. "Non-lamellar morphologies in bottlebrush diblock copolymers." *In preparation.*

C.L. Tsai, K.T. Delaney, G.H. Fredrickson. "Using particle swarm optimization and SCFT to agnostically determine the stable morphologies of block copolymer melts." *In preparation.*

Teaching Experience and Outreach

ScienceLine — UCSB Materials Research Laboratory, Outreach Program
ScienceLine Responder — 2012 – 2018

SciTrek — UCSB Chemistry, Outreach Program
Volunteer — 2012–2013

Teaching Assistant for Undergraduate Physical Chemistry — UCSB Chemistry Department
Physical Chemistry 112B (Statistical Mechanics) — Winter Quarter 2016

Teaching Assistant for Computational Chemistry — UCSB Chemistry Department
Physical Chemistry 145/245 (Computational Chemistry) — Spring Quarter 2015

Teaching Assistant for Physical Chemistry — UCSB Chemistry Department
Physical Chemistry 112A (Thermodynamics) — Fall Quarter 2013

Teaching Assistant for Introductory Voice Lessons — UCI Music Department
Music 15 (Intro to Voice) — Spring Quarter 2011

Tutor at UCI Learning and Academic Resource Center — UCI Music Department
LARC Organic Chemistry Tutor — Spring Quarter 2010–2011

Soprano Section Leader at All Saints by the Sea Church — 2017–2018

Soprano Section Leader at Santa Barbara Mission — 2013–2017

Soprano Section Leader at Santa Barbara Choral Society — 2012–2013

Abstract

Heuristic algorithms for agnostically identifying the globally stable and competitive metastable morphologies of block copolymer melts

by

Carol Leanne Tsai

Block copolymers are composed of chemically distinct polymer chains that can be covalently linked in a variety of sequences and architectures. They are ubiquitous as ingredients of consumer products and also have applications in advanced plastics, drug delivery, advanced membranes, and next generation nano-lithographic patterning. The wide spectrum of possible block copolymer applications is a consequence of block copolymer self-assembly into periodic, meso-scale morphologies as a function of varying block composition and architecture in both melt and solution states, and the broad spectrum of physical properties that such mesophases afford.

Materials exploration and discovery has traditionally been pursued through an iterative process between experimental and theoretical/computational collaborations. This process is often implemented in a trial-and-error fashion, and from the computational perspective of generating phase diagrams, usually requires some existing knowledge about the competitive phases for a given system. Self-Consistent Field Theory (SCFT) simulations have proven to be both qualitatively and quantitatively accurate in the determination, or forward mapping, of block copolymer phases of a given system. However, it is possible to miss candidates. This is because SCFT simulations are highly dependent on their initial configurations, and the ability to map phase diagrams requires a priori knowledge of what the competing candidate morphologies are. The unguided search for the stable phase of a block copolymer

of a given composition and architecture is a problem of global optimization. SCFT by itself is a local optimization method, so we can combine it with population-based heuristic algorithms geared at global optimization to facilitate forward mapping. In this dissertation, we discuss the development of two such methods: Genetic Algorithm + SCFT (GA-SCFT) and Particle Swarm Optimization + SCFT (PSO-SCFT). Both methods allow a population of configurations to explore the space associated with the numerous states accessible to a block copolymer of a given composition and architecture.

GA-SCFT is a real-space method in which a population of SCFT field configurations "evolves" over time. This is achieved by initializing the population randomly, allowing the configurations to relax to local basins of attraction using SCFT simulations, then selecting fit members (lower free energy structures) to recombine their fields and undergo mutations to generate a new "generation" of structures that iterate through this process. We present results from benchmark testing of this GA-SCFT technique on the canonical AB diblock copolymer melt, for which the theoretical phase diagram has long been established. The GA-SCFT algorithm successfully predicts many of the conventional mesophases from random initial conditions in large, 3-dimensional simulation cells, including hexagonally-packed cylinders, BCC-packed spheres, and lamellae, over a broad composition range and weak to moderate segregation strength. However, the GA-SCFT method is currently not effective at discovery of network phases, such as the Double-Gyroid (GYR) structure.

PSO-SCFT is a reciprocal space approach in which Fourier components of SCFT fields near the principal shell are manipulated. Effectively, PSO-SCFT facilitates the search through a space of reciprocal-space SCFT seeds which yield a variety of morphologies. Using intensive free energy as a fitness metric by which to compare these morphologies, the PSO-SCFT methodology allows us to agnostically identify

low-lying competitive and stable morphologies. We present results for applying PSO-SCFT to conformationally symmetric diblock copolymers and and a miktoarm star polymer, AB_4, which offers a rich variety of competing sphere structures. Unlike the GA-SCFT method we previously presented, PSO-SCFT successfully predicts the double gyroid morphology in the AB-diblock. Furthermore, PSO-SCFT successfully recovers the the A_{15} morphology at a composition where it is expected to be stable in the miktoarm system, as well as several competitive metastable candidates, and a new sphere morphology belonging to the hexagonal space group 191, which has not been seen before in polymer systems. Thus, we believe the PSO-SCFT method provides a promising platform for screening for competitive structures in a given block copolymer system.

Contents

Chapter 1

Introduction

1.1 Block Copolymers

The influence of synthetic polymers on modern life is extensive, pervasive, and nearly inescapable. Everywhere one looks, one can find their mark: teflon in non-stick cooking pans, nylons and polyesters in fabrics and textiles, the gel soles of shoes, the rubber treads of tires, plastics for packaging of all kinds of consumer products, the materials encasing electronic equipment, as constituents of drug delivery vehicles, and so much more. Within the realm of synthetic polymers exists a special category called block copolymers, which feature desirable properties that could lend themselves to next-generation, high performance materials.

Block copolymers are composed of chemically distinct polymer chains that can be covalently linked in a variety of sequences and architectures. They are ubiquitous as ingredients of consumer products and also have applications in advanced plastics [1, 2], drug delivery [3, 4], advanced membranes [5, 6] and next generation nanolithographic patterning [7, 8, 9, 10]. The wide spectrum of possible block copolymer applications is a consequence of block copolymer self-assembly into periodic, meso-

scale morphologies as a function of varying block composition and architecture in both melt and solution states, and the broad spectrum of physical properties that such mesophases afford[2]. In the past few decades, tremendous advances in synthetic polymer chemistry have allowed for the controlled synthesis of polymers with well-defined architectures. By tuning the chemical composition and architecture of block copolymers, one can tune the self-assembly of a diverse array of periodic, meso-scale ordered morphologies that can be exploited in a broad spectrum of applications, including those listed above. With each new chemical and architectural system comes the possibility of new morphologies, and with them a variety of possibly desirable materials properties. In theory, there is an endless playground of new materials to discover and characterize.

It is easy to see that there is an abundance of potential opportunity in block copolymers. In an age with so many technological advancements, it is imperative to use theoretical and computational tools in tandem with experimental approaches to understand block copolymers, and thereby facilitate and expedite material discovery. The studies which comprise this thesis cumulatively aspire to this goal.

Being able to predict which microphase is expected to be stable, or even which microphases are competitive candidates for a given composition and architecture of a block copolymer is an important part of materials design. We begin with a brief introduction to one of the current methods, Self-Consistent Field Theory, which people employ to perform this "forward mapping," then discuss challenges of this method, and suggest ways to address these issues through the development of heuristic algorithms that fall generally under the category of global optimization strategies, which we will see are necessary for solving the forward problem.

1.2 Self-Consistent Field Theory (SCFT)

There is a diverse collection of simulation methods available to study block copolymers. All of the methods involve some way of sampling statistical configurations accessible to the polymer chains, but vary in the mathematical approaches and physical interpretations they use to do so. One common simulation approach is to use particle-based methods, such as molecular dynamics (MD), in which one accounts for the interactions between all segments of all polymer chains [11, 12]. Another class of methods involves using field-based approaches such as Self-Consistent Field Theory (SCFT) or Complex Langevin (CL), which can be constructed from a variety of chain models [13]. Monte Carlo (MC) methods may be applied to both particle-based and field-based methods[14]. Coarse-grained molecular dynamics and Monte Carlo simulations are typically efficient for low molecular weight polymers, but are less so for the denser polymer melts we will be interested in simulating in this dissertation. Furthermore, even at a coarse-grained level MD and MC simulations can be prohibitively expensive, as they are often limited by the size and relaxation timescales necessary for studying polymer systems. Although SCFT is a non-fluctuating simulation method, it is quantitatively accurate for dense polymer melts. If necessary, fluctuation corrections beyond SCFT can be included using complex Langevin sampling [15, 16]. However, CL simulations are costly and often unnecessary for a simple exploration of candidate mesophases and their stabilities, particularly far from the Order-Disorder Transition (ODT), where we will be primarily concerned.

The results presented in any given study in this thesis will be based on SCFT simulations. Individual SCFT parameters used for each study will be delineated within the section or subsection to which it belongs. The remainder of this subsection will serve as an introduction for SCFT.

When we build models to study systems at the meso-scale, we are often interested in being able to accurately account for statistical mechanical quantities such as the partition function, which sums over all of the statistical configurations accessible to the system and allows us to connect microscopic quantities to thermodynamic variables of interest, such as free energies, pressures, or chemical potentials. The partition function of a block copolymer system is often prohibitively difficult to calculate in particle-based models because doing so requires summing over the interactions of all pairs of atoms or coarse-grained polymer segments between all pairs of polymer chains in a given system. The power of SCFT lies in the conversion of a particle-based model to a field theory via an exact mathematical transformation called a Hubbard-Stratonovich Transformation. This mathematical transformation allows us to decouple the many-body particle-particle interactions of the particle-based model by introducing instead one or more auxiliary fields. In a polymer field theory, we only need to consider the interaction of a single chain with these auxiliary fields, which can be interpreted as chemical potential-like fields that are generated by the presence of the chain of interest itself as well as all of the other chains present in the system.

The recipe for building a field theory begins with selecting a particle-based chain model. There are many chain models to choose from, including bead-spring chains, freely-jointed chains, worm-like chains, and gaussian chains, all of which are typically coarse-grained models that ignore the effects of individual atoms. This is not a problem, since the phase behavior of polymer systems of interest are reasonably captured using coarse-grained models. For the work presented in this thesis, we will consider SCFT based on a continuous gaussian chain model. The continuous gaussian chain can be interpreted as the continuum extension of the discrete gaussian chain, in which points along the backbone of the chain are connected using entropic springs. In the continuous gaussian chain, these springs become a continuous penalty for stretching

the chain along its contour.

For this introduction, we will construct a field theory using continuous gaussian chains for homopolymers in implicit good solvent (Model A). It is straightforward to extend this methodology to other models[13]. We begin by defining a pair-potential of mean force for the gaussian chain model is given by:

$$u(\mathbf{r}) = k_B T u_0 \delta(\mathbf{r}) \tag{1.1}$$

with u_0, the excluded volume parameter restricted to values greater than zero, thus restricting Model A to good solvents. If we consider a system of n chains, each of which is N segments long, we can keep track of the segment coordinates as the set of conformations of space curves, $\mathbf{r}^{nN} \equiv \mathbf{r}_j(s)s \in [0, N], j = 1, 2, ..., n$, with s being a contour variable tracing the segment position along a chain. The total interaction potential, $\mathcal{U}[\mathbf{r}^{nN}]$, of such a system can thus be constructed by summing over the contributions from all pairs of segments across all chains. We can partition this into two terms, $U_0[\mathbf{r}^{nN}]$ and $\bar{U}[\mathbf{r}^{nN}]$, which account for the contributions from intramolecular, short-range interferences and the intermolecular interactions between segments and long-range interferences, respectively.

$$\mathcal{U}[\mathbf{r}^{nN}] = U_0[\mathbf{r}^{nN}] + \bar{U}[\mathbf{r}^{nN}] \tag{1.2}$$

with

$$U_0[\mathbf{r}^{nN}] = \frac{3k_B T}{2b^2} \sum_{j=1}^{n} \int_0^N ds \left| \frac{d\mathbf{r}_j(s)}{ds} \right|^2 \tag{1.3}$$

$$\bar{U}[\mathbf{r}^{nN}] \approx \frac{1}{2} \sum_{j=1}^{n} \sum_{k=1}^{n} \int_0^N ds \int_0^N ds' \, u(|\mathbf{r}_j(s) - \mathbf{r}_k(s')|) \tag{1.4}$$

The canonical partition function associated with the continuous gaussian chain can thus be written as the integral over configurations which are weighted by the total interaction energy of the system:

$$\mathcal{Z}_C(n,V,T) = \frac{1}{n!(\lambda_T^3)^{nN}} \prod_{j=1}^{n} \int \mathcal{D}\mathbf{r}_j \exp(-\beta \mathcal{U}[\mathbf{r}^{nN}]) \tag{1.5}$$

At this point, it is useful to define a microscopic segment density operator, $\hat{\rho}$:

$$\hat{\rho}(\mathbf{r}) = \sum_{j=1}^{n} \int_0^N ds\, \delta(\mathbf{r} - \mathbf{r}_j(s)) \tag{1.6}$$

Using this definition, we can rewrite Eqn. 1.4 as:

$$\bar{U}[\mathbf{r}^{nN}] = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}'\, \hat{\rho}(\mathbf{r}) u(|\mathbf{r} - \mathbf{r}'|) \hat{\rho}(\mathbf{r}') - \frac{1}{2} nNu(0) \tag{1.7}$$

It is also useful to note the following representation of a delta functional:

$$\delta[\rho - \hat{\rho}] = \int \mathcal{D}w \, \exp(\, i \int d\mathbf{r}\, w(\mathbf{r})[\rho(\mathbf{r}) - \hat{\rho}(\mathbf{r})]) \tag{1.8}$$

We can use this representation and the definition of a delta functional, $\int \mathcal{D}\rho\, \delta[\rho - \hat{\rho}]F[\rho] = F[\hat{\rho}]$, to insert $F[\rho] = 1$ into to the version of Eqn. 1.5 which uses Eqn. 1.7. By doing so, we perform a Hubbard-Stratonovich Transformation and thus recast our particle-based partition function in terms of a partition function whose action is a functional of auxiliary fields that were introduced by the transformation:

$$\mathcal{Z}_C(n,V,T) = \mathcal{Z}_0 \int \mathcal{D}\rho \int \mathcal{D}w \, \exp(-\mathcal{H}_{eff}[\rho, w]) \tag{1.9}$$

where $\mathcal{Z}_0$ is the ideal gas partition function of n non-interacting continuous gaussian chains. w and ρ can be interpreted respectively as fluctuating chemical potential and density fields. The form of the effective hamiltonian of this field-based version of our

theory will change from one polymer system to another. For a homopolymer in good solvent (Model A), it takes the following form:

$$\mathcal{H}_{A,eff}[\rho, w] = -i \int d\mathbf{r}\, w(\mathbf{r})\rho(\mathbf{r}) + \frac{\beta}{2} \int \mathbf{r} \int \mathbf{r}'\, \rho(\mathbf{r}) u(|\mathbf{r} - \mathbf{r}'|)\rho(\mathbf{r}') - n \ln Q[iw] \quad (1.10)$$

The first term of the hamiltonian accounts for the interactions between the auxiliary fields introduced by the transformation and the second term accounts for particle-particle interactions. This first term also tells us that the chemical potential field w is thermodynamically conjugate to the density field ρ. The Hubbard-Stratonovich Transformation is mathematically exact, and has the effect of decoupling pair-wise interactions among polymer segments and replacing them with interactions between the polymer segments of a single chain with one or more auxiliary fields which are introduced by the transformation. This effect is manifested in the third term of the hamiltonian. This last term contains a quantity called the single-chain partition function, $Q[iw]$, which is a non-local quantity that accounts for the statistical configurations accessible to our gaussian chain in the presence of the chemical potential field generated by the presence of all the chains in the system. It is the fact that we have decoupled the particle (chain segment) degrees of freedom and replaced them with a single chain interacting with auxiliary fields which has made the problem of simulating polymers much more tractable.

However, calculating the single-chain partition function is still an involved process. Doing so requires us to integrate over all of the possible chain configurations, which are captured in a quantity called the chain propagator, q:

$$Q[iw] = \frac{1}{V} \int d\mathbf{r}\, q(\mathbf{r}, N - s; [iw]) q(\mathbf{r}, s; [iw]) \quad (1.11)$$

Here, the chain propagator $q(\mathbf{r}, s; [iw])$ can be interpreted as a reduced distribution function describing the statistical weight associated with a chain which begins with segment 0 and ends on segment s and experiences a chemical potential field iw. Similarly, $q(\mathbf{r}, N-s; [iw])$ describes the statistical weight associated with a chain beginning at segment N and ending at segment s. Here, the path integral for the single-chain partition function has been factored at an arbitrary contour position. By doing this, one can interpret the path integral as the sum over joint-probability distributions associated with the configurations accessible to the chain. Chain propagators are determined by solving the following modified diffusion equation:

$$\frac{\partial}{\partial s} q(\mathbf{r}, s; [iw]) = \frac{b^2}{6} \nabla^2 q(\mathbf{r}, s; [iw]) - iw(\mathbf{r}) q(\mathbf{r}, s; [iw]) \tag{1.12}$$

The modified diffusion equation is the continuum analog of the Chapman-Kolmogorov equation, which relates joint-probability distributions of chain configurations that are constructed by stochastically building the chain one segment at a time.

Thus far we have discussed the development of expressions for quantities of interest for homopolymer in implicitly treated good solvent (Model A), but we can extend these ideas to other systems. For a large part of the work presented in this dissertation, we use continuous gaussian chains for Model E, the incompressible AB-diblock copolymer melt. In Model E, we define the following microscopic segment density operators:

$$\begin{aligned} \hat{\rho}_A(\mathbf{r}) &= \sum_{j=1}^{n} \int_0^{fN} ds\, \delta(\mathbf{r} - \mathbf{r}_j(s)) \\ \hat{\rho}_B(\mathbf{r}) &= \sum_{j=1}^{n} \int_{fN}^{N} ds\, \delta(\mathbf{r} - \mathbf{r}_j(s)) \end{aligned} \tag{1.13}$$

where f indicates the fraction of statistical segments in the A block. By imposing an incompressibility constraint, $\hat{\rho}_+ = \hat{\rho}_A + \hat{\rho}_B$, and a local A-B interaction penalty that goes as $\chi_{AB}\hat{\rho}_A\hat{\rho}_B$, we can determine a field-based effective hamiltonian for Model E analogous to the one in Eqn. 1.10:

$$\mathcal{H}_{E,eff}[w_+, w_-] = \rho_0 \int d\mathbf{r}\,[(1/\chi_{AB})w_-^2 - iw_+] - n \ln Q[w_A, w_B] \tag{1.14}$$

where χ_{AB} is a Flory-Huggins interaction parameter between A- and B-type polymer segments. Now, the auxiliary fields for Model E are complex chemical potential fields w_A and w_B which act separately on A and B polymer segments. They are defined in terms of linear combinations of w_+ and w_-: $w_A \equiv iw_+ - w_-$ and $w_B \equiv iw_+ + w_-$. Here, w fields are still conjugate to microscopic densities ρ, but with a slightly different interpretation from Model A. Now, the w_+ field is conjugate to the sum of microscopic densities $\rho_A + \rho_B$ and enforces local incompressibility, and can be interpreted a pressure-like potential field. The w_- field is instead conjugate to the difference between microscopic densities $\rho_A - \rho_B$ and can thus be interpreted as an exchange chemical potential field. In Model E, the single-chain partition function can be constructed by integrating over the statistical configurations of two quantities respectively called the forward and complementary chain propagators: q and q_c. The forward chain propagator can be obtained by using the initial condition $q(\mathbf{r}, 0; [w_A, w_B])$ and solving the following modified diffusion equation:

$$\frac{\partial}{\partial s} q(\mathbf{r}, s; [w_A, w_B]) = \frac{b(s)^2}{6} \nabla^2 q(\mathbf{r}, s; [w_A, w_B]) - w(\mathbf{r}) q(\mathbf{r}, s; [w_A, w_B]) \tag{1.15}$$

with

$$w(\mathbf{r},s) \equiv \begin{cases} w_A(\mathbf{r}), 0 \leq s \leq fN \\ \\ w_B(\mathbf{r}), fN \leq s \leq N \end{cases} \tag{1.16}$$

In a similar fashion, but constructing the complementary propagator from the other end by using the initial condition $q_c(\mathbf{r}, N; [w_A, w_B])$ instead, q_c can be obtained by solving:

$$\frac{\partial}{\partial s} q_c(\mathbf{r}, s; [w_A, w_B]) = \frac{b_c(s)^2}{6} \nabla^2 q_c(\mathbf{r}, s; [w_A, w_B]) - w_c(\mathbf{r}) q_c(\mathbf{r}, s; [w_A, w_B]) \tag{1.17}$$

with

$$b_c(s) \equiv \begin{cases} b_B(\mathbf{r}), 0 \leq s \leq (1-f)N \\ \\ b_A(\mathbf{r}), (1-f)N \leq s \leq N \end{cases} \tag{1.18}$$

and

$$w_c(\mathbf{r},s) \equiv \begin{cases} w_B(\mathbf{r}), 0 \leq s \leq (1-f)N \\ \\ w_A(\mathbf{r}), (1-f)N \leq s \leq N \end{cases} \tag{1.19}$$

The single chain partition function for Model E can be calculated by solving either Eqn. 1.15 or Eqn. 1.17, since the forward or complementary construction of the chain

propagators should yield statistically equivalent weights for the overall chain:

$$\begin{aligned} Q[w_A, w_B] &= \frac{1}{V} \int d\mathbf{r}\, q(\mathbf{r}, N; [w_A, w_B]) \\ &= \frac{1}{V} \int d\mathbf{r}\, q_c(\mathbf{r}, N; [w_A, w_B]) \end{aligned} \tag{1.20}$$

However, calculating the density operators for Model E would require knowing both chain propagators, since it is necessary to take the joint-probability distributions afforded by the propagators in each direction for each block:

$$\begin{aligned} \rho_A(\mathbf{r}; [w_A, w_B]) &= \frac{1}{VQ[w_A, w_B} \int_0^{fN} ds\, q_c(\mathbf{r}, N-s; [w_A, w_B]) q(\mathbf{r}, s; [w_A, w_B]) \\ \rho_B(\mathbf{r}; [w_A, w_B]) &= \frac{1}{VQ[w_A, w_B} \int_{fN}^{N} ds\, q_c(\mathbf{r}, N-s; [w_A, w_B]) q(\mathbf{r}, s; [w_A, w_B]) \end{aligned} \tag{1.21}$$

The "self-consistent" aspect of SCFT comes from the fact that we are computing the chain configurations for a polymer experiencing w_A and w_B auxiliary fields (as determined by solving Eqn. 1.15 or Eqn. 1.17), while also calculating the configurations of the auxiliary fields themselves. In the work we present, we consider solving for the w fields using a mean-field approximation, rather than a fully fluctuating theory. This means that we enforce the following condition:

$$\left. \frac{\delta \mathcal{H}_{E,eff}[w]}{\delta w(\mathbf{r}} \right|_{w=w^*} = 0 \tag{1.22}$$

The requirement in Eqn. 1.22 amounts to two results: 1) the Hamiltonian we solve for will be at a stationary state with respect to variations in the field configuration $w(\mathbf{r})$, and 2) the evaluation of the fields w at a single field configuration w^* yields a mean-field approximation, which assumes that w^* dominates the functional integral

in the partition function. In the work presented in this thesis, simulations were run in the canonical ensemble, which means that the Helmholtz free energy is equal to the hamiltonian evaluated at the mean field configuration, since $\beta F = -\ln \mathcal{Z}_C = \mathcal{H}[w^*]$.

Fig. 1.1 provides a summary of the SCFT algorithm. An SCFT simulation is initialized with some guess for the w fields, either randomly or with a seed. Densities are updated by first solving the modified diffusion equation to obtain the chain propagators q, then determining the single chain partition function Q, and finally employing Eqn. 1.21. Fields are then updated according to the equations appropriate for a mean field approximation (Eqn. 1.22) or with a fully fluctuating theory (*e.g.*, complex Langevin). The process of updating densities and fields is iterated until convergence criteria are met.

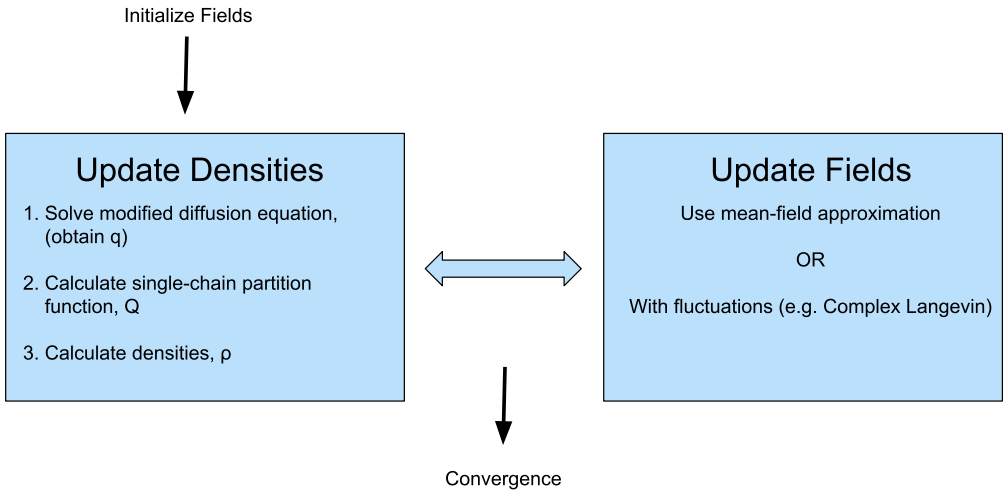

Figure 1.1: Overview of SCFT algorithm. SCFT simulations consist of two main parts: an inner loop which updates the densities, and an outer loop which updates the fields.

1.3 The Forward Problem

Increased complexity in the composition and architecture of block copolymer materials can lead to a wide variety of microstructures that are otherwise not seen in the traditional linear diblock[17, 18, 19, 20, 21, 22, 23]. Materials exploration and discovery has traditionally been pursued through an iterative process between experimental and theoretical/computational collaborations. This process is often implemented in a trial-and-error fashion, and from the computational perspective of generating phase diagrams, usually requires some existing knowledge about the competitive phases for a given system. SCFT simulations have proven to be both qualitatively and quantitatively accurate in the determination, or forward mapping, of block copolymer phases of a given system [24, 25]. However, it is possible to miss candidates. This is because SCFT's ability to map phase diagrams requires *a priori* knowledge of what the competing candidate morphologies are. Phase diagrams are usually constructed by gridding up the phase space of interest, performing seeded SCFT simulations of every competing morphology at every point in the phase space, then comparing the free energies of each morphology. If a candidate is unknown, it will not be entered into the running, and therefore it cannot be determined to be stable. An example of this can be seen even in the linear diblock. The theoretical phase diagram for the conformationally symmetric diblock was initially established in 1994 and originally showed spheres, hexagonally packed cylinders (HEX), double-gyroid (GYR), and lamellar morphologies as the competing candidates of this system [24, 26]. Over the next decade or so, hexagonally perforated lamellae (HPL) and the network phase O^{70} were added as competitors and the latter was even shown to have its own pocket of stability between DG and LAM near the order-disorder transition. Similarly, in recent years, many new exotic sphere morphologies, including Frank-Kasper and Laves phases,

have been observed for the conformationally asymmetric diblock [23, 21, 27, 28]. Enthusiasm for these exotic sphere phases, which are well-established in metallurgy, but are new to the polymer community, has spread and resulted in the hunt for different compositional and architectural systems which could afford these structures larger windows of stability.

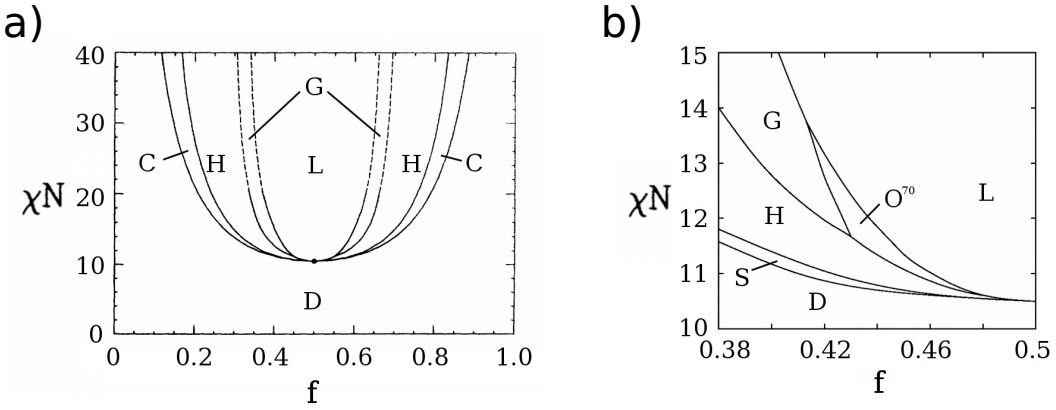

Figure 1.2: AB diblock phase diagram from (a) 1994[24] and b) 2005[26]. The phase diagram in panel (b) shows a window of stability for the Fddd morphology (belonging to the O^{70} space group), which had not previously been considered as a competing candidate.

It is clear then that unguided structure discovery with regard to expected candidates and also with regard to their expected unit cell sizes is an important step in materials discovery and design. Indeed, unguided large-cell real-space structure discovery has been feasible for two-dimensional morphologies since the the late 1990's [29]. Unguided spectral methods for discovering block copolymer phases have been developed as well and are successful at mapping phases for small, three-dimensional simulation cells [30, 31]. However, the unguided determination of globally stable phases with very large repeat units, where there is no intuition for what the lattice parameters and mesophase symmetry should be, remains a challenge [23, 32].

The difficulties in three-dimensional structure discovery emerge as a consequence

of a rough free energy landscape associated with numerous candidate structures even at fixed block copolymer composition and architecture. Local saddle points in such a free energy landscape correspond to defective states and metastable polymorphs of the globally stable structure. This has important ramifications with regard to the limitations of SCFT when used by itself, since an SCFT simulation is heavily dependent on the initial condition with which it is seeded. This means that SCFT simulations can often become trapped in one of the many metastable basins of attraction at a given point in phase space, and thus results in local optimizations, rather than global ones.

There are several methods for overcoming this phenomenon, including approaches such as harmonic filtering, or annealing [33, 34]. These approaches can broadly be categorized into "single-point" methods, which act to improve an individual SCFT field configuration. However, one might instead use population-based methods, which exploit an element of learning among multiple "members" or "agents" that carry information about previously visited configurations and the free energies associated with them. To the best of our knowledge, there has not been previous work that exploits iterative modification of many structures simultaneously to predict globally optimal morphologies and low-lying competitors of block copolymers.

1.4 Mimetic Algorithms

In this dissertation, we explore the use of mimetic algorithms, which combine a popular global optimization strategy of choice with an efficient local optimizer to expedite structure determination in systems of block copolymers of fixed composition and architecture. Specifically, we will discuss the development of two approaches: genetic algorithm + SCFT (GA-SCFT) and particle swarm optimization + SCFT

(PSO-SCFT). In both cases, the idea is to take advantage of SCFT as a local optimizer in conjunction with a global optimization strategy (either the GA or the PSO). In each case, an iteration of the entire algorithm looks like the following (as represented by Fig 1.3):

1. Launch a set of seeded SCFT simulations to relax to some set of local basins of attraction.
2. Compare the relative fitness (intensive free energy) of structures determined by SCFT.
3. Generate a new set of seeds according to either the GA or PSO moves.
4. Initialize the next iteration of SCFT simulations using these updated seeds, thus relaxing to a new set of local basins of attraction, and exploring the free energy landscape.
5. Repeat until the maximum number of iterations is met.

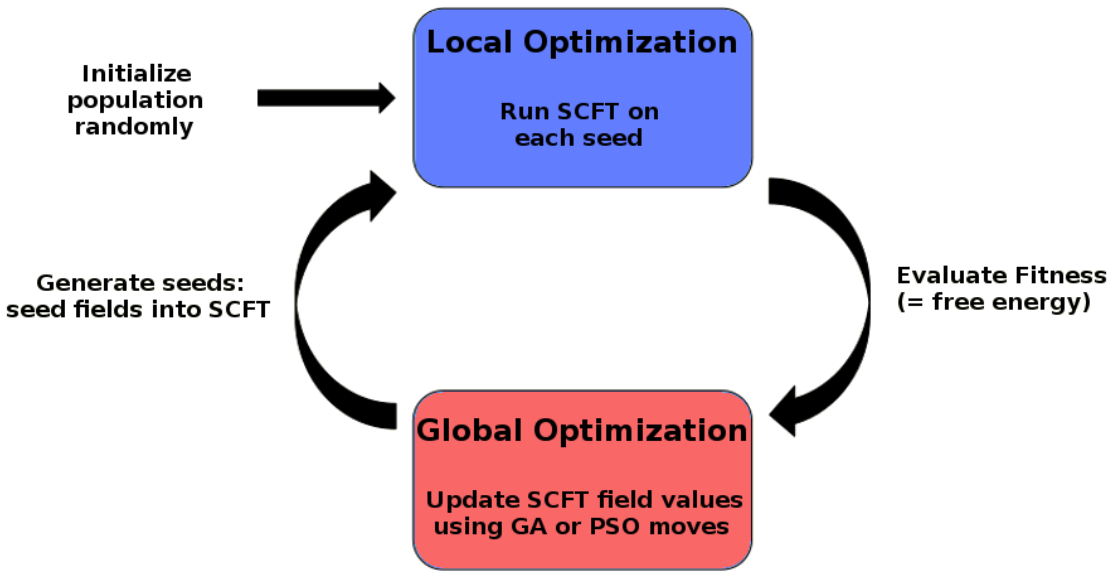

Figure 1.3: Framework for global+local optimizations to identify competitive morphologies of BCP melts.

SCFT is a deterministic simulator, which means that each seed corresponds to a unique structure that is being explored. The stochasticity in each method comes from the moves performed by the global optimizer.

In the GA-SCFT method, we manipulate the **real-space SCFT field values**. For a given SCFT simulation, we employed anywhere between 32^3 and 64^3 plane waves to spatially resolve the simulation, thus leading to on the order of 10^5 real space field values which were being optimized by the GA-SCFT algorithm, which explores the space of real-space SCFT seeds generated by combining the real-space field values from individual simulations.

One way to reduce the degrees of freedom being optimized is by instead manipulating peaks in reciprocal space. Specifically, in the PSO-SCFT method, we are concerned with optimizing the positions of a set of peaks that fall within a thin shell at a radius of roughly the primary wave vector away from the origin **in reciprocal space**. Such a set of peaks comprises "single-mode" seeds which can be locally relaxed using SCFT. Local manipulations of peaks in such a shell in reciprocal space correspond to changing non-local features in real-space. In the PSO-SCFT method, then, the optimization space is reduced to determining the optimal relative positions of peaks in the shell, their amplitudes, and the appropriate simulation cell size. The positions of the peaks are kept track in terms of two angles in spherical coordinates. Each peak also has an amplitude associated with it, and each agent has a value q^* which corresponds to some initial cell size for each SCFT simulation. Moves are applied to peaks in one hemisphere and then mirrored in the other hemisphere. Thus, the total number of degrees of freedom for using 6 peaks in a hemisphere is: $6(2) + 6 + 1 = 19$, since there are two angles per peak, one amplitude for each peak, and one initial cell size associated with each agent (set of peaks). This is good news, because 19 is much less than 10^5! However, we will find that even 19 is a very large

number of degrees of freedom to be optimizing simultaneously.

1.5 Summary of Chapters

This dissertation is divided into two main body chapters. In Chapter 2, we examine results from the GA-SCFT method. We begin by discussing the various elements which comprise a genetic algorithm: selection strategies, crossover methods, mutations, and other miscellaneous features that are unique to the problems that emerge from employing GA in conjunction with SCFT. Results are presented for benchmark testing the GA-SCFT method applied to the AB diblock, for which the theoretical phase diagram is well established. We will see that the difficulty of the forward problem, as well as the ability of the GA-SCFT method to be successful at recovering expected morphologies, will vary with the expected morphology, simulation cell size, and segregation strength.

In Chapter 3, we discuss the development of the PSO-SCFT approach. Switching the algorithmic manipulations to a reciprocal space basis in spherical coordinates gives rise to a variety of considerations which were not necessary to navigate in the real-space GA-SCFT method. We discuss some of these considerations in the early sections of the chapter, then present the PSO-SCFT algorithm itself. We then investigate the application of PSO-SCFT to two systems: the same diblock we used to benchmark the GA-SCFT method, and an AB_4 miktoarm system where exotic sphere morphologies of recent interest have been observed[23, 21, 27, 28]. We then discuss the implementation and results of employing $q*$ as a PSO-SCFT search variable as a means of optimizing simulation cell size. We finish by discussing the characterization of a mystery morphology which was found by the PSO-SCFT method in the miktoarm system.

Finally, we will conclude with some observations about the implications involved in attempting to solve the forward problem, as well as some considerations for future directions with these challenges in mind.

Chapter 2

GA-SCFT

Here we propose the use of a genetic algorithm (GA) in conjunction with SCFT to undertake the forward problem of determining the globally stable structures at fixed compositional and architectural parameters. In each section, we will explore some choices for the various components of a GA and their implementation in the context of GA-SCFT. We will then examine the successes of the GA-SCFT method from benchmark testing performed with the generic diblock copolymer, as well as cases where the method is unsuccessful.

Results for the first several sections will primarily consist of those for 2D SCFT simulations which were studied. Morphologies such as lamellae and hexagonally packed cylinders can be studied using 2D SCFT simulations. Much of the preliminary work towards GA move development was performed using such 2D simulations, since they are less computationally expensive, and many of the GA moves explored are easily extensible from two dimensions to three dimensions. Thus, a few examples of results for 2D simulations as proofs of concept will be provided in the early sections of this chapter. However, 3D SCFT simulations provide a much richer space of structures as well as a more realistic purview of the kinds of challenges one might

encounter in attempts at structure determination. Results in the later sections of this chapter will be presented for the fully-developed GA-SCFT method applied to 3D systems.

2.1 Background and Overview

2.1.1 What are Genetic Algorithms?

Genetic algorithms are a class of biologically-inspired global search heuristics called evolutionary algorithms, which were initially developed in the 1960's and 1970's [35, 36, 37]. GAs have been applied in various engineering, game theoretic, and biophysics optimization problems [35, 38, 39, 40]. More pertinent to our interests, however, is the successful application of GAs to solve crystallographic structure optimization problems in the solid state community. [41, 42, 43, 44]

Genetic algorithms can be used to evolve an initial population of candidate solutions toward fitter solutions. In the context of SCFT, improved fitness implies less defective (often lower free energy) structures. In this work, we only consider free energy as a measure of fitness. Improving fitness is accomplished by iteratively selecting fittest members to survive, reproduce, and undergo mutations to more extensively explore the free energy landscape [42]. By themselves, GAs have proven to be successful in many kinds of optimization problems, although they may take hundreds to thousands of iterations to reach optimal solutions [45]. Mimetic algorithms, in which a global optimization strategy such as a GA is implemented in coordination with an efficient local-optimization technique, appear to converge more rapidly to optimal solutions than unadorned genetic algorithms[46]. Because of this, as well as the robust successes observed in the prediction and discovery of atomic crystal structures, we propose that structure discovery of block copolymers can be acceler-

ated by implementing GAs as a wrapper around a local-optimization method. In our method, we iteratively cycle between performing local optimizations with SCFT and using GA moves on the real-space fields generated by SCFT simulations in order to move between metastable states, *i.e.*, basins of attraction.

2.1.2 GA-SCFT Approach

The hybrid SCFT-GA methodology proceeds by initializing a population of possible structures from random initial conditions using SCFT simulations. We then evaluate the relative fitness of each member of the population by comparing their intensive free energies. Lower-free energy structures are considered to be more fit. Fitter members are statistically more likely to be selected to crossover, or recombine, with one another to generate offspring. Mutations to the offspring involve perturbation of the real-space SCFT fields and occur with some tunable probability. Once the recombination and mutation steps are performed, each individual of this new generation of candidate structures undergoes an SCFT optimization, and the whole process is iterated until non-defective structures emerge. Fig. 2.1 provides a schematic of the algorithm.

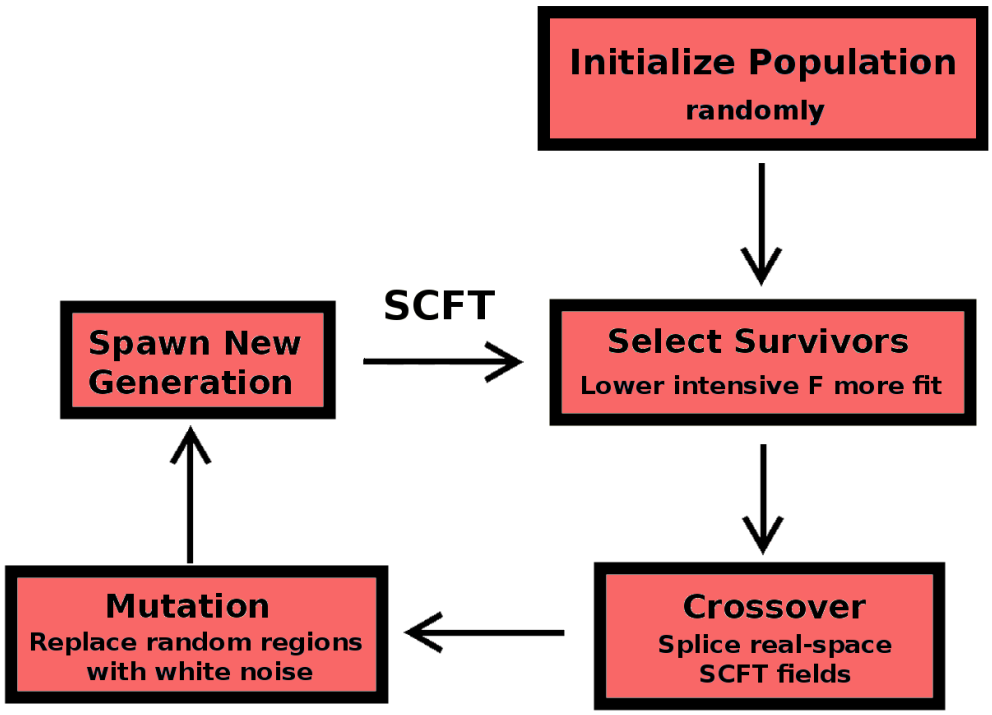

Figure 2.1: GA-SCFT flowchart.

Each SCFT local optimization was not always run to full self consistency; rather, we imposed an upper limit of 20,000 field update iterations. This number was selected from a cursory study performed using full GA simulations for the case of hexagonally packed cylinders in large, cubic cells. Intensive free energy typically changes on the order of 10^{-4} or $10^{-5}\, kT$/chain once coarse features of a morphology are resolved in an SCFT simulation. For hexagonally packed cylinders, 20,000 iterations was identified to strike a balance between capturing the important, larger features in a cell and the unnecessary extra computational expense of continuing to refine finer features which contribute very little to intensive free energy. Indeed, the fittest structure obtained at the end of the GA can be refined to arbitrary accuracy (up to numerical roundoff error in the algorithms) as a post-processing step.

2.1.3 Benchmark Testing Set-up

To test our GA-SCFT method, we applied the algorithm to a conformationally symmetric diblock copolymer melt, for which the theoretical phase diagram is well-understood [24, 47]. This allows us to assess our method, since the expected stable phase is known across various compositions and interaction strengths. In this chapter, we will eventually explore the application of the full GA-SCFT approach in two contexts in Section 2.7: large-cell and small-cell simulations, at high and low χN values. Small-cell simulations, which attempt to capture only a single primitive cell of the (unknown) periodic mesophase structure, are computationally less expensive than large-cell simulations that may contain many primitive cells. However, there are many structures whose primitive cells are very large, such as the recently discovered Frank-Kasper σ phase [23]. Using the GA-SCFT method to successfully discover previously unknown phases in the future would likely require large-cell simulations, since the primitive cell lattice vectors of such phases would not be known *a priori*.

As we show in Section 2.7.2, random SCFT quenches (without GA moves) are more likely to obtain the correct structure if a commensurate cell is used. However, for discovery of unknown mesophases, prior knowledge of the commensurate cell size is unavailable. To compensate for this lack of information, we use cubic large-cell simulations with a first generation initial lattice parameter, L_{init}, of $20\,R_g$ (radius of gyration) along each dimension for our initial GA studies. This number was chosen to capture several periods of the classical phases, while not being exactly commensurate with any morphology's unit-cell size at the compositions we chose for testing; this allowed us to ensure an agnostic starting point for GA-SCFT. Results for full GA simulations using initial cell sizes ranging from $15 - 30\,R_g$ are presented in Section 2.5.2. We will see that as long as the size of the simulation cell was large enough to capture more than one period, the choice of size did not affect whether the correct

χN	f_A	Expected Morphology
15	0.30	BCC
	0.35	HEX
	0.39	GYR
	0.50	LAM
20	0.24	BCC
	0.30	HEX
	0.36	GYR
	0.50	LAM

Table 2.1: GA-SCFT simulation benchmarking compositions for conformationally symmetric diblock copolymers used in the present study.

phase was obtained, but did affect how quickly convergence to the expected structure occurred. Fields for large-cell simulations were sampled with 64^3 collocation mesh points.

Specifically, in Section 2.7, we apply the GA-SCFT approach at two intermediate values of segregation strength, χN: 15 and 20. At each value of χN, the GA-SCFT is tested at diblock compositions deep in the regions where the following phases are expected: BCC-packed spheres (BCC), hexagonally-packed cylinders (HEX), double-gyroid (GYR), and lamellae (LAM). The specific compositions are listed in Table 2.1 and were extracted from Matsen and Schick's phase diagram for conformationally symmetric diblock copolymers [24].

Sections 2.2, 2.3, and 2.4 contain results primarily for large-cell simulations with χN=20 at compositions where HEX and GYR are expected, unless otherwise stated. We will also see that population size is an important feature that must strike a balance between computational expense and maintaining enough diversity in the population in order to have better exploration of the energy landscape.

2.1.4 Considerations for GA Moves

The following three sections delineate the various selection, crossover, and mutation strategies we employed. We performed a series of basic sweeps to determine which strategies to use. However, the methods we selected were not highly optimized via an exhaustive search of the many tunable parameters. It is important that we distinguished two important goals in the development of GA-SCFT: 1) effectiveness of the search, which concerns whether globally stable structures are actually recovered, and 2) efficiency of the search, where we were interested in minimizing computational cost. Ultimately, we would like to balance maximizing effectiveness and efficiency; we chose methods and parameters that appeared from our preliminary studies to provide reasonable effectiveness while also reducing computational cost.

2.2 Selection Methods

2.2.1 Purpose and Options

The method one uses for selection ultimately decides how "genes," or morphological features in the case of block copolymers, are transferred from one generation to the next. On one hand, it is important to select parents so that fitter members of each generation are allowed to "pass on" their "desirable" qualities to the next generation. However, as in the biological analogy, diversity in the genome is imperative for a more thorough exploration of the optimization landscape. One risks missing more fit structures if a population collapses to one configuration too quickly.

There is a wide variety of selection methods available to use in genetic algorithms. We present the results for three strategies in the following subsections: 1) elitist selection, 2) roulette wheel selection, and 3) tournament selection. Details for each method are described in their respective subsections.

Testing of the elitist and roulette wheel selection methods was performed using 2D SCFT simulations that were run at a composition where HEX is the expected morphology, using the features and parameters listed in Table 2.2, and with a variable cell method that allows the cell size to change over the course of an individual SCFT simulation. SCFT simulations using these parameters typically finished in under five minutes. Thus, running the GA-SCFT for 50 generations could be accomplished within a handful of hours using a batch queuing system. A population size of 20 was used for the 2D benchmarking results presented for the elitist and roulette wheel testing. Crossover (splicing of the parent SCFT fields to yield "children" seeds for the next generation of the GA) was performed using circular cuts as described in Section 2.3.

Parameter	Value
χN	20
f_A	0.3
expected morphology	HEX
number of plane waves	128×128 (2D simulation)
chain contour resolution	0.02
field timestep size	5
simulation cell size	40 R_g×40 R_g

Table 2.2: SCFT parameters used in selection methods studies.

The tournament selection strategy as explored as an alternative much later in the development of the GA-SCFT method. Testing for this strategy was applied to 3D SCFT simulations where GYR and BCC were the expected morphologies, consistent with the parameters used in Section 2.7 where the GA-SCFT used a roulette wheel selection strategy. More details on relevant variables tested are provided in the tournament selection subsection.

2.2.2 Elitist Selection

In the elitist selection method, members of a generation are sorted by fitness (intensive free energy) in descending order. Members can be chosen to "mate" with one another by pairing up each one with the next member in the list (*i.e.* 1 with 2, 3 with 4, and so on). Fig 2.2 displays a plot of the intensive free-energy of every member vs. GA-SCFT iteration (generation) (panel (a)), as well as density profiles of the lowest free energy member of generation 1 (panel (b) and that of generation 25 (panel (c)). From the plot in panel (a), we see that initially there is a distribution of intensive free energies, each corresponding to a different defective structure.

There are variations on the elitist method that include pairing members up randomly and either allowing for overlap (one member can mate with many others) or no

overlap (each member can mate with only one other member). Varying the way parents are chosen to recombine with one another does not seem to dramatically change the rate of convergence for the elitist strategy.

As another variation to the elitist selection procedure, a percentage of parents in each generation can be generated randomly by randomly initializing SCFT simulations instead of seeding them with recombined members from the previous generation. In Fig 2.3, we see that seeding 25% of each new generation randomly results in a slower convergence to non-defective structures. Even after 60 generations, we see that this implementation of the GA-SCFT does not converge to perfectly hexagonally-packed cylinders.

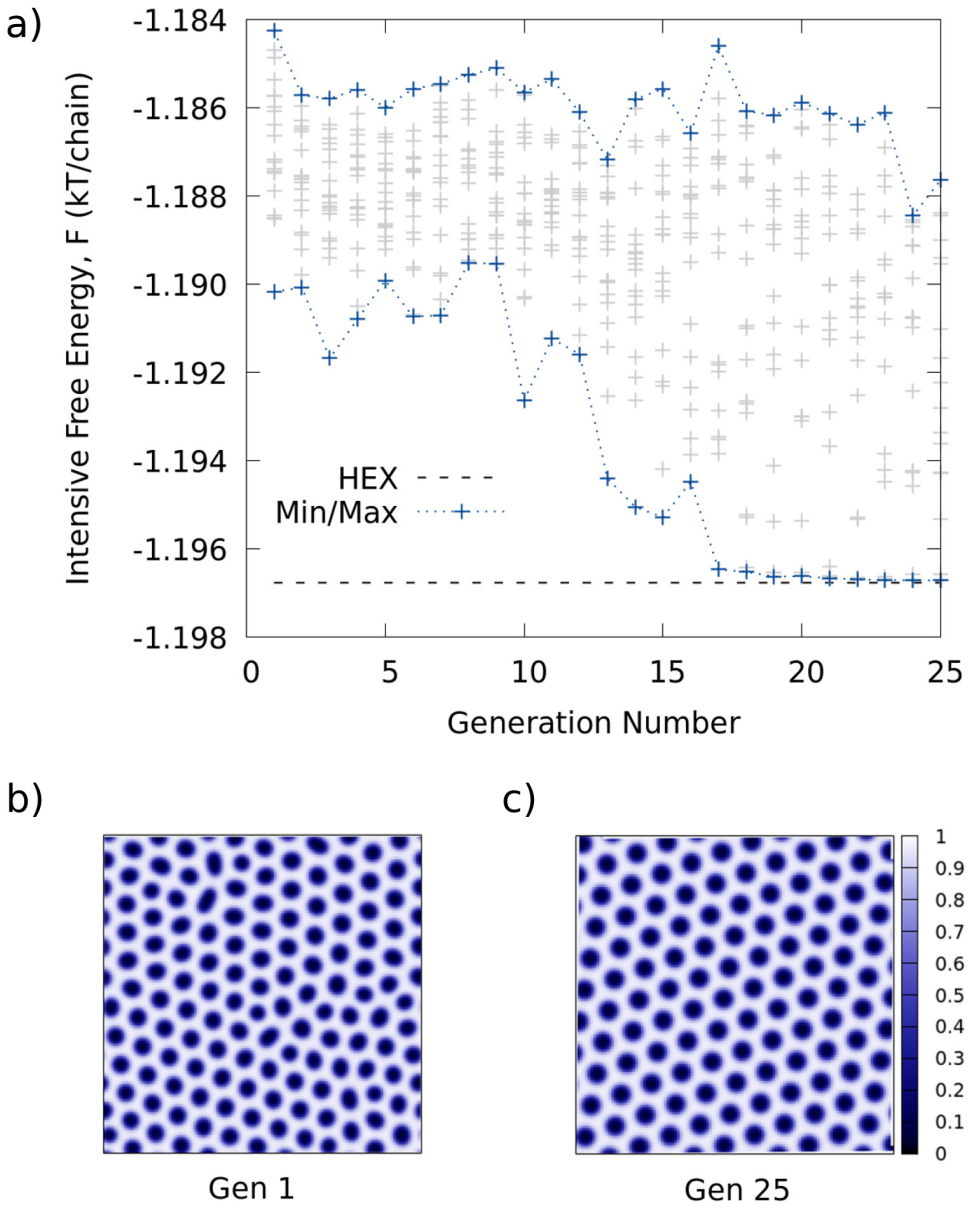

Figure 2.2: Results from 2D benchmark testing of elitist selection strategy for the AB diblock at $f_A = 0.3$, χN=20, where HEX is expected. (a) Intensive free energy vs. GA-SCFT iteration; the intensive free energy of each member corresponds to a gray point, and blue lines indicate the minimum, maximum, and average values of intensive free energy across iterations. (b) Density profile of the A-component for the lowest free energy structure in generation 1. (c) Density profile of the A-component for the lowest free energy structure in generation 25.

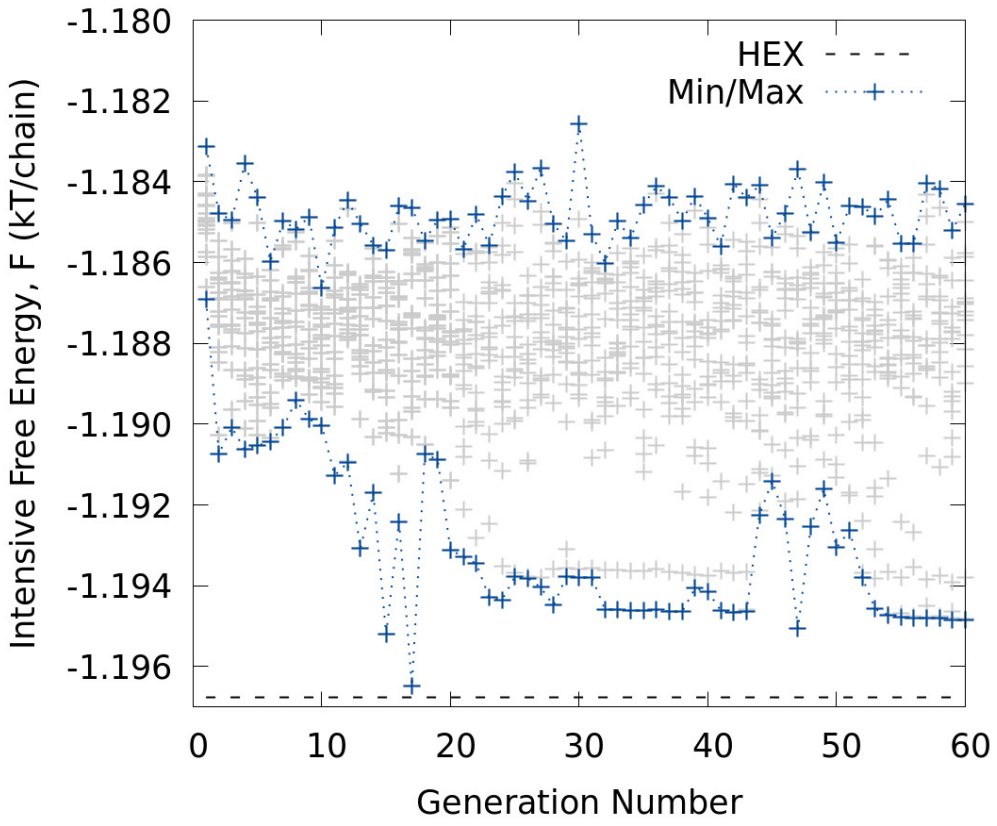

Figure 2.3: Intensive free energy vs. iteration for 2D benchmark testing of elitist selection strategy with 25% of each generation seeded from random initial conditions. Simulations were run at $f_A = 0.3$, χN=20, where HEX is expected to be stable.

2.2.3 Roulette-Wheel Selection

There are a variety of selection strategies other than the elitist method of the previous section which are available for GAs[45, 48]. In this section, we study a roulette-wheel method [49], which is a probabilistic, fitness-proportionate method, meaning that survivors are selected with some tunable probability according to their relative fitness to the rest of the individuals in their generation. The roulette-wheel approach allows very fit candidates to contribute to more than one child structure. However, even these fit structures do not pass through the GA unmodified. In our

method, we do not allow one individual to act as both parents (*i.e.*, there is no "asexual reproduction"), and every contributing parent is subject to the crossover step described in the next subsection.

Two parents are selected for each child member of a following generation. Each parent selection is performed independently (*i.e.*, for a population size of 50, there will be 50 parent selections performed in each generation). The following selection probability is assigned to individual i:

$$\mathcal{P}_i = \frac{e^{\frac{-F_i}{T}}}{\sum_j e^{\frac{-F_j}{T}}}, \tag{2.1}$$

where

$$T = \frac{-F_{low} + F_{high}}{\ln \mathcal{P}_{\text{ratio}}} \tag{2.2}$$

F_{low} and F_{high} refer to the SCFT intensive free-energies of the lowest- and highest-free energy members, respectively. $\mathcal{P}_{ratio}$ is an adjustable parameter that corresponds to a ratio between the probability that the lowest free energy member will be selected and the probability that the highest free energy member will be selected. Thus, in the limit that $\mathcal{P}_{ratio} \to 1$, all members would be equally likely to be selected as a parent in each selection, if one parent was allowed to contribute to the same child. Alternatively, for very large $\mathcal{P}_{ratio}$, the lowest free energy member would always be selected. We tested values of $\mathcal{P}_{ratio}$ ranging from 1 to 10 for a composition where HEX was expected at χN=20.

Preliminary studies performed for compositions where hexagonally packed cylinders should be most stable indicated that the roulette-wheel outperforms a simple elitist strategy. In the elitist strategy, the lowest-free energy member is always selected to survive and other parents are paired from an ordered list sorted by intensive

free energy. In the roulette-wheel strategy, members that are considered to have lower fitness are still allowed to contribute to the next generation, albeit with a low probability. Furthermore, with the roulette-wheel approach, individual members are allowed to engender multiple offspring for the next generation.

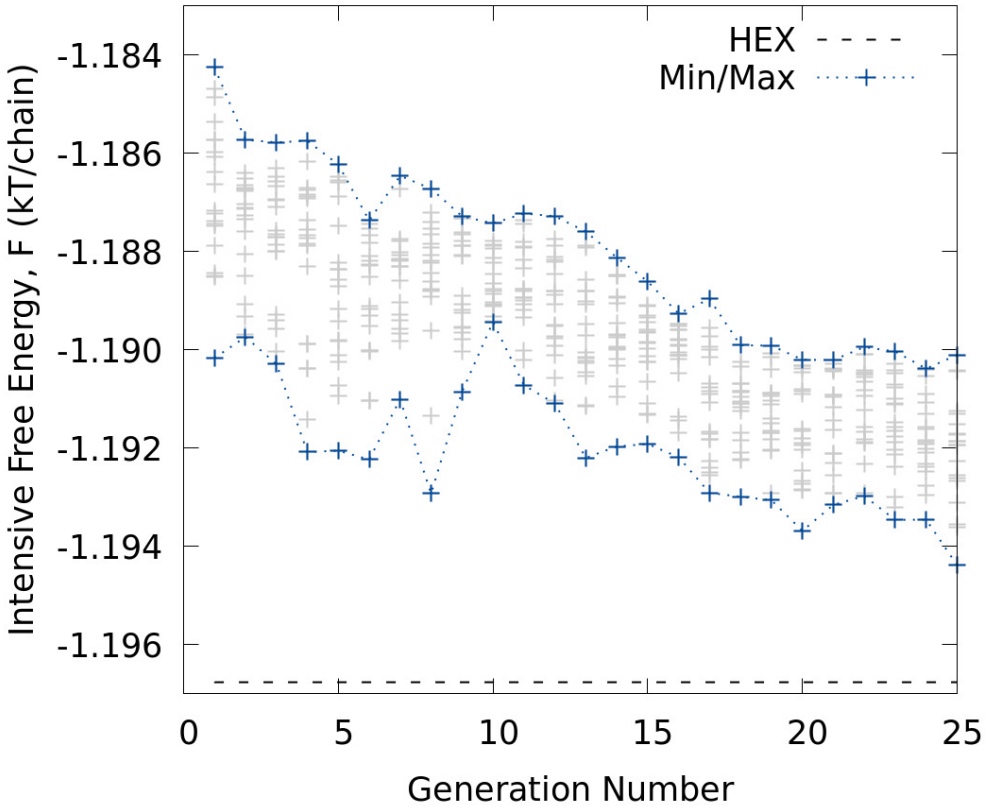

Figure 2.4: Intensive free energy vs. iteration for 2D benchmark testing of roulette-wheel selection strategy (with $\mathcal{P}_{ratio} = 1$) for the AB diblock at $f_A = 0.3$, χN=20, where HEX is expected.

Figures 2.4 and 2.5 show the plots of intensive free energy vs. generation for the GA-SCFT method applied with a roulette-wheel where $\mathcal{P}_{ratio} = 1.1$ and 5, respectively. The first generation for each implementation matches that of the sample for the elitist selection method shown in Fig. 2.2, and we can thus obtain a direct comparison between the three samples. Performing GA-SCFT using a roulette-wheel with $\mathcal{P}_{ratio}$

= 5 obtains defect-free structures by generation 10, which is faster than both the elitist selection implementation, which takes 15 generations, and the roulette-wheel with $\mathcal{P}_{ratio} = 1.1$, which takes more than 25 generations. This is consistent with the fact that in the case where $\mathcal{P}_{ratio} = 1.1$, the probability of selection any given member is roughly the equal, whereas with $\mathcal{P}_{ratio} = 5$, there is a preferential weighting given to fitter structures. Furthermore, even in the case of the elitist strategy, there is the bias of fitter members being allowed to recombine with other fitter members.

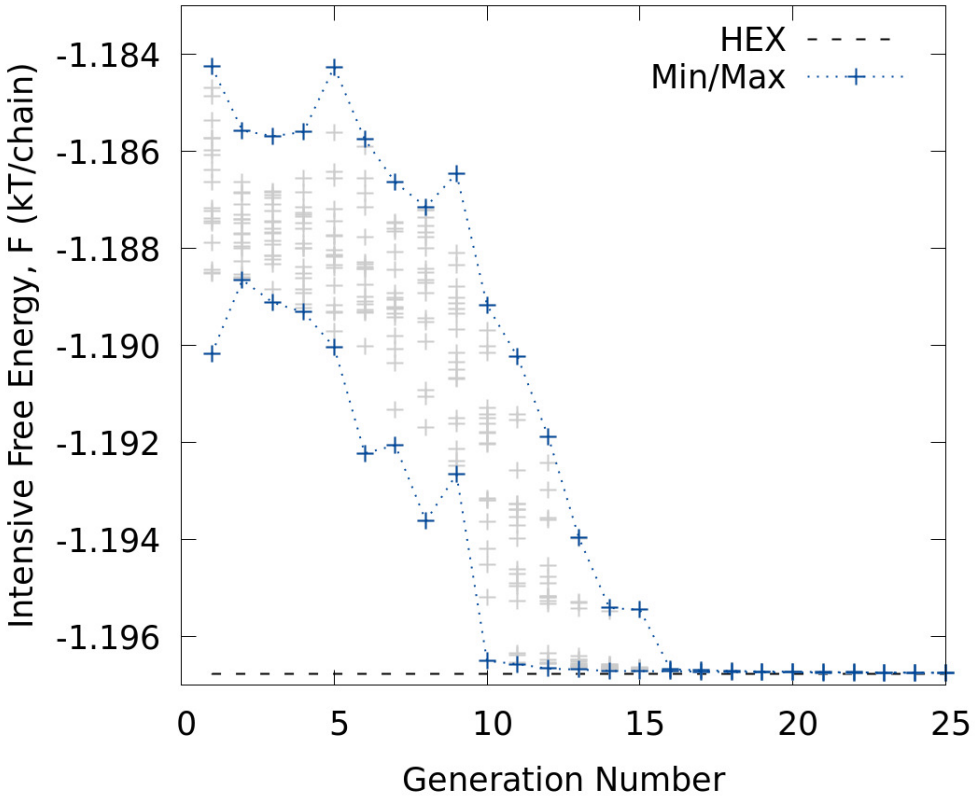

Figure 2.5: Intensive free energy vs. iteration for 2D benchmark testing of roulette-wheel selection strategy (with $\mathcal{P}_{ratio} = 5$) for the AB diblock at $f_A = 0.3$, χN=20, where HEX is expected.

$\mathcal{P}_{ratio} = 5$, seems to offer a balance between searches with premature population collapse and inefficient searches from preliminary three-dimensional studies. These

studies indicated that higher values resulted in premature population collapse to defective structures, whereas lower values of this ratio slowed the rate at which defect-free structures were obtained. The results presented in Section 2.7 therefore use the roulette-wheel with $\mathcal{P}_{ratio} = 5$.

2.2.4 Tournament Selection

The elitist and roulette wheel selection strategies were tested early in the development of the GA-SCFT method. The tournament selection method was explored as an alternative after the application of a fully-developed GA-SCFT at the compositions listed in Table 2.1. As we will see in Section 2.7, the GA-SCFT method often struggles with population collapse. One way to combat this is through the use of alternative selection strategies, such as tournament selection.

In the tournament selection method, k individuals are randomly selected from the population to form a pool of possible parents for the next generation. Parents are then selected by running a series of "tournaments" in which the fittest individual would be chosen with a probability p, the second fittest with a probability $p(1-p)$, the third fittest with $p(1-p)^2$, and so on. This task is performed until a number of parents equal to the population size has been selected, and parents are then paired off according to the order in which they were selected.

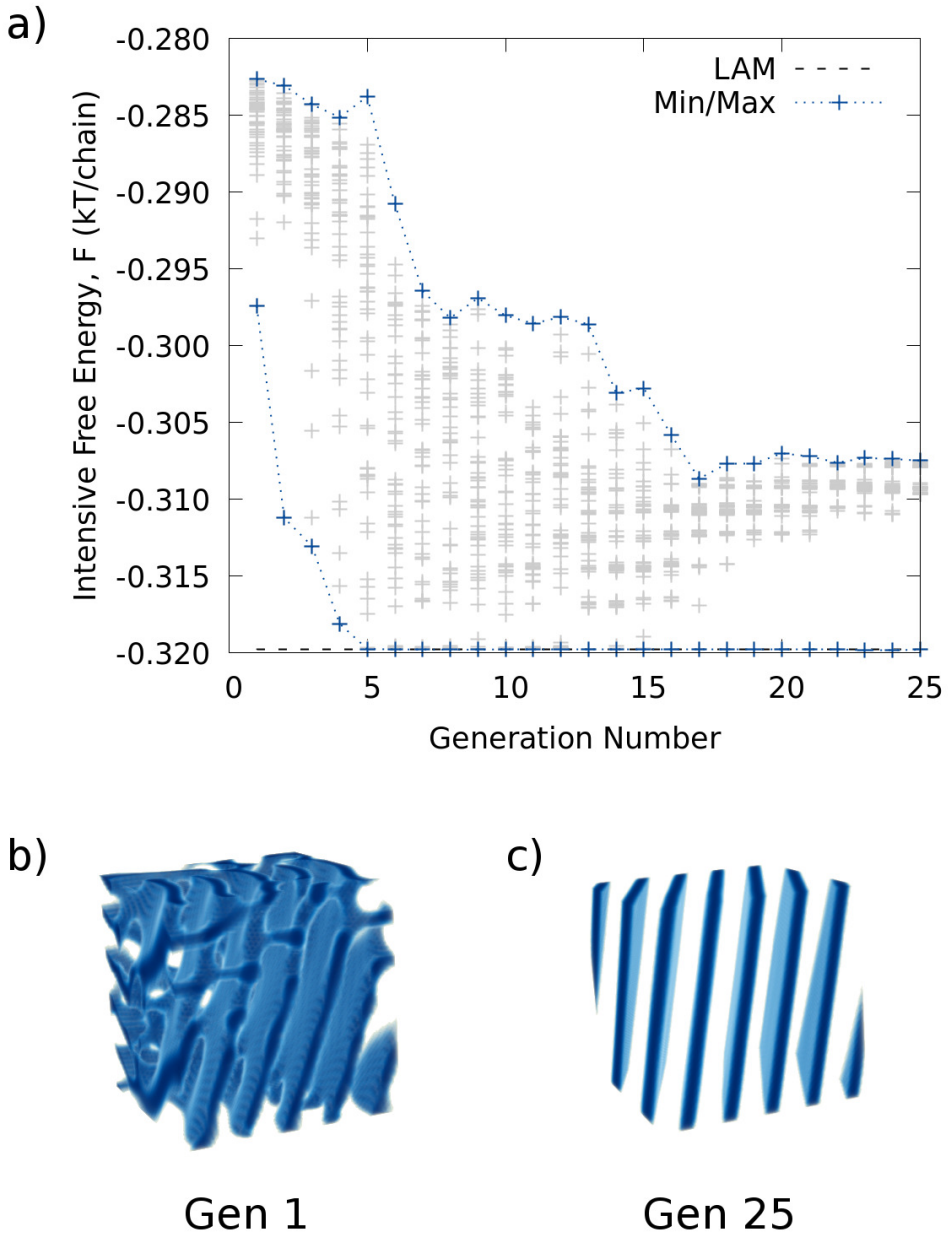

Figure 2.6: Results from 3D benchmark testing of tournament selection strategy using $p = 0.6$, 50 members in the population, and $k = 10$, for the AB diblock at $f_A = 0.50$, χN=15, where LAM is expected. (a) Intensive free energy vs. GA-SCFT iteration; the intensive free energy of each member corresponds to a gray point, and blue lines indicate the minimum, maximum, and average values of intensive free energy across iterations. (b) Density profile of the A-component for the lowest free energy structure in generation 1. (c) Density profile of the A-component for the lowest free energy structure in generation 25.

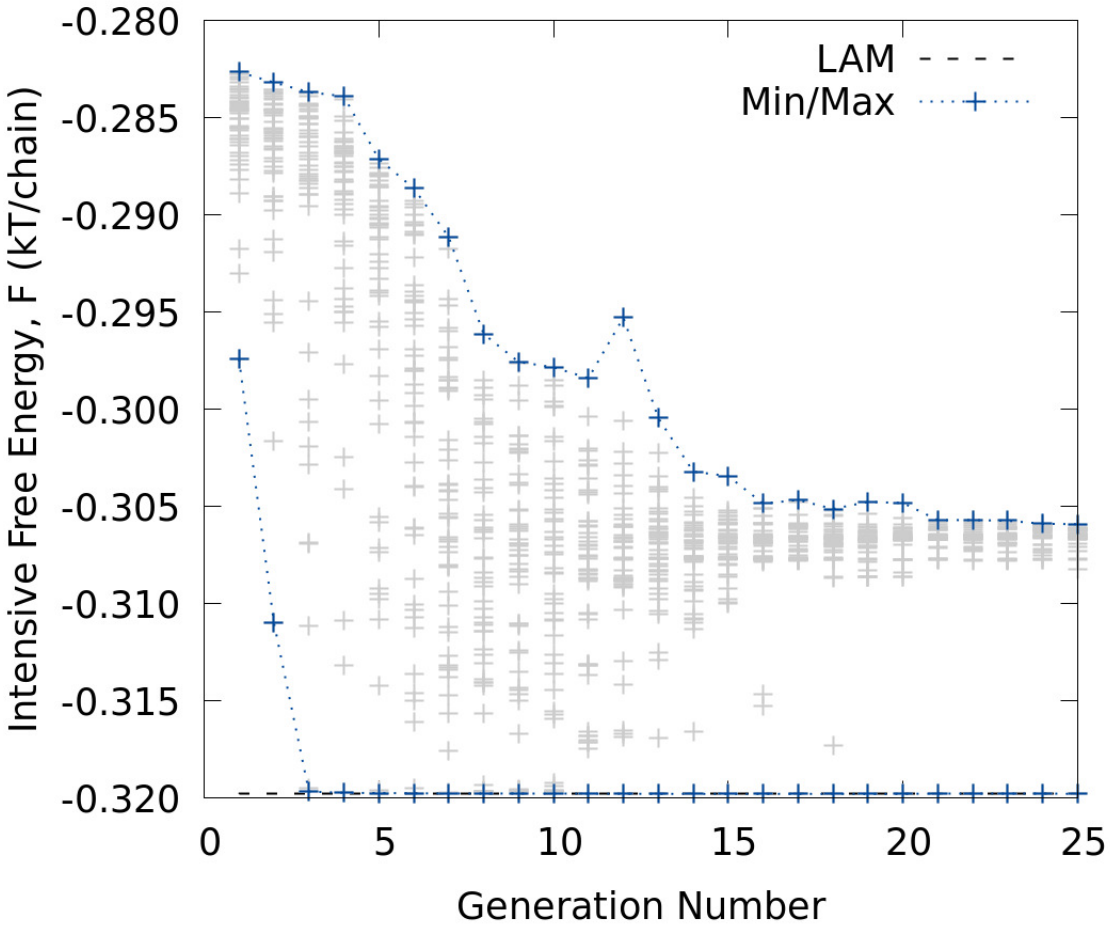

Figure 2.7: Intensive free-energy vs. iteration for 3D benchmark testing of tournament selection strategy using using $p = 0.8$, 50 members in the population, and $k = 10$, at $f_A = 0.5$, $\chi N = 15$, where LAM is expected to be stable.

Figures 2.6 and 2.7 show results for 3D benchmark testing for the tournament selection method, which was performed using $p = 0.6$ and $p = 0.8$, respectively. Simulations were run at $f_A = 0.5$ and $\chi N = 15$, where LAM is the expected morphology. Each implementation of the tournament selection used a population size of 50, with $k = 10$ as the tournament size. Having a larger probability that the lowest free energy structure be selected ($p = 0.8$ vs. $p = 0.6$) resulted in faster identification of non-defective structures (3 generations instead of 5). The distribution of structures by later generations does not seem to be significantly different between one value of

p and another.

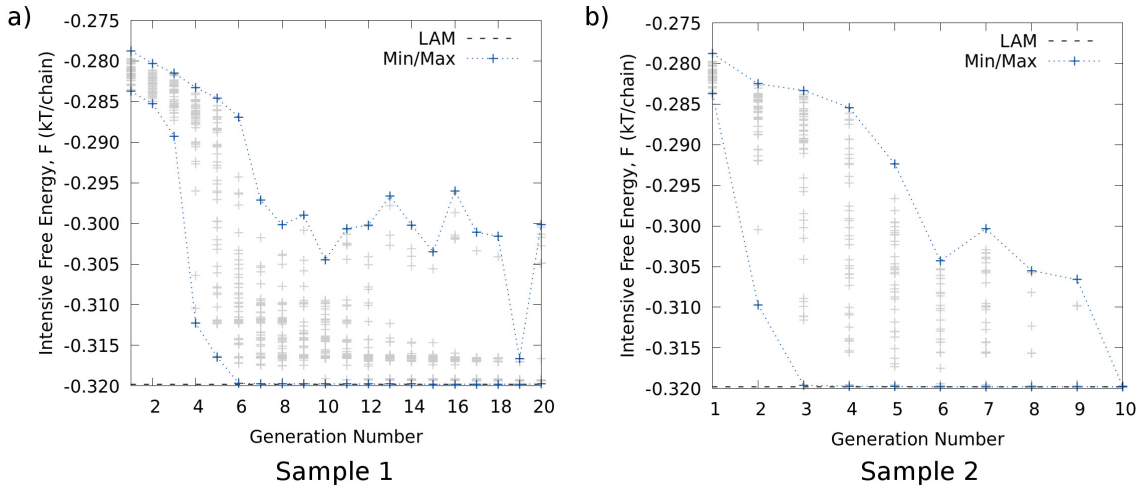

Figure 2.8: Intensive free-energy vs. iteration for 3D testing of roulette-wheel selection strategy using using $\mathcal{P}_{ratio} = 5$, and 50 members in the population, at $f_A = 0.5$, $\chi N = 15$, where LAM is expected to be stable. Panels (a) and (b) show results of two implementations of the GA-SCFT method using these parameters starting from the same population of structures for the first generation.

Figure 2.8 shows the intensive free energy vs. generation plots for two sample implementations of the GA-SCFT method using the roulette-wheel with $\mathcal{P}_{ratio} = 5$ also at $f_A = 0.5$ and $\chi N = 15$, where LAM is expected. As we can see from the results in panels (a) and (b), which both have the same distribution of first generation structures, the stochasticity of the method yields variation in the results. From these plots, as well as those of Figures 2.6 and 2.7, there does not seem to be an advantage to using the tournament selection method as opposed to the roulette wheel. In fact, we see that there are a larger number of structures at higher intensive free energies in later generations of the tournament selection method, as compared with the roulette-wheel. However, these structures seem to exist in some band of free energies, indicating that there has been a collapse to a non-ideal structure instead of LAM. Indeed, this is further corroborated by Fig. 2.9, which shows the density profiles for the

second lowest free energy structures in the tournament selection results from Fig. 2.6 (a) and the roulette-wheel results from Fig 2.8 (b). An exhaustive search through variations of p and k for the tournament selection method might show windows of these values where using the tournament selection produces better results than the roulette-wheel method. However, such studies are deferred to future work.

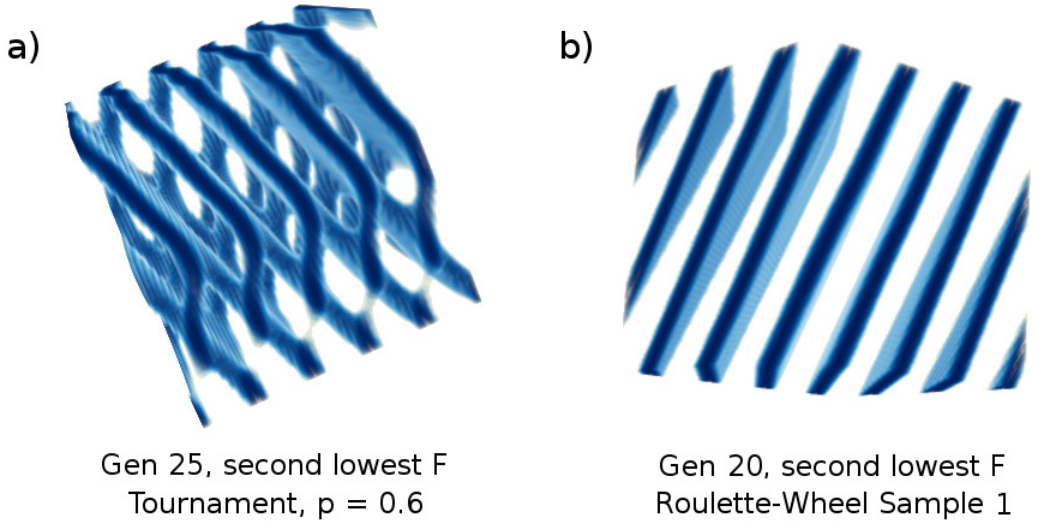

Figure 2.9: Density profiles of the second-lowest free energy structure from (a) generation 25 of the GA-SCFT using tournament selection with $p = 0.6$ (see Fig. 2.6) and (b) generation 20 of the GA-SCFT using roulette-wheel selection "sample 1" (see Fig. 2.8).

2.2.5 Comments

In this section, we explored three options for methods to select members of the GA-SCFT population to "recombine" their SCFT fields: 1) elitist, 2) roulette-wheel, and 3) tournament selection. We found that the roulette-wheel method out-performs the simple elitist method, as well as the tournament selection method for the few values of parameters for the tournament selection that we tested. It is important to recognize that there are many strategies for just the "selection" move alone, and this was by no means an exhaustive search, but rather a screening step to quickly identify a reasonable selection strategy. The application of the full GA-SCFT method in Sections 2.5, 2.6, and 2.7 uses a roulette-wheel selection with $\mathcal{P}_{ratio} = 5$.

2.3 Crossover Testing

2.3.1 Purpose and Options

Once survivors are selected, we perform crossover moves to initialize the next generation of members. The crossover step is applied to the real-space fields generated by the SCFT simulations. The child fields that come from the crossover step are then used as seeds for the next generation of simulations. Crossover is applied to two members at a time; their fields are spliced together. We divide each member's SCFT fields into two regions. Field values in the first region are swapped between the parents to generate the first child; the same is done for the second region to generate the second child. Figure 2.10 shows a schematic of this procedure. In this section, we investigate the effect of cut shape, edge hardness, and using disordered fields as parents.

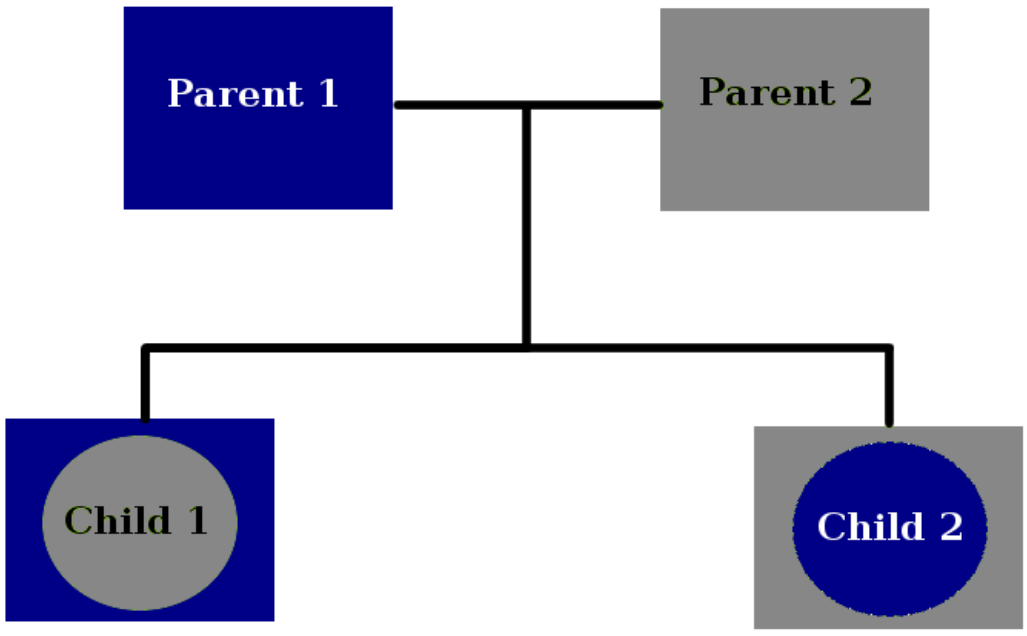

Figure 2.10: Schematic of child fields resulting from splicing two parent fields. Child fields are used as seeds for the next generation of SCFT simulations.

Figure 2.10 shows a schematic of what child fields look like after a crossover event between two parents. These child fields are used as seeds for the next generation of SCFT simulations. The Two cut types were tested: planar cuts and circular cuts.

In the sets of studies presented for cut shape and edge hardness in the following subsections, the GA-SCFT method was applied using an elitist selection strategy that randomly seeds 25% of members in each generation. Herein lies an example of the challenges faced in developing heuristic algorithms which have many modular steps. In the early stages of development, it made sense to choose a simple selection method and test the performance of various crossover moves against one another. We saw, however, from Section 2.2 that this version of the elitist strategy was actually the worst performing selection method we tested. It was fortunate that these 2D calculations ran quickly so testing could be performed fairly expediently, despite the inefficient selection strategy we were using. Developing the various selection, crossover, and mutation moves in the construction of the GA-SCFT method was thus an iterative process.

This iterative process is not totally unavoidable, since at some point, choices must be made to fix one set of parameters while doing a coarse exploration of another feature of the algorithm. It is important when designing such investigations to strike a balance between educating oneself about the various options available for each method, the cost to implement and study each option, and the fact that no choice is entirely isolated in its ability to affect the results of other choices. This kind of balance is difficult to achieve, and seems to require experience, as well as an understanding of the optimization problem at hand, which in turn requires its own level of experience.

We will also see that results from 2D studies do not always translate to the 3D case. This is particularly true for the effects of cut shape and using disordered fields as parents, and seem to primarily be dependent on what the target morphology was

for a given study.

As a final note, the SCFT relaxation step used in tandem with our GA algorithm includes not only field relaxation, but simultaneous relaxation of the cell-tensor that defines the shape and volume of the simulation cell. As part of a crossover event, we assign the average of the SCFT-relaxed cell-tensors of the two contributing parents as the initial guess of the cell-tensor for each child; we refer to this step as a "cell-tensor update."

2.3.2 Shapes: Planar vs. Circular Cuts

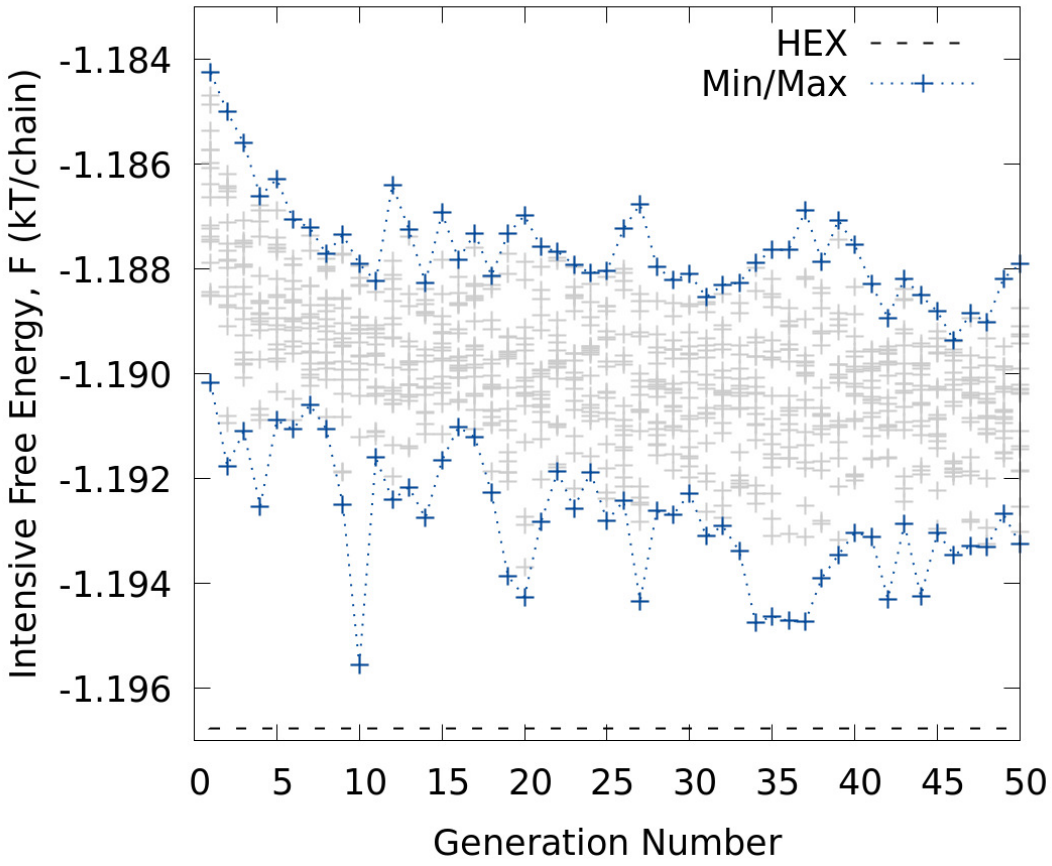

Figure 2.11: Intensive free energy vs. generation plot from GA-SCFT using planar cuts applied for 2D simulation cells at $f_A = 0.3$, $\chi N = 20$, where HEX is expected to be stable.

Figure 2.11 shows the results from performing GA-SCFT using planar crossover cuts in large, 2D simulation cells at a composition where HEX is the expected morphology. Even after 50 generations, we see that there is a failure to converge to defect-free structures. This is consistent with the fact that we were using the elitist strategy with randomly initialized fields.

Figure 2.12 shows the results from performing GA-SCFT using circular crossover cuts in large, 2D simulation cells at a composition where HEX is the expected morphology. Similarly to the planar crossover cut case, we see a failure of GA-SCFT to converge to defect-free structures. However, by tracking the minimum free-energy structure of each generation, we do see an improvement over the planar cuts in the rate at which lower free energy (less defective) morphologies are obtained. This, is actually in contrast with results from 3D studies, which saw better performance with planar crossover cuts.

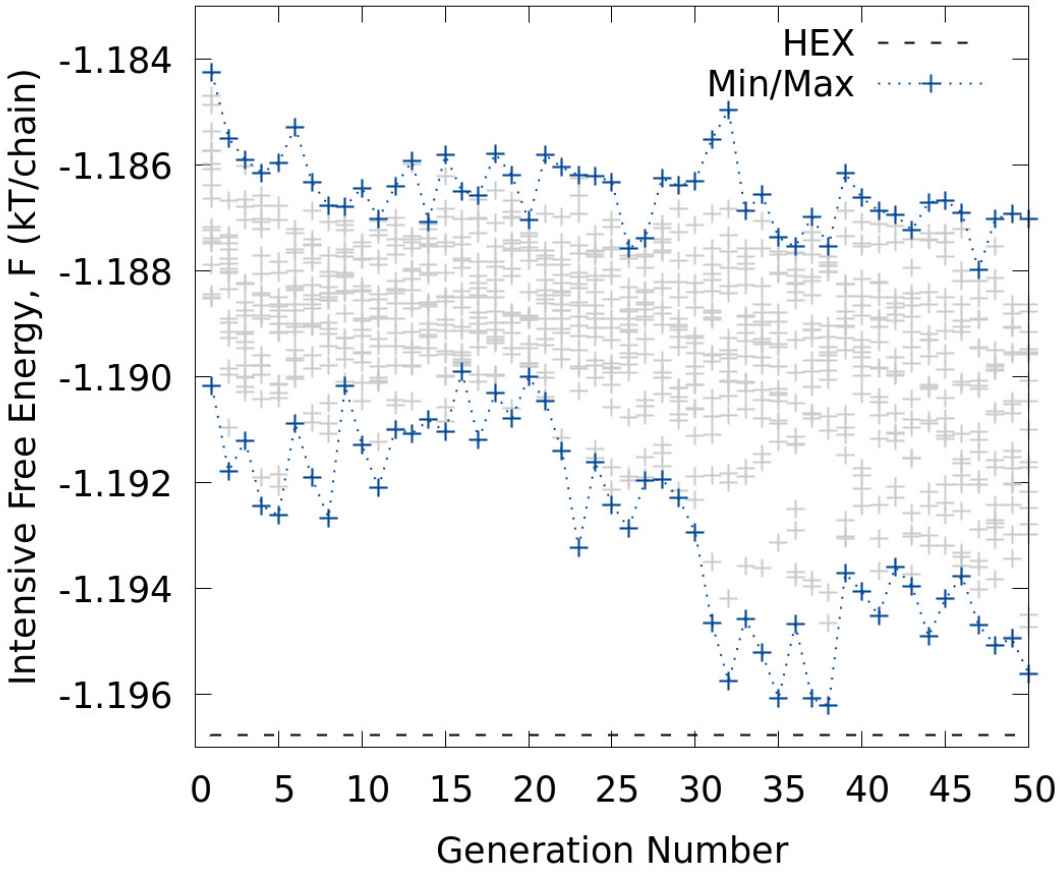

Figure 2.12: Intensive free energy vs. generation plot from GA-SCFT using circular cuts applied for 2D simulation cells at $f_A = 0.3$, $\chi N = 20$, where HEX is expected to be stable.

2.3.3 Edge Hardness

Splicing two disparate morphologies can often lead to discontinuities along the interface of the splice, which in turn resulted in numerical stability problems with the SCFT field relaxers. To mitigate this effect, we softened the interface of the cut by smearing with a tanh function. Values for fields for each child from a given set of two

parents, A and B, were written to SCFT field files in the following way:

$$\begin{aligned} \phi_{child1} &= x * \phi_A + (1 - x) * \phi_B \\ \phi_{child2} &= x * \phi_B + (1 - x) * \phi_A \end{aligned} \tag{2.3}$$

where ϕ_A and ϕ_B are the values of the fields from parent A and parent B, respectively. The value of x allows for an interpolation between the two regions separated by the cut, and x is determined using:

$$x = \frac{\tanh(a(\mathbf{r} - 1) + 1)}{2} \tag{2.4}$$

where $\mathbf{r}$ is the distance between the current coordinate in the simulation cell and the hard, unsmeared edge of the cut, and a determines the width of the smearing. Smaller values of a have the effect of smearing the edge of the cut more, where as larger values correspond to sharper interfaces. Preliminary tests using tanh function widths ranging from $0.005\,L$ to $0.1\,L$ were performed for a HEX composition at χN=20. Softer tanh functions had the effect of slowing the retrieval of defect-free structures, while sharper tanh functions resulted in divergent SCFT field relaxer trajectories. Using a width of 0.025 the simulation cell seemed to yield the best results for the ranges we observed.

Figure 2.13 displays results for GA-SCFT applied where HEX is the expected morphology using 2D simulations, an elitist strategy that randomly seeds 25% of members each generation, and sharp vs. smeared interfaces. The sample which used softened interfaces (blue data) shows an improvement in obtaining lower free energy structures than the GA-SCFT applied with sharp interfaces. This trend in performance, unlike the trend seen for the 2D studies for cut shape, seemed to carry to the 3D benchmarking.

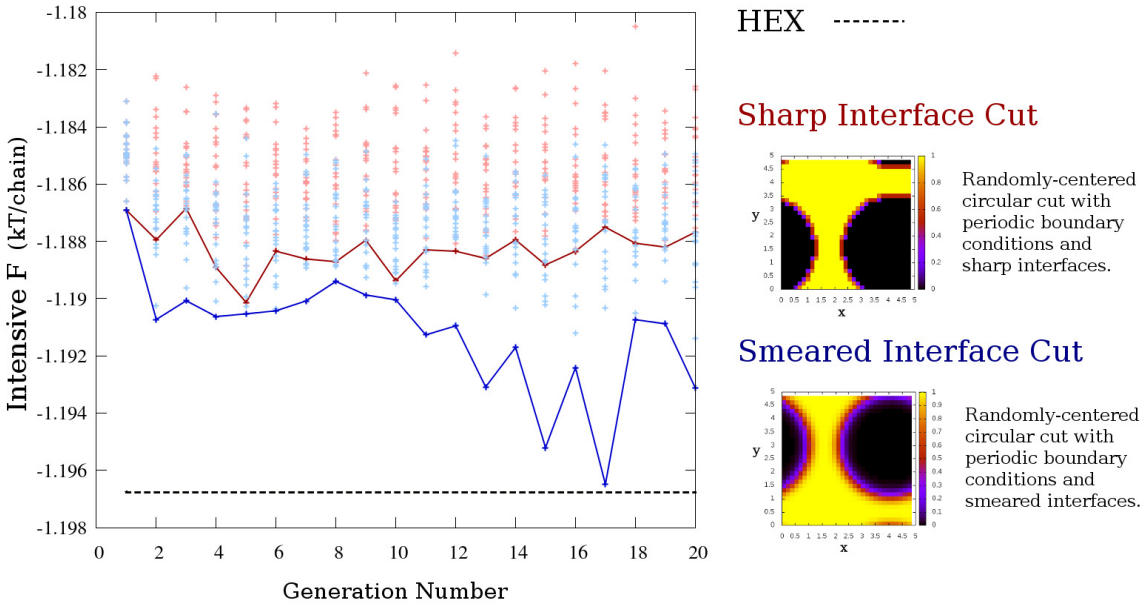

Figure 2.13: Intensive free energy vs. generation plots for GA-SCFT applied to a composition where HEX is expected using 2D SCFT simulations and sharp (red) and smeared (blue) interfaces.

2.3.4 Using Disordered Fields as Parents

Conceivably, allowing disordered fields to recombine with fit members of each generation could allow defect-free patterns to propagate through the homogeneous portion of the spliced fields, directed by the surrounding ordered regions (see Fig. 2.14). We probed this idea in 2D and 3D SCFT simulations, at compositions where LAM, HEX, BCC, and GYR were expected to be stable. In this section, we will only examine two case studies: 2D and 3D simulations at χN = 20 and $f_A = 0.3$, where HEX is the stable morphology.

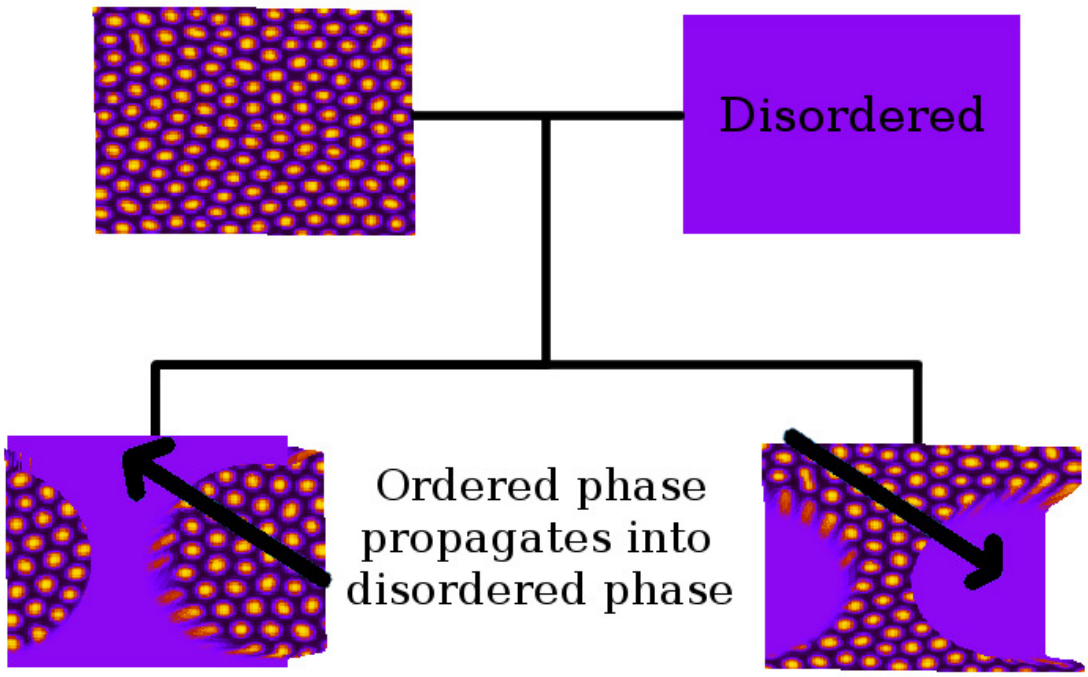

Figure 2.14: Strategy for using homogeneous fields as parents in the crossover step. Defect-free patterns might be able to propagate into the homogeneous, disordered fields.

Figure 2.15 shows the results from performing GA-SCFT using disordered fields as parents with circular crossover cuts (a) and planar crossover cuts (b) in large, 3D simulation cells at f_A=0.3, χN=20, a composition where HEX is the expected morphology. A roulette wheel with $\mathcal{P}_{ratio}$=5 was used to select parents, and homogeneous fields were used parents for every member of every generation (*i.e.*, for every pair of parents used to generate children, one of the parents was a homogeneous, disordered SCFT field, and the other was selected using the roulette-wheel). Using circular crossover cuts consistently resulted in more rapid recovery of defect-free structures, as well as collapse to these structures. Collapse to structures could be delayed by tuning the value of $\mathcal{P}_{ratio}$.

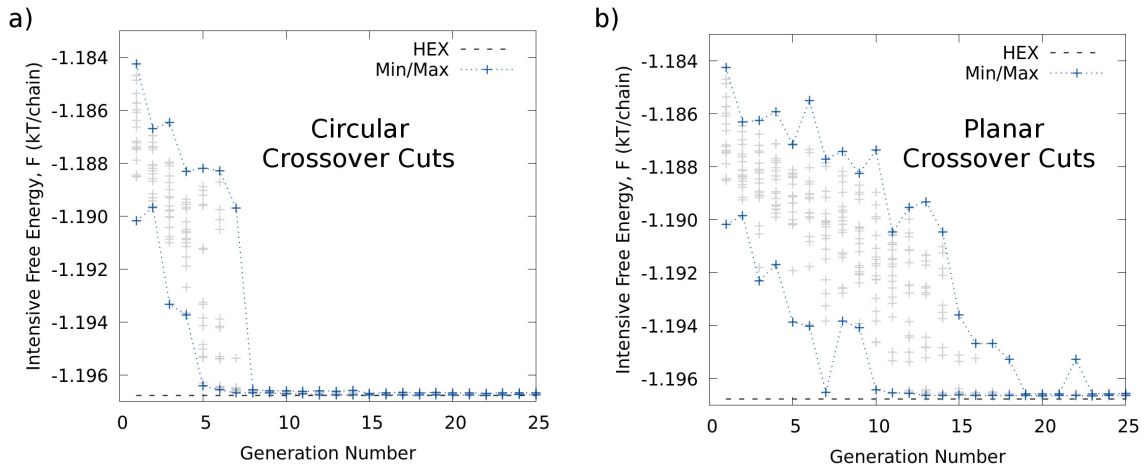

Figure 2.15: Intensive free energy vs. generation plot from GA-SCFT using homogeneous fields as parents and planar cuts applied for 2D simulation cells at f_A = 0.3, χN = 20, where HEX is expected to be stable.

Figure 2.16 shows the results from performing GA-SCFT using disordered fields as parents and planar crossover cuts in large, 3D simulation cells at f_A=0.3, χN=20, a composition where HEX is the expected morphology. Planar crossover cuts showed faster convergence to defect-free structures than spherical crossover cuts in the 3D studies, across the LAM, HEX, and BCC morphologies. This is in contrast to the 2D studies shown in the previous subsection on cut shape. In Fig. 2.16, we see that after only 11 generations, defect-free structures were already being obtained. Even by generation 9, the lowest free energy member has cylinders arranged in a hexagonal array; the difference in free energy between this structure and the defect-free HEX morphology comes from the branches that connect some of the cylinders to one another. Such defects were also commonly seen in the simulations where homogeneous fields were not introduced as parents. However, introducing the disordered fields as parents during crossover had the effect of facilitating the "clean-up" of these defects.

One would expect the search for 3D morphologies to be slower than the 2D morphologies. However, the combination of using a roulette-wheel selection strategy and

also introducing disordered fields as parents expedited the identification of defect-free hexagonally-packed cylinder structures. The values of the intensive free energies presented in this plot do not match those from previous plots for the same composition; this is simply an artifact of a change that was made to the in-house SCFT code some time in 2015, after the 2D benchmark tests had been performed. There is a flag in the code that can be turned on, called "UseLegacyOperators," which will provide comparable intensive free energies to previous studies. There is a small difference between the lowest free energy structures in generations 11-13 and the black dashed line depicted the expected free energy for HEX at this composition. This difference can be made up by further relaxing the SCFT fields of those individual members of the population, since in the GA-SCFT method, SCFT simulations are capped at 20,000 iterations to avoid unecessary computational cost associated with resolving the finer features in an SCFT simulation.

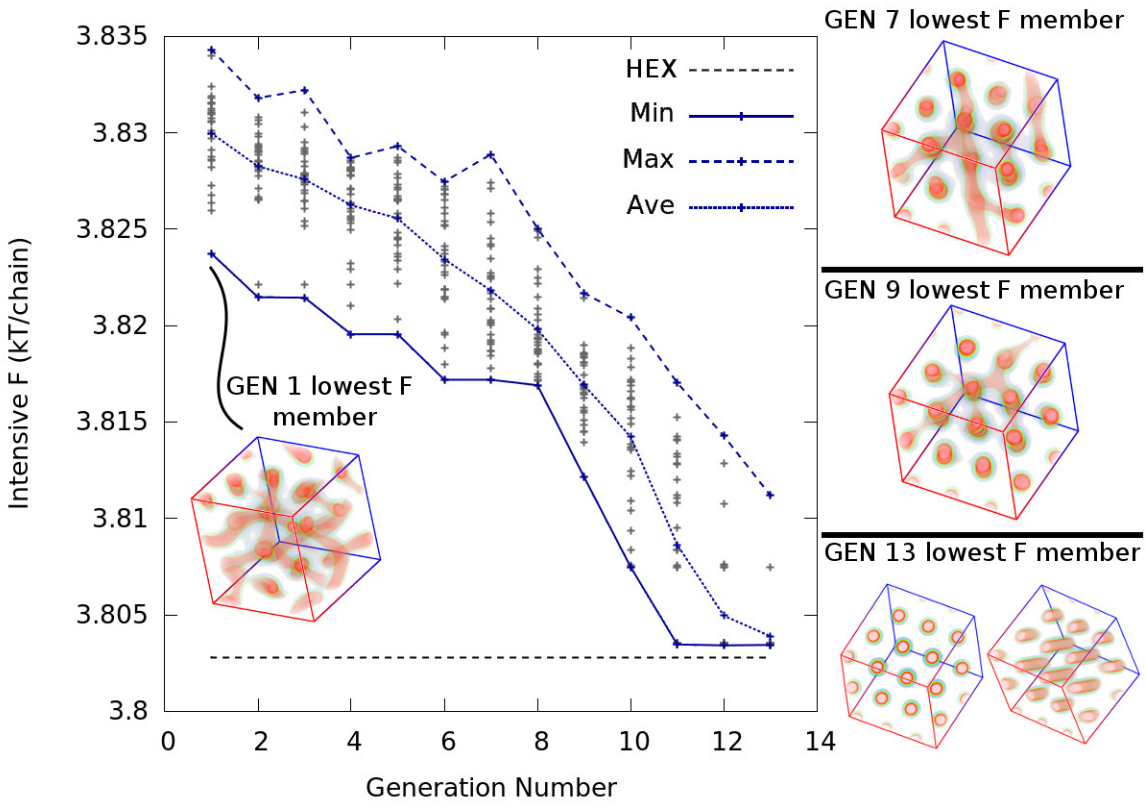

Figure 2.16: Intensive free energy vs. generation plot from GA-SCFT using disordered fields as parents and planar cuts applied for 3D simulation cells at $f_A = 0.3$, $\chi N = 20$, where HEX is expected to be stable.

The disordered parent method was very successful for accelerating the search for HEX. In some cases, half the number of generations were needed to obtain perfect HEX morphologies. However, the disordered parent crossover/mutation technique was unsuccessful for GYR simulations; the tubules in the simulation cell became elongated and eventually the GA-SCFT algorithm led to HEX or LAM structures for a variety of crossover cut shapes and relative volumes. The results look very similar to those shown in Figure 2.21, which shows the collapse of the population onto hexagonally-packed cylinders, rather than ever finding GYR.

2.3.5 Comments

In this section, we investigated the crossover, or recombination move, which comes immediately after members are selected as parents. In a genetic algorithm, crossover is a means to recombine the "genetic information" of parents which have been selected to reproduce, which helps to retain desirable features from fit parents, while also maintaning some diversity in the population. In the case of the GA-SCFT method, the genetic information of members is contained in the real-space SCFT field values from each simulation. This means that recombination moves involved the physical splicing of the real-space SCFT fields of parents.

We explored the effects of using different cut shapes: planar cuts vs. circular/spherical cuts. We also examined the effect of using various shapes at the interfaces, such as straight lines vs. wavy lines. These studies were not included in this dissertation, but for the 2D investigations performed, using straight rather than wavy lines did not seem to have much of an effect on the performance of the GA-SCFT method. We also looked at the effect of softening the interface with a tanh function. Lastly, we considered using disordered SCFT fields as parents, in the hope that doing so would allow "good features" to grow in at the boundaries. However, we saw that this had the effect of elongating the tubular regions rich in A-type polymer, and eventually collapsing the population into cylinder structures rather than network morphologies even at compositions where network phases were expected to be stable.

An interesting observation that came from the investigations surrounding the development of the crossover move is that some of the results of these studies changed drastically in going from 2D to 3D SCFT simulations. For instance, circular cuts performed better in the 2D simulation case, but planar cuts performed better in the 3D simulations. Some of this effect may stem from the fact that 1) the shape of the cut may guide the direction that A-rich or B-rich densities grow from the

cut interface, and 2) symmetries of the expected morphologies in each of the given case studies differ with different dimensionality. In the 2D simulations, we expected the HEX morphology, which has a radial symmetry in the cylinders. In the 3D simulations, the HEX morphology still exhibits the radial symmetry, but has an extra dimension which does not have radial symmetry. We found that defects branching across cylinders were more likely to remain in the spherical crossover cuts. This was also true for the application of GA-SCFT to compositions where GYR was expected. Introducing homogeneous fields as parents mitigated this effect in the case of HEX, but in the case where GYR was the target morphology, there was the added difficulty of tubules elongating in one direction.

The best performance of the GA-SCFT method for 3D simulations occured with planar crossover cuts that were softened with a tanh function and split the box into equal halves along one of the x, y, or z planes with equal probability. The application of the full GA-SCFT method in Sections 2.5, 2.6, and 2.7 uses these planar crossover cuts with softened interfaces.

2.4 Mutation

2.4.1 Overview

Once children are generated, mutations can be performed as perturbations to their real-space field values. This is important for maintaining diversity in the population to avoid premature population collapse to unoptimal structures.

Results are presented for mutations which consisted of replacing some portion of the simulation cell with random white noise. The idea behind this mutation is similar to the motivation for using disordered fields as parents (Section 2.3). In fact, employing disordered fields as parents with some probability was explored as a mutation as well, but resulted in a similar set of deficiencies seen in using disordered fields as part of the crossover move. That is, in the case of the application of GA-SCFT at a composition where GYR was expected, tubules end up elongating, and morphologies eventually collapsed into cylinders or lamellae.

In this section we will explore the performance of GA-SCFT for different mutation sizes, noise amplitudes in the field values, and mutation frequencies. We will also examine the use of a spectral filter, where low amplitude Fourier peaks of the reciprocal space fields are removed, following the methods discussed by Bosse *et al.* [33]

2.4.2 Random White Noise Mutations

The random white noise mutations were performed tested in several ways. One method involved replacing all original field values in a region of the simulation cell with zero, then adding random values to each point. Another version of the method involved adding random values to each field value in a particular region of interest in the simulation cell. In our initial work, we applied the noise mutation to the entire

simulation. In later studies, we applied the noise mutation to different shapes and sizes of regions. Regions would be selected randomly. Thus, for example, if we applied the mutation to half of the simulation cell, the noise could be applied to only the half of the cell that one parent contributed to, or it could apply to a half which spanned sides of the cell containing contributions from both parents. We also tried applying the mutation to an octant or multiple octants of the simulation cell, quadrants, and spheres of various radii.

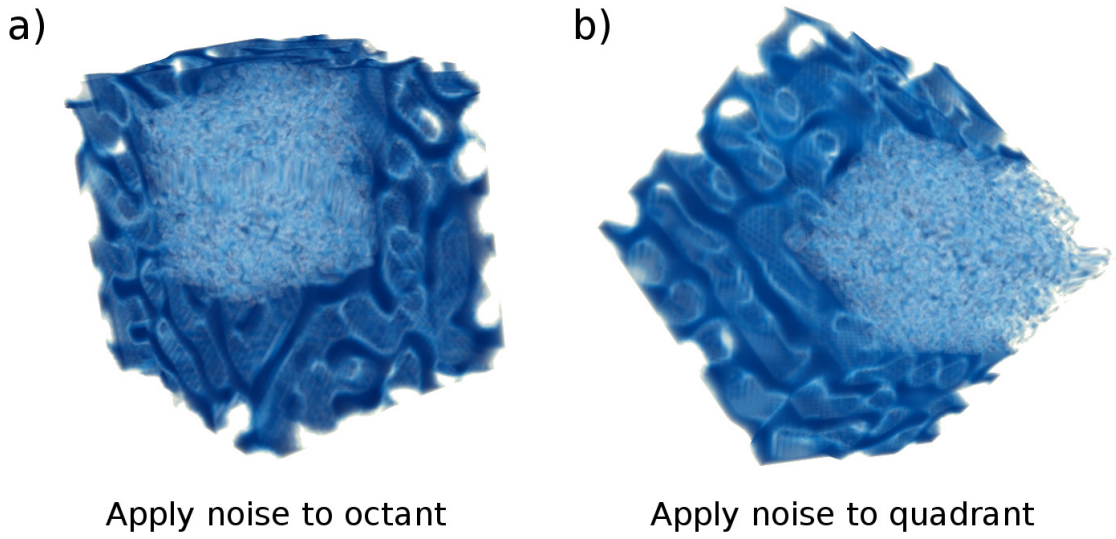

Figure 2.17: Examples of mutated child fields in GA-SCFT after random white noise has been added to a) an octant and b) a quadrant of the original child fields.

Applying random white noise to the children (input fields for the next generation) seemed to mitigate some of the population collapse we were observing. However, simply adding noise to the field values in each generation without bounding the value of the noise led to field values that grew with each successive generation. After several generations of this effect, SCFT simulations would result in divergent trajectories. To address this problem, we scaled the values of the noise so the values of the mutated fields would not exceed the magnitude of the original field values. In the final iteration

of the development of the noise mutation (*i.e.*, the version of the mutation used in Sections 2.6 and 2.7, we performed the following tasks:

1. Randomly select the region of the real-space fields to mutate (*e.g.*, half the simulation cell, an octant, or a sphere)
2. For every coordinate in the selected region:
   - Choose a random number, r, between $((-a, a))$. Here, $a = \Delta A * C$, $\Delta A = A_{max} - A_{min}$ is the difference between the maximum and minimum amplitudes of the original fields, and C is a prescribed percentage of the maximum magnitude of the original field values.
   - Use $A_{noise} = \frac{r+A}{s}$ as the new field value. A is the original field value at the current coordinate, and s is a scaling variable given by: $\frac{|A_{max}|+a}{|A_{max}|}$

Figure 2.17 displays examples of what the random white noise mutation might look like when applied to an octant (a) or a quadrant (b) of a 3D simulation cell. The images shown in this figure are visual representations of fields which were used as seeds for a subsequent generation (iteration) of the GA-SCFT. Using octants or quadrants as regions for the application of the mutation seemed to yield better performance than applying noise to the entire field, or applying noise in spherical regions. In the case of applying noise to the entire field, GA-SCFT took more generations to identify the expected morphology when the target was HEX than when noise was only applied to an octant or a quadrant. When the target was GYR, GA-SCFT never recovered GYR. This was no different from any of the other implementations of GA-SCFT applied at a composition where GYR was the target morphology. In the case of applying the noise mutation to spherical regions of varying radii, GA-SCFT would often converge to cylinder morphologies, even when the target was GYR. This result is similar to the ones seen for using disordered, homogeneous fields in crossover and mutation, as

well as applying a harmonic filter as a mutation.

2.4.3 Applying a Mutation which Used Disordered Fields as Parents

One move which was studied was the employment of disordered fields as parents as a mutation. We tested a number of different strategies:

- Apply mutation every generation, to some subset of members (vary size of subset).
- Always apply before crossover.
- Always apply after crossover.
- Vary frequency of application (not necessarily performed every generation).
- Vary size of disordered fields introduced to child fields.
- Overlay disordered fields coincidentally with crossover cut.
- Overlay disordered fields non-coincidentally (and randomly) after crossover cut.
- Apply in addition to random white noise

Figure 2.18 provides a schematic for the application of the DIS mutation after crossover has already been applied. Homogeneous fields may be overlayed on child fields either coincidentally or non-coincidentally with the original crossover cut. The advantage of overlaying the homogeneous, disordered fields coincidentally with the original crossover cut is that the DIS crossover move seems to accelerate the recovery of defect-free morphologies. However, it has the effect of obliterating the "genetic information" from one of the parent fields, which may be unproductive when those fields contain desirable features.

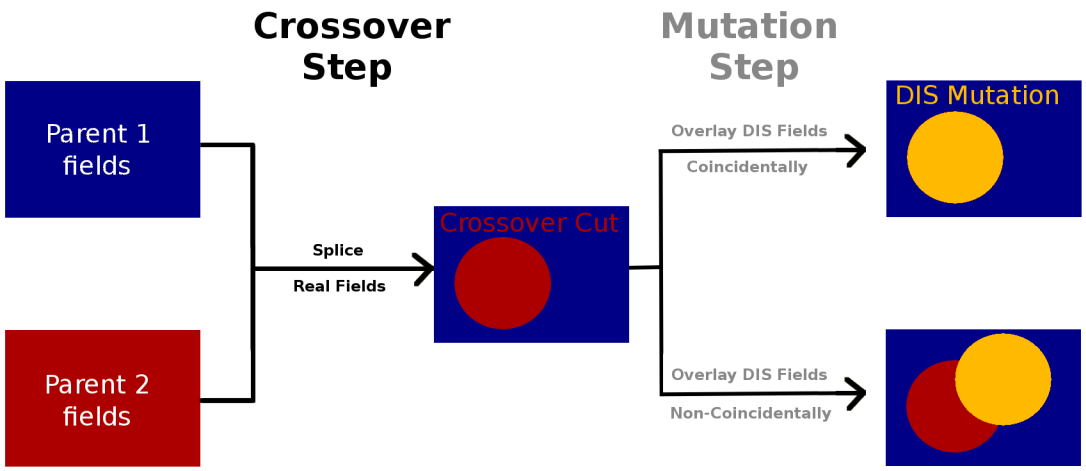

Figure 2.18: Schematic for applying disordered fields as a mutation after crossover. Once real-space fields are spliced from parents during the crossover step, disordered fields may be overlayed on the child fields coincidentally with the original crossover cut, or non-coincidentally.

Figure 2.19 provides examples of input field files one might encounter when performing the homogeneous DIS fields mutation. In this particular example, the DIS mutation was applied to every member of every generation (100% frequency). From both panels, we see a variety of possible input fields. Of particular interest are the fields shown in red (second row, second from the left in panel (a) and third row, farmost right square in panel (b)). The simulations that were seeded using these fields actually resulted in divergent trajectories. For the 2D testing where the target morphology was HEX, the GA-SCFT algorithm was surprisingly robust to the existence of divergent trajectories, as long as it was limited to one or two per generation. The members which experienced these trajectories would simply not get selected as parents for the following generation. Even in 3D testing where HEX was the target morphology, this effect seemed to hold.

In 3D testing where GYR was the expected morphology, the existence of divergent trajectories seemed to aggravate the problems we were already seeing with the

application of the DIS mutation: the frequent collapse to HEX or LAM structures we observed seemed to be somewhat accelerated (occuring within 10 generations instead of 15, for instance). We did add a feature which would remove divergent trajectories by creating new children from the selected pool of parents, or by spawning a new individual by using random initial conditions, which seemed to somewhat extend the number of generations to collapse. However, it is not entirely clear to what extent this effect was due to the stochastic nature of the algorithm. Running a full GA-SCFT for the 3D simulations was fairly expensive; in many cases we would only run one or maybe two samples of each trial we were investigating. However, population collapse to LAM and HEX occurred across the board for the various implementations of the DIS fields mutation.

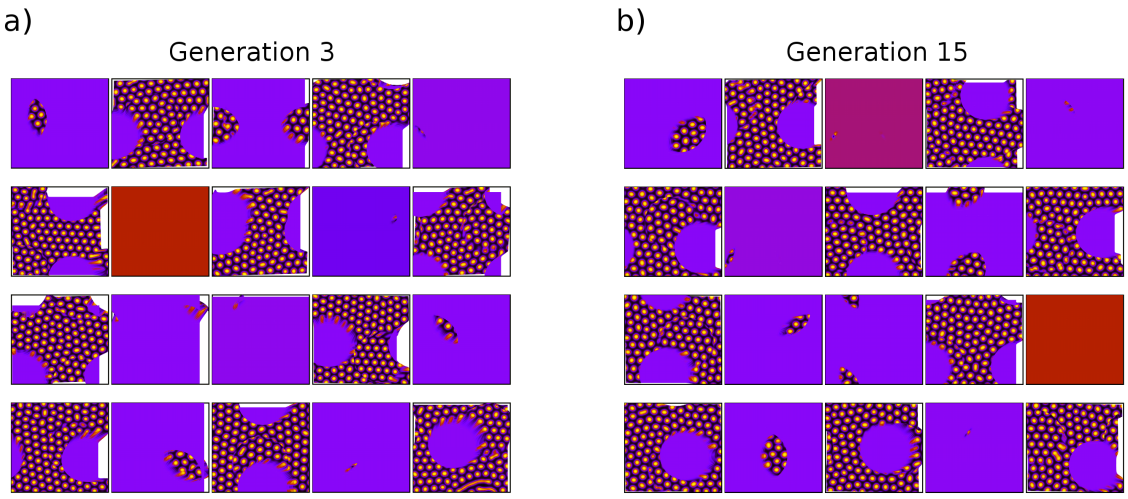

Figure 2.19: Examples of input fields for generation 3, members 1-20 (a) and generation 15, members 1-20 (b) for GA-SCFT applied to a composition where HEX is the expected morphology. A mutation strategy where disordered parents were applied to every member of every generation after the crossover step, at positions which were not coincident with the original crossover cuts.

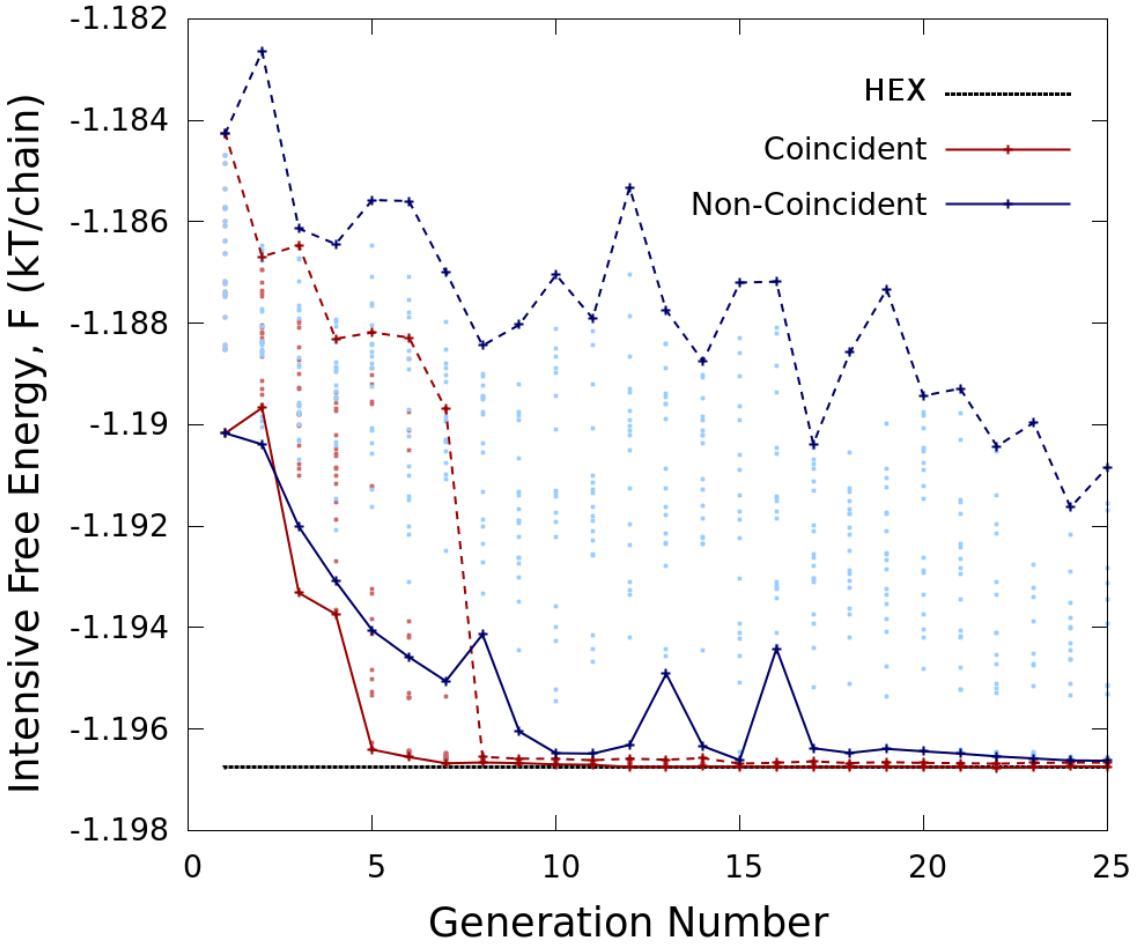

Figure 2.20: Intensive free energy vs. generation plot for GA-SCFT applied to 2D Hex

2.4.4 Spectral Filter

In the spectral amplitude filter mutation, low amplitude Fourier components of the power spectrum of parent fields are removed. A range of values for the percentage of Fourier modes to be removed was investigated over the range of 50% to 90% of the lowest amplitude components in 10% increments. Similar to the case of the disordered parent crossover/mutation, GA-SCFT applied to compositions where HEX was expected were accelerated by the application of the filter. However, GA-SCFT with spectral filtering for GYR was unable to recover GYR; again we encountered

the elongated tubules that eventually yielded HEX or LAM morphologies.

Thus, both the disordered parent method and spectral amplitude filter mutation led to improvements in cases where the unembellished GA-SCFT was already successful, but was unsuccessful for GYR, for which the basic GA-SCFT method failed anyway.

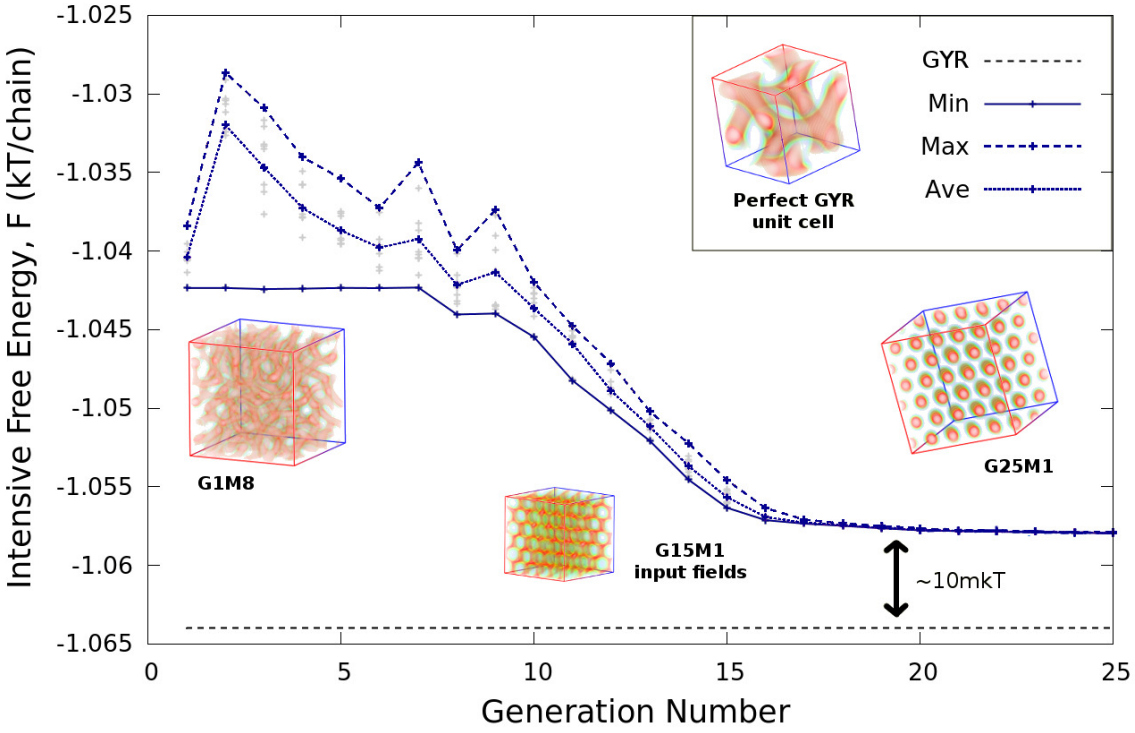

Figure 2.21: Intensive free energy vs. generation plot for GA-SCFT using 3D simulations and the spectral filter mutation at f_A=0.37 and χN=20, where GYR is the expected morphology. By about 15 generations, the population of structures collapses into hexagonally-packed cylinders, whose intensive free energy is about 10mk$_B$T/chain higher than the GYR morphology.

2.4.5 Comments

In this section, we studied a variety of mutations, including 1) the addition of random white noise to field values, 2) swapping in disordered, homogeneous fields to

the original fields either before or after crossover, and 3) the application of a spectral filter. We saw that the application of disordered, homogeneous field and the spectral filter mutation were similar in the results, both for the implementation of GA-SCFT at compositions where HEX was the target morphology and compositions where GYR was expected. The identification of HEX as a stable morphology was accelerated by the use of these mutations, while GYR was never observed, but rather the GA-SCFT algorithm often collapsed into competing morphologies such as HEX or LAM.

For the results presented in Sections 2.6 and 2.7, we used a version of GA-SCFT where noise was applied to a single, randomly selected octant of every member after crossover was performed. Field values in these random octants of the simulation cell were replaced with white noise of amplitudes selected from a uniform random distribution between $(-A, A)$, where A is the maximum magnitude real-space-field value observed in a given simulation.

2.5 Miscellaneous Features

In this section, results are presented for the application of GA-SCFT which used the moves listed in Table 2.3.

GA Move	Parameter Value
roulette-wheel selection	$\mathcal{P}_{ratio} = 5$
planar crossover	half box, tanh width = 0.025
white-noise mutation	single octant, 100% of members

Table 2.3: GA moves used in the application of the full GA-SCFT method to the diblock for studies on the effects of population size, SCFT simulation cell size, and χN (see Sections 2.5, 2.6, and 2.7.

2.5.1 General Commentary

In the context of the combination of genetic algorithms with SCFT, there are a variety of important considerations that fall beyond the scope of the basic GA moves of selection, crossover, and mutation. For instance, it is conceivable that the choice of initial cell size might affect the results of the GA-SCFT. We will see in Section 2.7.2 that how commensurate an SCFT simulation cell is with the cell size of the expected, stable morphology affects how accessible that morphology is to our group's implementation of the SCFT algorithm. Thus, in our benchmark testing of the GA-SCFT, it is important to choose large and small simulation cell sizes which are not exactly commensurate with a multiple of the unit cell of any of the individual competing candidate morphologies, to unintentionally favoring one structure over another.

Another feature which seemed worth exploring was the question of whether simulation cell sizes should be averaged between each pair of parents. It was also important to tune individual SCFT simulations to reduce computational cost, since the SCFT simulations themselves take many orders of magnitude longer than the GA moves,

which are applied with a series of simple scripts.

2.5.2 Initial Cell Size

Here we present some findings with regard to the choice of the initial lattice parameter, L_{init}. The results presented here are given for a diblock composition where HEX is expected to be stable: χN=15 and f_A=0.35, where f_A is the block fraction of monomer type A in the diblock. Figure 2.22 depicts plots of the value of the intensive free energy per chain of each member vs. generation number in $L_{init} = 15\,R_g$ (panel A) and $L_{init} = 20\,R_g$ (panel B) simulation cells. Grey dots represent the intensive free energy for each member at the end of their respective SCFT relaxation. The blue dotted lines track the highest and lowest free energy members of each generation, while the solid blue line is the arithmetic mean and the black dashed line is the defect-free (target) morphology.

There is a small difference in intensive free energy seen between the dotted black line representing the expected value and the intensive free energy to which the GA-SCFT method converges. This discrepancy arises from a difference in the spatial resolution of the simulations run to calculate the expected value and the simulations in the GA-SCFT samples, and has been independently verified by using the fields from the lowest intensive free energy members as seeds in simulations with spatial resolutions below $0.2\,R_g$ for both the $L_{init} = 15\,R_g$ and $L_{init} = 20\,R_g$ cases. The unit-cell reference free energies plotted in all figures were run at spatial resolutions in a range of about $0.25\,R_g$ to $0.26\,R_g$, depending on the expected morphology of interest. In the GA-SCFT simulations themselves, collocation meshes of 48 and 64 collocation points in each dimension were used for the $15\,R_g$ and $20\,R_g$ initial lattice parameters, respectively. Upon SCFT relaxation during the GA, lattice vectors ranged from $14.5\,R_g$ to $17.5\,R_g$ in the $L_{init} = 15\,R_g$ simulations, implying a range of

spatial resolutions from $0.302\,R_g$to $0.365\,R_g$. In the $L_{init} = 20\,R_g$ simulations, lattice vectors ranged from $19.8\,R_g$ to $22.2\,R_g$; the range of spatial resolutions for this set of simulations was $0.309\,R_g$ to $0.347\,R_g$.

The appropriate defect-free, 3D hexagonally-packed cylinder morphology was ultimately achieved for both $L_{init} = 15\,R_g$ and $L_{init} = 20\,R_g$ within a modest number of generations. However, the simulations that were run in the smaller cell size obtained the defect-free structures in fewer generations. In both cases, we observe the removal of defects as the search progresses (Figure 2.22), and the populations eventually collapse onto a single morphology due to the relatively large selection bias employed.

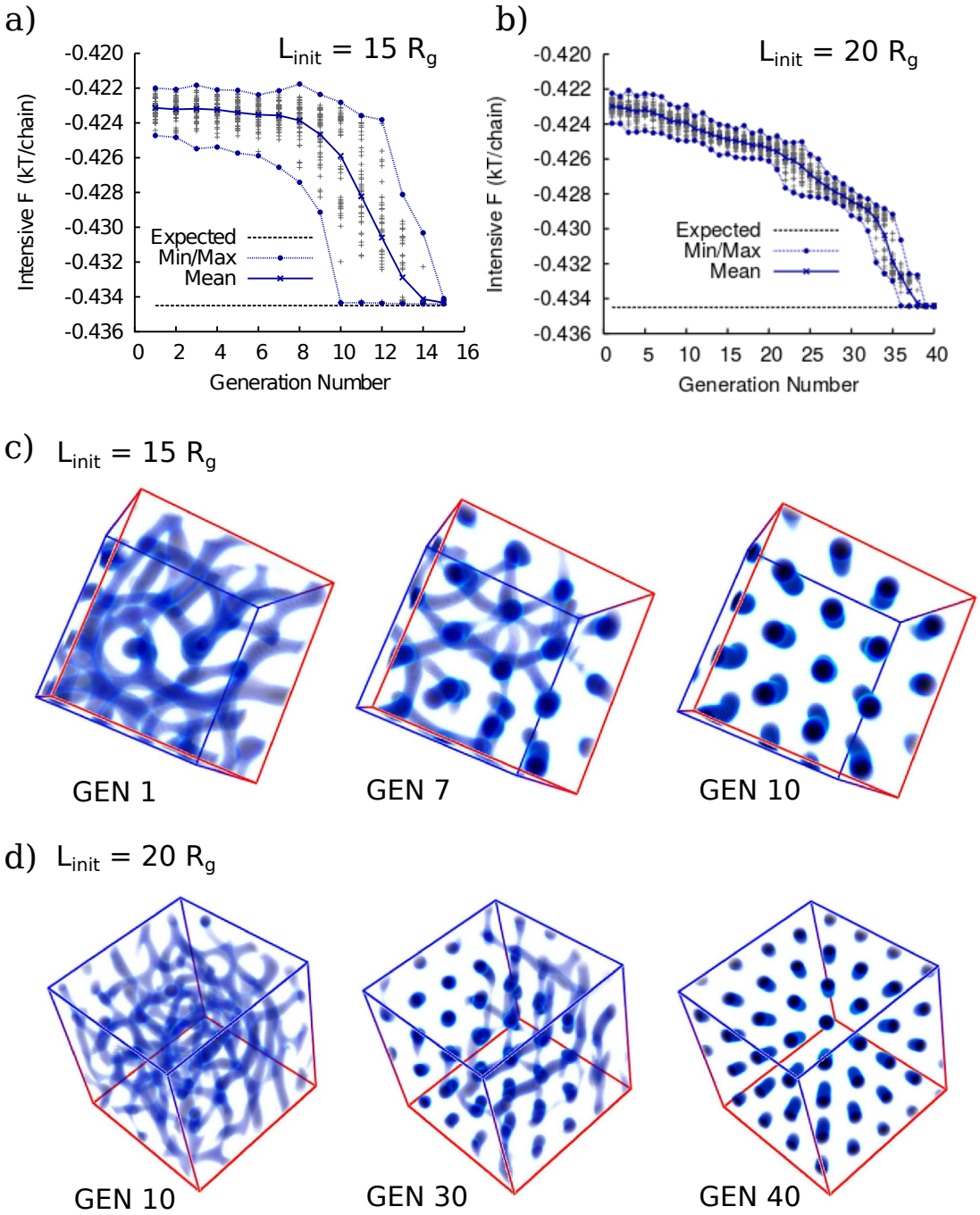

Figure 2.22: Effect of simulation cell size on GA-SCFT performance. GA-SCFT applied to a diblock copolymer melt at χN=15 and f_A=0.35, where the HEX morphology is expected. (a) Plot of the value of intensive free energy of all members vs. generation for members with $L_{init} = 15\,R_g$. (b) Corresponding plot for $L_{init} = 20\,R_g$. The black dashed line in each plot indicates the value of the intensive free energy of the expected, defect-free structure. Plots of the density of A monomers of the lowest free energy member for the indicated generations of the $L_{init} = 15\,R_g$ (c) and $L_{init} = 20\,R_g$ (d) simulations.

Population collapse is evident in (Fig. 2.22) – eventually all members collapse onto the same structure. In this example, population collapse is not a problem because the most stable phase is found first. However, to confidently apply GA-SCFT for structure discovery in new systems, it is critical to maintain population diversity in order to reduce the change of false convergence to a locally stable but globally metastable structure. An important factor that can affect whether population collapse will occur is the choice of the number of members used in each generation. The effect of varying population size will be explored in Section 2.6.

The results we present for large-cell simulations in Section 2.7 $L_{init} = 20\,R_g$ simulation cells, which provide representative large-cell tests for our method.

2.5.3 Averaging Parent Cell Sizes

In our initial work, we took the average of parent cell sizes during the crossover step. This seemed reasonable, since the idea with crossover is to "genetically recombine" the individuals who were selected as parents. Averaging the simulation cell sizes of parents was not a problem with the 2D benchmarking we performed, and even in the case of 3D benchmarking for LAM, HEX, and BCC there were no notable issues.

Figure 2.23 displays a plot of final simulation cell size of each member vs generation for GA-SCFT benchmarking at a composition where GYR is the target morphology. From the figure, we see that when cell averaging is included (red), we encounter the problem of the simulation cell size growing over the course of successive generations of GA-SCFT. This is problematic if we recall that the individual SCFT simulations are run with a spatial resolution of 64 plane waves in each dimension. A simulation cell size of 60 would mean that the resolution, Δx, would be roughly $1\mathrm{R}_g$, which would yield totally inaccurate intensive free energies. Intensive free energy is the fitness met-

ric for the GA-SCFT method, which means if they are inaccurate, "good" structures are not going to be accurately selected, with this effect being exacerbated with the progression of generations. By not averaging the cell sizes of the parents in each generation, but rather resetting the initial cell size of all simulations to some prescribed value, and simply turning on the "variable cell" feature in individual simulations to allow for stresses to relax, we are able to maintain cell sizes to reasonable values for the spatial resolution of the SCFT simulations (blue). We found that in tracking the cell sizes of the other benchmark cases, cell sizes also grow over the course of many generations, but much more slowly than in the case of the GYR benchmarking. It is unclear why this was happening, but also contributed to elongated tubules in later generations.

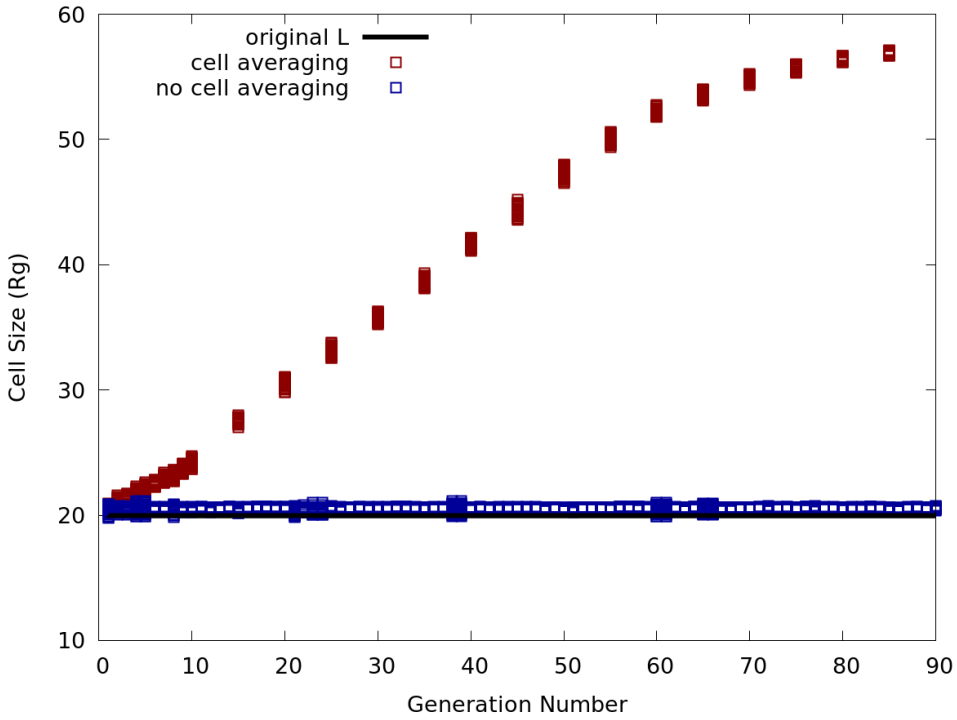

Figure 2.23: Final simulation cell size of each member of the population vs. generation of GA-SCFT applied to a composition where GYR is stable with cell-averaging as part of the crossover move (red) and without cell-averaging as part of crossover (blue).

2.5.4 Removing Divergent Trajectories

In some cases, especially for the trials of GA-SCFT with the unscaled additive random white noise, and also for some of the instances where homogeneous, disordered fields were used, individual SCFT would exhibit divergent trajectories. One way of dealing with this is to simply ignore those members as competitors for the parent pool. This could be problematic, since having divergent trajectories means the parent pool for a given generation would be reduced, thus decreasing diversity and opportunity to fully explore the free energy landscape. To address this issue, we removed SCFT simulations which had divergent trajectories by replacing the input fields with random seeds.

2.5.5 Tuning SCFT Iterations

The most computationally expensive step of the GA-SCFT method is the set of SCFT simulations that must be run for each generation. Because of this, it is imperative to tune parameters to minimize individual SCFT run-time. For AB-diblock melts, 3D SCFT simulations seemed to run most efficiently, while also generating accurate results, with the parameters listed in Table 2.4.

Parameter	Value
number of plane waves	32^3 (small cell) or 64^3 (large cell)
chain contour resolution	0.02
modified diffusion equation solver	SOS
field updater	SIS
field timestep size	5
total iterations	20,000
SCFT force stopping tolerance	10^{-4}

Table 2.4: SCFT parameters used in 3D benchmark testing for GA-SCFT.

2.6 Population Size

2.6.1 Importance

Population size is another important consideration for the success and/or efficiency of a GA optimization [45, 50, 51, 52]. It is necessary to employ enough members to maintain diversity in the population for the duration of the search, to more thoroughly explore the configuration space and increase the likelihood of discovering the correct defect-free ground-state structure. However, using more members makes the GA-SCFT method more computationally expensive, resulting in fewer generations completed within a fixed runtime. Thus, a balance must be struck to obtain defect-free structures as quickly as possible. We tested using population sizes of 10-100 members per generation for a composition where HEX was expected at χN=15. A range of 40-200 member populations were tested at a GYR composition. For the studies we have performed, using 50 members per generation appears to be optimal; using fewer members results in populations that are not very diverse, and using more members per generation only adds computational expense with little or no benefit with respect to population diversity and/or how quickly defect-free structures are recovered.

2.6.2 Some 2D studies

Figure 2.24 displays the intensive free energy vs. generation plots for GA-SCFT applied for a range of population sizes for 2D simulation benchmarking with HEX as the target morphology. For all cases, we used a roulette wheel with $\mathcal{P}_{ratio} = 5$ for selection, circular crossover cuts, and mutations employing scaled random white noise chosen from a range of values up to 90% of the original field values. From the plots, we see that increasing member size results in a larger distribution of intensive

free energies, and thus morphologies, across iterations of GA-SCFT. This effect is still present in the 3D simulation cases seen in the next subsection in Figures 2.25 and 2.26, but to a less dramatic extent. With greater diversity comes the opportunity to find more structures. However, using a large population size results in more computational expense, since each member needs to undergo an SCFT simulation in each generation of GA-SCFT. This is one of the recurring themes of exploiting heuristic algorithms such as the GA: the necessity of finding values of relevant parameters, such as population size, which leverage computational expense and optimal results.

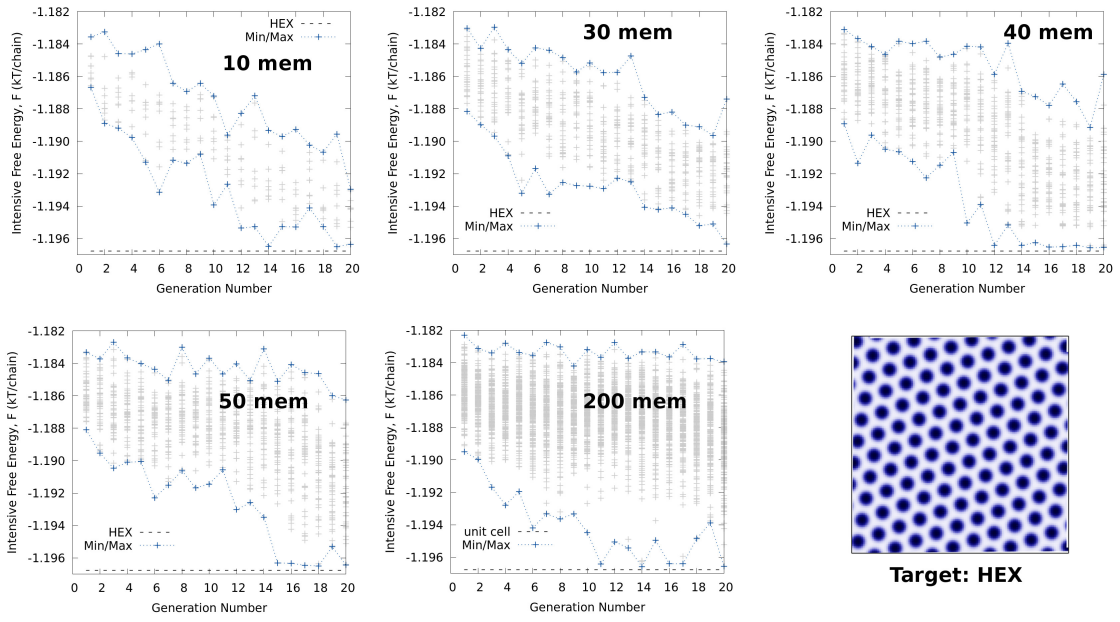

Figure 2.24: Plots of intensive free energy vs. GA-SCFT generation for 2D benchmark testing where HEX was the target morphology. GA-SCFT was applied using a roulette wheel with $\mathcal{P}_{ratio} = 5$, random white noise mutations applied to every member of the population, and a series of population sizes. The population size used in each case is indicated on the plots.

2.6.3 Application to Compositions where HEX and GYR are Expected

We have investigated the effect of population size on the performance of the GA-SCFT algorithm. There is a compromise between computational cost and improving population diversity. For this set of tests, we studied the case of χN=15 at f_A=0.35 (HEX) and f_A=0.38 (GYR). Population sizes ranging from 10-200 were investigated. A number of preliminary population size tests indicated that 1) optimal parameters for the GA-SCFT scheme are not always transferable between studies conducted in 2D and 3D (as seen in comparison with the results from the previous subsection), and, 2) below 50 members per generation, population collapse to defective structures occurred in the GYR simulations, and the spread of free-energies was smaller for HEX simulations. We chose to compare the efficiency of the algorithm using 50 and 200 members per generation for each test morphology.

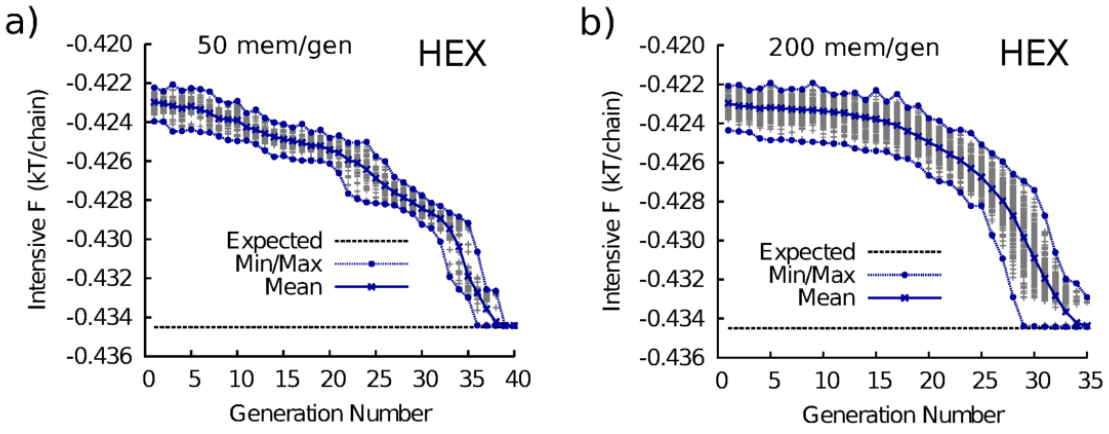

Figure 2.25: GA-SCFT applied at χN = 15, f_A=0.35, where HEX is expected to be stable. Plots display intensive free energy of all members (grey) versus generation number using populations sizes of (a) 50 members per generation and (b) 200 members per generation.

At the HEX composition, the GA-SCFT was able to find HEX as the stable structure within 35 generations, for either population size. In Figure 2.25, we show that the 200 population-size search required almost as many iterations as the search using 50, but the cost per iteration is four times higher. Population collapse is evident in both cases, and is clearly more of a function of strong selection bias than population size.

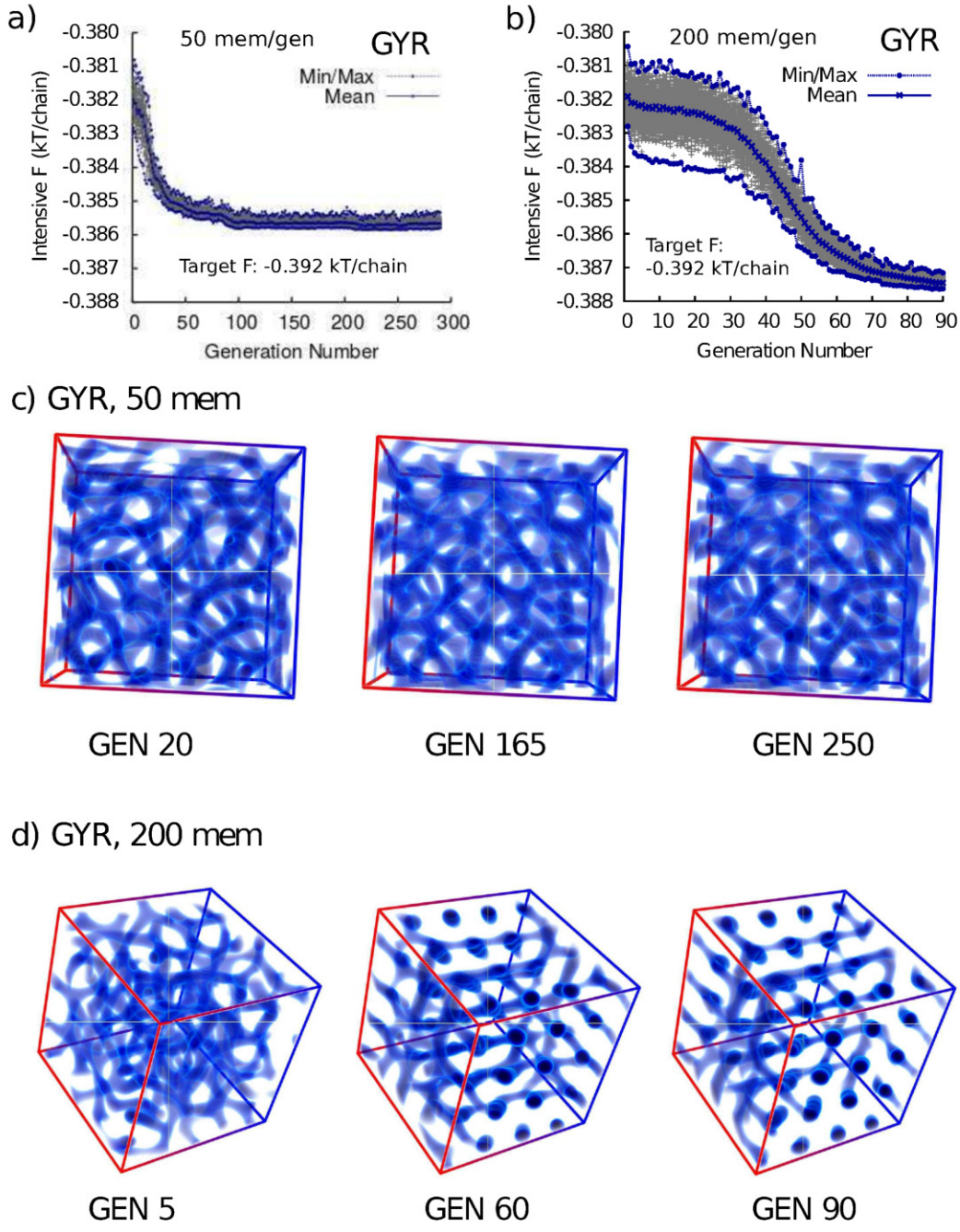

Figure 2.26: GA-SCFT applied at χN = 15, f$_A$=0.38, where GYR is expected to be stable. Plots in (a) and (b) display intensive free energy of all members (grey) versus generation number for 50 and 200 members per generation, respectively. Panels (c) and (d) show volumetric renders of the A-monomer density of the fittest candidate at the indicated generation for 50 and 200 member GYR simulations.

For the studies at compositions where GYR is the target, neither the 50 or 200 member test predicted GYR to be stable (Fig. 2.26); we limited the latter to 90 generations. The morphologies selected in this figure at indicate that the search may have been converging to HEX instead of the more stable GYR. In contrast, the 50 member search did not find any ordered structures within 250 generations. Moreover, the population's free energy spread stabilized for the final 150 iterations, indicating a stagnant search trapped in the space of diverse and highly defective structures.

For all compositions tested, we allowed individual SCFT simulations to relax to stress-free structures with concentration-conserving relaxation [53]. Most SCFT relaxations began with initial cell-tensor guesses produced by averaging those of the parents, but for GYR searches we reset the initial guess to L_{init} for each member of each generation. We found that isolated SCFT simulations performed on highly-defective network structures resulted in lower free-energies for increasingly larger cell-sizes. Thus, defective structures that emerged in the GA-SCFT algorithm would tend to grow within the SCFT local relaxation portion of the algorithm, causing unbounded cell-volume growth as the GA-SCFT iterations progressed. The unbounded growth in the cell-tensors we see may be a complicated interplay between the interface generated during the crossover step for each member and how the GA explores the free energy landscape, but we do not fully understand these interactions. Moreover, we see no evidence that the GYR structure would eventually emerge, regardless of whether or not we average the cell-tensors of parents. Clearly, more development towards an efficient exploration of candidate network phases is needed.

In this study, we only examined the effect of population size on the ability of GA-SCFT to maintain diversity. However, several other factors can affect population collapse. We employ a roulette-wheel selection method, but this can be too aggressive in its preference of lower-free energy members as parents. Preliminary work suggests

that retaining memory of all-time high and low values of intensive free energy for the determination of $\mathcal{P}_{ratio}$ can reduce the selection bias and assist in preventing population collapse. Other selection methods, such as tournament selection, may provide a less aggressive alternative without sacrificing search efficiency, as one would encounter with simply reducing $\mathcal{P}_{ratio}$ [54]. Further investigation of this method is not pursued here, but would be an interesting direction for future work.

2.7 The Effects of Simulation Cell Size and χN

The results presented in this section for large-cell GA-SCFT are based on simulations that were performed at the diblock compositions listed in Table 2.1 in $L_{init} = 20\,R_g$ cells. To mitigate computational cost, and noting the diminishing returns of using more members from Figures 2.25 and 2.26, 50 members per generation were used for the remaining tests. Moves for this application of the GA-SCFT method follow the prescriptions from Table 2.3.

2.7.1 large-Cell Simulation Results

"Low" χN

At χN=15, BCC, HEX, and LAM morphologies emerge at the appropriate block fractions where they are expected to be stable; for BCC and LAM, this occurs within a few generations (Figure 2.27).

As we saw in Figure 2.26, the application of the GA-SCFT method in large-cells at compositions where GYR was expected resulted in the algorithm becoming trapped in the space of diverse and highly defective structures. Hexagonally packed cylinders were not the only structures obtained from simulations where GYR was expected. Simulations at χN=14 and f_A=0.4, where GYR is maximally stable relative to other candidate phases, were also performed. At this composition, the Fddd phase (O^{70}) and perforated LAM were both found using our GA-SCFT scheme (Figure 2.28). This indicates that algorithm is able to access a variety of candidate defect-free structures. However, once a cleaner structure emerges, it appears the algorithm is unable to maintain any diversity, presumably due to strong selection bias, and population collapse occurs.

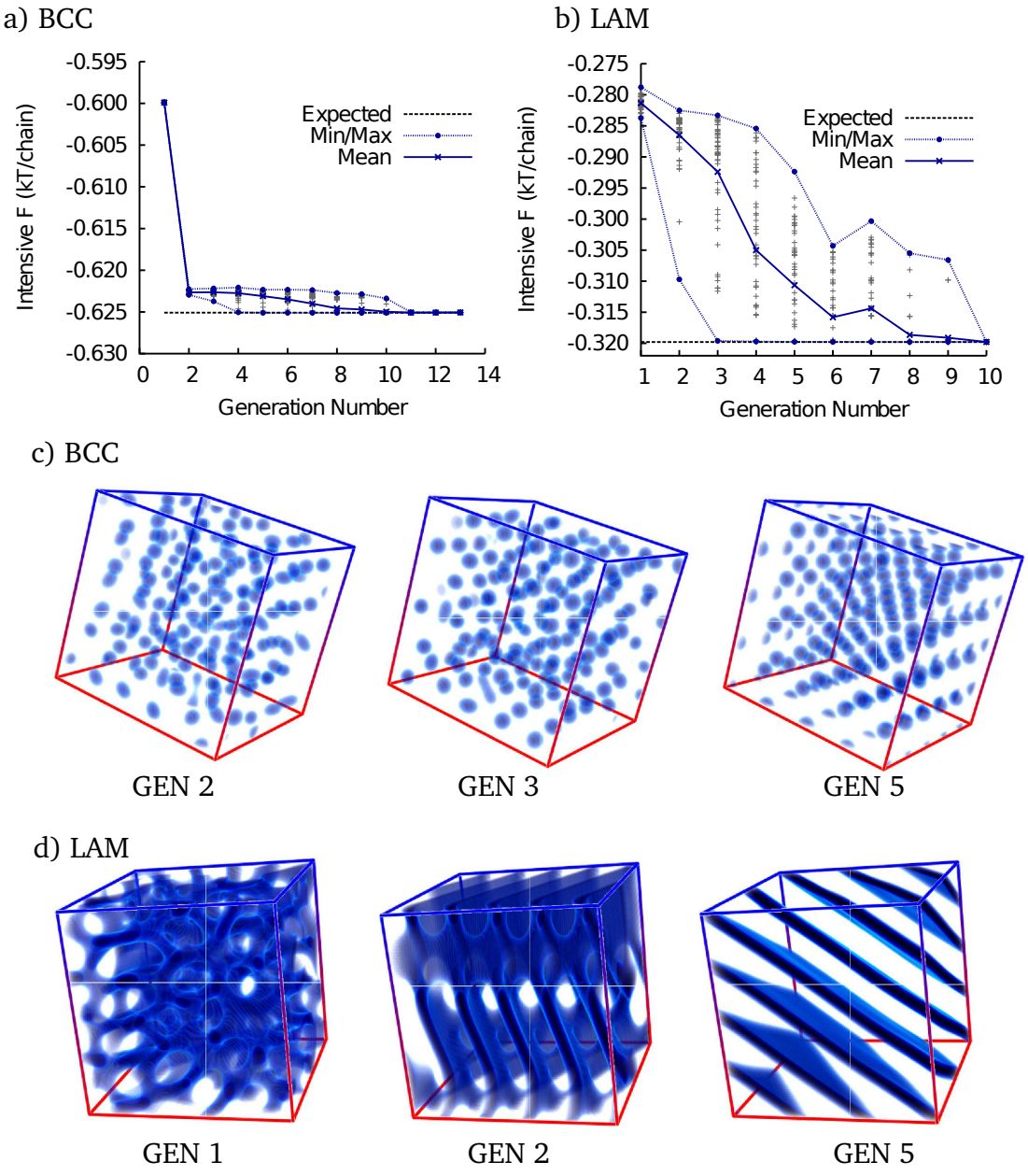

Figure 2.27: χN=15 large-cell GA-SCFT. GA-SCFT applied to a diblock at χN = 15, f_A=0.30 (a) and f_A=0.50 (b), where BCC and LAM morphologies are expected, respectively. (c) Plots of A monomer density of the lowest intensive free energy member at the indicated generations from simulations where BCC is expected. (d) Plots of A monomer density for the simulations where LAM is expected.

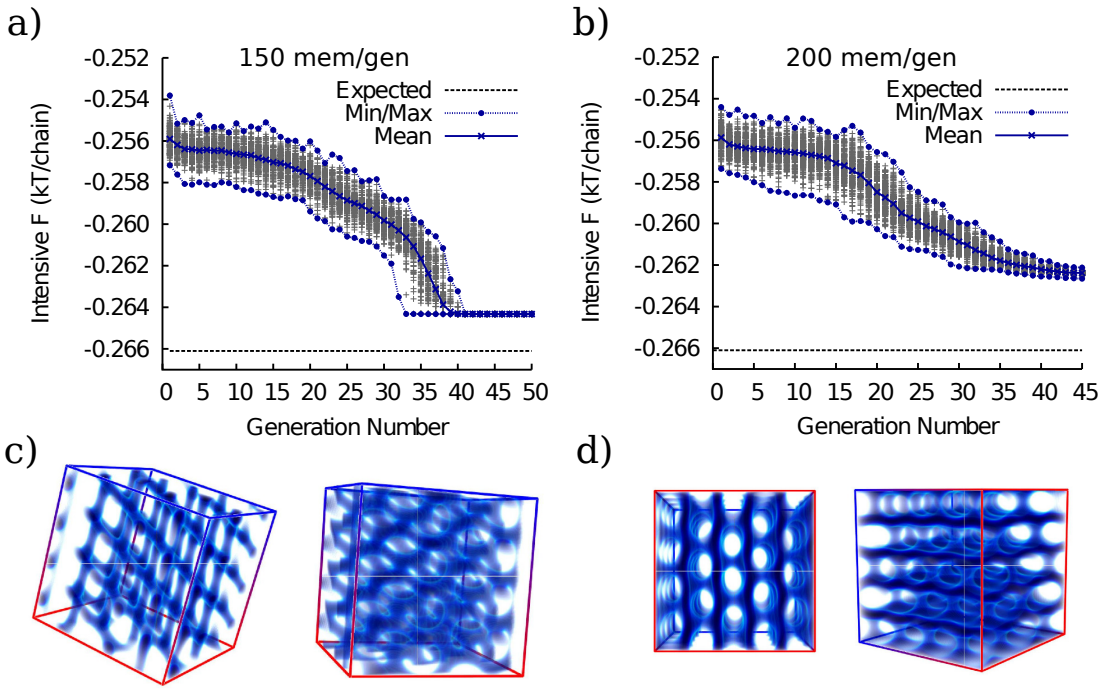

Figure 2.28: Other morphologies obtained where GYR is expected. GA-SCFT applied to a diblock at χN = 14, f$_A$=0.40, using 150 members per generation (a) and 200 members per generation (b). O^{70} structures were obtained for the 150 member simulations; alternative views of the densities are given in panel (c). The 200 member GA-SCFT yielded perforated LAM; plots of A monomer density are provided in panel (d).

"High" χN

At χN=20, GA-SCFT applied to the HEX and LAM compositions successfully produces the target morphology, but was unable to obtain a defect-free BCC structure (Figure **??**). Furthermore, at this higher χN, the algorithm required more generations to produce the expected structures. This increasing difficulty for the GA-SCFT scheme to find ordered structures at larger χN is perhaps not surprising because it is likely that the roughness of the free energy landscape increases with segregation strength.

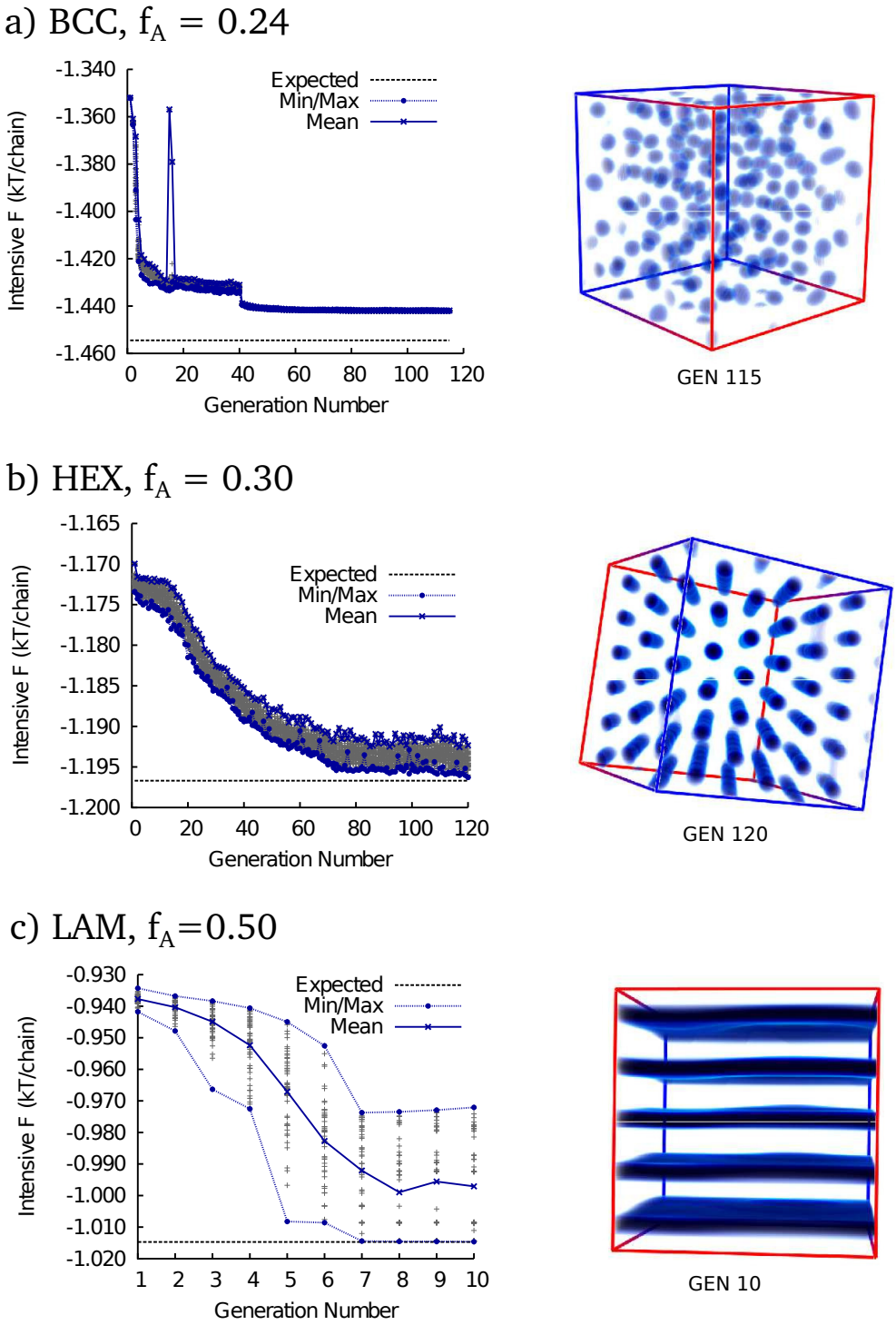

Figure 2.29: χN=20 large-cell GA-SCFT. GA-SCFT applied to a diblock at χN = 20, f_A=0.24 (a), f_A=0.30 (b), and f_A=0.50 (c), where the BCC, HEX, and LAM morphologies are expected, respectively. Plots of A monomer density of lowest free energy member of final generation are shown in the right column for each morphology.

2.7.2 Challenges of Large-Cell Simulations and High χN: High Density of Defective States

An interesting and related question is whether the density of defective states also increases with χN. To explore this issue and also establish differences in the accessible states between morphologies, we performed 100 isolated, SCFT-only (no GA) 3D simulations with fields seeded from uniform random white noise – for HEX and GYR at χN=14 and χN=20 (Figures 2.30 and 2.31). We then generated a histogram of the per chain intensive free-energies. This is an indirect probe of the density of defective states, because it only counts states that were obtained from random quenches using our numerical field relaxation algorithm. One can view these free energy histograms as a convolution of the actual density of states and the non-uniform probability that our SCFT update scheme relaxes into any given state. The lattice vectors in each direction of the simulation cells were fixed to be commensurate with one or two times the primitive unit-cell of the expected phase.

From the histograms, we make two observations: (1) fewer defect-free structures emerge from random initial conditions at higher values of χN, suggesting either a rougher free energy landscape with a higher density of defect states, or a stronger tendency for our SCFT relaxation algorithm to converge onto defective structures, and (2) defective structures are more accessible to our field relaxation algorithm in the larger cells, but the dispersion in free energy of these defective structures is confined to a smaller range.

No defect-free structures were obtained for either HEX or GYR in the large-cell, higher χN simulations. We believe this high density of accessible defective states is manifest in the observation that the GA-SCFT technique becomes trapped among numerous defective GYR states in the large-cell simulations (Figure 2.26).

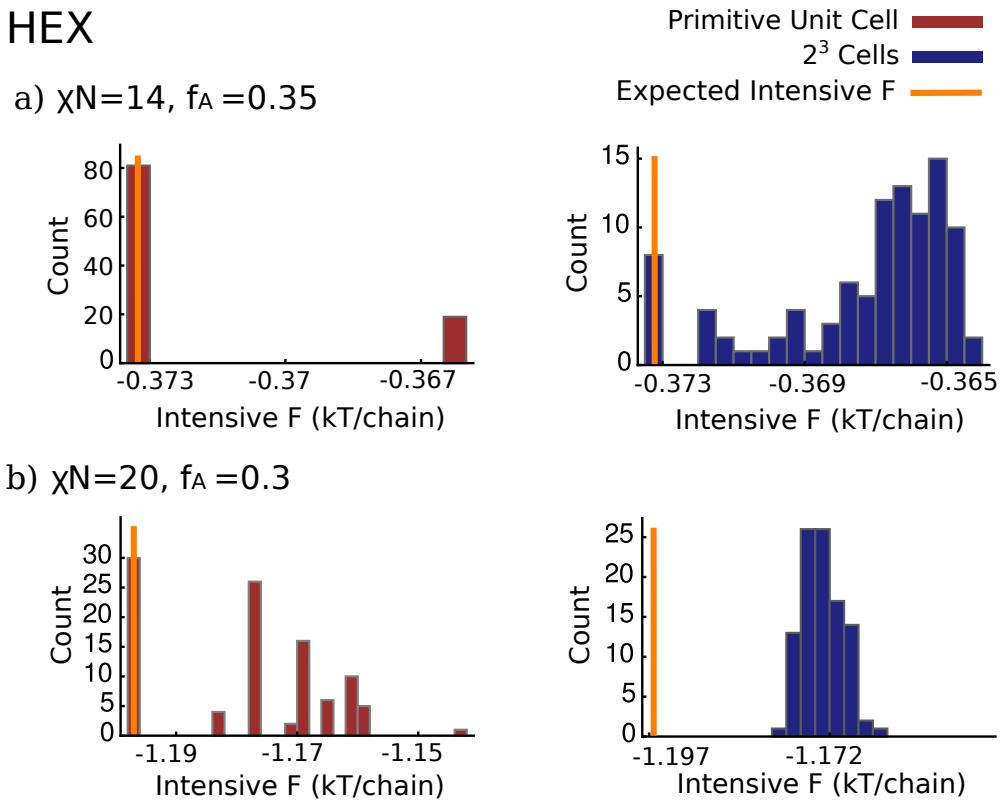

Figure 2.30: Histograms of intensive free energy for SCFT simulations from random initial conditions at the indicated compositions consistent with HEX. Simulations were performed in fixed-cell boxes commensurate with unit-cell (red) and two times the unit-cell lattice vector in each dimension (8 primitive cells, blue). The orange line in each histogram denotes the expected intensive free energy (kT/chain) for the corresponding perfect state. 100 independent simulations were run for each histogram.

In this subsection we observed the lowest number of accessible defective states for the small-cell and low χN systems. However, similar quenches performed for off-commensurate simulation cells have larger numbers of accessible defective states in all cases than their corresponding commensurate counterparts. This effect has some important ramifications. The goal of the GA-SCFT method is to find the globally stable morphologies of given compositions and architectures. Commensurate cell lattice vectors for these morphologies are not established *a priori* since the morphologies

themselves are not known. From the free energy histograms in Figs. 2.30 and 2.31, we see that the problem of having more accessible defective states at higher χN values is exacerbated in larger simulation cells. Recognizing this and the fact that commensurability also influences the distribution of states suggests that the GA-SCFT method applied to large-cell simulations is faced with the daunting challenge of thoroughly exploring an enormous number of defective states and cell-tensor variations on the way to locating the singular globally stable state. It is reasonable to infer an improved success and efficiency of finding the globally stable structure in a smaller simulation cell.

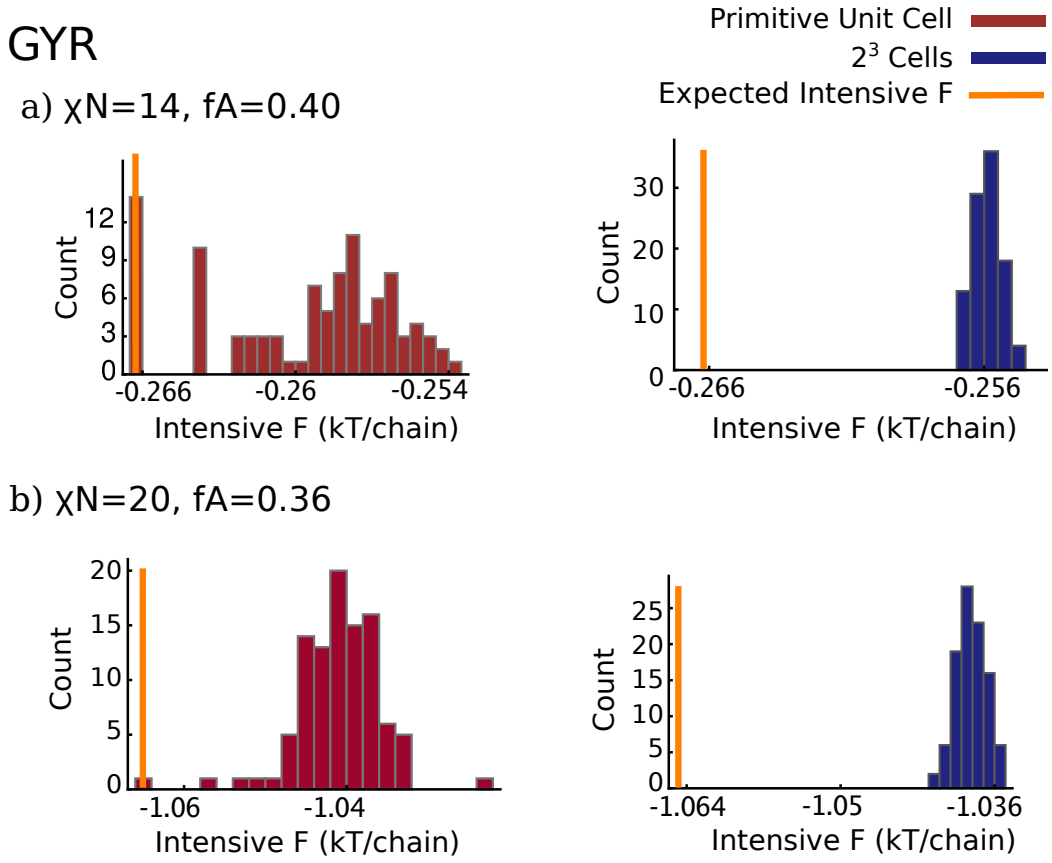

Figure 2.31: Histograms of intensive free energy for SCFT simulations from random initial conditions at the indicated compositions consistent with GYR. Simulations were performed in fixed-cell boxes commensurate with unit-cell (red) and two times the unit-cell lattice vector in each dimension (8 primitive cells, blue). The orange line in each histogram denotes the expected intensive free energy (kT/chain) for the corresponding perfect state. 100 independent simulations were run for each histogram.

2.7.3 Small-Cell Simulation Results

In this section, we present studies of the GA-SCFT algorithm applied to small-cell simulations, which are computationally less costly than their large-cell counterparts and we expect also have a smaller density of defective states to trap the search. Small-cell simulation results presented here used initial lattice parameters that are close, but not commensurate, to the unit-cell size of the largest of the classical phases

(gyroid, GYR) at the composition and interaction parameters selected for testing. We present results using an initially cubic simulation cell of side length $9\,R_g$. SCFT fields for small-cell simulations were sampled with 32^3 collocation mesh points.

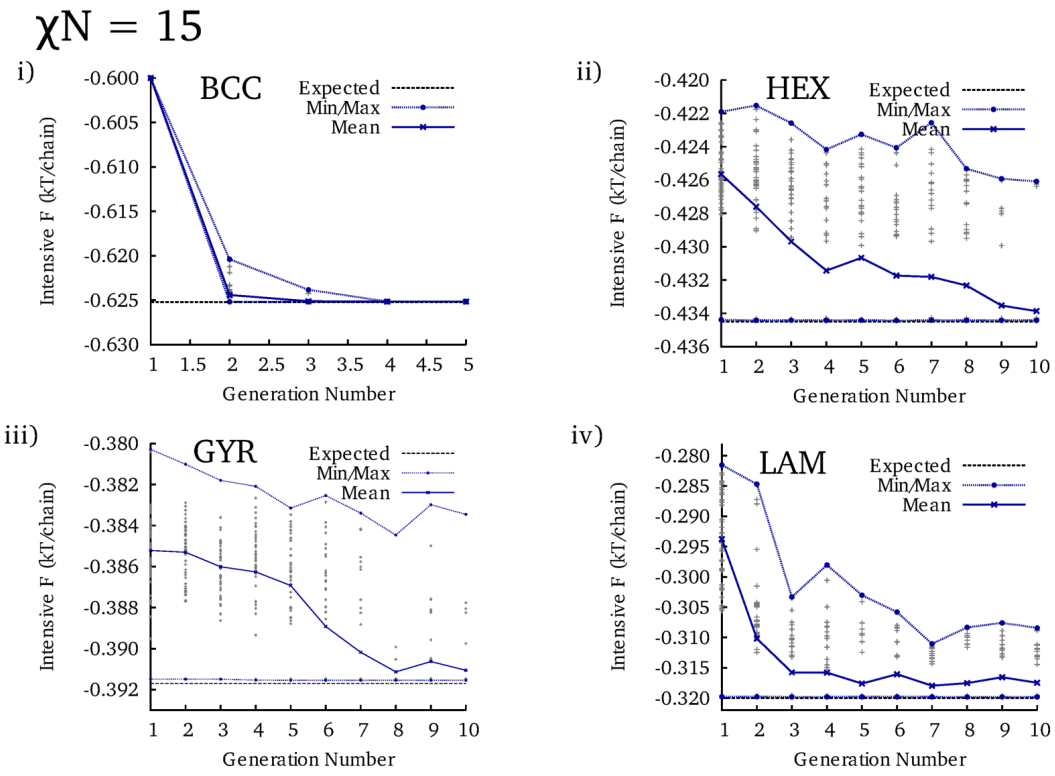

Figure 2.32: Small-cell GA-SCFT. Simulations were run in a variable cell with $L_{init} = 9\,R_g$ and 32 collocation mesh points in each dimension for a diblock at the compositions listed in Table 2.1. Plots of intensive free energy for the following expected morphologies at χN=15: (i) BCC, (ii) HEX, (iii) GYR, (iv), LAM.

Indeed, our small-cell simulations obtain morphologies that are consistent with the correct microphases for all of the expected phases at χN=15 (although random quenches often succeed in this limit, and indeed the perfect structure is often present in the first generation, before GA moves have been executed), and every expected morphology at χN=20 except for GYR (Figures 2.32 and 2.33). This is compatible with findings from the density-of-states study in the previous section. However, ex-

clusively enforcing small-cell simulations makes discovering new morphologies more difficult; there are possible structures that can undoubtedly not be found by initializing the algorithm from a small cubic cell. Even with cell relaxation methods, there is no guarantee that such structures will be observed; initial cell sizes and shapes still play a complicated role in the the morphologies obtained by SCFT or GA-SCFT.

GYR and the sphere phases seem to be particularly sensitive to initial cell sizes, even for the small-cell GA-SCFT simulations. GA-SCFT did not obtain the expected GYR morphology when $L_{init} = 9\,R_g$, which is not commensurate with the lattice parameter for GYR at χN=20, f_A=0.36 (Figure 2.33). However, using initial lattice parameters commensurate ($L_{init} = 9.5\,R_g$) with the expected GYR morphology at this composition yields GYR within 3-4 generations. Similarly for compositions where BCC is expected, only when GA-SCFT is applied to near-commensurate cells are BCC morphologies obtained ($L_{init} = 9.2\,R_g$ for χN=15, f_A=0.3, and $L_{init} = 9.3\,R_g$ for χN=20, f_A=0.24). When GA-SCFT is run at a BCC composition, but with an initial lattice parameter that is $\gtrsim$ 10% off-commensurability, BCC was not found. Instead, a competing phase such as cylinders or a defective sphere/micelle structure emerges. An ostensible solution to this would be to seed the first generation with a range of initial lattice parameters to ameliorate the sensitivity we observe. When this approach was attempted, the competing cylinders disappeared. However, perfect BCC structures were not obtained; only defective sphere morphologies appeared.

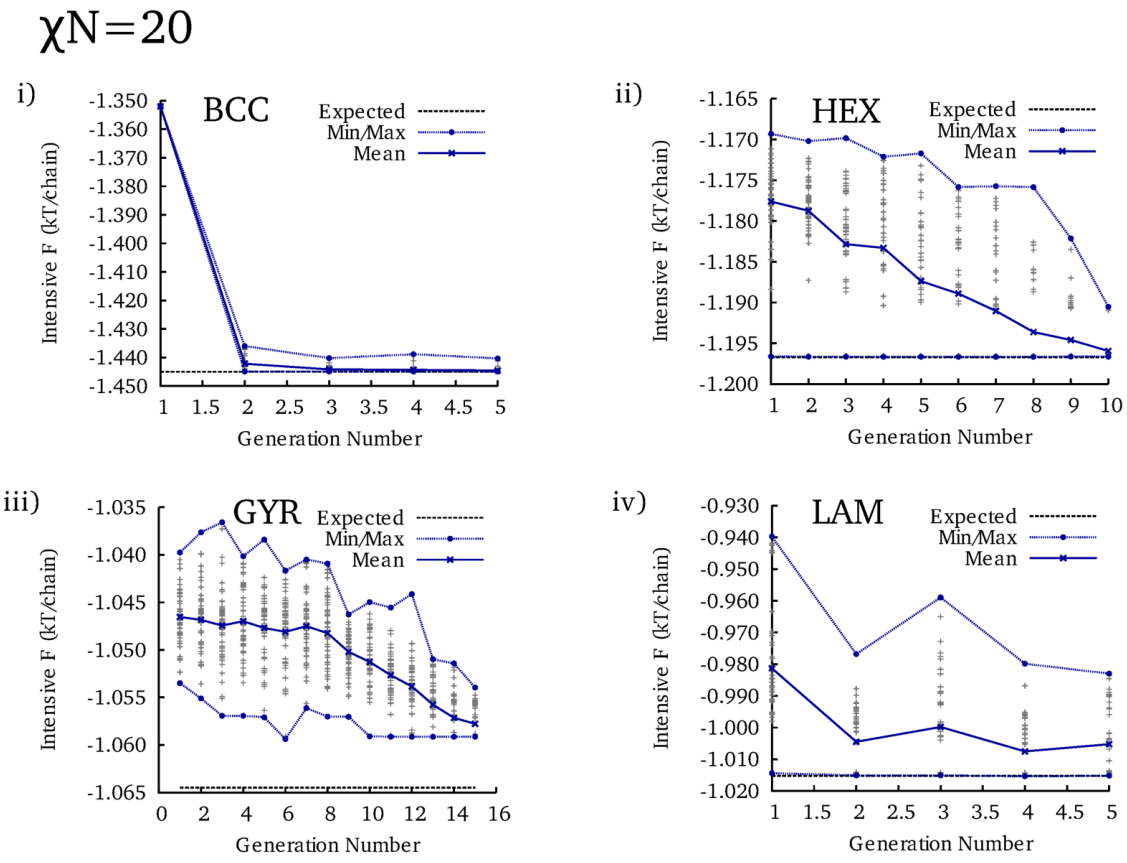

Figure 2.33: Small-cell GA-SCFT. Simulations were run in a variable cell with $L_{init} = 9\,R_g$ and 32 collocation mesh points in each dimension for a diblock at the compositions listed in Table 2.1. Plots of intensive free energy for the following expected morphologies at χN=20: (i) BCC, (ii) HEX, (iii) GYR, (iv), LAM.

2.8 Concluding Remarks

For a fixed composition and segregation strength, an unguided SCFT search in three dimensions is not likely to find the globally stable mesophase structure; the system can become trapped in one of numerous metastable states that correspond to defective structures or other metastable polymorphs. The GA-SCFT method presented here represents a significant advance in the challenge of mesophase discovery, *i.e.*, identifying the globally stable phase of a specified block copolymer.

The GA-SCFT method is a real-space, global optimization method. It has successfully recovered the canonical non-network diblock phases HEX, LAM, and BCC at the compositions where they are expected to be stable in both large- and small-cell 3D simulations from random initial conditions. At a higher χN=20 value, large-cell determination of the correct structure was slower, but still successful for HEX and LAM, while it did not recover the BCC phase. GYR was not obtained in any of the large-cell investigations, but was achieved for χN=15 (from a random quench in generation 1) with a small, near-commensurate initial lattice parameter, $L_{init} = 9\,R_g$. The failure of the GA-SCFT method at higher segregation strengths for BCC in large simulation-cells and generally for GYR likely arises from an interplay between two factors: 1) more numerous defective states accessible in larger simulation cells that are incommensurate with the unit cell, and 2) higher barriers at stronger segregation strengths. It is probably the case that free energy alone does not provide sufficient contrast to determine the fitness of individual members.

It remains a challenge to determine more efficient ways of exploring this vast configurational space and accurately predict the global stability of GYR and other network phases. A critical element toward this end is the identification of an economical and effective way to prevent population collapse. Our preliminary studies

indicate that employing a less aggressive selection method than the roulette-wheel to maintain population diversity and retaining memory of all-time low and high free energy members instead of generational low and high free energy members may increase fitness diversity. We might also be able to tune GA parameters more meticulously to obtain improved results. Furthermore, though we explored a variety of options for selection, crossover, and mutation strategies, the list was by no means an exhaustive one. Methods like genetic algorithms have a huge parameter space of their own to explore and optimize. Their power, but also their shortcomings, lie in the fact that they are so modular and extensible. They appear to be more problem-dependent than one would hope, and so it is important that one frames a problem very carefully when designing such an algorithm.

Nevertheless, we believe the results we have observed with the current GA-SCFT technique constitute a significant step in solving the forward problem of determining globally stable phases of block copolymers. GA-SCFT has the exciting potential to be employed for structure discovery in multi-species, multi-component systems [32, 55]. In a multi-species system, the primary change is the addition of more fields and interaction terms to the Hamiltonian. Since the GA only requires a measure of fitness (*e.g.*, the scalar-valued free energy), and the ability to manipulate and combine the fields that gave rise to that fitness value, the method immediately extends to multi-species systems. The addition of different molecular constituents and architectural complexity changes the SCFT Hamiltonian by modified single-molecule statistics terms. It does not modify the number of fields nor the nature of the fitness metric. Thus, extending GA-SCFT to accommodate more complex molecular constituents and architectures would be straightforward. Expediting the discovery of new morphologies and the composition-spaces in which they may be found could have broad-reaching impacts in terms of material applications, and furthermore is an

important element in realizing a true inverse approach to materials design [56, 57].

Ultimately, it seems necessary that one explores methods which can reduce the dimensionality of the configuration space that must be searched for candidate structures. For example, a large-cell calculation on a $128 \times 128 \times 128$ grid represents each density or potential field using over 2×10^6 degrees of freedom, each of which can be varied continuously. Strategies that invoke a significantly smaller number of collective variables such as Fourier modes or blocked field values could potentially focus the search in optimally productive regions for pattern evolution, thereby accelerating heuristic algorithms aimed at global optimization. In the next chapter of this thesis, we explore such options.

Chapter 3

PSO-SCFT

3.1 Introduction and Overview

In the previous chapter of this dissertation, we saw that the GA-SCFT methodology was successful in the agnostic identification of non-network morphologies of conformationally symmetric diblock copolymers in large simulation cells. However, we also observed that GA-SCFT failed quite spectacularly at recovering the GYR network morphology when it was expected to be stable for the same benchmarking system. Particularly, the algorithm would become "stuck" in numerous defective metastable morphologies which visually resembled a vat of tangled spaghetti. Otherwise, the population of a given instantiation of the method would collapse prematurely into a competing morphology such as HEX or LAM.

Let us remind ourselves of our objective: We would like to be able to determine the globally stable morphology and its low-lying metastable competitors (*i.e.*, solve the forward problem) agnostically. In this chapter, we explore an alternative method for solving the forward problem. In the last chapter, we were concerned with being able to do this in a large simulation cell, to facilitate the identification and discovery

of morphologies of which we do not know the unit cell sizes *a priori*. However, in the approach we will discuss in this chapter, we will frame the problem differently, so that the unit cell size is a dimension we add to the parameter space we are attempting to optimize. We will also manipulate a small set of reciprocal space variables, rather than the entire real-space SCFT field values, which greatly reduces the degrees of freedom we need to optimize. A schematic representation of the reduced search space can be observed in Figure 3.1. The idea is to search through a space of "single-mode" SCFT seeds, which consist of peaks that populate a thin shell whose radius is approximately the primary wave vector, q^*, away from the origin in reciprocal space. The width of this shell is set by the nearest neighbor distances between grid points in our simulation cell, and the value of q^* can either be estimated by following the recipe specified by Leibler[58] via a Random Phase Approximation (RPA), or by incorporating is as a variable which the PSO-SCFT algorithm optimizes. By varying the relative positions and amplitudes of these peaks, as well as q^*, we can search through a variety of morphologies, since SCFT is a deterministic simulator and each seed corresponds to a specific structure. These tasks will be accomplished by combining SCFT with the framework of a Particle Swarm Optimization (PSO), which is another heuristic algorithm that can be used for global optimization problems.

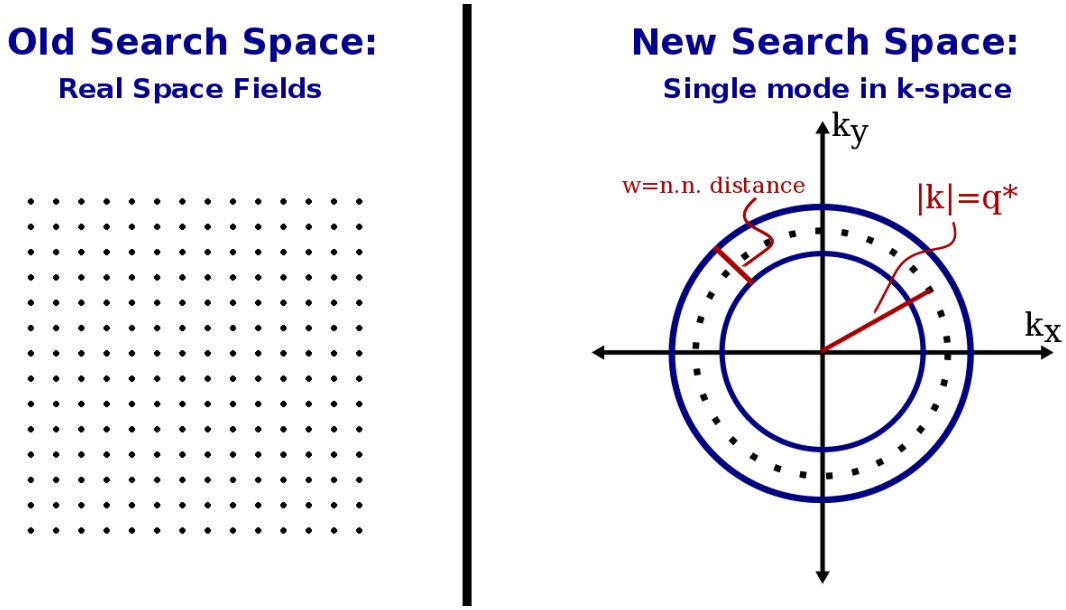

Figure 3.1: Schematic depiction of the reduced search space in the reciprocal space PSO-SCFT method. Instead of optimizing all field values in real space as in the GA-SCFT method, only the positions of a set of peaks in a single mode in reciprocal space are optimized.

3.1.1 What is Particle Swarm Optimization?

Particle Swarm Optimization is a type of biologically-inspired, population-based heuristic algorithm that was first developed by Kennedy and Eberhart in 1995[59]. These algorithms have been successfully implemented for a variety of optimization problems, including those in economics, various fields of engineering, protein structure determination, and crystallographic structure prediction [60, 61, 62, 63, 64, 65]. Recently, PSO has also been used in conjunction with SCFT towards solving the inverse design problem: determining compositions and architectures that would yield a desired pattern (morphology) [56, 57]. However, this inverse problem relies on being able to accurately solve the forward problem: determining the stable morphology and low-lying metastable competitors that emerge for a given block copolymer composition and architecture. Optimization algorithms such as the PSO are versatile; their

success depends largely on how well a problem is framed within the context of how such an algorithm works. Features of a given algorithm itself are largely modular, and can be extended to include more elaborate embellishments. However, for the work we present in this chapter, we employed a basic PSO algorithm.

In a PSO algorithm, a number of particles, or agents, are initialized with random positions and velocities at various points in a search space of interest. In each iteration of the PSO, each agent's fitness is assessed at their current position. Agent positions and velocities are then updated using information about each agent's personal fittest (best) position and the swarm's best known positions. For our work, we use the intensive free energy determined by SCFT as the metric to compare agent fitness. Typical PSO convergence times occur on the order of hundreds to thousands of iterations for various test functions. However, mimetic algorithms, which combine a global optimization strategy such as a PSO with an efficient local-optimization technique, appear to converge more rapidly to optimal solutions than the unembellished PSO. With this in mind, as well as inspiration drawn from the crystallographic community where PSO has been used in conjunction with DFT packages, we propose the implementation of PSOs as a wrapper around SCFT to facilitate structure discovery.

3.1.2 PSO-SCFT Approach

We initialize the PSO-SCFT method with a population of possible structures generated from randomly initialized SCFT seeds. The seeds themselves are the agents of the PSO, and are comprised of peaks (field values) in reciprocal space which are placed within a thin shell at radius q^* away from the origin. The idea is to take advantage of the fact that many morphologies of interest in block copolymer systems can be constructed from a single-mode approximation, which relies on the proper relative placement and phase relationship between peaks on the surface of a sphere in reciprocal space. In a single-mode approximation, polymer densities, ρ, are constructed by retaining only the first term in a sum over basis functions, ϕ that contribute to the density: $\rho = \sum_n c_n \phi_n$, where $\phi_n = \sum_k a_k \exp(2\pi i k x)$. That is, only ϕ_1 is retained to approximate the density.

The number of peaks and their proper relative placement to one another correspond to the symmetry elements represented in a particular space group of interest. We will also see that the relative sign of the peak amplitudes matters, in that they contribute to either cosines (same phase relationship) or sines (opposite phase relationship) in the basis function expansion for the single-mode approximation. For example, the hexagonally-packed cylinders morphology can be constructed by placing six peaks at 60^o angles from one another on the perimeter of a circle whose radius is q^*. q^* is a quantity analogous to the primary wave vector seen in scattering plots, such as those generated by SAXS or SANS experiments. Figure 3.2 shows examples of the results from SCFT simulations which were initialized with single-mode seeds at compositions where the indicated morphologies are expected. The orange peaks in the single-mode seed which resulted in the GYR morphology indicate the opposite phase from the blue peaks.

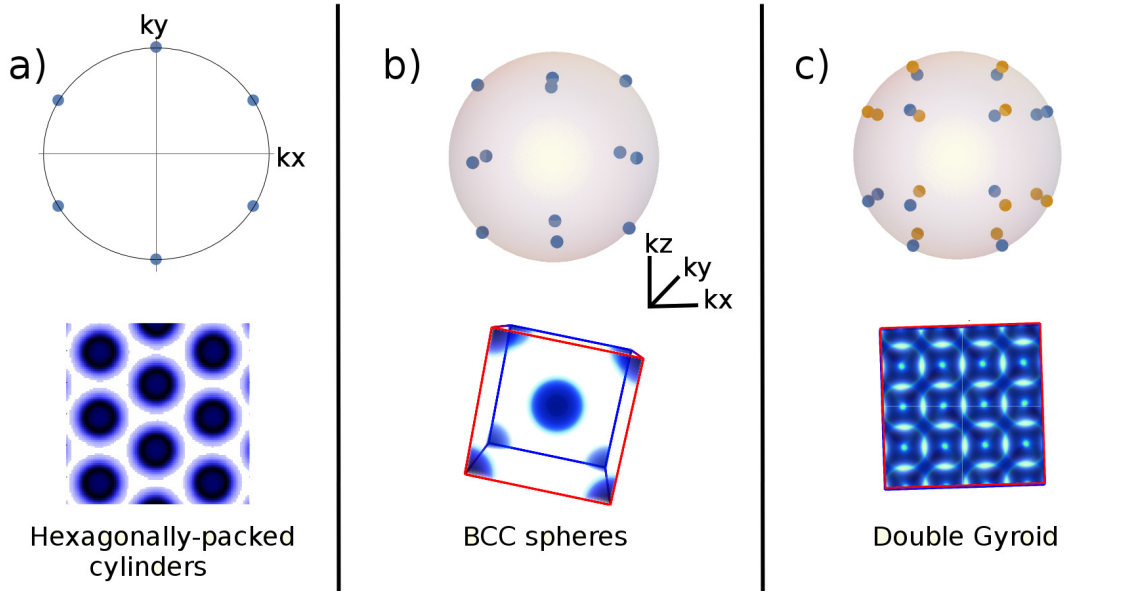

Figure 3.2: Morphologies from SCFT simulations initialized with single-mode seeds at compositions where a) HEX, (b) BCC, and c) GYR are the expected morphologies.

Once the single-mode seeds (agents) have been relaxed using SCFT, we evaluate their relative fitness by comparing their intensive free energies as determined from each SCFT simulation. Here we would like to point out that SCFT simulations are deterministic, and therefore result in field configurations which are specific to the initial conditions (seed) of the simulation. This means that each different single-mode seed corresponds to a different structure, each with its own intensive free energy, or fitness, in the context of PSO-SCFT. **Thus, the space which the PSO-SCFT method searches through is comprised of *sets* of single-mode seed peaks which correspond to the various structures accessible through SCFT.** That is: each agent, or particle, consists of a set of peaks on the surface of a sphere of radius q^*.

The SCFT simulations initialized with single-mode seeds thus act as a local optimization step. The PSO update equations described in the subsection below serve in turn to provide stochastic, global optimization moves. Each iteration of the PSO-

SCFT method then consists of the following steps:

1. Relax agents (a set of single-mode seeds) locally with SCFT.
2. Evaluate all agents' fitnesses (intensive free energies).
3. Apply PSO update equations to the peak positions, peak velocities, peak amplitudes, and simulation cell size via a variable called q^* (discussed in the next subsection) to generate new single-mode seeds.
4. Repeat steps 1)-3) until maximum PSO-SCFT iterations have been attained.

This process is shown in the schematic in Figure 3.3. Inputs to the PSO are the number of agents, number of peaks per agent, peak amplitudes, and q^*, which is related to the simulation cell size by $q^* = \frac{2\pi n}{L}$, where n is an even integer that determines the number of periods of the primary wave vector that exist in a simulation cell of length L.

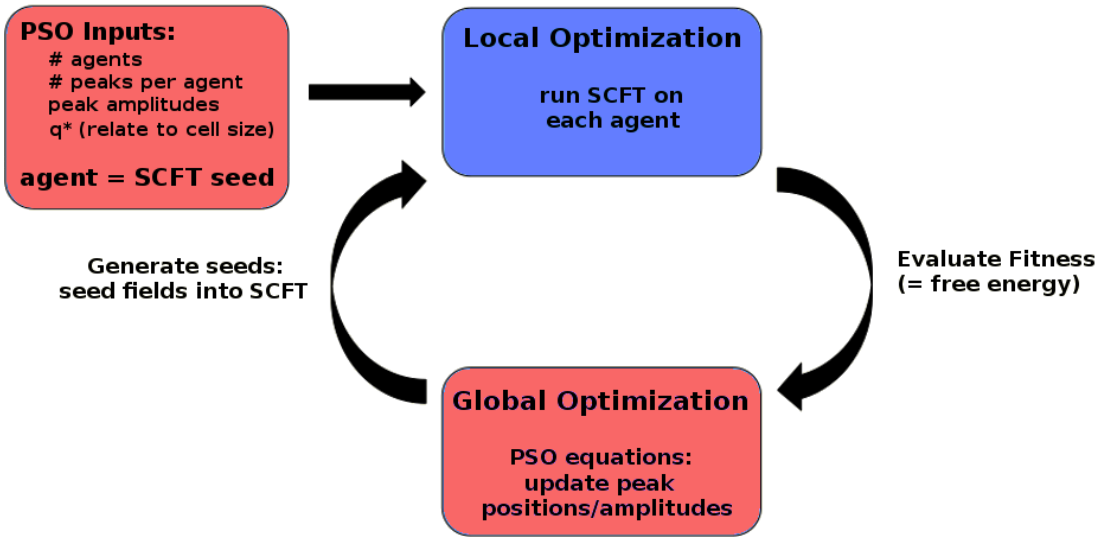

Figure 3.3: Schematic of the PSO-SCFT method.

3.1.3 Benchmark Testing Set-up

In this chapter, we will investigate the application of the PSO-SCFT methodology to two systems: the AB-diblock which we had studied using the GA-SCFT approach, as well as an AB_4 miktoarm star system, which yields a variety of interesting sphere microphases, including A_{15} and the Frank-Kasper σ phase[27]. For both cases, we explore the application of PSO-SCFT with (Section 3.7) and without (Sections 3.4 and 3.5) q^* as a PSO optimization variable.

For the diblock, we present results for applying PSO-SCFT to the compositions listed in Table 2.1 at χN=15, and use the parameters listed in Table 2.4 for "small cell" simulations. For the miktoarm system, we apply PSO-SCFT at χN = 40 with $f_A = 0.32$ and the SCFT parameters listed in Table 3.1.

Parameter	Value
number of plane waves	32^3 (q^* fixed) or 48^3 (vary q^*)
chain contour resolution	0.02
modified diffusion equation solver	SOS
field updater	SIS
field timestep size	4
total iterations	20,000
SCFT force stopping tolerance	10^{-4}
λ_+ (mobility parameter for pressure-like fields)	1.0
λ_- (mobility parameter for exchange fields)	0.5

Table 3.1: SCFT parameters used in 3D benchmark testing for GA-SCFT.

When including q^* as a PSO-SCFT variable to optimize, simulation cell sizes are allowed to vary over a pre-determined window specified by setting the range on q^*. This means that the spatial resolution of individual SCFT simulations may vary quite widely for a given number of plane waves used in each dimension of the simulation cell. Achieving accurate relative values of intensive free energies is imperative, since

free energy is the fitness metric we assess in the algorithm to update agent positions and velocities. Thus, for the SCFT simulations in these samples, it is important to set a number of plane waves which will account for the variation and still result in spatial resolutions that will yield accurate relative values of the intensive free energy between PSO-SCFT agents. A good rule of thumb for obtaining accurate relative free energies is to use simulations with a spatial resolution of $\Delta x \lessapprox 0.35$. Thus, to avoid incurring prohibitive computational expense, we used 48^3 plane waves in SCFT simulations and chose ranges of q^* for PSO-SCFT to explore that corresponded to spatial resolutions which remained less than $\Delta x \approx 0.35$.

3.1.4 Chapter Outline

In the remainder of this chapter, we will first discuss the development of the PSO as applied to simple test functions that can be mapped onto a spherical manifold (Section 3.2). We will see from these studies that the unadorned PSO has the problem of agent trajectories which bunch at the poles of the sphere. We can instead achieve uniform sampling of the sphere by applying rotations to the agent positions so that their updates are always performed from a specified location on the equator.

In Section 3.3, we discuss the PSO-SCFT algorithm in more detail. We then explore the application of PSO-SCFT to diblock (Section 3.4) and miktoarm (Section 3.5) systems. In these sections, we do not include the primary wave vector q^* as a PSO search variable, which means that the unit cell size is not something that is being optimized.

We address this issue separately in Section 3.7, where we do include q^* as a variable for PSO-SCFT to optimize, and discuss the challenges and successes of this version of the method, which is an important step in actually solving the forward problem, since the unit cell size of a given morphology is not known *a priori*. We conclude

with a characterization of a competitive sphere morphology which was predicted by PSO-SCFT in the miktoarm system.

3.2 Addressing Challenges of the Spherical Search Space: Basic PSO applied to Test Functions

3.2.1 The Basic PSO Algorithm

In this section, we will investigate the application of a basic PSO algorithm to test functions which can be mapped to a spherical manifold. In this series of tests, the PSO agents are the azimuthal and polar angles of an individual position, rather than a set of peaks on the surface of the sphere. This is because in this case, we are simply trying to determine the angles which extremize a given test function of interest. This differs from the full PSO-SCFT methodology, which is looking for the *set* of peaks which corresponds to a minimal free energy block copolymer morphology.

The following equations of motion are used to update the positions $\mathbf{x}_i$ and velocities $\mathbf{v}_i$ of i PSO agents[59]:

$$\begin{aligned} \frac{d\mathbf{v}_i}{dt} &= c_p\xi_p(t)(\mathbf{x}_i^{pbest} - \mathbf{x}_i) + c_g\xi_g(t)(\mathbf{x}_i^{gbest} - \mathbf{x}_i) - w\mathbf{v}_i \\ \frac{d\mathbf{x}_i}{dt} &= \mathbf{v}_i \end{aligned} \tag{3.1}$$

Here, the $\mathbf{x}_i$ are the set of azimuthal and polar angles (ϕ_i and θ_i, respectively) being explored by the PSO. $\mathbf{x}_i^{pbest}$ is the personal best position found by an individual agent and $\mathbf{x}_i^{gbest}$ is the global best position found by any agent. The first two terms in the first line of Equation 3.1 can be interpreted as springs which tether agent positions back to the personal and global best positions, respectively. c_p and c_g are scalar values which can be used to tune the relative strengths of the springs, and $\xi_p(t)$ and

$\xi_g(t)$ are random uniform variables between [0,1] that contribute to the stochasticity of the algorithm. The last term in the first line of Equation 3.1 is an inertial term with inertial weight w. The formulas in Equation 3.1 are integrated using a first order Euler-Maruyama method with timestep $\Delta t = 1$ to yield the equations in Equation 3.2.

$$\begin{aligned} \mathbf{v}_i^{t+1} &= w\mathbf{v}_i^t + [c_p\xi_p^t(\mathbf{x}_i^{pbest,t} - \mathbf{x}_i^t) + c_g\xi_g^t(\mathbf{x}_i^{gbest,t} - \mathbf{x}_i^t)]\Delta t \\ \mathbf{x}_i^{t+1} &= \mathbf{x}_i^t + \mathbf{v}_i^t\Delta t, \quad \mathbf{x} = (\phi, \theta) \end{aligned} \tag{3.2}$$

The pseudocode for the basic PSO algorithm is provided in Algorithm 1. For any given test function, the fitness of an agent is simply the value of the function evaluated at the angles found by a given agent. Agents themselves are 2-vectors, comprised of the azimuthal and polar angles: (ϕ, θ). Thus, the fitness of an agent i for test function f is given by $f(\phi_i, \theta_i)$.

In order to achieve a uniform random sampling for the initialization of each angle in spherical coordinates, the polar and azimuthal angles should be selected in the following way:

1. Choose u and v randomly from (0,1).
2. Determine polar angle $\theta = \pi u$.
3. Determine azimuthal angle $\phi = \arccos(2v - 1)$.

---

initialize agent positions x_i^0 and velocities v_i^0 randomly;

for *each timestep* $t > 0$ **do**

  for *each agent* i **do**

    update position x_t according to Eqn. 3.2;

    calculate fitness $f(x_t)$;

    update x_p^t;

  end

  update x_g^t;

end

Algorithm 1: Update process for basic PSO (no SCFT).

We applied the basic PSO to a variety of test functions, including the spherical harmonic $Y_2^0 = \frac{1}{4}\sqrt{\frac{5}{\pi}}(3\cos^2(\theta) - 1)$, which is shown in Figure 3.4. In the initial investigations applying the basic PSO to a spherical search space, we found that the basic PSO failed to converge to the target maximal value of $Y_2^0 = 2$. This can be seen in Figure 3.9, which shows a plot of the value of Y_2^0 vs. PSO iteration for every agent. From the plot, we see that agents actually found angles corresponding to the target value within a couple of iterations, but would not converge to the appropriate values after successive iterations of the PSO. This failure to converge was observed consistently for a range of values of swarm size, w, c_p, and c_g.

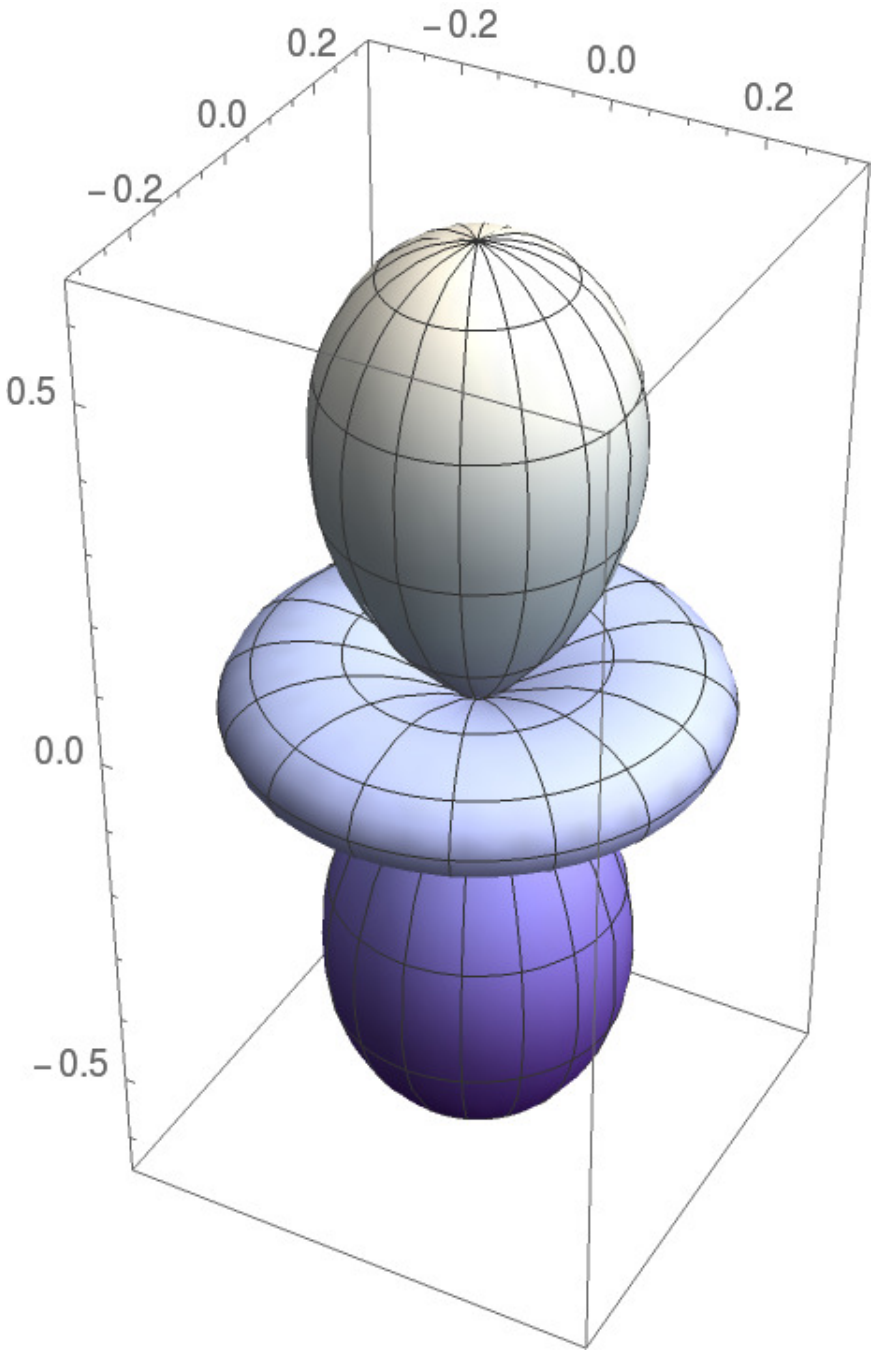

Figure 3.4: Plot of the spherical harmonic Y_2^0, which was used as a test function for the basic PSO.

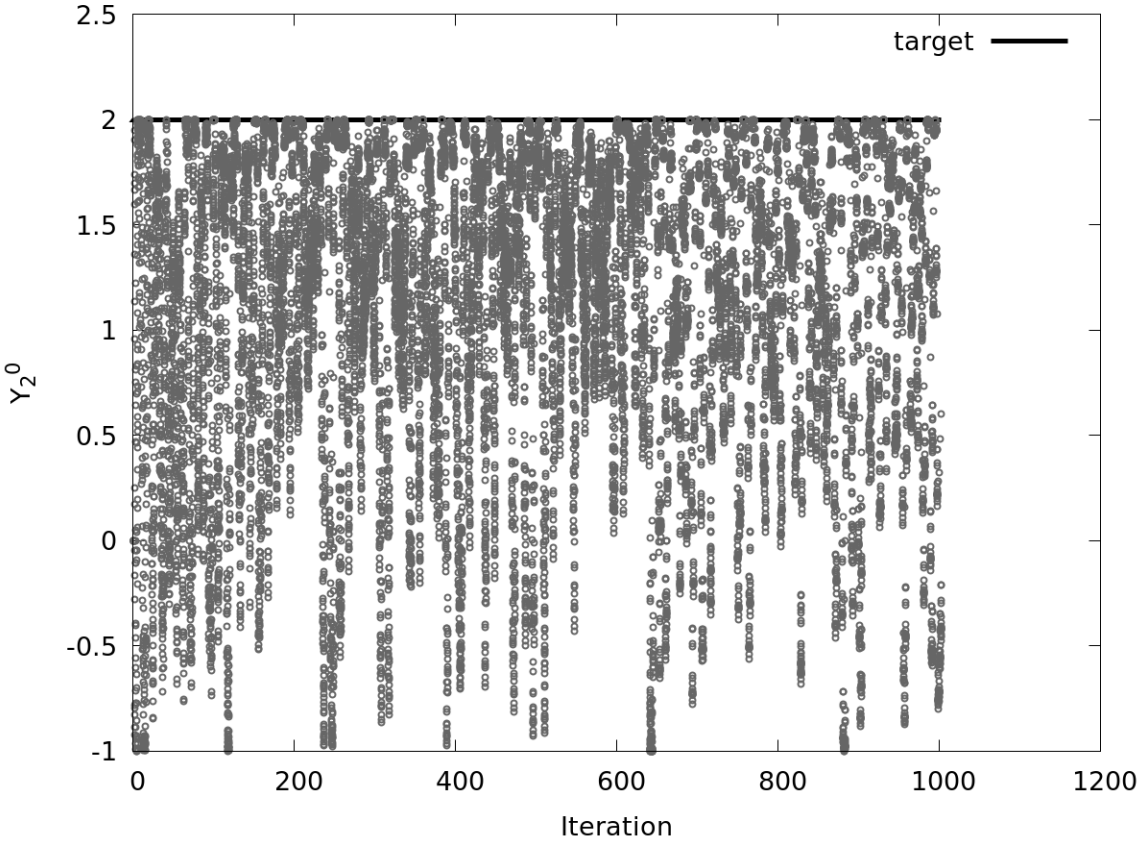

Figure 3.5: Plot of the value of spherical harmonic Y_2^0 vs. PSO iteration for all agents, represented as gray circles.

In an attempt to understand this behavior, we performed a series of tests which isolated each term in the PSO update equations. We applied the PSO to the same test function, and worked through the permutations of allowing either one or two of w, c_p, and c_g to be non-zero for a given sample, while also setting ξ_p and ξ_g to 1 to eliminate stochastic kicks (*e.g.*, $w = 0, c_p = 0$, and $c_g \neq 0$, or $w = 0, c_p \neq 0$, and $c_g \neq 0$). Some results from these experiments are shown in Figure 3.6. We found that the inertial term attached to the coefficient w behaved in an expected way. That is, when it was the only term that was turned on, the agents would begin at some position with some velocity which would eventually be drained away and agent positions would stagnate after several iterations. Having only c_p turned on seemed to also produce expected

behavior: agents would simply remain at their original positions (as seen in panel (b)). Furthermore, having only c_g turned on seemed to behave as expected as well, since all agents eventually collapse to the target fitness (panel (c)). This left us very perplexed as to why the PSO was failing to converge on good positions.

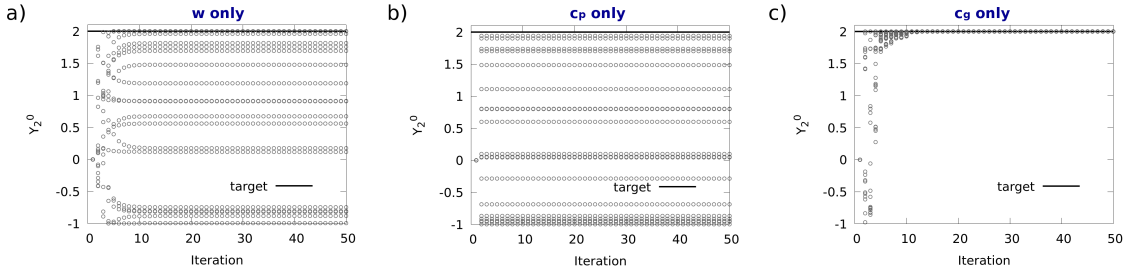

Figure 3.6: Plots of the value of spherical harmonic Y_2^0 vs. PSO iteration for all agents, represented as gray circles, for cases where the only non-zero term was (a) w, (b) c_p, and (c) c_g.

We decided to further simplify the problem and look at the PSO's exploration of a uniform function. The idea was to verify that the PSO would sample the sphere uniformly without a bias. This was achieved by modifying the algorithm to turn off the tethering terms and simply provide random kicks from the current position at each iteration, to mimic brownian dynamics. If the PSO did indeed exhibit brownian dynamics in its trajectory, after enough iterations we would expect to see a uniform distribution in the azimuthal angle, and a distribution that looked like a cosine function in the polar angle. However, this was not what we observed with the basic PSO algorithm written as it was. The distributions for the azimuthal (ϕ) and polar (θ) angles visited by a single agent whose position was updated by adding a random vector to the current position at each iteration of the simulation are shown in Figure 3.7. From the plot in the second panel, we can see that the particle tends to bunch at the poles.

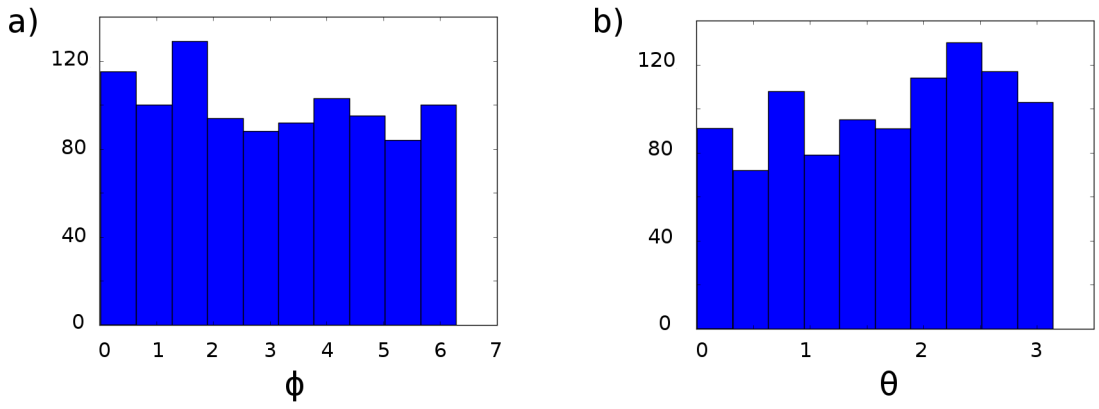

Figure 3.7: Distributions of the (a) azimuthal and (b) polar angles visited by a single agent whose position was updated by adding a random vector to its current position at each iteration of the simulation.

3.2.2 Rotation Matrices

To circumvent the issue of particles bunching at the poles, we can implement the PSO update equations in a local coordinate frame at the equator of the sphere. This can be achieved by applying a rotation matrix which will rotate the peaks from where ever they currently are in the global coordinate frame (the space we are searching) to a point on the equator, where the particle's position and velocity can be updated. The inverse of the original rotation matrix can then be applied to return the particle to the global coordinate system. This idea was inspired by the work presented in Fredrickson's 1988 paper discussing a theory describing results from rotational fluorescence depolarization experiments[66]. Our PSO+RM algorithm would then read as shown in Algorithm 2.

---

initialize agent positions x_i^0 and velocities v_i^0 randomly;

for *each timestep* $t > 0$ **do**

  for *each agent* i **do**

    apply $\mathbf{R}_{ge}$to current position;

    update position x_t using to Eqn. 3.2 in local, equatorial coordinates ;

    apply $\mathbf{R}_{eg} = (\mathbf{R}_{ge})^{-1}$ to updated position to rotate back to global coordinates;

    apply boundary conditions (periodic in ϕ, reflecting in θ);

    calculate fitness $f(x_t)$;

    update x_p^t;

  end

  update x_g^t;

end

Algorithm 2: Update process for the basic PSO+RM strategy (PSO with rotation matrices) to avoid particle bunching at the poles.

The rotation matrix which takes particles from the global frame to the local, equatorial frame takes the following form:

$$\mathbf{R}_{ge} = \begin{bmatrix} \sin\theta\cos\phi & \sin\theta\sin\phi & \cos\theta \\ -\sin\phi & \cos\phi & 0 \\ -\cos\theta\cos\phi & -\cos\theta\sin\phi & \sin\theta \end{bmatrix} \tag{3.3}$$

This matrix was obtained by multiplying the rotation matrices $\mathbf{R}_z$ and $\mathbf{R}_y$: $\mathbf{R}_{ge} = \mathbf{R}_z\mathbf{R}_y$. $\mathbf{R}_z$ yields a rotation about the z-axis to get to the xz-plane, and $\mathbf{R}_y$ yields a rotation about the y-axis to get to the xy-plane. Thus, the rotation matrix $\mathbf{R}_{ge}$ always results in a rotation to the same point on the equator: the prime meridian at $(\pi, \frac{\pi}{2})$. The updates to positions are performed in this local, equatorial frame according to the equations in 3.2, and new positions are rotated back using the inverse of $\mathbf{R}_{ge}$.

Brownian motion and a uniform sampling of the sphere were recovered when coordinates were updated in the local frame of reference, as seen in Figure 3.8. This is in comparison with Figure 3.7, where angles were updated without rotating to a local frame of reference, and the distribution for the polar angle does not take on the expected cosine shape associated with a uniform exploration of the sphere.

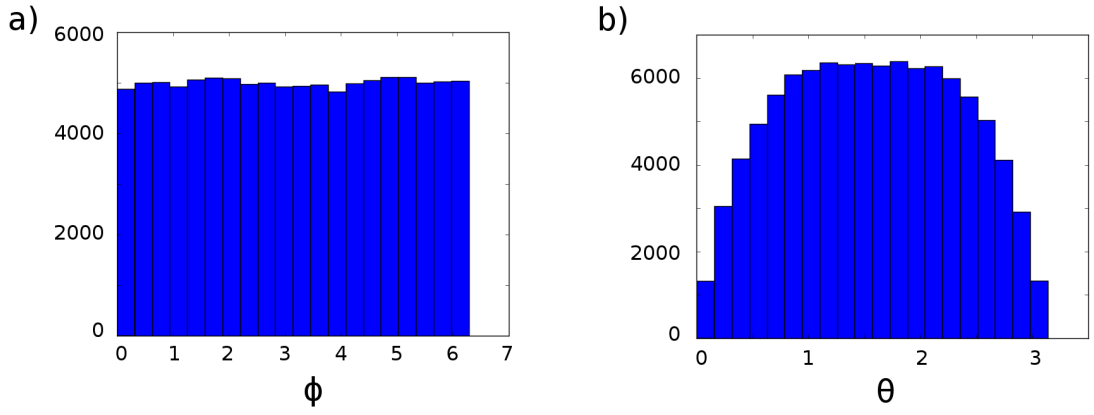

Figure 3.8: Distributions of the (a) azimuthal and (b) polar angles visited by a single agent whose position was updated by adding a random number to its current position in a local, equatorial frame of reference at each iteration of the simulation.

The basic PSO+RM strategy also fixed the problem of convergence to target value of the Y_2^0 spherical harmonic. From Figure 3.9, we see that PSO+RM actually recovers the maximal value of Y_2^0 within 2 iterations, and converges to this value within less than 10 iterations. This is remarkable improvement over the failure of the basic PSO to recover the target value for this test function. Thus, the use of the rotation matrices to update velocities in the equatorial coordinate frame seems necessary for the successful application of the PSO in a spherical space.

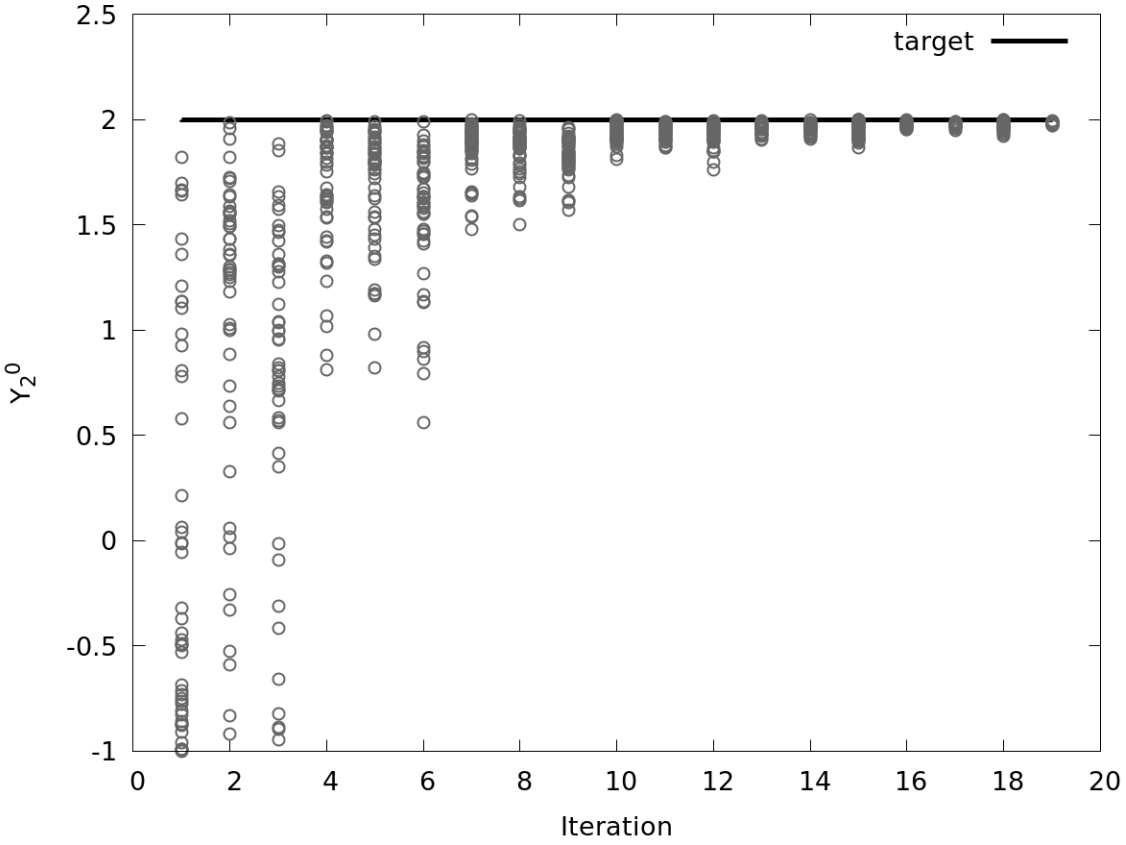

Figure 3.9: Plot of Y_2^0 vs iteration for all agents in a basic PSO+RM applied to the Y_2^0 spherical harmonic as a test function.

As with the basic PSO, we applied the basic PSO+RM strategy to a variety of test functions. One example can be seen in Equation 3.4. The function obtains its maximal value of 3.79 at $\phi = 1.6$, $\theta = 0.785$.

$$f_2(\phi, \theta) = 1 + 4\sin(5\phi)\sin(10\theta)/5 + \sin(\phi) + \sin(2\theta) \tag{3.4}$$

A contour plot of this test function is shown in panel (a) of Figure 3.10. The PSO recovers the angles associated with the maximum value of f_1 after about 50 iterations, as seen in panel (b). For this implementation of the PSO, we used 20 agents (particles), $\Delta t = 1$, $w = 0.9$, $c_p = 0.1$, and $c_g = 0.5$. We found that increasing the ratio of c_g to c_p dramatically decreased the number of iterations it took to obtain and converge to the optimized angles. For example, using $c_g = 0.5$ and $c_p = 0.5$ resulted in the PSO taking on the order of 1000 iterations to optimize f_1 rather than 50. The plots in Figure 3.11 display the trajectories followed by four agents from the 20-agent test. From the plots, we can see that while each agent begins at a different location on the sphere, they eventually converge to the same pair of angles.

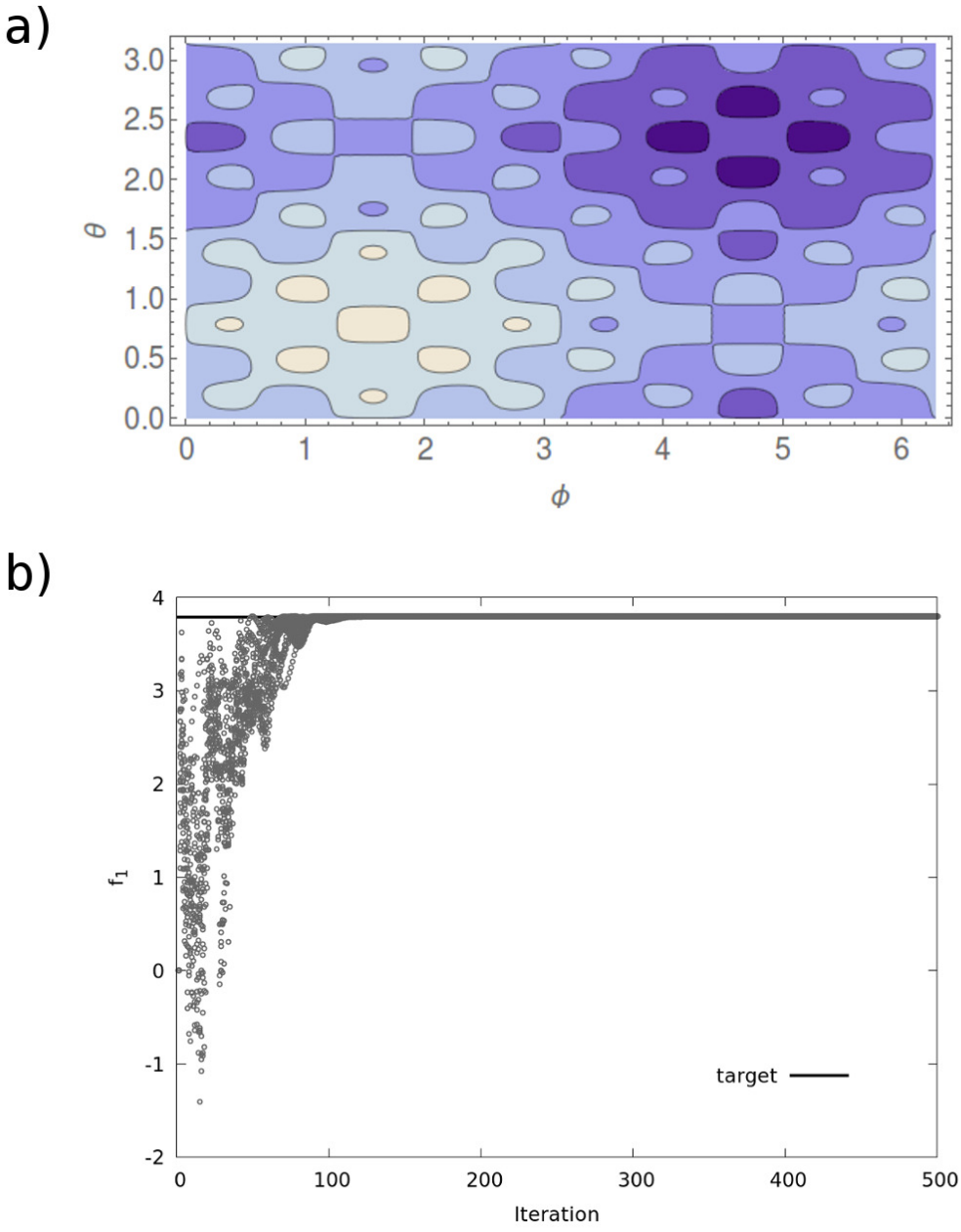

Figure 3.10: (a) Contour plot of the test function, f_2, in Equation 3.4. The function obtains its maximal value of 3.79 at $\phi = 1.6$, $\theta = 0.785$. (b) f_2 vs. basic PSO iteration.

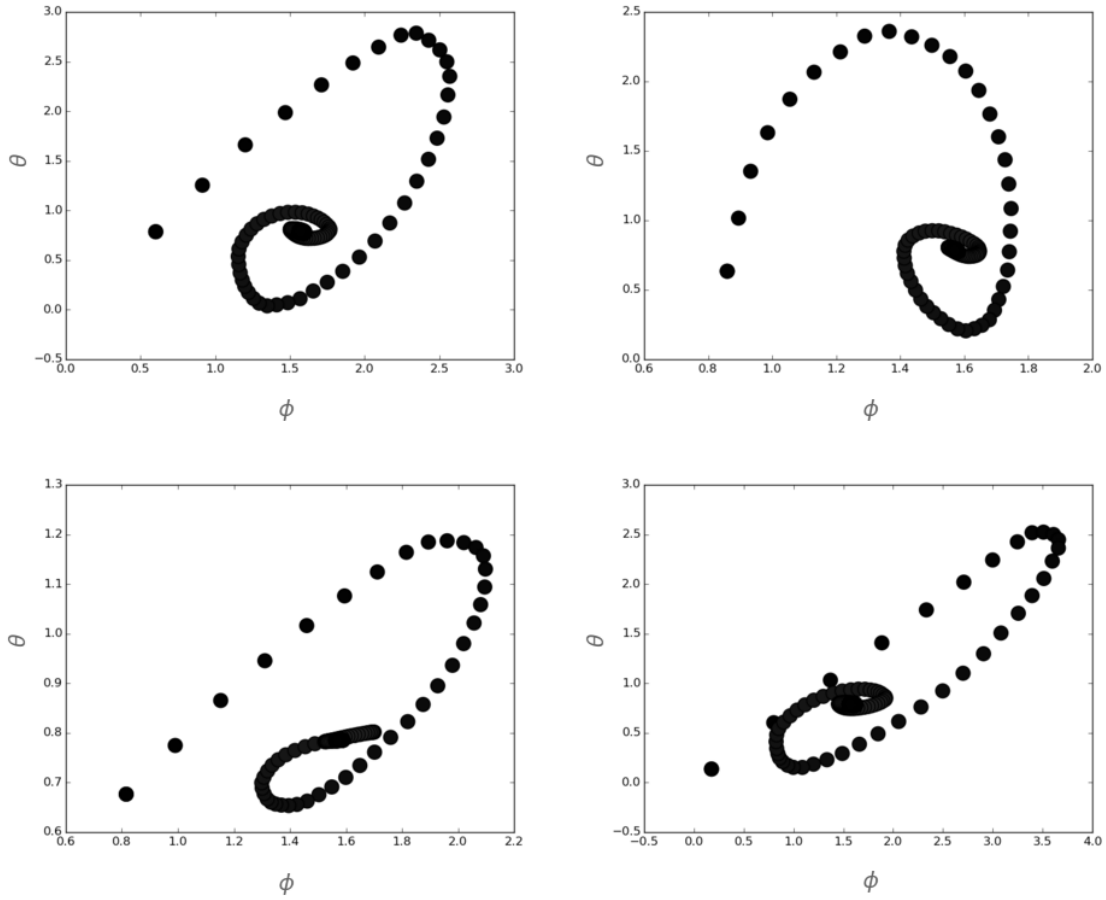

Figure 3.11: Samples of trajectories followed by four of the twenty agents in the PSO applied to optimizing f_2.

3.2.3 Comments

In this section, we tested the application of a basic PSO algorithm to functions which could be represented in a spherical coordinates. We saw that in order to achieve uniform sampling of a uniform function, as well as convergence to optimal coordinates for non-uniform functions, it was necessary to update the coordinates at each iteration of the PSO in a local, equatorial frame of reference. This result was consistent across several test functions.

We also saw that it was important to tune the PSO parameters to values that differed from the commonly seen literature values of $c_p = c_g \approx 2$ [67]. From a back of the envelope calculation, one can see that allowing c_p and $c_g > 1$ would mean that velocities would be larger than one. Since we fix $\Delta t = 1$ in Equation 3.2, this means positions would be changing by order 1. Since the entire space which we are optimizing falls in $\phi \in (0, 2\pi)$ and $\theta \in (0, \pi)$, a change in position of order 1 would quite a substantial change relative to this range; essentially, particles would be whizzing around the sphere, rather than being tethered to previously known good solutions. We found that using c_p and c_g between (0.01,0.5) performed best, depending on the test function.

3.3 PSO-SCFT Algorithm

After performing testing of a basic PSO on simple test functions that could be mapped to a spherical space and developing the strategy for using rotation matrices to update coordinates in a local, equatorial frame, we focused on developing the PSO-SCFT. The following equations of motion are used to update the positions $\mathbf{x}_{i,j}$ and velocities $\mathbf{v}_{i,j}$ of i PSO-SCFT agents with j peaks each:

$$\begin{aligned} \frac{d\mathbf{v}_{i,j}}{dt} &= \chi c_p \xi_p(t)(\mathbf{x}_{i,j}^{pbest} - \mathbf{x}_{i,j}) + \chi c_g \xi_g(t)(\mathbf{x}_{i,j}^{gbest} - \mathbf{x}_{i,j}) - (1-\chi)\mathbf{v}_{i,j} \\ \frac{d\mathbf{x}_{i,j}}{dt} &= \mathbf{v}_{i,j} \end{aligned} \tag{3.5}$$

where $\chi \in (0,1)$ is a constriction factor intended to facilitate convergence tuning, c_p and c_g are positive scalar influence weights, $\xi_p(t)$ and $\xi_g(t)$ are random uniform variables between [0,1], and $\mathbf{x}_{i,j}$ are the sets of azimuthal angles $\phi_{i,j}$ and polar angles $\theta_{i,j}$, where each of these quantities represents the angles of each peak j for the ith agent. The positions $\mathbf{x}_{i,j}^{pbest}$ refers to the best position visited by agent i, peak j in its history, and the positions $\mathbf{x}_{i,j}^{gbest}$ refers to the best position for peak j found globally by any agent. The first two terms in Equation 3.5 can be viewed as springs which tether the agents to their personal best and global best positions found. The ratio scalar weights c_p and c_g can be tuned to favor $\mathbf{x}_{i,j}^{pbest}$ or $\mathbf{x}_{i,j}^{gbest}$, and thus effectively can be interpreted as spring constants that mediate the strength of the tethering of PSO agents to these positions. The last term in the equation for particle acceleration provides inertia to the particles. The formulas in Equation 3.5 are integrated using a

first order Euler-Maruyama method with timestep $\Delta t = 1$:

$$\begin{aligned} \mathbf{v}_{i,j}^{t+1} &= \chi \mathbf{v}_{i,j}^{t} + [\mathbf{F}_{pbest} + \mathbf{F}_{gbest}]\Delta t \\ \mathbf{x}_{i,j}^{t+1} &= \mathbf{x}_{i,j}^{t} + \mathbf{v}_{i,j}^{t}\Delta t, \quad \mathbf{x} = (\phi, \theta) \end{aligned} \tag{3.6}$$

where

$$\begin{aligned} \mathbf{F}_{pbest} &= \chi c_p \xi_p^t (\mathbf{x}_{i,j}^{pbest,t} - \mathbf{x}_{i,j}^{t}) \\ \mathbf{F}_{gbest} &= \chi c_g \xi_g^t (\mathbf{x}_{i,j}^{gbest,t} - \mathbf{x}_{i,j}^{t}) \end{aligned} \tag{3.7}$$

It should be stressed that the PSO-SCFT algorithm differs from the basic PSO or PSO+RM algorithms we have examined thus far, not only in the fact that the PSO is now wrapped around the local optimizer SCFT, but also because agents are no longer comprised of a pair of angle coordinates. Instead, **in PSO-SCFT, agents consist of a *set* of peaks, each with their own azimuthal and polar angles associated with them**.

Furthermore, only half of the total number of peaks are actually manipulated. This reduces the number of degrees of freedom that is necessary for the PSO to optimize by a factor of two, which is important, since PSO and other similar heuristic algorithms perform better with less variables to optimize. It is possible to do this as a first approach to performing PSO-SCFT since for many of the morphologies we are interested in, the symmetries they obey can be captured in one hemisphere. To generate an entire SCFT seed, the peak amplitude at a given coordinate (k_x, k_y, k_z) is copied to the coordinate corresponding to a reflection through the origin, $(-k_x, -k_y, -k_z)$.

Thus, "pairs of peaks" in single-mode seeds are related to one another by this reflection through the origin. Another relevant consideration is that PSO-SCFT updates coordinates on the entire sphere. This means that even though half the total number of peaks are initialized, their trajectories may cover the entire sphere. Their counterparts are always simply updated to be placed at the corresponding point via a reflection through the origin.

SCFT seeds are written in terms of cartesian reciprocal space coordinates, while updates to the peak positions are performed in spherical coordinates. The grid spacing of an SCFT seed is dictated by the spatial resolution of the SCFT simulation for which it is intended. However, the PSO searches through a space of continuous angles, which means that some rounding must occur when converting an updated position from spherical to cartesian coordinates; this is accomplished by simply choosing the nearest grid point in the simulation cell. The PSO-SCFT algorithm is shown in Algorithm 3.

---

initialize peak positions $x_j^0 = (\phi_j, \theta_j)$ and velocities v_j^0 randomly for all peaks j of all agents i;

for *each timestep* $t > 0$ **do**

  for *each agent* i **do**

    for *each peak* j **do**

      apply $\mathbf{R}_{ge}$ to current position (rotate to local frame);

      convert from cartesian coordinates to spherical coordinates;

      update position x_j^t using Eqn. 3.2 in local, equatorial frame ;

      apply $\mathbf{R}_{eg} = (\mathbf{R}_{ge})^{-1}$ to updated position (rotate to global frame);

      apply boundary conditions (periodic in ϕ, reflecting in θ);

    end

    generate SCFT seed consisting of all $2j$ peaks (reflect through origin);

    run SCFT simulation using seed & get fitness (intensive free energy), f;

    update x_p^t of all j peaks;

  end

  update x_g^t;

end

Algorithm 3: Update process for the PSO-SCFT algorithm which optimizes sets of peak positions.

3.4 Application to AB Diblock

3.4.1 Experimental Set-up

In this section, we will discuss the application of the PSO-SCFT algorithm to the same conformationally symmetric diblock we used to test the GA-SCFT method using the compositions listed in Table 2.1. Specifically, we apply PSO-SCFT at the block fractions corresponding to χN=15 for HEX, BCC, and GYR as the target morphologies. It is important to recognize that as with GA-SCFT, there are two relevant sets of parameters that must be tuned: those associated with individual SCFT simulations, and those for the PSO. Parameters for SCFT simulations must be finely tuned, since they comprise the majority of the computational expense for a given implementation of the PSO-SCFT method. Thus, minimizing the runtime of any individual SCFT simulation is imperative for an efficient search with the PSO. In our initial work, we focused primarily on optimizing SCFT runtime (values shown in Table 2.4. Using such values, individual SCFT simulations could be completed within 5-15 minutes, and a whole 50-iteration implementation of PSO-SCFT could be performed in a day or two, depending on the queue on the high-performance clusters we used. We chose $\chi = 0.73, c_p = c_g = 2.05$ for the PSO, which were "standard" values used by Bratton and Kennedy[67, 56]. For the examples shown in this section, agents had 6 peaks, each with an azimuthal and polar angle to optimize. This means there was a total of 12 peaks in a single-mode seed, since seeds are constructed by generating pairs of peaks that are related to one another via a reflection through the origin.

3.4.2 Results

Application to Composition where HEX is expected

Figure 3.12 displays a plot of intensive free energy vs. iteration for PSO-SCFT applied to a composition where HEX was the target morphology, which resulted in the immediate identification of HEX and LAM as competitive structures within the first iteration of the algorithm. In this particular case, 5 agents were used, but in some instances as few as 2 agents could immediately recover HEX. It is unclear from this particular study whether PSO-SCFT outperforms GA-SCFT, since GA-SCFT is also able to immediately recover HEX in 3D simulation cells when L_{init} is set to $4.4\,R_g$. However, we will see the advantage of PSO-SCFT later in Section 3.7 when we implement simulation cell size as a parameter that must be optimized by the algorithm.

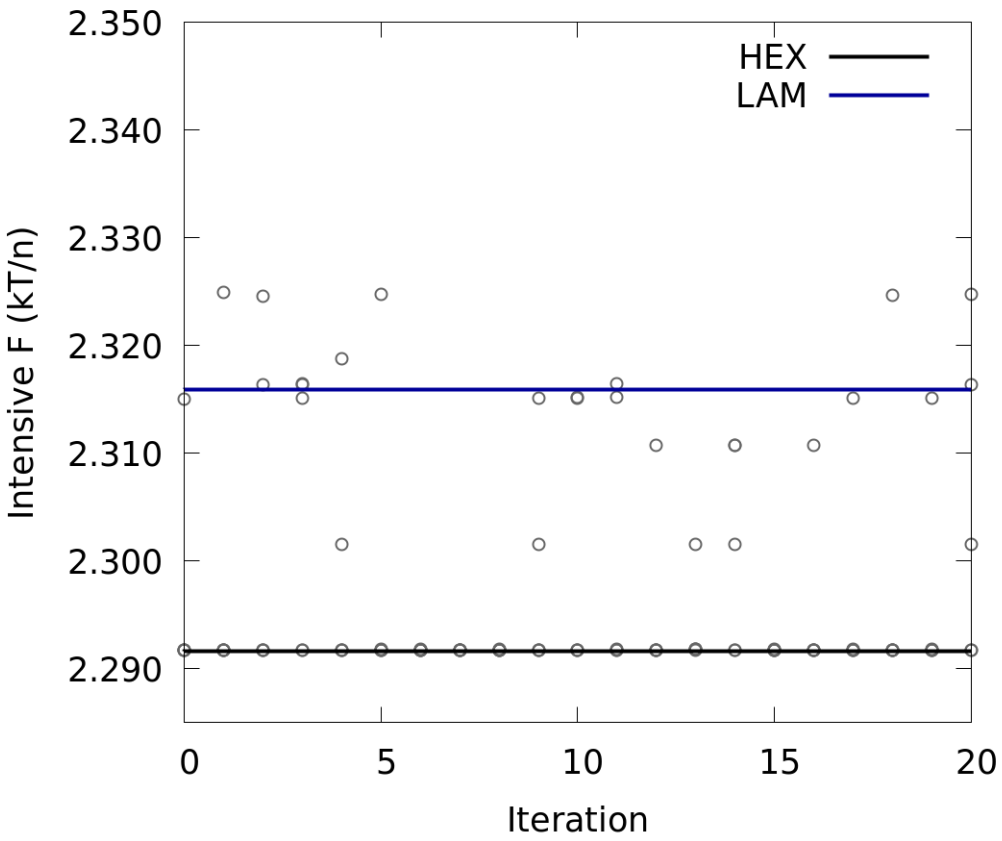

Figure 3.12: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT using 5 agents for a diblock at $f_A = 0.35$, $\chi N = 15$, where HEX is expected to be stable. The algorithm optimized 6 peaks, each with their own azimuthal and polar angles.

Application to Composition where BCC is expected

Similarly to the case for HEX, BCC is recovered immediately by PSO-SCFT. Figure 3.13 shows a plot of the intensive free energy vs. iteration of all agents for a case where PSO-SCFT was applied at χN=15, f_A=0.3, where BCC is the expected morphology. In this example 20 agents were used, but BCC could also be identified immediately with as few as 5 agents. All of the known competitive structures for this composition (LAM, HEX, and FCC) are also identified by PSO-SCFT, whereas GA-SCFT was not always able to do this.

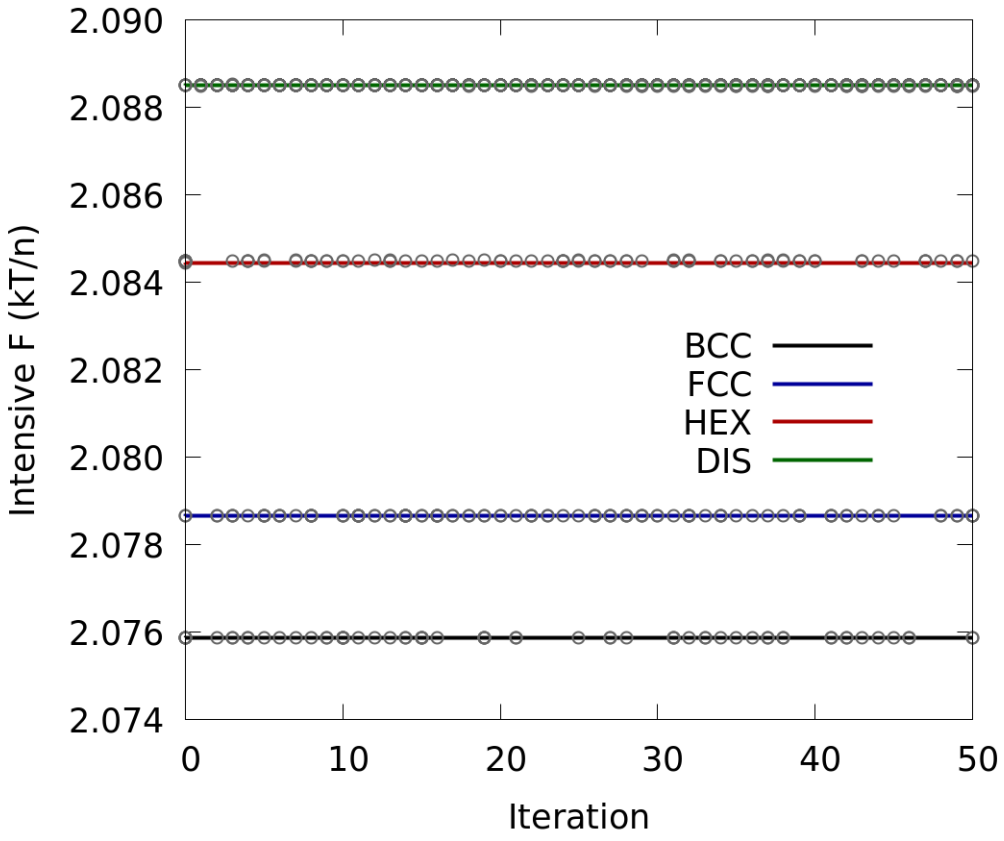

Figure 3.13: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT using 20 agents for a diblock at $f_A = 0.30$, $\chi N = 15$, where BCC is expected to be stable. The algorithm optimized 6 peaks, each with their own azimuthal and polar angles.

Looking at Figure 3.13, we also see that there is a failure to converge to the known best structure. We will see that this can partly be attributed to the PSO parameters χ, c_p, and c_g not being properly tuned. However, it is likely that a different network topology for communication between agents of the swarm may further facilitate convergence. This was something we realized much later in the development of PSO-SCFT and did not have time to explore.

Application to Composition where GYR is expected

In Figure 3.14, we have a plot of the intensive free energy vs. iteration for all agents in a sample where PSO-SCFT was applied at χN=15, f_A=0.39, where GYR is the expected morphology. For this example, 20 agents were used. Additional tests were performed using 10, 30, and 50 agents with similar results. In all cases, GYR was never identified, although Fddd, HEX, and HPL were recovered. It is interesting to note that the orthorhombic Fddd morphology was found by the PSO, given that the peaks in the "single mode" were confined to a sphere rather than an ellipsoid. It further remarkable that our SCFT optimizer seems to land more easily in the Fddd structure than GYR, as has now been observed in benchmark testing for both the GA-SCFT and PSO-SCFT methodologies. It is unexpected and unclear as to why this occurs, since at this composition, GYR is stable to Fddd by approximately $2\,mkT$/chain.

Using agents with 6 peaks in a hemisphere (12 total peaks) or 12 peaks in a hemisphere (24 total peaks) did not make a difference in the results. We originally thought using more peaks might be necessary, since the single-mode approximation for GYR is constructed using 24 total peaks on the sphere. However, what we did not include in these studies was the ability for the PSO to search through amplitude as a variable to optimize. In the single-mode approximation for GYR, the 24 peaks on the sphere have specific positions relative to one another, as well as signs, which contribute to either cosines or sines in the plane wave expansion for the basis function ϕ_1 in the single-mode approximation.

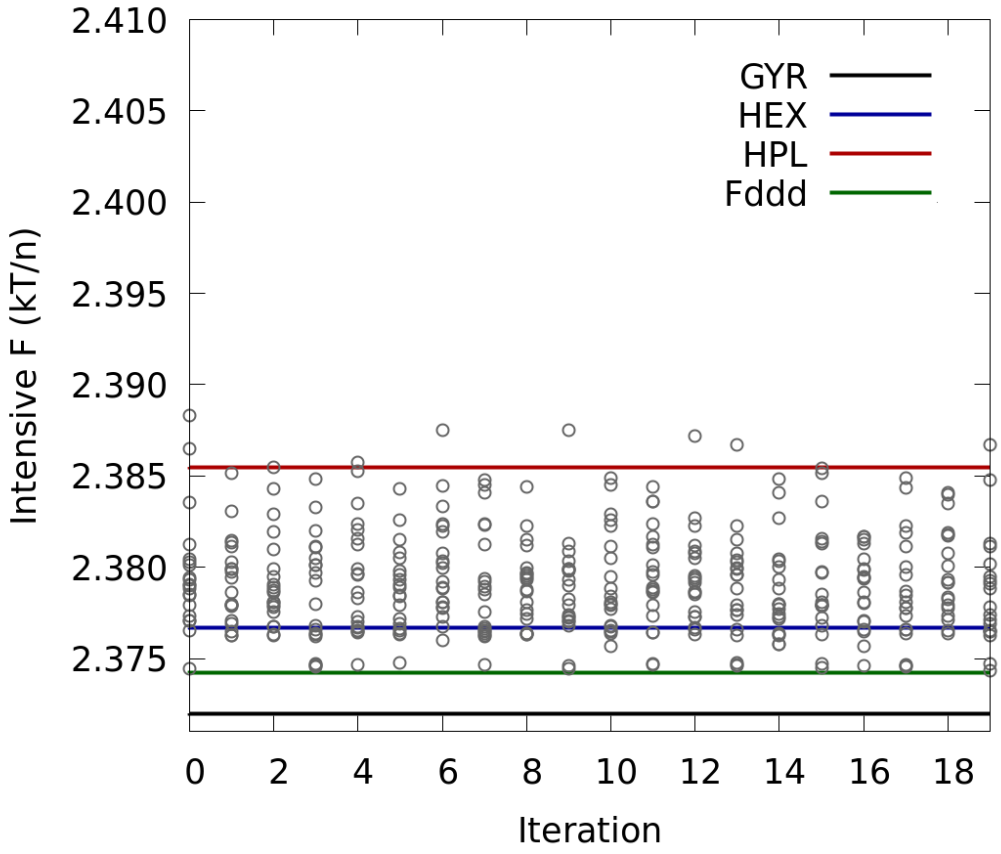

Figure 3.14: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT using 20 agents for a diblock at $f_A = 0.39$, $\chi N = 15$, where GYR is expected to be stable. The algorithm optimized 6 peaks, each with their own azimuthal and polar angles.

3.4.3 The Role of Peak Amplitude

In this subsection, we explore the role of the magnitude and signs of peak amplitudes in the single-mode approximation. In our investigation of these variables, we set up a series of tests in which independent SCFT simulations were run from single-mode seeds for a variety of morphologies: HEX, BCC, and GYR. For each morphology, we used single-mode seeds which consisted of peak amplitudes in a range of (0.1,10). For the case of GYR, where half of the signs are positive (contributing to cosines) and half of the signs are negative (contributing to sines), we tested the following three

cases: 1) make all signs positive, 2) make all signs negative, 3) retain the appropriate relative signs.

Independent SCFT Study for Peak Amplitude

For the studies in this subsection, SCFT simulations were run at χN = 15, using the block fractions listed in Table 2.1 for each morphology. The following magnitudes for all amplitudes were used: 0.1, 0.2, 0.5, 1.0, 2.0, 5.0, 10.0. All magnitudes resulted in the recovery of the expected morphology for a given simulation.

Figure 3.15 provides density profiles for A-rich regions from two independent SCFT simulations initialized with single-mode seeds at a composition for which GYR was the expected morphology. Panel (a) shows the result of using a single-mode seed in which all peaks were placed at appropriate positions in the spherical shell, but assigning their signs to be all positive or all negative. The morphology which arose was the "Plumber's Nightmare," rather than the expected GYR morphology. In contrast, panel (b) shows that using a single-mode see which has the proper relative placement and sign allows one to recover GYR.

Another important result is the fact that half of the peaks can be removed in the single-mode seed and SCFT can still recover GYR. Removing all peaks of one sign in a seed, for instance, would still result in an SCFT simulation which found GYR. However, completely removing pairs of peaks (those related by a reflection through the origin) resulted in non-GYR structures. Superficially this may seem inconsistent with the result from Figure 3.15, panel (a), since all peaks had the same sign in that study. However, what the combination of these two findings suggests is that it is the proper relative sign having at least one peak from a pair of peaks that matters. This is good news for us, since it means that 1) our in-house SCFT solver is good enough to "fill in" some blanks, and 2) the idea of PSO-SCFT searching through only half

of the peaks should not be an obstruction to actually identifying "good" structures. However, the caveat is that we must include enough peaks for the PSO to optimize. This is not known *a priori*, but in principle could be added another parameter for the PSO to optimize, limited to some range of values which seems to correspond to physical systems.

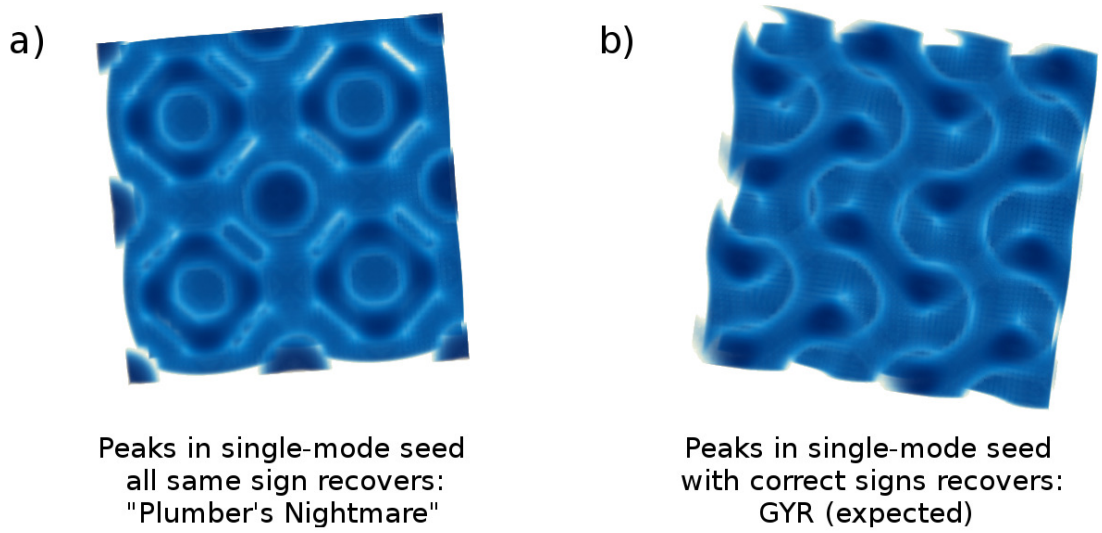

Figure 3.15: Density profiles from independent SCFT simulations run for a diblock at χN=15, f$_A$=0.39, where GYR is the expected morphology, using a cell size of 17.6 R_g in each dimension, which is commensurate with twice the unit cell size of GYR at this composition. Simulations were initialized with single-mode seeds where (a) all peaks had the same sign, resulting in the "Plumber's Nightmare" morphology, (b) peaks had relative signs that were consistent with those comprising a GYR single-mode approximation, resulting in the GYR morphology.

From these studies, we observe that while the magnitude of the peak amplitude does not seem to matter for recovering the appropriate morphology when performing an SCFT simulation from a single-mode seed, using the proper relative signs does. This means that if we want the PSO-SCFT algorithm to be able to agnostically identify structures, we must include the sign of the amplitude as a PSO search variable.

Set-up for Peak Amplitude as PSO Variable

From the previous subsection, we saw that it may be necessary to employ the sign of the amplitude as a variable which is optimized by PSO-SCFT. For the results shown in the remainder of this chapter, we implement this in a similar fashion to the peaks. Speaking in terms of Python's object oriented programming, the language in which PSO-SCFT is implemented, each peak has azimuthal position, polar position, and amplitude value attributes. A peak also has azimuthal velocity, polar velocity, and amplitude velocity. These values are all updated according to the integrated equations of motion in Equation 3.6. Amplitudes, however, are different from the peak position variables, in that their values are initialized from a random uniform distribution between (-1,1), and their values may only remain within that range. This range is enforced by using a tan function which maps onto a space in which amplitudes are updated, and then mapped back to the search space: $a_{mapped} = \tan(\frac{pi}{2}a)$, where a_{mapped} is the value of the amplitude in the space which is mapped to the real number line where updates are performed. To return to the confined range, one takes the arctan of the updated a_{mapped}.

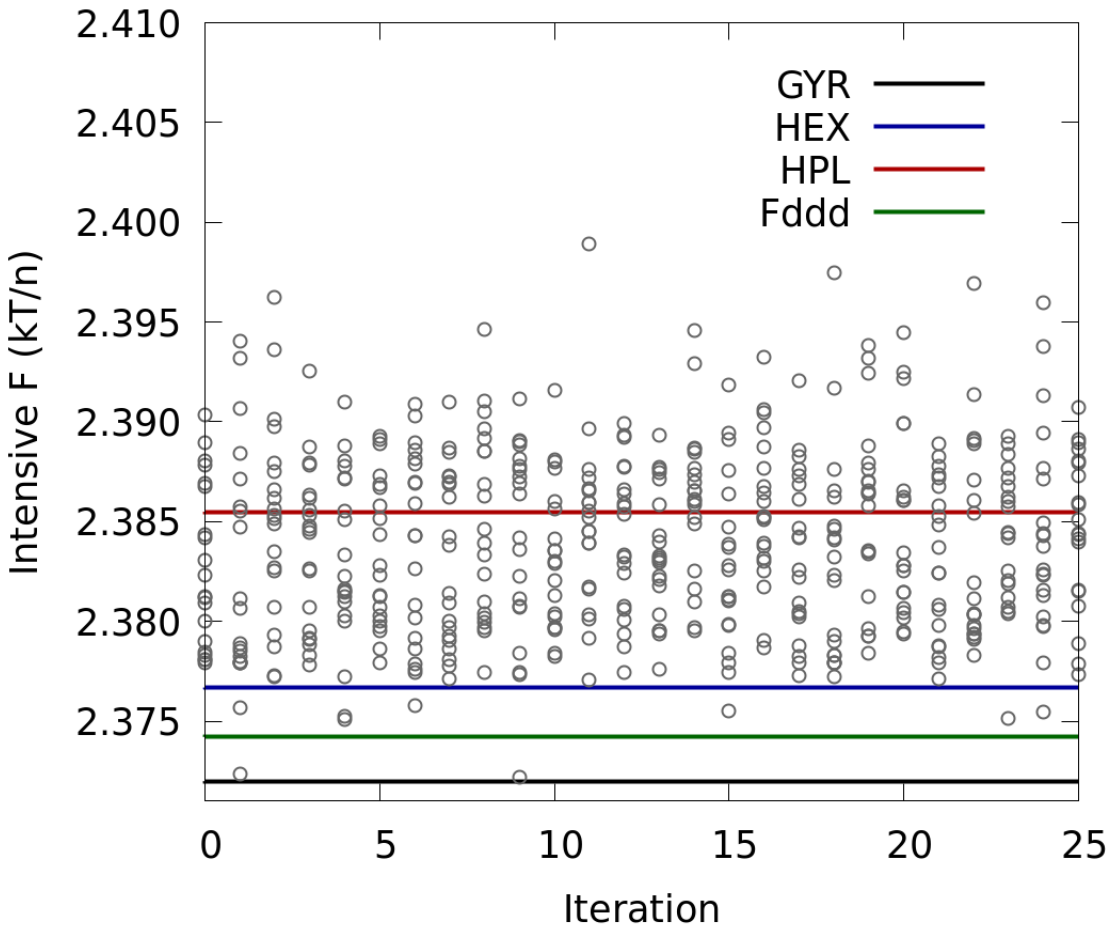

Figure 3.16: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT with amplitude as a search variable, using 20 agents for a diblock at $f_A = 0.39$, $\chi N = 15$, where GYR is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles.

Figure 3.16 displays the intensive free energy vs. iteration plot for PSO-SCFT with amplitude as an optimization variable applied to a diblock at f_A=0.39, $\chi N = 15$, where GYR is the target morphology. Comparing to Figure 3.14, we see that when amplitude is included as a variable to search, PSO-SCFT is able to identify GYR. Again, we see the problem of the failure of PSO-SCFT to converge to fit structures, as was the case with the benchmark testing for BCC and HEX. This can be mitigated to some extent by tuning χ, c_p, and c_g.

3.4.4 Comments

For these tests, we used variable SCFT simulations that were initialized in each iteration of the PSO with a prescribed cell size that fell within 5-10% of the commensurate cell size for a given target morphology. For example, SCFT simulations for the benchmarking where HEX was the target morphology were initialized with $L_{init} = 4.4\,R_g$. The commensurate cell size for HEX at f_A=0.35 and χN=15 is $4.65\,R_g$, which is a 5% difference from the initial cell sizes used for these simulations. Changing the initial cell size to as little as $3.5\,R_g$ (24.7% off-commensurability) did not drastically affect the ability of PSO-SCFT to recover HEX. Similar results were seen for BCC, but not for benchmarking for GYR. Implementing PSO-SCFT more than a few percent away from the commensurate cell size for GYR would result in a failure of the algorithm to identify GYR, even with signed peak amplitude as a search variable. We will see that this is the case in the other system we present results for in the next section: the AB_4 miktoarm. Ultimately, we will need to include a way for PSO-SCFT to optimize cell sizes; simply using variable cell in individual SCFT simulations is not enough. Before we take on this task, however, we explore the application of PSO-SCFT to the miktoarm system, where several sphere morphologies are competitive with one another.

3.5 Application to AB_4 Miktoarm

3.5.1 Experimental Set-up

In this section, we discuss the application of the version of the PSO-SCFT algorithm which optimizes a set of peak positions and their amplitudes to an AB_4 miktoarm system, benchmarking against the phase diagram provided by Xie *et al.*[27]. Particularly, we look at a composition of $f_A 0.32$ at χN=40, where the A_{15} sphere phase is expected to be stable, but other sphere morphologies are also competitive. It was important to first verify that A_{15} was indeed the expected morphology at the composition at which we applied PSO-SCFT. Verification for this can be seen in Table 3.2.

Seeded Morphology	Intensive Free Energy
BCC	-3.37
FCC	-3.44
σ	-3.47965
A_{15}	-3.47973

Table 3.2: Intensive free energies obtained from SCFT simulations seeded with the indicated morphologies for an AB_4 miktoarm star polymer at χN=40 and f_A=0.32.

Individual SCFT simulations were tuned to have a total runtime of 5-30 minutes. The large variation in runtime resulted from differences in individual single-mode seeds. An implementation of 50-iterations of PSO-SCFT could be accomplished in 1-3 days, depending on SCFT simulation time and queue wait time on the clusters.

For the results presented in this section, PSO parameters were modified to $\chi = 0.73$, $c_p = 0.2$, and $c_g = 0.2$, which is closer to the values which appeared to give optimal results in the studies of the basic PSO algorithm applied to various test functions in Section 3.2. The logic is the same here: values of c_p and c_g which are greater than 1 would result in velocities which would send peaks whizzing around the

sphere. This would render the tethering terms to personal and global best positions in Equations 3.6 effectively useless.

3.5.2 Results

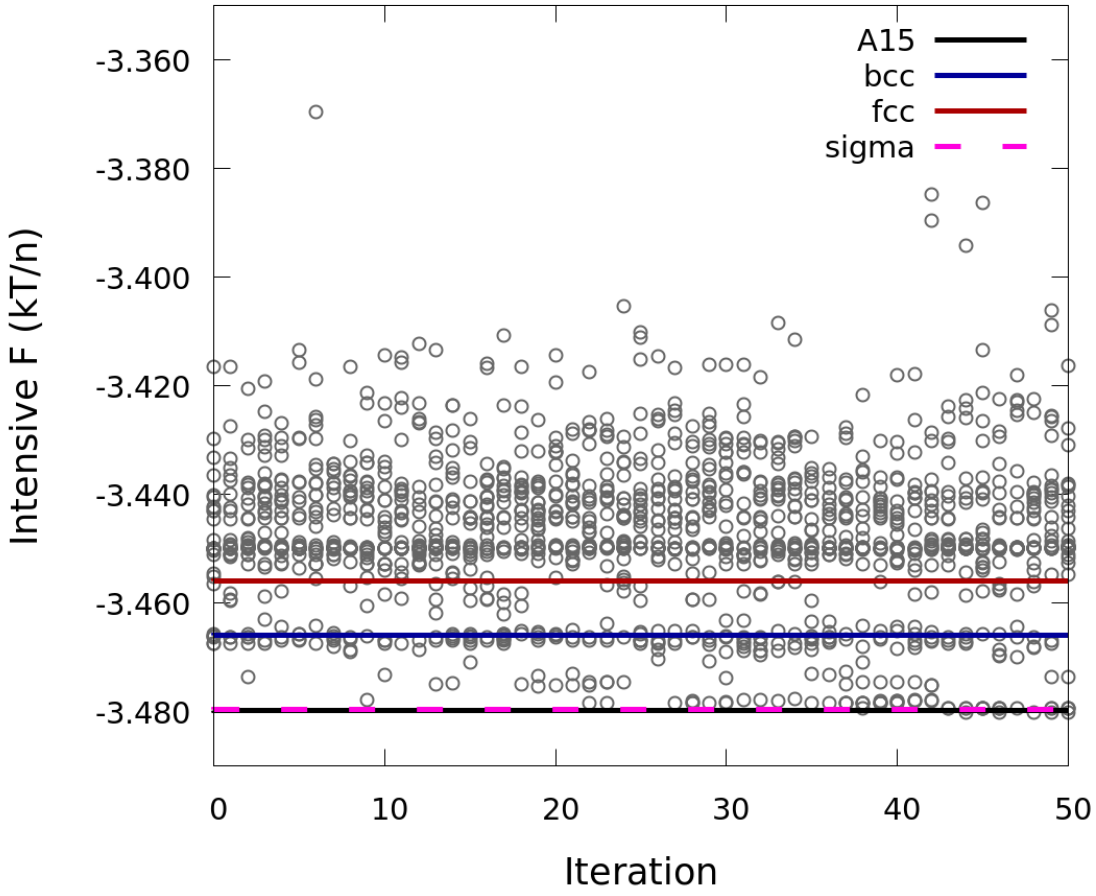

Figure 3.17: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT with amplitude as a search variable, using 30 agents for an AB_4 miktoarm star polymer at $f_A = 0.40$, $\chi N = 40$, where A_{15} is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles. Individual variable-cell SCFT simulations were initialized in cubic cells with $L_{init} = 7.39\, R_g$.

A plot of intensive free energy vs. iteration for the application of PSO-SCFT with amplitude as a search variable to an AB_4 miktoarm star polymer with A_{15} as

the target morphology is shown in Figure 3.17. In this test, 30 agents were used, and individual variable-cell SCFT simulations were initialized in cubic simulation cells with side lengths $L_{init} = 7.39\,R_g$. At this composition, $7.39\,R_g$ on a side is commensurate with the unit cell of the expected A_{15} morphology. From the plot, we see that PSO-SCFT is actually able to identify A_{15} within about 10 iterations, and over the course of successive iterations more agents return to the expected morphology.

A wide variety of structures was observed throughout all the iterations in this sample. Some examples of such morphologies can be seen in Figure 3.18. Panel (f) shows the A_{15} morphology. Other morphologies not shown include various defective sphere packings, spheres mixed with cylinders, and defective cylinders, all of which were seen throughout various implementations of the PSO-SCFT algorithm for this AB_4 miktoarm system. Additionally, a given intensive free energy might actually correspond to a few different structures, particularly in the window between (-3.46,-3.42) kT/chain. This was consistently observed in various implementations of the PSO-SCFT algorithm at this composition and architecture.

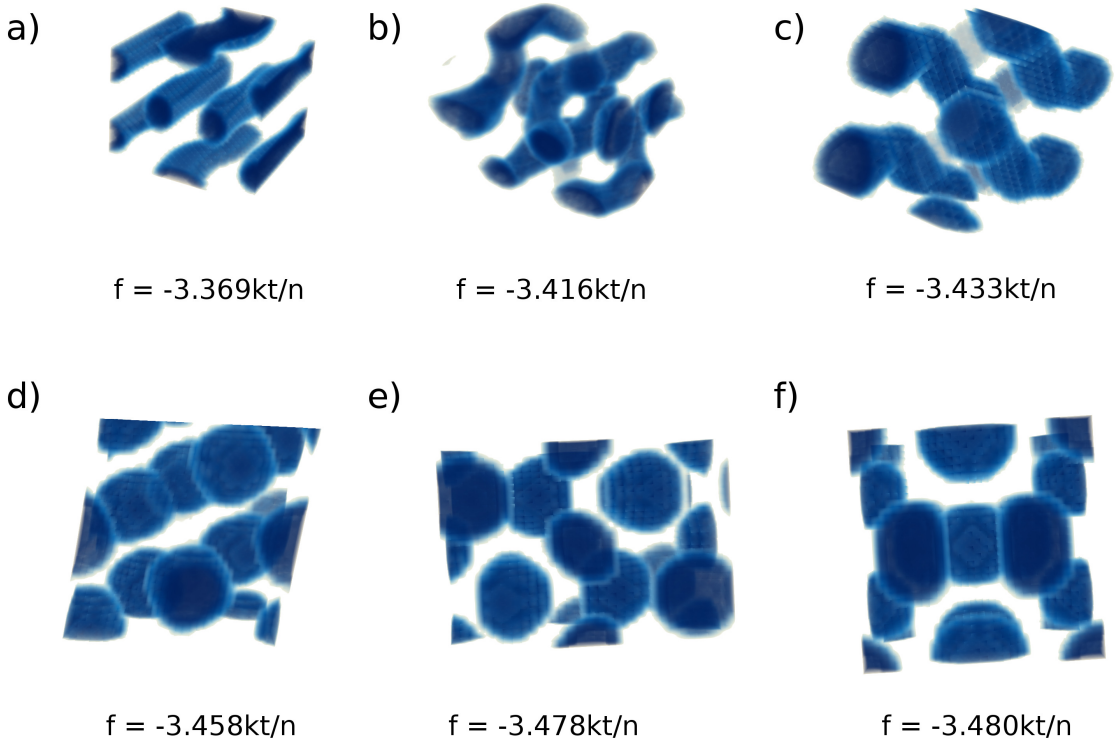

Figure 3.18: Examples of morphologies and their corresponding intensive free energies that were identified by the implementation of PSO-SCFT described in Figure 3.17.

From the plot in Figure 3.17, the rate of convergence to previously known good structures appears to be very slow. However, it appears that the PSO-SCFT is "working." This is partly evidenced by the distribution of intensive free energies observed in the first iteration as compared with the last iteration of this particular implementation, which can be seen in panels (a) and (b) respectively in Figure 3.19. The black line in each histogram indicates the intensive free energy of the A_{15} morphology at our benchmark composition. In panel (a), we see that the initial distribution of the 30 agents has no structures corresponding to the A_{15} free energy, whereas by iteration 50, 4 of the 30 agents yield A_{15} morphologies.

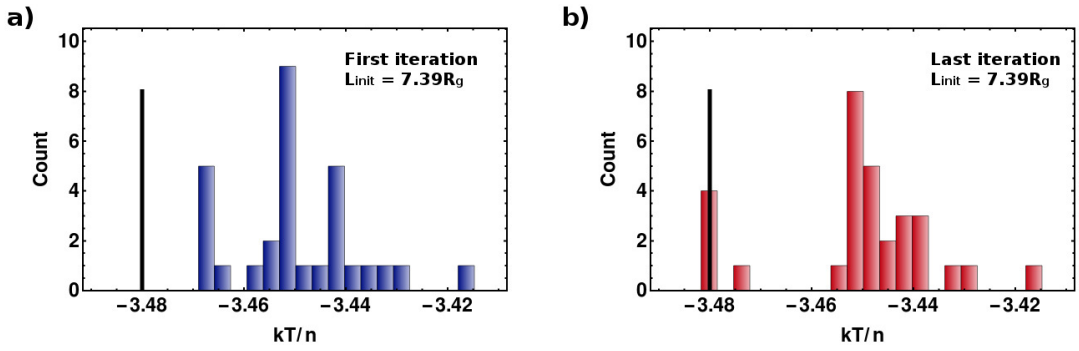

Figure 3.19: Histograms of the intensive free energies of all agents in the (a) 0^{th} and (b) 50^{th} iterations of PSO-SCFT applied to an AB_4 miktoarm. These distributions correspond to the test described in Figure 3.17.

The results shown in the previous three plots were all associated with PSO-SCFT applied to the miktoarm using initial cell sizes which were commensurate with the expected morphology. However, since the entire point of these heuristic algorithms is to agnostically identify structures, this is not particularly impressive. We would not expect to know *a priori* what the unit cell sizes of competitive morphologies are, and furthermore, they may vary quite substantially. This is certainly the case for this system: the unit cell of BCC has a lattice vector of $4.6\,R_g$, whereas A_{15}'s lattice vector is $7.39\,R_g$. Before we discuss the development of a version of PSO-SCFT which will account for variation in unit cell sizes among competing structures, it is an important first step to check that the current version of PSO-SCFT is still capable of identifying the expected morphology when initial cell sizes are some percentage away from commensurate.

Figure 3.20 displays a plot of intensive free energy vs. iteration for PSO-SCFT applied to the same system as Figure 3.17 (f_A=0.40, χN=40, target = A_{15}), but with individual variable-cell SCFT simulations initialized using $L_{init} = 6.64\,R_g$. This is about 10% less than the commensurate cell size of $7.39\,R_g$. From the plot, we see that A_{15} is eventually identified, but it takes more PSO-SCFT iterations to find it

than the case where commensurate cell sizes were used (*i.e.*, about 30 iterations instead of 10). This effect is reproducible, with a variation of a few iterations for both cases.

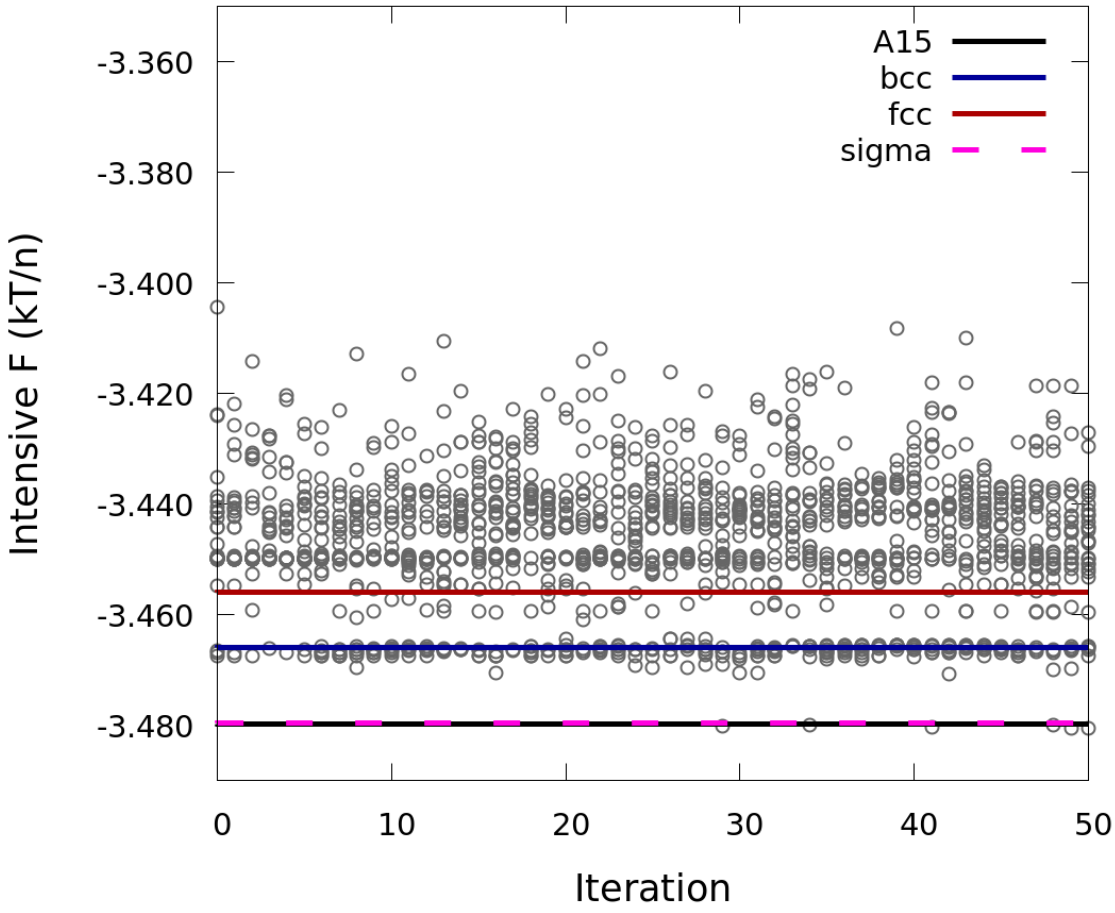

Figure 3.20: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT with amplitude as a search variable, using 30 agents for an AB_4 miktoarm star polymer at $f_A = 0.40$, $\chi N = 40$, where A_{15} is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles. Individual variable-cell SCFT simulations were initialized in cubic cells with $L_{init} = 6.64\, R_g$.

Figure 3.21 shows histograms of the intensive free energy of all agents in the first

iteration (panel (a)) and last iteration (panel (b)) of the the implementation of PSO-SCFT with $L_{init} = 6.64\,R_g$. The results are similar to those observed from the case where $L_{init} = 7.39\,R_g$, in that initially there are no free energies corresponding to A_{15}, but by the 50^{th} iteration there are. However, convergence to "good" structures is slow in this case as well.

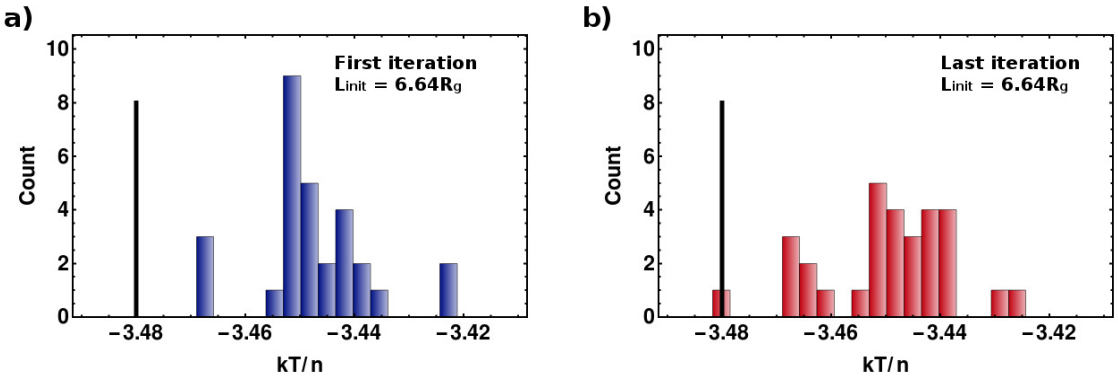

Figure 3.21: Histograms of the intensive free energies of all agents in the (a) 0^{th} and (b) 50^{th} iterations of PSO-SCFT applied to an AB_4 miktoarm. These distributions correspond to the test described in Figure 3.20.

3.5.3 Comments

In this section we applied PSO-SCFT to an AB_4 miktoarm star polymer. We saw that initializing the individual variable-cell SCFT simulations to be commensurate with the expected A_{15} morphology's unit cell size resulted in a faster identification of A_{15} than using a non-commensurate initial cell size. Initializing cell sizes within 10% *larger* than that of the commensurate cell size resulted in the failure of the PSO-SCFT algorithm to identify A_{15}. We did not try using L_{init} that were even larger than 110% of the commensurate cell size, but it seems unlikely that PSO-SCFT would be successful in that case, either. This seems to stem from a peculiarity of the SCFT solver we use in the group. We explore this phenomenon more deeply in the following

section.

Of further interest is the fact that the Frank-Kasper σ phase was not observed. σ's intensive free energy is only a little over $0.1\,mkT/chain$ higher than that of A_{15} at this composition. Furthermore, PSO-SCFT applied at f_A=0.24, χN=40, in the center of a large window of stability for σ, fails to recover σ across a range of 12-48 total peaks per agent, initial cell sizes, and values of the PSO parameters (χ, c_p, and c_g). There are a few factors contributing to the difficulty of recovering the σ phase. First of all, the unit cell size for σ is quite large, and its intensive free energy is quite close to one of its competitors, A_{15}. To even resolve the free energies properly, it is necessary to run SCFT simulations with at least 128 plane waves in each dimension, which becomes prohibitively expensive for PSO-SCFT. We originally thought, as with the other test cases we have investigated so far, that we could use a lower spatial resolution as a preliminary screen to identify competitive structures then perform a closer inspection of individual morphologies which looked promising. However, this approach becomes problematic if a competitor is simply never observed, which was the case for the studies where σ was competitive of even expected.

Furthermore, the σ morphology does not belong to a cubic space group, but rather to a tetragonal one, which means that the length of one of its lattice vectors does not equal the other two. Our current implementation of PSO-SCFT only technically accounts for cubic cells. It would be possible to generalize the algorithm to include unequal lattice vectors, but this would inherently add to the dimensionality of the degrees of freedom being optimized by PSO-SCFT. We do report results for including the cell size as a search variable via a parameter q^*, which will be described in Section 3.7.

It is curious that non-cubic morphologies have been observed, and even preferred, over cubic ones, in both the GA-SCFT and PSO-SCFT methodologies (*i.e.*, the or-

thorhombic Fddd phase). It may be the case that it is enough for the SCFT solver to have one of the lattice vectors correct, and the rest can be "filled in" as long as the seed carries enough information (*e.g.*, the correct placement of a large enough subset of the Fourier peaks in the single-mode seed). However, σ contains 30 spheres in its unit cell, and even more Bragg reflections, which means that it may require a much larger number of peaks, a non-spherical single-mode seed, *and* a decent guess for the cell size, all of which are hefty requirements that are not within the capabilities of the PSO-SCFT algorithm as presented thus far.

3.6 The Important of Initial Cell Size

The initial cell size chosen for any given SCFT simulation, L_{init}, plays a considerable role in retrieving expected morphologies. We have already seen hints of this in our studies of the density of defective states in Chapter 2 (see Figures 2.30 and 2.31). However, this notion has been a recurring theme – yea, nightmare! – throughout the course of this project of solving the forward problem. Another example of the importance of initial cell sizes can be seen in the case study presented in this section. Here, we will see that independent variable-cell SCFT simulations which have different initial cell sizes but start with the same single-mode seed generate different morphologies.

To test the robustness of our SCFT solver to different initial cell sizes, we performed two sets of variable-cell SCFT simulations which swept across initial cell size for a diblock at χN=15, f_A=0.39, where GYR is expected to be stable. In both sets of simulations, the total number of SCFT iterations was capped at 20,000, to mimic the conditions of PSO-SCFT.

In the first study, simulations were seeded using the *output fields* corresponding to the lowest free energy structure in iteration 1 of Figure 3.16. These output fields correspond to a GYR structure which had a unit cell size of $8.9\,R_g$. A plot of the intensive free energy vs. L_{init} from SCFT simulations which were initialized with this seed is shown in Figure 3.22. This plot looks like a typical plot of free energy vs. cell size for SCFT simulations which are run using the same seed; it is fairly smooth, and has a minimum at the commensurate cell size. The density profile shown in the inset looks like the GYR morphology. Each of the SCFT simulations initialized using the cell sizes resulted in structures which looked like GYR, but the free energies were different because the cell sizes had not grown or shrunk to the expected value over

the course of only 20,000 SCFT iterations. If instead we performed fixed-cell SCFT simulations where the cell size remained at the value prescribed by L_{init}, we would see larger free energy differences, but would still recover GYR structures through the range of L_{init} values reported (7,10) R_g. Thus far, everything which has been stated about this study does not seem like it would raise any concern. In fact, the robustness of SCFT results to variations in the initial cell size employed when using the same "perfect" GYR structure seems like it should be a reassurance.

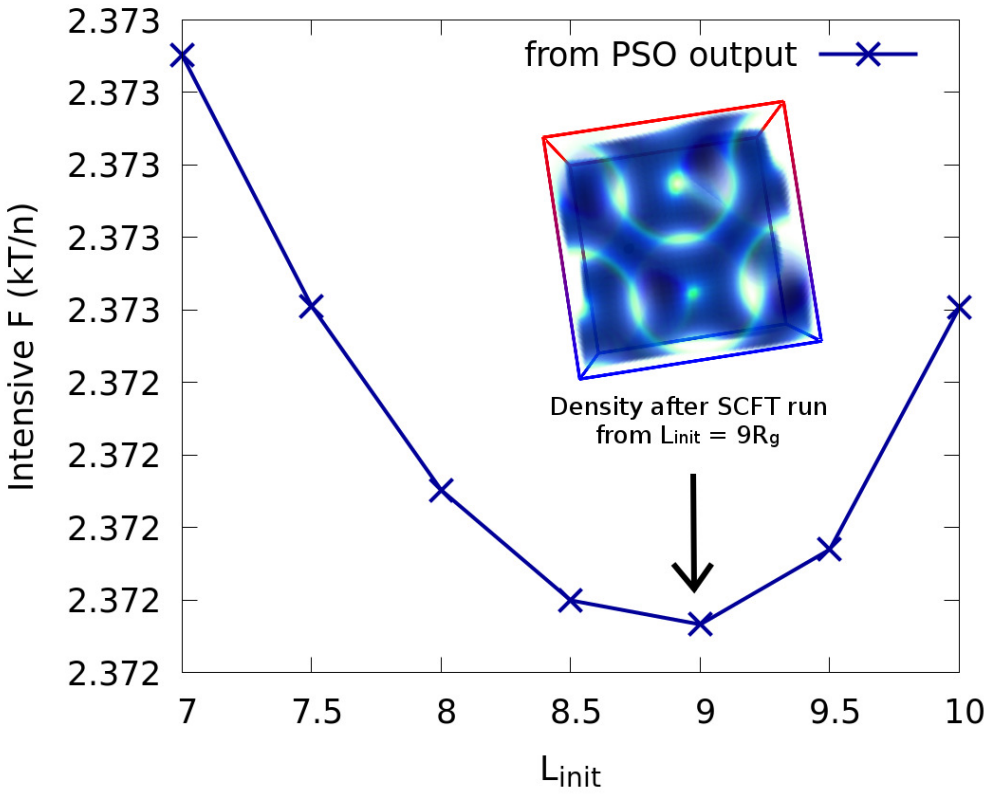

Figure 3.22: Plot of intensive free energy vs. initial cell size, L_{init} for SCFT simulations which were all seeded with the output fields corresponding to the lowest free energy structure in iteration 1 of Figure 3.16 (GYR structure). The inset shows the density profile of the A-rich polymer region after an SCFT simulation was run using the seed and $L_{init} = 9\,R_g$.

The disturbing part of the story comes when we instead use the single-mode seed which generated the GYR fields we used to initialize the simulations in Figure 3.22.

This single-mode seed was the PSO-generated seed corresponding to the first identification of GYR in Figure 3.16 (*i.e.*, the agent associated with the lowest free energy structure in iteration 1). Figure 3.23 displays the free energy vs. L_{init} plot of this set of SCFT simulations.

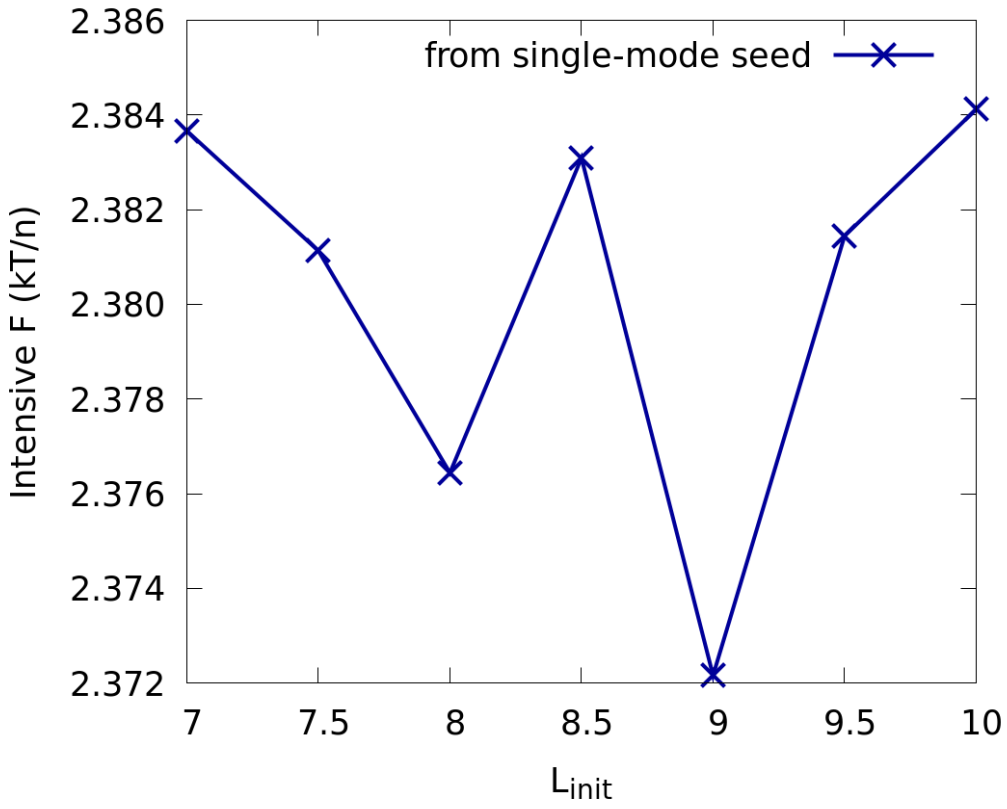

Figure 3.23: Plot of intensive free energy vs. initial cell size, L_{init} for SCFT simulations which were all seeded with the PSO-generated single-mode seed which corresponds to the lowest free energy structure in iteration 1 of Figure 3.16.

As we can see, the plot is no longer smooth, but does still have a minimum occurring at $L_{init} = 9\,R_g$. Figure 3.24 displays density plots of individual SCFT simulations which were initialized using this same single-mode seed and the indicated initial cell sizes. Here, we see that SCFT is no longer robust to using a range of initial cell sizes. This means that if we want to use the PSO-generated single-mode seeds, we

had better have a good guess for the initial cell size for SCFT as well!

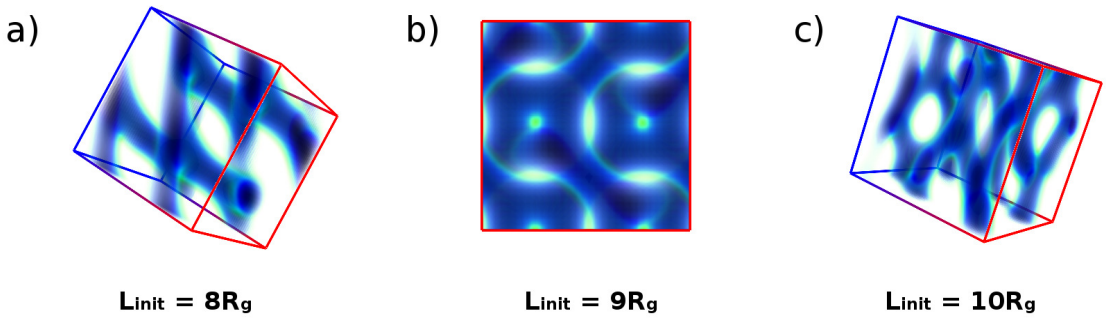

Figure 3.24: Density plots from variable-cell SCFT simulations using initial cell sizes of (a) $L_{init} = 8\,R_g$, (b) $L_{init} = 9\,R_g$, and (c) $L_{init} = 10\,R_g$. The SCFT simulations were all initialized using the PSO-generated single-mode seed which corresponds to the lowest free energy structure in iteration 1 of Figure 3.16.

This conclusion provokes two thoughts: **1) the forward problem is difficult, because SCFT is a local optimizer which depends heavily on the initial conditions used for a simulation, and 2) the initial conditions of an SCFT simulation in the context of a reciprocal space, single-mode seed include the relative peak positions, their relative phase relationships, *and* the initial cell tensor (characterized by a single lattice parameter in cubic cells)**.

3.7 q^* as a PSO-SCFT Search Variable

In this section, we discuss the development of a version of the PSO-SCFT algorithm which includes the cell size as a variable that must be optimized, in addition to the existing search variables of peak positions and amplitudes. We restrict ourselves to implementing this for simulations which begin in cubic cells, since these can be described with a single lattice parameter that indicates the length of cell. However, we will find that working with reciprocal space seeds lends itself to using a quantity we call q^*, which is related to simulation cell size by $q^* = \frac{2\pi n}{L}$, where n is an even integer that determines the number of periods of the primary wave vector that exist in a simulation cell of length L.

3.7.1 What is q^* and its Possible Relevance?

In the context of PSO-SCFT, q^* is the magnitude of the wave vector which dictates the lengthscale of the features in the density profile of a morphology generated by SCFT. In a sense, it is analogous to the primary wave vector seen in scattering experiments such as SAXS or SANS. We first began thinking about this quantity upon the realization that the unit cells of different morphologies of interest contain different numbers of periods of the primary wave vector. For instance, the commensurate lattice vector for BCC is related to q^* by $L_{BCC} = \frac{4\pi}{q^*_{BCC}}$, but the commensurate lattice vector for GYR is related to q^* by $L_{GYR} = \frac{8\pi}{q^*_{GYR}}$. This means that in a BCC unit cell, there are 2 periods of the primary wave vector, whereas for GYR there are 4.

To better understand the relationship between q^* and the commensurate cell size of a morphology, one can study the analysis performed by Leibler in his 1980 paper [58]. Specifically, one can look at Section IV of his manuscript, where he determines the critical point and order-disorder phase boundary for a diblock by constructing

correlation functions from the Debye function. From the plot in Figure 2 of his paper, one can estimate the value of q^* at a given block fraction. For $f_A \approx 0.4$, we see that $q^* \approx 2.8$ and domain spacing $D = \frac{2\pi}{q^*} \approx 2.2$. This means that for GYR at χN=15 and $f_A = 0.39$, where the commensurate cell size is about $8.9\,R_g$, there are about 4 domain spacings (4 periods of the primary wave vector) in a unit cell.

We have seen that the initial cell size we choose to initialize an SCFT simulation is important, especially when starting from a single-mode seed. Thus, we need to find some way to incorporate simulation cell size as part of the space which the PSO-SCFT algorithm explores. It seems like a sensible way to do this is by exploring the q^* space. We are interested in generating reciprocal space, single-mode seeds as agents. The other variables being optimized by PSO-SCFT live in reciprocal space: peak positions in terms of their azimuthal and polar angles, and amplitudes. L_{init} is a real-space quantity, where as q^* is a reciprocal space quantity. Thus, for consistency across PSO-SCFT optimization variables to remain in the same sort of "thematic space", we chose to make q^* the quantity that varies in the search.

3.7.2 RPA Calculation of q^* for an AB_4 Miktoarm

Before we incorporate q^* as a variable for PSO-SCFT to optimize, it is important to verify that using such a scheme would be applicable to systems besides the diblock. In this subsection, we follow the analysis performed by Leibler discussed above, but for the AB_4 miktoarm. The idea is to see if we can accurately determine q^* for the miktoarm, and if so, what kinds of ranges for the value of q^* we might expect to observe.

The first step in obtaining a curve analogous to the one in Figure 2 from Leibler's

paper is to define Debye-like functions for the miktoarm:

$$\begin{aligned} g_1(f,x) &= \frac{2}{x^2}(fx + \exp(-fx) - 1) \\ h_d(f,x) &= \frac{1}{x}(1 - \exp(-fx)) \end{aligned} \tag{3.8}$$

Here, $g_1(f,x)$ was taken from the Leibler paper (Equation IV-3), and $h_d(f,x)$ was determined by modifying the work in Fredrickson's book [13] for a homopolymer star (Equations 3.140, 3.141). g_1 and h_d provide expressions for the correlations for segments on the same arm and correlations with different arms, respectively. From here, we write down correlation functions for the two types of polymer in our system, A and B:

$$\begin{aligned} S_{AA} &= n\, n_A[g_1(\alpha_A,x) + (n_A - 1)h_d(\alpha_A,x)^2] \\ S_{BB} &= n\, n_B[g_1(\alpha_B,x) + (n_B - 1)h_d(\alpha_B,x)^2] \\ S_{AB} &= n\, n_A\, n_B h_d(\alpha_A,x)h_d(\alpha_B,x) \end{aligned} \tag{3.9}$$

where n is the total number of arms, n_A and n_B are the number of A- and B-type arms, and α_A and α_B are the volume fractions of the A- and B-type polymers, respectively. S_{AA}, S_{BB}, and S_{AB} can be interpreted as correlations for A-type segments with other A-type segments, B with B, and A with B, respectively. As a quick verification that these definitions are correct, one can set $n_A = n_B = 1$ and choose $\alpha_A \mapsto \mathrm{f}_A$, $\alpha_B \mapsto$ 1-f$_A$. By doing this, we should recover $S_{AB} = S_{12}$, the correlation function for the diblock in Equation IV-2 of Leibler's work, which does indeed happen.

The correlation functions S_{AA}, S_{BB}, and S_{AB} can be added together to define $S(q)$, a density-density correlation function which sums over all correlations in the

system: $S(q) = S_{AA} + S_{BB} + 2S_{AB}$. $S(q)$, in turn, can be used to determine $\tilde{S}(q)$, the structure factor. Equating Leibler's equations IV-1 to IV-5, we determine that we can obtain:

$$F(x) = \frac{N\,\tilde{S(q)}}{W} \tag{3.10}$$

where W is the determinant of the matrix composed of the correlation functions of the ideal independent copolymer chains $||S_{ij}||$. q^* can be obtained by determining the q which minimizes $F(x)$, with $q^* = \sqrt{x^*}/R_g$. For the AB_4 miktoarm at $f_A = 0.32$, we obtain $q^* = 2.6$. From comparisons with SCFT simulations to identify the commensurate cell size for A_{15} at χN=40, f_A=0.32, we find that this is a gross overestimate of q^*, which should be approximately 1.7. The difference may be accounted for by the fact that such RPA analysis is accurate close to the spinodal line, but breaks down with the divergence of the structure factor at higher segregation strengths. Nevertheless, we can make a coarse estimate from what we observed in this exercise that the range of values accessible to q^* for the AB_4 miktoarm would most likely be (1,3).

3.7.3 Application in the Context of PSO-SCFT

Implementing q^* as a variable for PSO-SCFT to optimize requires only some slight modifications to Algorithm 3. The algorithm for the full version of PSO-SCFT which optimizes peak positions and amplitudes, as well as q^* is shown in Algorithm 4.

initialize peak positions $x_j^0 = (\phi_j, \theta_j)$, peak velocities v_j^0, peak amplitudes a_j^0, and peak velocities $v_{a,j}^0$ randomly for all peaks j of all agents i;

initialize $q_i^{*,0}$ and its velocities $v_{q^{*,0},i}$ randomly for all agents i;

for *each timestep* $t > 0$ **do**

  for *each agent* i **do**

    for *each peak* j **do**

      apply $\mathbf{R}_{ge}$ to current position (rotate to local frame);

      convert from cartesian coordinates to spherical coordinates;

      update position x_j^t using Eqn. 3.2 in local, equatorial frame ;

      apply $\mathbf{R}_{eg} = (\mathbf{R}_{ge})^{-1}$ to updated position (rotate to global frame);

      apply boundary conditions (periodic in ϕ, reflecting in θ);

    end

    generate SCFT seed consisting of all $2j$ peaks (reflect through origin);

    run SCFT simulation using seed & get fitness (intensive free energy), f;

    update x_p^t and a_p^t of all j peaks;

    update $q_i^{*,t}$ according to Eqn 3.6 and update $q_p^{*,t}$;

  end

  update x_g^t and a_g^t;

  update $q_g^{*,t}$;

end

Algorithm 4: Update process for the PSO-SCFT algorithm which optimizes sets of peak positions, their amplitudes, and q^*.

For this version of the algorithm, each agent consists of a *set of peaks*, each with their own amplitudes and azimuthal/polar angles, as well as *one* value of q^*. This is because each agent is used to generate a single-mode SCFT seed. Values of q^* are restricted to a prescribed range that can be assigned upon initialization of the PSO-SCFT algorithm. The restriction is accomplished by using a tanh function which maps q^* onto the entire number line, in a similar fashion to what was done with peak amplitudes.

The forward mapping to the space where q^* values are updated is given by: $\frac{1}{p}\tanh^{-1}\frac{(a+b-2x)}{(a-b)}$, where a and b are the bounds on the range to which q^* is restricted, and p sets the width of the tanh function. The backward mapping is simply given by the inverse of this function: $\frac{b-a}{2}(\tanh(p\,x)+1)+a$, and is applied to the value of q^* which would have been updated in the mapped space. We found that using $p = 0.005$ resulted in a soft enough tanh function that the entire range could be explored. For larger values of p, agents would have q^* values which bunched at the extremes of a given range.

Values for q^* for each agent are updated according to the integrated equations of motion described in Equation 3.6, and can be assigned χ, c_p, and c_g values that are independent of the analogous values in the equations updating peak positions or amplitudes. We will not present a discussion of the tuning of these parameters in this thesis. However, values we use for the results presented in Sections 3.7.4 and 3.7.5 are listed in Table 3.3.

Search Variable	PSO Parameter	Value
$\{\phi, \theta,$ amplitudes$\}$	χ	0.5
	c_p	0.2
	c_g	0.2
q^*	χ	0.7
	c_p	2
	c_g	2

Table 3.3: PSO-SCFT parameters for optimizing peak position and q^*.

3.7.4 Results for AB Diblock

In this subsection, we discuss the results of applying the "full" version of PSO-SCFT which optimizes peak positions, their amplitudes, and q^*, to the conformationally symmetric diblock we have been using for benchmarking throughout this thesis. Specifically, we look at results for χN = 15 with $f_A = 0.35$ (target is HEX) and $f_A = 0.39$ (target is GYR).

Figure 3.25 displays a plot of the intensive free energy vs. iteration for PSO-SCFT applied to the diblock at χN = 15 and $f_A = 0.35$, where HEX is the target morphology. 6 peaks were used, which means there was a total of 12 peaks on the sphere. q^* was restricted to values between (1.5,3.5), which corresponded to a range of L_{init} falling between (3.58,8.37) R_g. From the plot, we see that even with only 5 agents, HEX is identified within the first iteration of the algorithm. We also see an envelope on the distribution of values, suggesting that convergence to HEX is occurring over successive iterations.

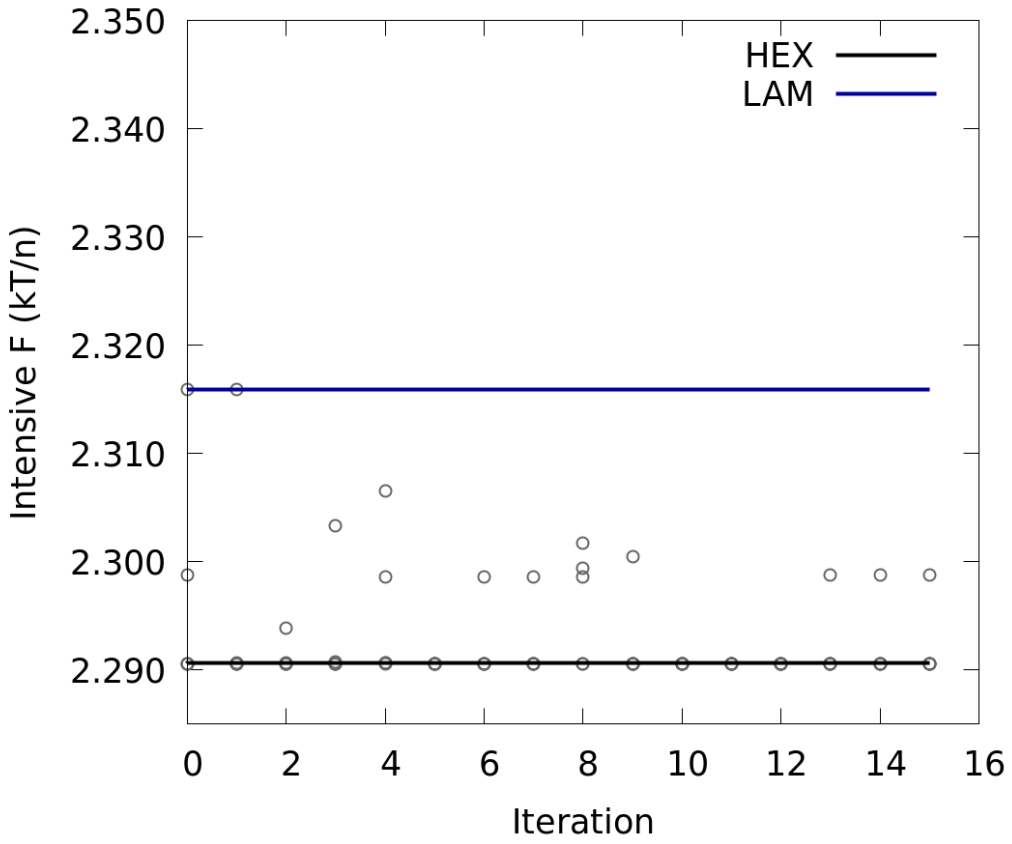

Figure 3.25: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT using 5 agents for a diblock at $f_A = 0.35$, $\chi N = 15$, where HEX is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles. Individual variable-cell SCFT simulations were initialized in cubic cells over a range of $q^* \in (1.5, 3.5)$, which corresponds to $L_{init} \in (3.58, 8.37)\, R_g$.

We can verify convergence to HEX by examining plots of the values of q^* and v_{q^*} vs. iteration, shown in panels (a) and (b), respectively, in Figure 3.26. From the plot in panel (a), we see that there is a distribution of values for q^* in the first iteration, but by the last iteration, 3 of the 5 agents have swarmed toward a value of approximately $q^* = 1.9$, which corresponds to $L_{init} = 6.6\, R_g$. This makes sense, since we are initializing simulations in cubic cells, rather than hexagonal cells. In a rectangular cell spanning the height of two hexagonal unit cells, the length of the rectangular cell would be $a\sqrt{3}$, where a is the side length of the hexagonal cell. $\frac{6.6\, R_g}{\sqrt{3}} = 3.8\, R_g$,

which is the unit cell size for HEX in a hexagonal lattice at this composition. Visual inspection of the structures and the intensive free energy provide further verification of the identification of HEX by the full version of PSO-SCFT.

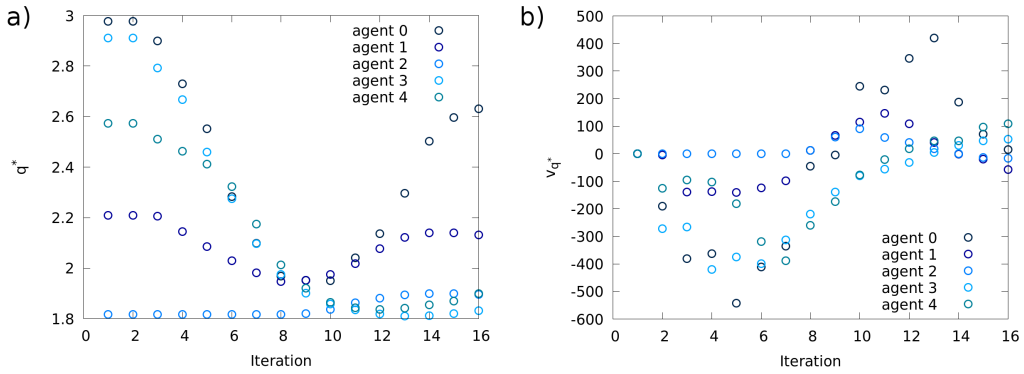

Figure 3.26: Plots of (a) q^* and (b) v_{q^*} vs. iteration for all agents in an implementation PSO-SCFT at $f_A = 0.35$, χN = 15, where HEX is the target morphology. Individual variable-cell SCFT simulations were initialized in cubic cells over a range of $q^* \in (1.5, 3.5)$, which corresponds to $L_{init} \in (3.58, 8.37)\, R_g$.

The fact that PSO-SCFT can identify HEX in 3D simulation cells with only 5 agents in a total runtime of less than an hour without any *a priori* knowledge of the symmetry or the expected cell size is a significant improvement over the GA-SCFT algorithm, which did not have the capability to screen cell sizes at all. However, HEX is a simple morphology and straightforward to identify, even without the help of a mimetic algorithm. We are interested in determining whether PSO-SCFT is able to recover GYR under these conditions.

We present a plot of intensive free energy vs. iteration for all 50 agents of a full PSO-SCFT applied for the diblock at χN = 15, f$_A$ = 0.39, where GYR is the target morphology in Figure 3.27. The range to which q^* values were restricted was smaller than that of the example for the case for HEX, but as we saw in our studies of the density of states of defective structures in Section 2.7.2 and the studies of the robustness

of SCFT to changing L_{init} using the same single-mode seed in Section 3.6, even small differences in cell sizes can produce drastically different structures at this composition.

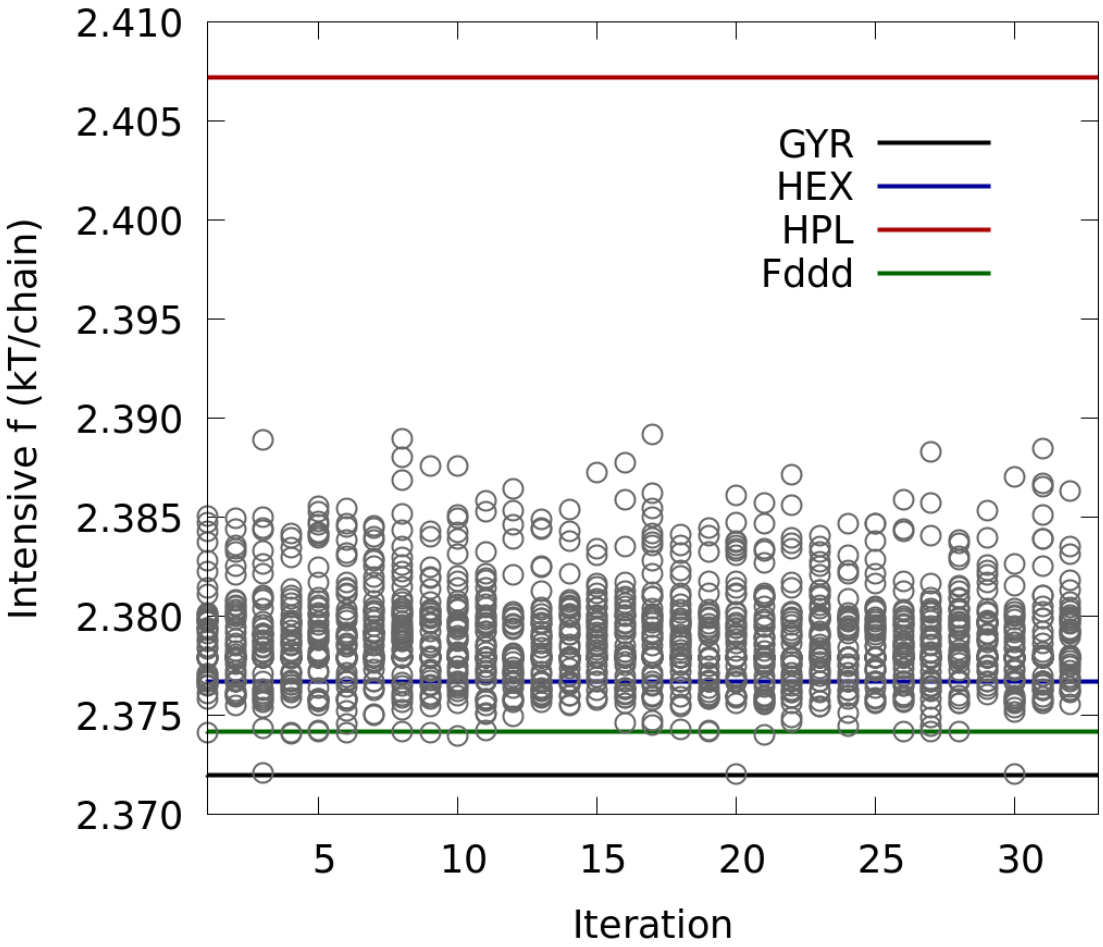

Figure 3.27: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT with amplitude as a search variable, using 50 agents for a diblock at $f_A = 0.39$, $\chi N = 15$, where GYR is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles. Individual variable-cell SCFT simulations were initialized in cubic cells over a range of $q^* \in (2.7, 3.1)$, which corresponds to $L_{init} \in (8.1, 9.3)\, R_g$.

From the plot, we see that we were actually able to identify GYR within a handful of iterations. PSO-SCFT is slow to converge to this structure, but agents do eventually return to it over the course of successive iterations. Convergence may be helped

with finer tuning of the PSO parameters, and possibly by using a different network topology. Nevertheless, this is a promising result, and shows marked progress towards identifying competitive structures from an agnostic standpoint.

3.7.5 Results for AB_4 Miktoarm

The last case study in this section is for the full PSO-SCFT algorithm applied to the AB_4 miktoarm at χN = 40 and $f_A = 0.32$, where A_{15} is the target morphology. Figure 3.28 shows the intensive free energy vs. iteration plot for all 30 agents in the swarm. A_{15} is successfully identified after 14 iterations, and agents return to sets of peak positions and q^* values which correspond to the A_{15} morphology in later iterations. As in previous examples, we observe a wide variety of structures through successive iterations. For this example, it was important to use SCFT simulations which had 48 plane waves in each direction, since the cell sizes could grow to be quite large, and we wanted to maintain the spatial resolution $\Delta x \lessapprox 0.35$. This resulted in individual SCFT simulations having a runtime of up to 4 hours, which drastically slows down the implementation of the full PSO-SCFT. An 80-iteration run, for instance, could take up to 5 days.

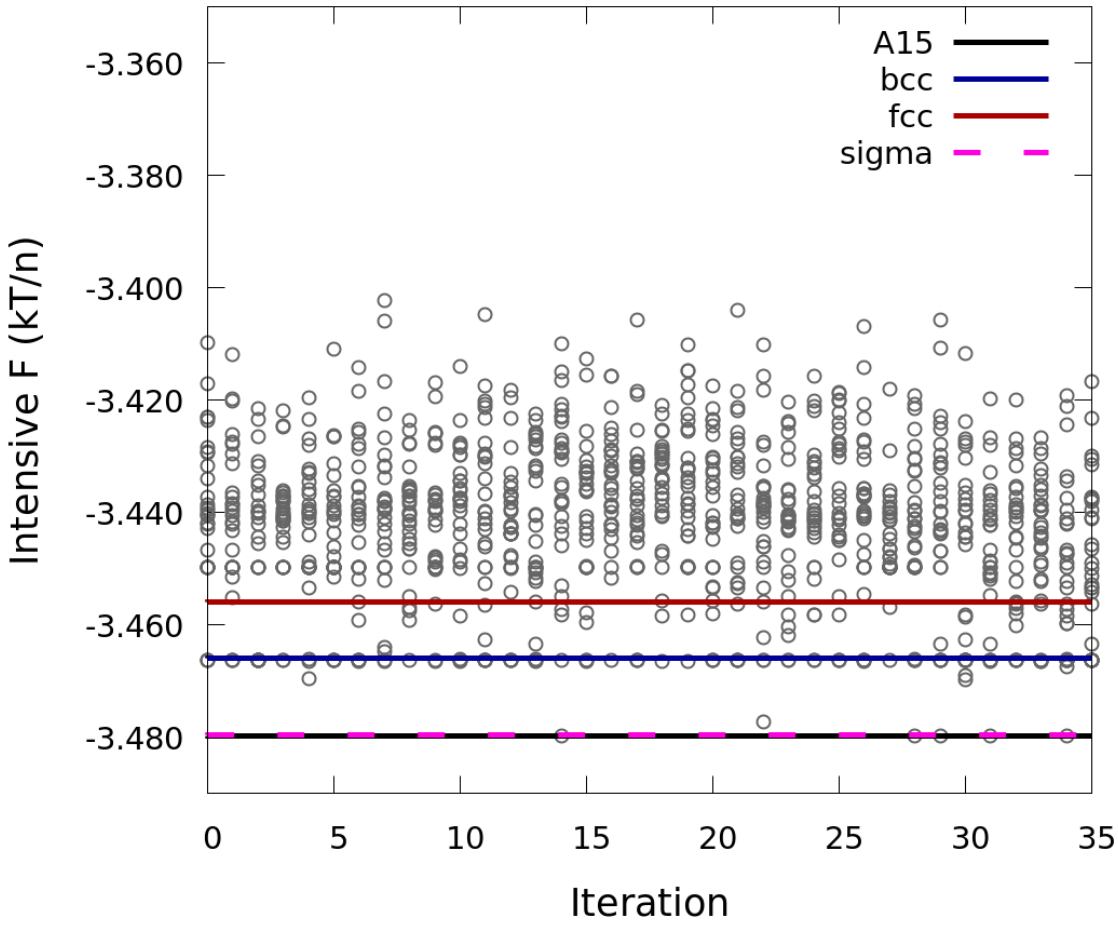

Figure 3.28: Intensive free energy vs. iteration for benchmark testing of PSO-SCFT with peak positions, amplitude, and q^* as search variables, using 30 agents for an AB_4 miktoarm star polymer at $f_A = 0.32$, $\chi N = 40$, where A_{15} is expected to be stable. The algorithm optimized 6 peaks, each with their own amplitude and azimuthal and polar angles. Individual variable-cell SCFT simulations were initialized in cubic cells over a range of $q^* \in (1.2, 2)$, which corresponds to $L_{init} \in (6.28, 10.47)\, R_g$.

3.7.6 Comments

In this section, we applied the full PSO-SCFT algorithm to the conformationally symmetric diblock, at compositions where HEX and GYR were expected to be stable, as well as to the AB_4 miktoarm at a composition where A_{15} is the target morphology. The full version of the PSO-SCFT algorithm was successful at identifying the target morphologies in all cases.

It should be noted, however, that it is important to set the correct number of periods of the q^* wave vector when employing the PSO-SCFT algorithm. As previously mentioned, the GYR unit cell has 4 periods of the primary wave vector, which means $q^* = \frac{2n\pi}{L}$, with $n = 4$. If the PSO-SCFT algorithm is performed using $n = 2$, instead, GYR is never found. Indeed, when PSO-SCFT is performed using $n = 2$ at the same composition, the algorithm seems to favor the Fddd structure instead, even though it is more than $2\,mkT/chain$ higher in intensive free energy than GYR, which is an order of magnitude greater than the differences between some of the competitive sphere morphologies seen in asymmetric AB systems.

A brief explanation for this phenomenon may lie in the fact that the plane wave expansion for the single-mode approximation is comprised of a linear combination of the plane waves in the basis. What that means is that there can be different coefficients on the plane waves which contribute to the single-mode seed that would yield a morphology of interest. For instance, for the case of FCC, a "single-mode" seed can be constructed by populating peaks at k-indices of (2,0,0), (0,2,0), (0,0,2) and various plus/minus permutations of (2,2,2). However, the FCC morphology cannot be retained by populating only the (1,0,0), (0,1,0), (0,0,1) and permutations of (1,1,1) positions, even though as a cubic morphology it takes (1,0,0), (0,1,0), (0,0,1) as its primitive lattice vectors. Furthermore, we know that $\sqrt{2} \neq \sqrt{1^2 + 1^2 + 1^2}$, so there appears to be more than a single "radius" associated with the peaks comprising the

single-mode seed.

It may be helpful to recall at this point that in a single-mode approximation, polymer densities, ρ, are constructed by retaining only the first term in a sum over basis functions, ϕ that contribute to the density: $\rho = \sum_n c_n \phi_n$, where $\phi_n = \sum_k a_k \exp(2\pi i k x)$. That is, only ϕ_1 is retained to approximate the density. In other words, in order for ϕ_1 to span the symmetries (simultaneously satisfy all of the symmetry elements belonging to a space group of interest), a linear combination of the plane waves comprising ϕ_1 must be used. This is consistent with what we see in our example above with FCC.

The fact that n must be assigned correctly has some important ramifications in terms of the development and capabilities of PSO-SCFT. It is yet unclear whether it is simply enough to prescribe n to be some large enough even integer that will capture enough periods of the primary wave vector in the search for competitive structures, or if it is necessary to set it to its exact value for the unit cell of a given possible morphology. The latter constraint is obviously more strict, and would require either some kind of manual sweep or implementation of n as a variable for PSO-SCFT to optimize. This study, however, is deferred to future work.

3.8 Mystery Morphology

In the final section of this chapter, we discuss the characterization of a "mystery" morphology that was observed in several independent samples of PSO-SCFT applied for the miktoarm system. After using the fields file associated with this structure to seed a longer, more highly resolved SCFT simulation, we determined that this structure was indeed metastable at χN = 40 and $f_A = 0.32$, with an intensive free energy which was $2\,mkT/chain$ higher than that of A_{15}. This mystery structure thus has

a lower intensive free energy than BCC and FCC at this composition, but a higher intensive free energy than σ and A_{15}. Figure 3.29 shows images of the density profile of the "mystery" morphology shown from three different perspectives.

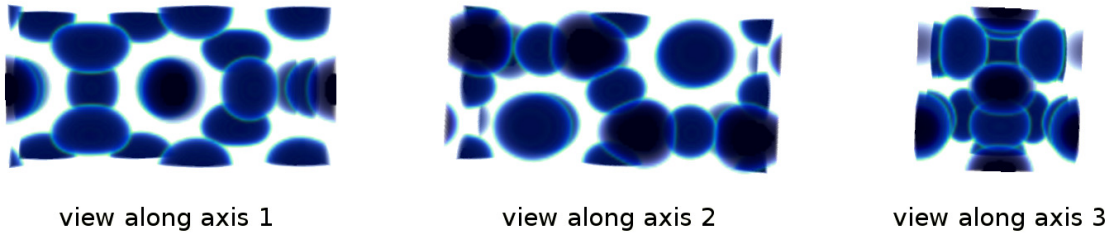

Figure 3.29: Density profile shown from three different perspectives of the "mystery" morphology observed in multiple samples where PSO-SCFT was applied for the AB_4 miktoarm at χN = 40 and $f_A = 0.32$.

Characterizing an unknown structure presents a slew of difficulties. Determining the space group to which the morphology belongs is one of these challenges, but can be accomplished using online tools such as FindSym, as long as the coordinates of the sphere centers are known within some tolerance. The coordinates of the sphere centers were identified both manually and via a script written by group mate Joshua Lequieu, yielding the same values. It was interesting to see that the spheres had slightly different volumes and were not exactly spherical. This seems to be the case for many of the exotic sphere morphologies which have been observed in asymmetrical block copolymer systems. This seems to stem from the Frank-Kasper phases are comprised of different combinations of slightly "deformed" voronoi cells that allow for perfect tiling of the cells. For our "mystery" morphology, there were three sphere "types" and their associated volumes, corresponding to three different Wigner-Seitz cells in a 2:2:3 ratio.

From visual inspection we originally thought the "mystery" phase belonged to

the tetragonal space group, since two of the sides were very similar and unequal to the third, longer side. However, from our work to characterize the morphology using FindSym, it appears that the structure belongs to Space Group 191 (P6/mmm), which is a hexagonal space group. This can be verified by entering the coordinates of the sphere centers into a program such as VESTA, acquiring the Wyckoff positions, and going to the International Crystallographic Tables to confirm that the correct linear combination of the Wyckoff positions exist in the structure[68]. Indeed, the Wyckoff positions corresponding to the symmetry elements $c\,\bar{6}m2$, $d\,\bar{6}m2$, and $fmmm$ occur in the correct 2:2:3 ratio seen for the spheres, with matching fractional coordinates those listed in the table for Space Group 191.

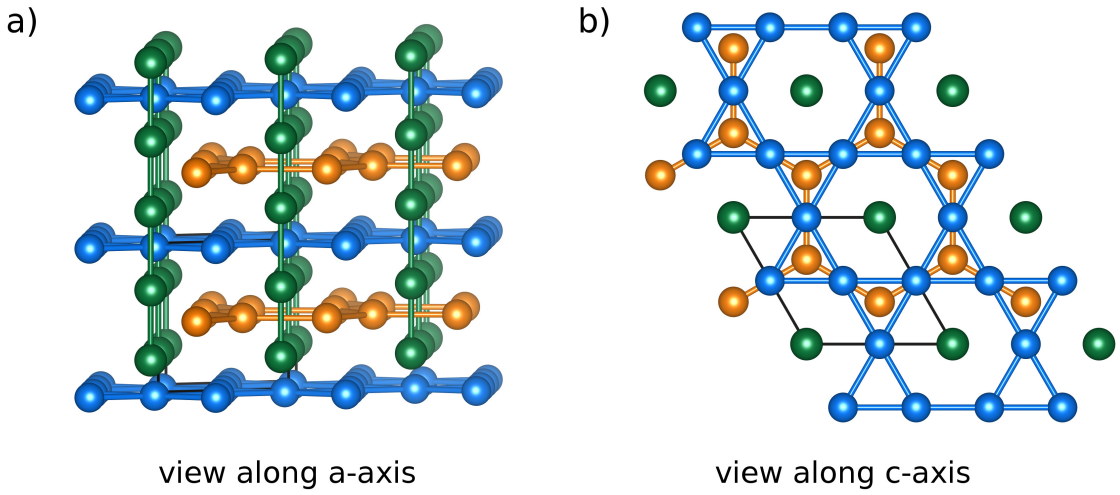

Figure 3.30: Crystal structure of Al_3Zr_4 viewed along (a) a-axis and (b) c-axis.

The next challenge in characterizing a structure is the fact that a given space group can have thousands of structures which belong to it. This would have been akin to searching for a needle in a haystack. However, Professor Ram Seshadri offered tremendous assistance in identifying the "mystery" morphology as the crystal structure of Al_3Zr_4. Figure 3.30 shows the structure of Al_3Zr_4. It is characterized by

alternating graphite sheets separating kagome sheets with vertical rods in a triangular lattice.

At first glance, it is difficult to see how the structure of Al_3Zr_4 is the same as the "mystery" morphology from our PSO-SCFT algorithm. However, we can see from side-by-side images of the hexagonal unit cells for each structure that they look quite similar (see Figure 3.31). The image for the unit cell in panel (b) is colored according to the three different types of Wigner-Seitz cells observed in the mystery morphology. The fact that the structures of the hexagonal unit cell match so closely is indicative that they are indeed the same.

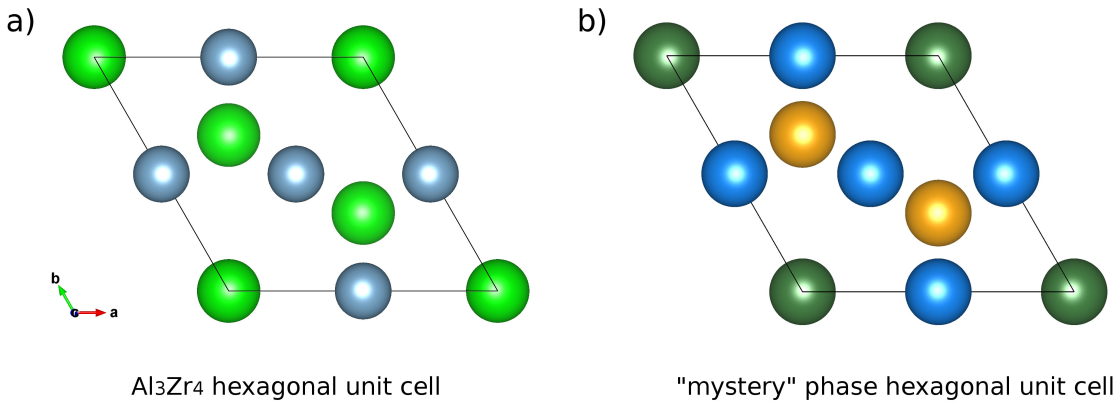

Figure 3.31: Hexagonal unit cells for (a) Al_3Zr_4 and (b) the "mystery" morphology observed in multiple samples where PSO-SCFT was applied for the AB_4 miktoarm at $\chi N = 40$ and $f_A = 0.32$.

To the best of our knowledge, this structure has not been observed in the block copolymer community. Experimental verification of this morphology is needed. Such verification, however, would mark non-trivial progress toward solving the forward problem and more predictive capabilities on the part of theory and computation.

3.9 Concluding Remarks

In this chapter, we discussed the development of a reciprocal space global optimization method, PSO-SCFT, with the goal of agnostically determining stable and low-lying metastable polymorphs accessible to a given block copolymer system. The full version of the PSO-SCFT algorithm seeks to optimize the positions and amplitudes of a set of peaks, as well as the value of the primary wave vector, q^*, which sets the lengthscale for morphological features and is related to the unit cell size by $q^* = \frac{2n\pi}{L}$. The combination of all these variables were used to generate single-mode seeds which were used to initialize SCFT simulations that yielded a variety of morphologies. Intensive free energy was used as a metric to determine the fitness of each seed which guided the trajectory of PSO-SCFT agents through the space of possible sets of peak positions, amplitudes, and q^* values.

PSO-SCFT successfully identified the expected stable morphology and low-lying known competitors for the conformationally symmetric diblock and an AB_4 miktoarm system, both of which have established phase diagrams [24, 26, 27]. In the case of the AB_4 miktoarm, PSO-SCFT was even able to identify a previously unobserved structure.

We found that for the case where GYR was the target morphology, it was important include the sign of the peak amplitudes as a variable which the PSO-SCFT algorithm could search. Furthermore, when q^* was included as an optimization variable, the identification of GYR required assigning the correct number of periods of q^*, n. This provides a small complication, which can be addressed by performing a brief sweep of n, which realistically would likely be restricted to even numbers less than 8 or 10.

The full version of PSO-SCFT demonstrated slow convergence to good structures

which were identified. This could possibly be addressed by using a different network topology and more aggressively tuning the coefficients in the integrated equations of motion for updating agent positions [60, 62].

The current version of the PSO-SCFT algorithm is far from perfect; it is fairly basic and does not always work. It can also become prohibitively expensive, since the runtime of SCFT simulations can vary quite drastically from one polymer system to another. However, PSO-SCFT does seem like a promising step towards addressing the forward problem, and as a proof-of-concept works as a tool which can be used to screen for competitive morphologies.

Chapter 4

Conclusions and Outlook

4.1 Summary

In this dissertation, we considered the development of two heuristic algorithms for the agnostic identification of globally stable and low-lying metastable competing morphologies of block copolymer melts. This is an important problem which needs to be solved in order for true inverse design of materials to be possible, since being able to determine the structure and composition of a block copolymer which would yield desired material properties requires knowledge of the morphologies accessible to such systems in the first place. However, the forward problem is a difficult one. The current computational tools for mapping phase diagrams of block copolymers rely on experimental guidance to even know which structures might be competitive and *a priori* knowledge of the symmetry group and therefore the symmetry elements spanned by a given candidate morphology. This is a consequence of the fact that for a given block copolymer system, there is a rugged free energy landscape associated with the possible morphologies accessible to system. However, current computational methods such as SCFT rely heavily on the initial conditions used to seed simulations,

and thus result in local optimizations, rather than global ones.

Mimetic algorithms, which combine heuristic global optimization algorithms with local optimization methods, have been successfully applied to structure determination problems in biophysics and metallurgical systems [41, 42, 63, 61]. These algorithms seek to explore the landscape associated with various structures accessible to their respective systems. The combination of global and local optimization methods is achieved by developing algorithms which iterate through global optimization moves, which generate "perturbed" configurations which can be relaxed by the local optimizer.

We developed two approaches toward addressing the forward problem in block copolymers, drawing inspiration from mimetic algorithms in other applications: GA-SCFT (Genetic Algorithm + SCFT) and PSO-SCFT (Particle Swarm Optimization + SCFT). Both algorithms we discuss were benchmarked by applying them to the diblock, for which the theoretical phase diagram is well-established. The PSO-SCFT method was also benchmarked on the AB_4 miktoarm star polymer, where sphere morphologies are competitive. The idea with both methods was to explore the free energy landscape associated with the numerous defective and competitive metastable states by generating seeds which SCFT would relax into local basins of attraction. Structures with better fitness, which was measured by comparing their intensive free energies, would inform the exploration of the optimization space.

In Chapter 2, we discussed the GA-SCFT algorithm, which uses a Genetic Algorithm as a wrapper around SCFT. In the GA-SCFT method, we manipulated real-space SCFT fields. By selecting fit members to go on to "reproduce" by exchanging regions of their SCFT fields and then allowing them to undergo a variety of mutations, we were able to generate many seed fields which were used as initial configurations for SCFT to relax. Each seed corresponded to a different structure. We saw that the

GA-SCFT approach was successful at identifying non-network morphologies, but was never able to recover GYR when it was the target structure. One of the challenges with this method was that it would become "stuck" in the numerous defective states accessible to the system, or would prematurely collapse into a competing candidate morphology such as HEX or LAM.

In Chapter 3, we developed the PSO-SCFT approach, which addressed some of the challenges we faced in Chapter 2 by reframing the problem to reduce the degrees of freedom that we were manipulating. Instead of recombining entire real-space SCFT fields, we chose to move to reciprocal space to generate "single-mode" seeds. These seeds are effectively the agents of the PSO-SCFT swarm. Agents have positions and velocities in the space of parameters the PSO-SCFT algorithm attempts to optimize. Thus, an agent consisted of a set of peaks, their amplitudes, and an initial cell size which was determined by the primary wave vector, q^*. We found that it was important to include signed peak amplitudes in the search space, since the relative signs of the peaks carry information about the symmetries spanned by a morphology. This became evident in an independent study of independent SCFT simulations which were initialized with various single-mode seeds. Furthermore, we observed that cell size plays a non-trivial role in the identification of structures, and that it was important to assign the appropriate value of n, the number of periods of the primary wave vector when including q^* in the search space. **The full version of PSO-SCFT was able to successfully recover the expected target morphology in all cases, identify known competitors, and in one case, even found a previously unknown structure in the block copolymer community.**

4.2 Concluding Remarks

With PSO-SCFT, we observed a pronounced improvement over the abilities of the GA-SCFT method, which was never able to successfully identify the GYR morphology, even when applied using small simulation cells. By reducing the degrees of freedom which needed to be optimized, we were able to more effectively search through the space of possible seeds. This is necessary because SCFT depends heavily on the initial conditions used for a simulation. For the real-space GA-SCFT approach, the initial configurations used to seed SCFT simulations were comprised of the entire real-space SCFT fields. However, for PSO-SCFT, the initial configurations SCFT simulations were single-mode seeds which contained information about the relative peak positions, their relative phase relationships, and the initial cell tensor that is characterized by a single lattice parameter in cubic cells. Furthermore, PSO-SCFT allows us to search through seeds which correspond to range of unit cell sizes, whereas GA-SCFT did not effectively incorporate this consideration.

It is important to note that the reason for the success of PSO-SCFT as compared with GA-SCFT is not because we used a different algorithm, but because we changed the way we were looking at the problem in a fundamental way by looking at reciprocal space, single-mode seeds. The choice of using a PSO as opposed to a GA made sense in this context for a few reasons: PSO's are better suited to manipulating continuous degrees of freedom, whereas GA's are better for discrete manipulations. Furthermore, other members in the group had seen relative success with the implementation of a PSO toward solving the inverse problem, and there are less tuneable knobs that must be optimized in a PSO as compared with a GA.

Solving the forward problem is far from complete. PSO-SCFT "worked" in the benchmark studies we performed, insofar as it was able to reliably and agnostically

predict a given target morphology and its competitors. However, the algorithm does not converge in a timely manner to "good" structures. More aggressive tuning of the parameters of the PSO, or changing the network topology may facilitate more efficient convergence. The issue of convergence to high fitness structures may not actually be considered that much of an obstacle, since one can effectively use PSO-SCFT as a screening tool. PSO-SCFT can become very expensive. Generating phase diagrams to facilitate inverse design would not be efficiently performed by using PSO-SCFT at every point in phase space. Rather, PSO-SCFT can be used to perform "spot checks" at various points in phase space, to see what kinds of morphologies may be accessible to a new system of interest. In such a case, it would be necessary to "zoom in" to particular structures the method identifies and perform more rigorous SCFT simulations to thoroughly characterize structures, determine their intensive free energies, and generate "clean" seeds which can be used in other simulations. Once a good seed is identified by PSO-SCFT and undergoes such post-processing, it can be used to perform the typical types of simulations necessary to generate phase diagrams for specific block copolymer architectures and compositions of interest.

There is a further complication that is important to acknowledge: SCFT simulations themselves can become prohibitively expensive, depending on the system. For instance, blended systems or block copolymer brushes may result in significantly swollen domains which require individual SCFT simulations to be run at higher spatial resolutions. This coupled with the fact that initial cell size is part of the set of variables being optimized by the PSO means that it may not be possible in all cases to realistically perform PSO-SCFT.

Both Chapters 2 and 3 are presented primarily in a chronological way. This on many levels reflects the process involved in developing such heuristic methods, since there are so many tunable knobs to explore that doing so in an exhaustive way is

an optimization problem unto itself. What this means ultimately was that in many cases, we probed some options for methods and parameters, and then made choices according to the limited scope of results we saw, as well as our intuition and understanding of the forward problem as they matured over the course of the project. This path for developing a method is in contrast with the more comprehensive approaches taken in projects which investigate various implementations of global optimization methods against the same, well-established, standard set of test functions. Therein lies the difference in how we can approach the forward problem in the context of block copolymer systems: our benchmarking is being performed on a system where full SCFT simulations must be run as an inner loop to our global optimizer of choice. The performance of different global optimization strategies ends up being quite problem dependent[69, 70, 71, 60]. Furthermore, it is not always clear that all algorithms which are being compared are themselves fully optimized in terms of values of their specific parameters. What is clear is that there are many considerations which must be entertained in the development of these algorithms.

Nevertheless, the successful results seen in the application of the PSO-SCFT approach have important ramifications for materials design. It is a modular, straightforwardly extensible method which can be applied to a wide range of block copolymer systems, since the only requirements for its use are a measure of fitness (*i.e.*, intensive free energy), and the capacity to manipulate the degrees of freedom which generate single-mode seeds: peak amplitudes/positions and q^*. Being able to reliably predict possible candidates in a given block copolymer system in a day or a handful of days by simply "pressing go" and waiting for simulations to run on a high-performance cluster exhibits a clear advantage over performing a series of experimental syntheses and characterizations. PSO-SCFT method should not necessarily replace experiments, but instead facilitate and guide them. We believe the success of PSO-SCFT

marks significant progress toward solving the forward problem, expediting structure discovery, and ultimately, the ability to achieve inverse design of materials.

Bibliography

[1] F. S. Bates, G. H. Fredrickson, D. Hucul, and S. F. Hahn, *Pche-based pentablock copolymers: Evolution of a new plastic*, *AIChE Journal* **47** (2001) 762–765.

[2] F. S. Bates and G. H. Fredrickson, *Block copolymers - designer soft materials*, *Physics Today* **52** (1999) 32–38.

[3] F. Meng, Z. Zhong, and J. Feijen, *Stimuli-responsive polymersomes for programmed drug delivery*, *Biomacromolecules* **10** (2009) 197–209.

[4] L. Yin, M. Dalsin, A. Sizovs, T. Reineke, and M. Hillmyer, *Glucose-functionalized, serum-stable polymeric micelles from the combination of anionic and raft polymerizations*, *Macromolecules* **45** (2012) 43224332.

[5] M. Ulbricht, *Advanced functional polymer membranes*, *Polymer* **47** (2006) 22172262.

[6] E. A. Jackson and M. A. Hillmyer, *Nanoporous membranes derived from block copolymers*, *ACS Nano* **4** (2010) 35483553.

[7] R. Ruiz, H. Kang, F. A. Detcheverry, E. Dobisz, D. S. Kercher, T. R. Albrecht, J. J. de Pablo, and P. F. Nealey, *Density multiplication and improved lithography by directed block copolymer assembly*, *Science* **321** (2008), no. 5891 936–939, [http://science.sciencemag.org/content/321/5891/936.full.pdf].

[8] S. Kim, H. Solak, M. P. Stoykovich, N. J. Ferrier, J. J. de Pablo, and P. F. Nealey, *Epitaxial self-assembly of block copolymers on lithography*, *Nature* **424** (2003) 411414.

[9] A. Baruth, M. D. Rodwogin, A. Shankar, M. J. Erickson, M. A. Hillmyer, and C. Leighton, *Non-lift-off block copolymer lithography of 25 nm magnetic nanodot arrays*, *ACS Appl. Mater. Interfaces* **3** (2011) 34723481.

[10] S. Kim, P. F. Nealey, and F. S. Bates, *Decoupling bulk thermodynamics and wetting characteristics of block copolymer thin films*, *ACS Macro Lett.* **1** (2012) 1114.

[11] G. Srinivas, D. E. Discher, and M. L. Klein, *Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics*, *Nature Materials* **3** (2004), no. 638.

[12] M. Murat, G. S. Grest, and K. Kremer, *Statics and dynamics of symmetric diblock copolymers: a molecular dynamics study*, *Macromolecules* **32** (1999), no. 3 595–609, [https://doi.org/10.1021/ma981512p].

[13] G. H. Fredrickson, *The Equilibrium Theory of Inhomogenous Polymers*. Oxford Science Publications, Oxford, UK, 2006.

[14] K. Binder and W. Paul, *Monte carlo simulations of polymer dynamics: Recent advances*, *Journal of Polymer Science Part B: Polymer Physics* **35** no. 1 1–31.

[15] K. T. Delaney and G. H. Fredrickson, *Recent developments in fully fluctuating field-theoretic simulations of polymer melts and solutions*, *The Journal of Physical Chemistry B* **120** (2016), no. 31 7615–7634, [http://dx.doi.org/10.1021/acs.jpcb.6b05704]. PMID: 27414265.

[16] M. C. Villet and G. H. Fredrickson, *Efficient field-theoretic simulation of polymer solutions*, *The Journal of Chemical Physics* **141** (2014), no. 22 224115.

[17] J. Qin, F. S. Bates, and D. C. Morse, *Phase behavior of nonfrustrated abc triblock copolymers: Weak and intermediate segregation*, *Macromolecules* **43** (2010) 5128–5136.

[18] C. Tyler, J. Qin, F. Bates, and D. Morse, *Renormalized one-loop theory of correlations in polymer blends*, *Macromolecules* **40** (2007) 4654.

[19] M. Liu, B. Xia, W. Li, F. Qiu, and A. Shi, *Self-assembly of binary mesocrystals from blends of babcb multiblock copolymers and abc triblock copolymers*, *Macromolecules* **48** (2015) 3386–3394.

[20] H. Liu, C. Huang, and A. Shi, *Self-assembly of linear abcba pentablock terpolymers*, *Macromolecules* **48** (2015) 6214–6223.

[21] N. Xie, M. Liu, H. Deng, W. Li, F. Qiu, and A. Shi, *Macromolecular metallurgy of binary mesocrystals via designed multiblock terpolymers*, *Journal of the American Chemical Society* **136** (2014) 2974–2977.

[22] F. Zuo, C. Alfonso, and F. Bates, *Structure and dynamics of elastomeric multiblock terpolymers containing glassy, rubbery and semicrystalline blocks*, *Macromolecules* **44** (2011) 8143–8153.

[23] S. Lee, M. Bleumle, and F. Bates, *Discovery of a frank-kasper sigma phase in sphere-forming block copolymer melts*, *Science* **330** (2010) 349–353.

[24] M. W. Matsen and M. Schick, *Stable and unstable phases of a diblock copolymer melt*, *Physical Review Letters* **72** (1994) 2660–2663.

[25] P. Tang, F. Qiu, H. Zhang, and Y. Yang, *Morphology and phase diagram of complex block copolymers: Abc star triblock copolymers*, *J. Phys. Chem. B* **108** (2004) 8434–8438.

[26] C. A. Tyler and D. C. Morse, *Orthorhombic fddd network in triblock and diblock copolymer melts*, *Phys. Rev. Lett.* **94** (May, 2005) 208302.

[27] N. Xie, W. Li, F. Qiu, and A.-C. Shi, *phase formed in conformationally asymmetric ab-type block copolymers*, *ACS Macro Letters* **3** (2014), no. 9 906–910, [https://doi.org/10.1021/mz500445v].

[28] K. Kim, M. W. Schulze, A. Arora, R. M. Lewis, M. A. Hillmyer, K. D. Dorfman, and F. S. Bates, *Thermal processing of diblock copolymer melts mimics metallurgy*, *Science* **356** (2017), no. 6337 520–523, [http://science.sciencemag.org/content/356/6337/520.full.pdf].

[29] F. Drolet and G. H. Fredrickson, *Combinatorial screening of complex block copolymer assembly with self-consistent field theory*, *Phys. Rev. Letters* **83** (1999) 4317–4320.

[30] Z. Guo, G. Zhang, F. Qiu, H. Zhang, Y. Yang, and A. Shi, *Discovering ordered phases of block copolymers: New results from a generic fourier-space approach*, *Phys. Rev. Lett.* **101** (2008) 028301.

[31] W. Xu, K. Jiang, P. Zhang, and A. Shi, *A strategy to explore stable and metastable ordered phases of block copolymers*, *J. Phys. Chem. B* **117** (2013) 5296–5305.

[32] F. S. Bates, M. A. Hillmyer, T. P. Lodge, C. M. Bates, K. T. Delaney, and G. H. Fredrickson, *Multiblock polymers: Panacea or pandora's box?*, *Science* **336** (2012) 434–440.

[33] A. W. Bosse, S. W. Sides, K. Katsov, C. J. Garca-Cervera, and G. H. Fredrickson, *Defects and their removal in block copolymer thin films*, *J. Pol. Sci. B: Pol. Phys.* **44** (2006) 2495–2511.

[34] S. P. Paradiso, K. T. Delaney, H. D. Ceniceros, C. J. Garcia-Cervera, and G. H. Fredrickson, *Block copolymer self assembly during rapid solvent evaporation: Insights into cylinder growth and stability*, *ACS Macroletters* **3** (2012) 16–20.

[35] J. H. Holland, *Adaptation in Natural and Artificial Systems*. MIT Press, Cambridge, Massachusetts, 1992.

[36] I. Rechenberg, *Evolutionsstrategie - Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. PhD thesis, Technical University of Berlin, 1971.

[37] H. P. Schwefel, *Numerical Optimization of Computer Models*. Birkhuser Verlag, Basel, Switzerland, 1977.

[38] T. Riechmann, *Genetic algorithm learning and evolutionary games*, *Journal of Economic Dynamics and Control* **25** (2001) 1019–1037.

[39] M. A. Nowak and K. Sigmund, *Evolutionary dynamics of biological games*, *Science* **303** (2004) 793–799.

[40] S. Forrest, B. Javornik, R. Smith, and A. Perelson, *Using genetic algorithms to explore pattern recognition in the immune system*, *Evolutionary Computation* **1** (1993) 191–211.

[41] N. Abraham and M. Probert, *A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction*, *Phys. Rev. B* **73** (2006) 224104.

[42] A. R. Oganov and C. W. Glass, *Crystal structure prediction using ab initio evolutionary techniques: Principles and applications*, *J. Chem. Phys.* **124** (2006) 244704.

[43] Q. Zhu, A. R. Oganov, A. O. Lyakhov, and X. Yu, *Generalized evolutionary metadynamics for sampling the energy landscapes and its applications*, *Phys. Rev. B* **92** (2015) 024106.

[44] A. N. Kolmogorov, S. Shah, E. R. Margine, A. F. Bialon, T. Hammerschmidt, and R. Drautz, *New superconducting and semiconducting fe-b compounds predicted with an ab initio evolutionary search*, *Phys. Rev. Letters* **105** (2010) 217003.

[45] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, Massachusetts, 1989.

[46] F. Neria and C. Cottab, *Memetic algorithms and memetic computing optimization: A literature review*, *Swarm and Evolutionary Computation* **2** (2012) 1–14.

[47] M. Matsen, *Self-consistent field theory for melts of low-molecular-weight diblock copolymer*, *Macromolecules* **45** (2012) 8502.

[48] N. M. Razali and J. Geraghty, *Genetic algorithm performance with different selection strategies in solving tsp*, *Proceedings of the World Congress on Engineering* **2** (2011).

[49] A. Lipowski and D. Lipowska, *Roulette-wheel selection via stochastic acceptance*, *Physica A* **391** (2011).

[50] J. Alander, *On optimal population size of genetic algorithms*, in *CompEuro '92. 'Computer Systems and Software Engineering',Proceedings*, pp. 65–70, IEEE, May, 1992.

[51] K. A. D. Jong and W. M. Spears, *An analysis of the interacting roles of population size and crossover in genetic algorithms*, *Lecture Notes in Computer Science* **496** (1991) 38–47.

[52] J. J. GREFENSTETTE, *Optimization of control parameters for genetic algorithms*, *IEEE Transactions of Systems, Man, and Cybernetics* **SMC16** (1986) 122–128.

[53] J. Barrat, G. Fredrickson, and S. J. Sides, *Introducing variable cell shape methods in field theory simulations of polymers.*, *Phys. Chem. B* **109** (2005) 6694–6700.

[54] D. Goldberg and K. Deb.

[55] S. Chanpuriya, K. Kim, J. Zhang, S. Lee, A. Arora, K. D. Dorfman, K. T. Delaney, G. H. Fredrickson, and F. S. Bates, *Cornucopia of nanoscale ordered phases in sphere-forming tetrablock terpolymers*, *ACS Nano* **10** (2016), no. 5 4961–4972, [http://dx.doi.org/10.1021/acsnano.6b00495]. PMID: 27055118.

[56] S. P. Paradiso, K. T. Delaney, and G. H. Fredrickson, *Swarm intelligence platform for multiblock polymer inverse formulation design*, *ACS Macro Letters* **5** (2016), no. 8 972–976, [https://doi.org/10.1021/acsmacrolett.6b00494].

[57] M. R. Khadilkar, S. Paradiso, K. T. Delaney, and G. H. Fredrickson, *Inverse design of bulk morphologies in multiblock polymers using particle swarm optimization*, *Macromolecules* **50** (2017), no. 17 6702–6709, [https://doi.org/10.1021/acs.macromol.7b01204].

[58] L. Leibler, *Theory of microphase separation in block copolymers*, *Macromolecules* **13** (1980), no. 6 1602–1617, [https://doi.org/10.1021/ma60078a047].

[59] J. Kennedy and R. Eberhart, *Particle swarm optimization*, in *Proceedings of ICNN'95 - International Conference on Neural Networks*, vol. 4, pp. 1942–1948 vol.4, Nov, 1995.

[60] Eberhart and Y. Shi, *Particle swarm optimization: developments, applications and resources*, in *Proceedings of the 2001 Congress on Evolutionary Computation (IEEE Cat. No.01TH8546)*, vol. 1, pp. 81–86 vol. 1, May, 2001.

[61] A. Băutu and H. Luchian, *Protein structure prediction in lattice models with particle swarm optimization*, in *Swarm Intelligence* (M. Dorigo, M. Birattari, G. A. Di Caro, R. Doursat, A. P. Engelbrecht, D. Floreano, L. M. Gambardella, R. Groß, E. Şahin, H. Sayama, and T. Stützle, eds.), (Berlin, Heidelberg), pp. 512–519, Springer Berlin Heidelberg, 2010.

[62] M. Meissner, M. Schmuker, and G. Schneider, *Optimized particle swarm optimization (opso) and its application to artificial neural network training*, *BMC Bioinformatics* **7** (2006).

[63] J. Liu, L. Wang, L. He, and F. Shi, *Analysis of toy model for protein folding based on particle swarm optimization algorithm*, in *Advances in Natural Computation* (L. Wang, K. Chen, and Y. S. Ong, eds.), (Berlin, Heidelberg), pp. 636–645, Springer Berlin Heidelberg, 2005.

[64] Y. Wang, J. Lv, L. Zhu, and Y. Ma, *Crystal structure prediction via particle-swarm optimization*, *Phys. Rev. B* **82** (Sep, 2010) 094116.

[65] J. Lv, Y. Wang, L. Zhu, and Y. Ma, *Particle-swarm structure prediction on clusters*, *The Journal of Chemical Physics* **137** (2012), no. 8 084104, [https://doi.org/10.1063/1.4746757].

[66] G. H. Fredrickson, *Concentration depolarization of fluorescence in the presence of molecular rotation*, *The Journal of Chemical Physics* **88** (1988), no. 9 5291–5299, [https://doi.org/10.1063/1.454587].

[67] D. Bratton and J. Kennedy, *Defining a standard for particle swarm optimization*, in *Proceedings of the 2007 IEEE Swarm Intelligence Symposium*, (Washington, DC, USA), pp. 120–127, IEEE Computer Society, 2007.

[68] IUCr, *International Tables for Crystallography, Volume A: Space Group Symmetry*. International Tables for Crystallography. Kluwer Academic Publishers, Dordrecht, Boston, London, 5. revised edition ed., 2002.

[69] D. W. Boeringer and D. H. Werner, *Particle swarm optimization versus genetic algorithms for phased array synthesis*, *IEEE Transactions on Antennas and Propagation* **52** (2004).

[70] Y. Duan, R. G. Harley, and T. G. Habetler, *Comparison of particle swarm optimization and genetic algorithm in the design of permanent magnet motors*, in *2009 IEEE 6th International Power Electronics and Motion Control Conference*, pp. 822–825, May, 2009.

[71] S. Panda and N. P. Padhy, *Comparison of particle swarm optimization and genetic algorithm for facts-based controller design*, *Applied Soft Computing* **8** (2008), no. 4 1418 – 1427. Soft Computing for Dynamic Data Mining.