EMERGENT PHENOMENA AND DEFECT PROPERTIES ACROSS LENGTH SCALES: ATOMS, SMECTICS, COLLOIDS, AND HUMANS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Matthew Keenan Bierbaum

August 2016

EMERGENT PHENOMENA AND DEFECT PROPERTIES ACROSS
LENGTH SCALES: ATOMS, SMECTICS, COLLOIDS, AND HUMANS

Matthew Keenan Bierbaum, Ph.D.

Cornell University 2016

We investigate the properties and dynamics of defects in several disparate physical systems spanning a factor of 10^{15} in length scale and develop theoretical and experimental tools to facilitate their study. In particular, we present results from dislocations in atomic systems all the way to the creation of non-equilibrium vortex states in human crowds. At the smallest scales, we develop a framework to use interatomic models to study material properties using molecular dynamics of atoms. We describe the complex microstructure that forms when atomic materials are plastically deformed using continuum dislocation dynamics. Using similar techniques we simulate focal conic defects using a continuum model for smectic-A liquid crystals. At slightly larger scales we develop an experimental technique to measure the nonlinear stresses around individual vacancies, dislocations, and grain boundaries in colloidal crystals. On human length scales we describe the formation of vortex-like states at heavy metal concerts and the spread of zombism across the continental United States. Finally, we develop a new method to extract particle sizes and positions from confocal images at a resolution $10 - 100\times$ higher than current methods, reaching 1 nm in precision.

BIOGRAPHICAL SKETCH

Matt Bierbaum was born in Maryland, where he attended high school, played in numerous musical ensembles, and participated in swimming, soccer, and basketball. In 2005 he moved to Chicago and graduated from Northwestern University in 2009 with a Bachelor of Arts in Integrated Science and Physics. In 2012 he received his Masters of Science in Physics from Cornell University under his thesis advisor James Sethna.

Aside from science, Matt actively participates in cycling of many varieties including road and mountain biking with his 5 bicycles, Papa Smurf, The Wiz, Stump Thumper, The King, and Bike. Additionally, he plays classical piano in his free time, particularly Chopin and Debussy. Occasionally he directs and edits short films, mainly for the Cornell physics holiday party. He has built numerous websites that interactively demonstrate topics of his research. These sites have garnered extensive attention with unique visitors reaching into the several millions.

To my family.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The description of the world that physics provides is surprisingly simple and yet it admits a wide variety of wonderfully rich behaviors (see Fig. 1.1). While we are taught in school that there are only four fundamental forces (up to debate with our high energy colleagues), it would be extremely intractible to describe the complex swirling of turbulent fluids using only the strong nuclear force and quantum electrodynamics. Just as Newtonian gravity is an effective description of general relativity, in condensed matter physics we often turn to *effective descriptions* of the world to try to understand complex phenomena. In the case of turbulence in fluids, we use continuum conservation laws to describe all of the underlying microscopic degrees of freedom. Even using this simplified theory we are able to recover all macroscopic fluid properties and behaviors. Furthermore, by employing these effective descriptions we actually gain generality and perspective into many different areas of physics since most details in the system are typically unimportant.

In this thesis we investigate the properties of defects, a general class of such effective descriptions which often dominate many material behaviors, and emergent phenomena which result from microscopic rules. Using the properties of defects, one can simplify many degrees of freedom into only a few which capture the underlying complex physics. For example, the turbulence in fluids can be further simplified into the interaction of vortices which has a set of simple rules for their equations of motion. In the case of dislocations in metals, a large volume of atoms can be described by a one dimensional line that has a line tension and a long range strain field associated with it. We can then describe the deformation of

the material using only the interactions between dislocations in the crystal. In this way, defects prove extremely useful and important in studying many types of phenomena.
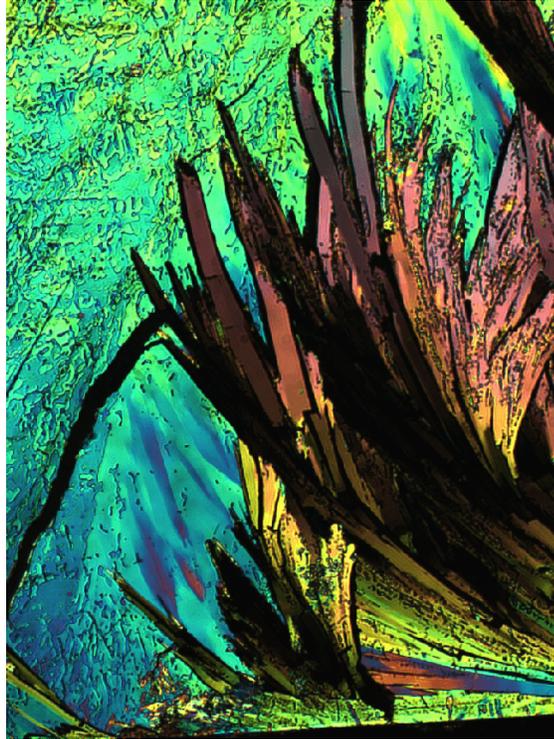
Figure 1.1: **Crystallized 4-(heptyloxy)benzoic acid viewed through cross-polarizers**. Even within a single sample, the crystal structure of this solidified mesogenic material (one which has liquid crystal phases) displays many different forms. The colors here illustrate the local orientation of the molecules, which are related to the crystal orientation and amount of local deformation. While we will not describe images of this type in this thesis, it provides a example of the rich behaviors afforded by defects in materials. Image provided by Brian D. Leahy.

Here we describe many types of defects and collective phenomena in both materials as well as human systems. We measure their detailed properties, how they collectively move, and the properties that they imbue in their material. At the smallest scales we investigate how effective descriptions of quantum mechanics, interatomic potentials, can be used to simulate different types of defect and material properties using the OpenKIM project. By comparing the results of simulation

to reference data created with DFT or measured experimentally, we can choose which effective potential is best able to describe particular material properties. Experimentally, we measure these same properties in colloids, looking at the nonlinear stress fields around vacancies, dislocations, and grain boundaries in colloidal crystals. The details of these stress fields determine basic behaviors of defect interaction such as the formation of voids from vacancy aggregation.

At larger scales, we investigate the collective motion of dislocations by studying the formation of complex fractal structures in metals as they are plastically deformed. We apply a similar technique to study how strange geometric structures, focal conic defects, behave in smectic liquid crystals. At human scales we study how gas-like and vortex-like defects in crowds at metal concerts (mosh and circle pits respectively) form spontaneously from local interactions between 'intelligent' agents. Finally, we describe the spread of the disease zombism across the United States and study the critical nature of the epidemic transition on a lattice. Along the way, we improve techniques of particle featuring in confocal microscopes to the theoretic maximum precision of 1 nm.

CHAPTER 2

LOCATING COLLOIDAL SPHERES AT THE MAXIMUM

THEORETICAL RESOLUTION[1]

2.1 Introduction

We increase the resolution of standard confocal microscopy to the maximum theoretic precision by the Parameter Extraction from Reconstruction of Images (PERI). In this technique we model the physics of image formation, describing all of the complexities present in typical confocal microscopes, in order to extract all available information. We provide a summary of the technique and generative model in the main manuscript and give extensive details and a software implementation in the supplementary information.

Super-resolution microscopy techniques have pushed the limits of optical microscopy well below the diffraction limit, enabling unprecedented views into processes that occur in biological systems. These advances were made experimentally by isolating fluorescent point sources through detailed control of emission from fluorophores in space (STED, RESOLFT [109, 108]) or by fluctuating in time (STORM, PALM [214, 31, 111]) to make their diffracted emitted light non-interacting [221]. From these images, computational techniques are used to extract the positions of point sources at the information limit as determined by the Cramér-Rao bound (CRB) [236]. This combination of techniques has allowed detailed measurement of the stepping behavior of myosin motors and Ising correlation function of lipids in cell membranes. In this work, we focus on the computational side of feature extraction from more traditional microscope images. Without modification

---

[1]Matthew Bierbaum, Brian D. Leahy, Alexander A. Alemi, Itai Cohen, & James P. Sethna
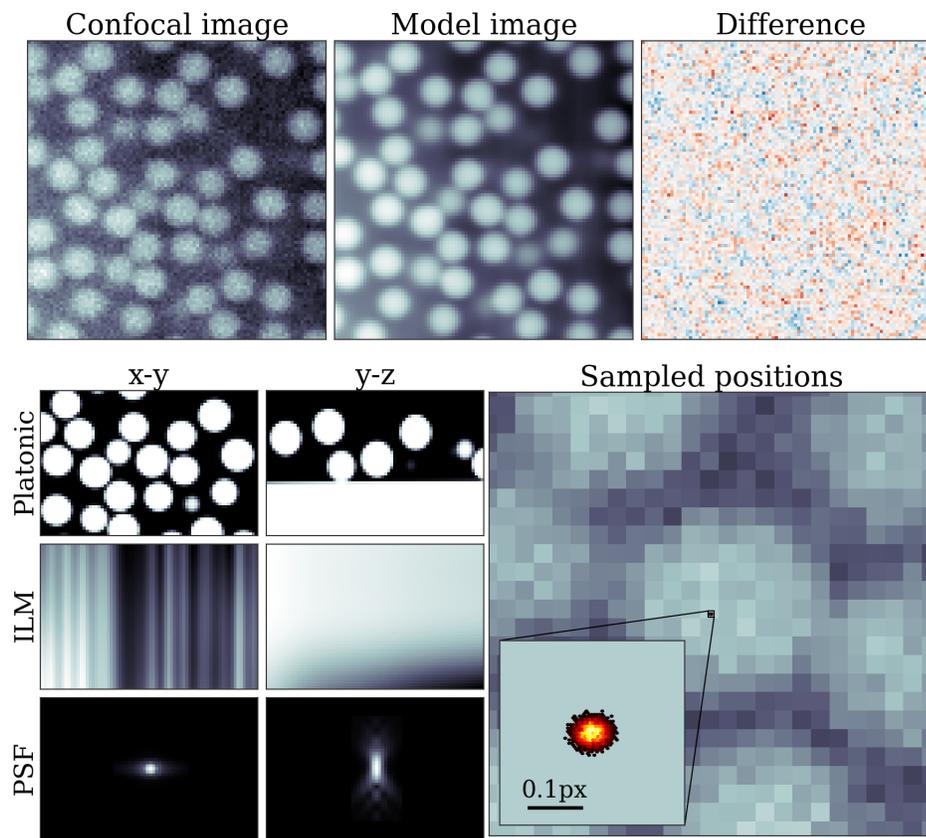
Figure 2.1: **PERI overview** – A demonstration of model information recovered from real confocal microscope images of $\langle a \rangle = 1.343(8)$ μm colloidal spheres at a volume fraction of $\phi = 0.130(5)$. On the top row, we compare (left) the true microscope image including CCD noise (middle) the reconstructed model image (right) the difference between the true image and model image. Notice that most of the structure in the difference image is uncorrelated white noise. In the lower left panels, we show reconstructed global parameters of the image along two different slices of the 3D confocal image, perpendicular to the scanning direction and including the scanning direction. We show the platonic (perfect) spheres as well as the coverslip (top), the spatially varying illumination field produced by the confocal scanning laser (middle), and the fitted point spread function of the microscope (bottom). Finally, in bottom right we show a histogram of x-y positions sampled from a single particle displaying a variance of 2 nm in each coordinate.

to standard confocal microscopes, we extend the accuracy of position and structure determination in dense, complex confocal microscope images by $10 - 100\times$, down to ~ 1 nm resolution by detailed reconstruction of the entire image. Our technique provides accurate estimations of errors associated with these measurements and can be used to characterize microscope details such as the point spread function, dirt on the lens, and correlated noise.

In this work, we apply our technique to colloidal spheres, a simple but rich system that has proven to be a useful experimental tool in understanding many diverse phenomena including glassy dynamics, microfluidics, and microrheology. In each case, the experiment is concerned with finding particle positions (and potentially radii) and tracking them in time. Current featuring methods are based on heuristics which are designed to quickly find sub-pixel positions so that the proper statistics can be acquired [62, 193, 29, 17]. Currently, these techniques achieve an accuracy around 30 nm (0.2 px) and have rudimentary error estimates. However, much of the interesting physics in colloidal suspensions – from interparticle interactions to packing questions in glasses and shear-thickening suspensions – occurs at separations smaller than this length scale. As such, locating colloidal particles more accurately seems of critical importance.

How precisely can a particle be located in an image? The fundamental limitation in imaging resolution arises from noise in the image formation, and not diffraction or optical limitations [205]. The fundamental limit on localization accuracy in the presence of noise is known as the Cramer-Rao bound (CRB). The CBR states that the covariance matrix of the estimated parameters $\boldsymbol{\theta}$ is always larger than the inverse of the Fisher information matrix: $\mathrm{cov}\,\theta_{ij} \geq (I^{-1})_{ij}$, where $I_{ij} \equiv -\langle \partial_i \partial_j \log \rho \rangle$ depends on derivatives of the noise distribution ρ. For a deter-

ministic image $\mathcal{I}$ given by parameters $\boldsymbol{\theta}$ and sampled at points $\mathbf{x}_k$ with signal-to-noise ratio SNR, this bound simplifies to

$$\operatorname{cov} \theta_{ij} \geq \frac{1}{\mathrm{SNR}^2} \left(\sum_k \frac{\partial \mathcal{I}(\mathbf{x}_k)}{\partial \theta_i} \frac{\partial \mathcal{I}(\mathbf{x}_k)}{\partial \theta_j} \right)^{-1}. \tag{2.1}$$

We can use Eq. 2.1 to estimate the minimum uncertainty of a colloidal particle's position and radius. If the particle has a radius R and is blurred by diffraction by an amount w, in pixels, the radial derivative $\partial \mathcal{I}/\partial R$ will affect approximately $4\pi R^2 w$ pixels, and the positional derivative will affect $\sim \pi R^2 w$ pixels. As each of these pixels will change by an amount $\sim 1/w$, this gives a minimum uncertainty in a particle's radii as $\sigma_R \sim \sqrt{4\pi R^2/w}/\mathrm{SNR}$, and in its position as $\sigma_x \sim \sqrt{\pi R^2/w}/\mathrm{SNR}$. For a typical colloidal particle of diameter 1 μm, imaged with a confocal microscope with pixel size of 100 nm and diffractive blur of 200 nm at an SNR = 20, these correspond to about $\sigma_R \sim 1$ nm and $\sigma_x \sim 1$ nm, a fantastically high resolution compared to current reconstruction errors.

2.2 Generative Model

To achieve this precise localization, we create a generative model of the microscope image based on the physics of the light interacting with the sample and the optical train, and we then fit every parameter in the model. This model can be simplified to a small list of knowledge about a microscope and the sample inside: (1) the colloidal particles are spheres, (2) the suspension fluid is illuminated unevenly by the laser, (3) the resultant image is blurred due to diffraction, and (4) the final image is noisy.

Platonic image: To reconstruct the image, we start with the continuous distribution of the fluorescent dye in the sample, which we call the Platonic image. For the samples we analyze in this work, the Platonic image is a collection of perfect spheres or spherical lacunae representing the colloidal particles. To represent the spheres in the continuous Platonic image on a pixelated grid, we implement these spheres using a real-space function that is tuned to match the exact Fourier representation of a sphere, parameterized by their positions and radii (see SI for an extensive, detailed discussion of this and the rest of the generative model). While we focus on featuring only spheres in this work, in order to reach the CRB, we must include the microscope slide in our description of the Platonic image. In general, PERI is flexible enough to include any parameterizable object in the generative model, such as ellipsoidal, rod-like, and dimerized particles.

Illumination field and background: The 3D images formed in confocal microscopes are formed through an illumination and detection process. However, due to imperfections and dirt in the optics, this illumination is not uniform but instead can vary strongly in space. For instance, for our line-scanning confocal, the illumination field (ILM) is highly anisotropic with large correlations in intensity perpendicular to the scan direction caused by disorder in the optics train being dragged across the field of view. We describe the ILM as a Legendre polynomial in the direction perpendicular to the coverslip and a Barnes interpolant in the plane of the coverslip to account for the small wavelength features of our line-scan confocal. Additionally, even when not receiving input light, the microscope registers a non-zero signal. We parameterize this background signal (BKG) similarly to the ILM and add it to our model as well.

Point spread function: Due to diffraction, a single dye molecule in the sample projects a comparatively large blur, known as the point-spread function (PSF) onto the imaging camera. As a result, the image captured on the camera is not an exact representation of the illuminated platonic image, but instead is that image convolved with the PSF. While complicated, this PSF has been calculated exactly by many researchers for different geometries [106, 258, 271, 181]. For microscope samples which have a different index of refraction than the optical train's design, the PSF worsens with depth, becoming both significantly broader and more aberrated. We use a modification of these exact PSF calculations, adapted for our line-scanning confocal, as our PSF model, optimizing over parameters such as the acceptance angle or numerical aperture of the lens and the index mismatch of the sample to the optics.

Putting these three processes together, the model image can be written as

$$\mathcal{M}(\mathbf{x}) = B(\mathbf{x}) + \int d^3\mathbf{x}' \; [I(\mathbf{x}')(1 - \Pi(\mathbf{x}'))]P(\mathbf{x} - \mathbf{x}'; \mathbf{x}) \tag{2.2}$$

where I is the ILM, B is the BKG, Π is the platonic image, and P is the PSF. Each pixel value is found by sampling this function at points $M_i = \mathcal{M}(\mathbf{x}_i)$ (see SI for a discussion of pixel sampling).

Noise: Finally, the image recorded on the camera is degraded by noise, either from the finite number of photons collected or from simple electronic noise in the camera. We accurately incorporate this noise with a Bayesian view of the generative model, forming the likelihood of our model given the experimental image. As the noise is empirically Gaussian (see SI), the log-likelihood can be written as the sum of the squared differences between the image and model normalized by the

estimated noise σ,

$$\log \mathcal{L}(\{M_i\} \mid \{\theta\}, \{d\}) = -\sum_i \frac{(M_i(\theta) - d_i)^2}{2\sigma_i^2} \tag{2.3}$$

where the sum over i now indicates the sum over all pixel values in the image. It is important to note that this sum is over all of the pixels in the image – to get a meaningful extraction of parameters, every pixel must be described just as well as the others. While this makes it easy to identify mis-featured regions, it also makes the optimization much more difficult – dimerized particles and PSF leakage from particles outside the image can affect the entire image reconstruction through their influence on global variables.

A typical microscope image contains a few times 10^4 particles, each with 4 fit parameters (x, y, z, R). In addition, there are a few dozen global parameters to optimize, such as the illumination and PSF parameters and the z-step size z_{scale} of the lens, resulting in $\approx 10^5$ parameters per image. To optimize these parameters, we begin by using standard particle locating techniques on the experimental image [62] and minimize the log-likelihood for this set of particles and the global variables. From here, we ensure that we have identified every particle in the image by adding and subtracting particles probabilistically based on the log-likelihood. For performance, there are many algorithmic optimizations made in the computation which are detailed in the SI. After optimization, we sample from the log-likelihood using standard MC techniques, primarily slice sampling [182] to estimate the errors associated with the image reconstruction.
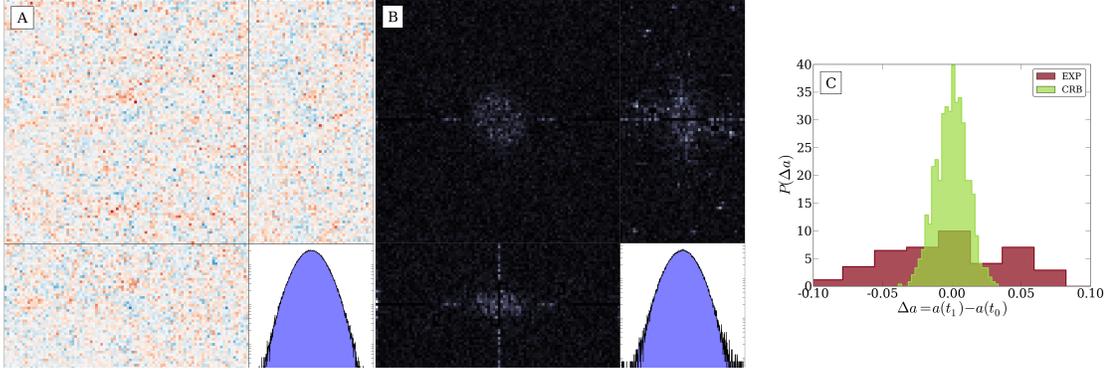
Figure 2.2: **Fitting the generative model to experimental data** – We experimentally verify our featuring algorithm by looking at the residuals in real and Fourier domains as well as comparing the radii of tracked particles between two different frames of a movie. (A) We plot three cross sections of the real space residuals showing nearly Gaussian white noise. The 'shadows' of spheres can be seen due to complications of the PSF. However, the probability distribution of all residuals (lower right) is very nearly Gaussian as shown as a quadratic on a logarithmic scale. (B) The Fourier power spectrum of the same residuals displayed along q_x, q_y, or $q_z = 0$. In all panels, excess power is visible at scales larger than the particles themselves but smaller than the features given by the ILM. These residuals are associated with the approximations we have made in the point spread function, particularly the difficulty in calculating the long tails of the PSF and the cutoff we employ to speed up numerical computations. The q-space histogram is also very nearly Gaussian with slight deviation from quadratic in the tails. (C) We plot the difference in radii across frames (red) as compared to the difference that we estimate using the radii CRB (green). The difference in peak high is proportional to the distance our experimental measurements are from the theoretic limit, roughly $3\times$ larger.

2.3   Experimental data

To verify the PERI method on experimental data, we prepare a sample of $\sim 1.3~\mu$m silica spheres suspended in a water and glycerol mixture. We take several images of the particles as a function of time as the particles diffuse across the field of view. We then feature these images using PERI with a model given by a line-scan confocal PSF, Barnes ILM, Legendre polynomial BKG, and initial particles

positions given by Crocker-Grier (CG) [63]. Optimization was first performed using Levenberg-Marquardt and then all parameters were sampled to obtain estimates of the fit covariance matrix. Each frame was analyzed independently in order to verify that convergence is robust across images.

First, we analyze the residuals of our fits to the experimental data. In Fig. 2.2(a,b) we show these residuals in both real- and Fourier-space. While the overall probability distribution is nearly Gaussian in both domains, we are able to still see slight hints of particles in the real-space difference image. In Fourier-space, there are distinct wave vectors which are above the noise floor which represent excess signal at length scales larger than the particle but smaller than typical ILM features. We believe that these features arise from our approximate line-scanning point spread function, excess aberrations in the microscope, and the artificial cutoff value we introduce in our PSF calculation that we use to speed up optimization. Additionally, scanning noise can be seen in one slice of the Fourier-space residuals with sharp peaks present at high q values. Overall, the remaining unexplained signal is very small, comprising of roughly 10^{-3} of the power in the model residuals. The remaining question is how much these residuals affect the parameters of interest, the particle positions and radii.

Since we are able to extract radii with PERI, we can directly compare positions and radii of particles from frame to frame. Physically, the particle radii do not fluctuate in time and particles cannot overlap, providing us with two tangible measurements that we can use to verify our method. In Fig. 2.2(c), we see that the variation of radii between frames is roughly 3× larger than that predicted by the CRB. This corresponds to a radius variation of 2 nm between consecutive images in the movie. Additionally, we find that the number of overlaps is greatly reduced

by using PERI compared to conventional methods. In the sample presented in Fig. 2.1 and Fig. 2.2 we find an overlap only every few frames, which is consistent with the radii fluctuations discussed earlier.

With this increase in resolution, we compare several standard measurements employed in colloidal science as measured by both CG and PERI. First, we measure the radial pair correlation function $g(r)$, the density-normalized probability of two particles in a sample being separated by a distance r. We prepare a sample of silica spheres at a volume fraction of $\phi \approx 0.45$ and compute the pair correlation as shown in the top panel of Fig. 2.3. Focusing on the first peak, we can see that there is roughly a factor of three increase in practical resolution given by the relative heights of the peak at $g(1)$ as measured by CG and PERI. In particular, the values of $g(r)$ that occur in the region $r < 1$ are mainly attributed to featuring uncertainty and represent overlaps in the sample. However, part of the spread in the first peak of $g(r)$ is due to polydispersity, the fact that colloids are synthesized at slightly different sizes. We attribute most of the weight of $g(r)$ for $r < 1$ in the PERI measurement to polydispersity.

In the bottom panel of Fig. 2.3 we show a second common experimental sample characterization, the mean-squared-displacement (MSD) which gives ensemble information about caging, interactions, and other deviations from Brownian motion. Here, we see the CG measurement diverges from the linear prediction given by simple Brownian motion at 2 sec and 50 nm. Featuring with PERI, we are able to utilize the full time resolution of the confocal microscope, following the linear prediction down to 10 nm at 100ms, the maximum scan rate of our line-scanning confocal.

Finally, even seemingly simple volume fraction measurements are made precise
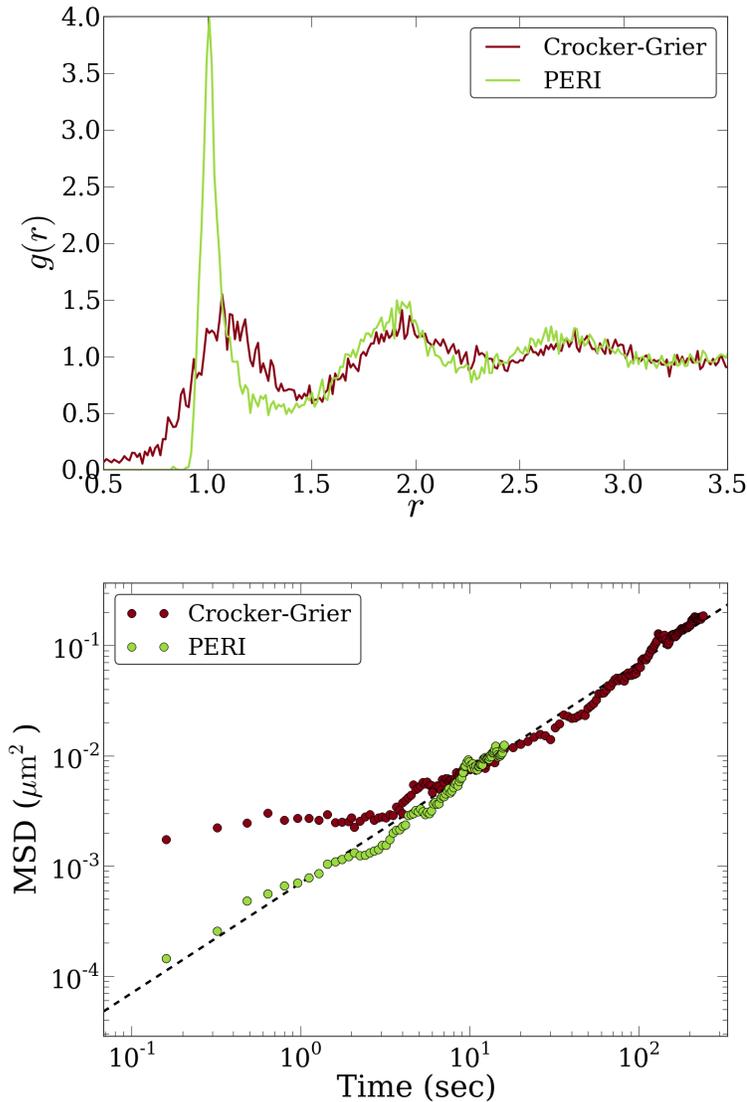
Figure 2.3: **Bulk characterization comparison of PERI and CG** – We compare the pair correlation (top panel) and mean squared displacement (bottom panel) as measured by both PERI (green) and Crocker-Grier (red) featuring algorithms. In both probes we can see evidence of a marked improvement in resolution by fully reconstructing the experimental images. In the pair correlation function $g(r)$, the number of overlaps (values of $r < 0$) is significantly diminished – most overlaps present in the PERI measurement are accounted for by polydispersity (3.2%). In the MSD, we see a departure from a linear relation (dashed line) in the CG measurements (red) at approximately 50 nm whereas the trend continues to the time resolution of the microscope with PERI (green).

by PERI due to the need for complete image reconstruction. For a dense colloidal crystal, we measure the volume fraction $\phi = 0.6734(3)$ for a single frame with PERI, where the volume fraction takes into account all particles within the field of view, including edge particles. Conventional methods often miss upwards of 10% of particles, measuring $\phi = 0.579$ for this same sample. With conventional methods, PSF overlap between particles causes some close neighbors to be skipped and most edge particles are ignored due to low total mass.

2.4 Conclusions

In this paper we present a method for extracting nearly all information from confocal images of colloidal spheres, leading to precise localization and size measurements, down to the resolution of 1 nm. We verify that these measurements are correct for experimental images by comparing particle radii between adjacent frames of a movie, ensuring that their physical sizes do not fluctuate much larger than the theoretical minimal estimate. We show pronounced improvement in several standard measurements including the pair correlation function $g(r)$, mean-squared-displacement, and the volume fraction.

This method of recovering parameters using generative models is quite generic and can be applied to many other types of microscope images. Currently we are extending PERI to work for common 2D brightfield images [75] with promising initial success. Additionally, we have recently used similar ideas to extract currents from magnetic flux images measured by scanning SQUID microscopy [185], improving current localization two fold. Finally, we are beginning to expand PERI to scanning transmission electron microscope (STEM) images. Here, while the objects

of interest are much smaller in pixel extent (dramatically increasing the CRB), the role of aberrations becomes much more important. If we can directly model these aberrations and include them in our featuring then many types of systematic biases can be reduced or eliminated from the extracted atomic positions.

2.5   Supplemental information

2.5.1   Overview

In this supplemental material we describe the details of our method for extracting parameters from experimental confocal images at the highest resolution possible without modifying the microscope itself. To achieve maximal resolution, we build a generative model which aims to describe the value of every pixel in the experimental image. That is, we create simulated images by explicitly modeling every relevant aspect of image formation including particle positions and sizes, the location of dirt in the optics, amount of spherical aberration in the lens, and the functional form of the point spread function. We describe each of these model components in detail in Section 2.5.3 and how we decided on these particular components in Section 2.5.4. In order to fit this model to the experiment, we adjust all model parameters until the features present in the true experimental image are duplicated in the simulated one. We decide when the fit is complete and create a fair sample of the underlying parameters by using a traditional Bayesian framework which is described in general terms in Section 2.5.2. This high dimensional optimization is in general very difficult and so we describe our algorithmic improvements and particular techniques in Section 2.5.5. Finally, we assess the accuracy of this

method in extracting underlying parameters and compare its performance with traditional featuring methods in Section 2.5.6.

Overall, this document is meant to provide a roadmap for other researchers to follow when adapting this technique to other types of microscopy and other types of samples in order to extract the maximal amount of information from their experimental images.

2.5.2   Bayesian framework

When fitting a model to noisy data, it is useful to adopt a Bayesian framework in which we rigorously treat the noise as part of our model. In the case of our featuring method, we fit a model of each image pixel M_i to experimental data d_i, which can be described as a combination of signal and noise $d_i = S_i + \eta_i$. This noise is present due to the detection of a finite number of photons by the microscope sensor, noise in the electronics, etc. and can be well described for our system by uncorrelated $\langle \eta_i \eta_j \rangle = 2\sigma^2 \delta_{ij}$, Gaussian noise $\eta_i \sim \mathcal{N}(0, \sigma)$ (see Section 2.5.3).

In a Bayesian framework, the likelihood that an individual pixel is correctly described by our model is given by the Gaussian likelihood,

$$\mathcal{L}(M_i \mid d_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-(M_i - d_i)/(2\sigma_i^2)} \tag{2.4}$$

For uncorrelated pixel noise, the entire likelihood of the model given the image is given by the product over all pixels, $\mathcal{L}(\mathbf{M} \mid \mathbf{d}) = \prod_i \mathcal{L}(M_i \mid d_i)$. We are ultimately interested in the probability of the underly parameters given the image we record.

According to Bayes' theorem, we can write this as

$$P(\boldsymbol{\theta} \,|\, \mathbf{d}) \propto P(\mathbf{d} \,|\, \boldsymbol{\theta})P(\boldsymbol{\theta})$$

$$\propto \mathcal{L}(\mathbf{M}(\boldsymbol{\theta}) \,|\, \mathbf{d})P(\boldsymbol{\theta})$$

where $P(\boldsymbol{\theta})$ are priors that allow us to incorporate extra information about the parameters $\boldsymbol{\theta}$. These priors can be as simple as the fact that the particle radius is positive definite or that a group of images share similar PSFs. For example, an overlap prior $P_{\text{overlap}}(\mathbf{x}_i, \mathbf{x}_j, a_i, a_j) = \text{H}(a_i + a_j - |\mathbf{x}_i - \mathbf{x}_j|)$, where H is the Heaviside step function, can be used to impose the physical constraint that particles cannot overlap. However, we found that the overlap prior only becomes relevant when the free volume of a particle is small compared to the average sampling error volume (when a particle is caged by ~ 1 nm on all sides) and so we ignore it most of the time.

We primarily work with the log-likelihood function $\log \mathcal{L}$ because the number of pixels in the image can be very large, on the order 10^7. For Gaussian noise, the log-likelihood for Gaussian noise is precisely the square of the L_2 norm between the model and the data. Therefore, we are able to maximize this log-likelihood using a variety of standard routines including linear least squares and a variety of Monte-Carlo (MC) sampling techniques. After optimizing, we use MC algorithms to sample from the posterior probability distribution to extract full distributions of the model parameters. In this way, any quantity of interest that is a function of particle distribution can be calculated using MC integration by

$$\langle \mathcal{O}(\boldsymbol{\theta}) \rangle = \int \mathcal{O}(\boldsymbol{\theta}) P(\boldsymbol{\theta} \,|\, \mathbf{d}) \, \mathrm{d}\boldsymbol{\theta}$$

$$= \frac{1}{N} \sum_i^N \mathcal{O}(\boldsymbol{\theta}^i)$$

Here, $\boldsymbol{\theta}^i$ is a parameter vector sampled fairly from the posterior probability

distribution and $\mathcal{O}(\boldsymbol{\theta})$ is an observable such as the pair correlation function, packing fraction, or mean squared displacement. Higher order moments can also be calculated to estimate errors for these observables by evaluating $\sqrt{\langle\mathcal{O}^2\rangle - \langle\mathcal{O}\rangle^2}$. This is one of the more powerful aspects of this method – one can generate a probability distribution for each parameter and directly apply these distributions to any observable that can be inferred from the parameters.

Given this Bayesian framework, the main idea of this work is to create a full generative model for confocal images of spherical particles and provide algorithmic insights in order to implement the model on commodity computer hardware.

2.5.3   Generative model

Most of the difficulty in our method lies in creating a generative model that accurately reproduces each pixel in an experimental image using the fewest number of parameters possible. Our model is an effective description of how light interacts with both the sample and the microscope optics to create a distribution of light intensity as measured by the microscope sensor and rendered as an image on the computer. In this section we describe the model which we use to create images similar to those generated by line-scanning confocal microscopy of spherical particles suspended in a fluorescent fluid.

Our generative model aims to be an accurate physical description of the microscope imaging; it is not a heuristic. Creating this model requires a detailed understanding of image formation of colloidal spheres in a confocal microscope. In the simplest view, our samples consist of a continuous distribution of dye distributed throughout the image. If the fluid is dyed (as for the images in this work),

due to diffusion the dye is uniformly distributed through the fluid. The fluid-free regions, such as those occupied by the particles, are perfectly dye-free. The sample is illuminated with a laser focused through an objective lens. This focused laser excites the fluorescent dye only in the immediate vicinity of the lens's focus. An objective lens captures the dye's emitted light, focusing it through a pinhole to further reject out-of-focus light. The collected light passes through a long-pass or band-pass filter, which eliminates spurious reflected laser light before collection by a detector. This process produces an image of the sample at the focal point of the lens. Finally, rastering this focal region over the sample produces a three-dimensional image of the sample.

However, the actual image formation is more complex than the simple view outlined above. Excessive laser illumination can cause the dye to photobleach. Due to dirt and disorder in the optical train, the sample is not illuminated uniformly. Diffraction prevents the laser light from being focused to a perfect point and prevents the objective lens and pinhole from collecting light from a single point in the sample. Moreover, additional aberration is present if the sample's refractive index is not matched to the design of the objective lens, broadening the diffractive blur deeper into the sample. Both the illuminating and fluorescing light can scatter off refractive index heterogeneities in the sample due to the particles.

Some of these complications can be eliminated by careful sample preparation. In practice, we eliminate photobleaching by using an excessive amount of dye in our samples and illuminating with a weak laser light. We eliminate scattering by matching the refractive index of the particles to the suspending fluid. It is fairly easy to match the refractive indices to a few parts in 10^3. Since the scattering is quadratic in the index mismatch, the effect of turbidity due to multiple-scattering

is very weak in our samples. However, the rest of these complications must be accurately described by the generative model.

Based on this physical setup, we can describe the confocal images through three main generative model components:

- *Platonic image* $\Pi(\mathbf{x})$ – the physical shape of the dye distribution in the sample (unmodified by perception of light).

- *Illumination field* $I(\mathbf{x})$ – the light intensity as a function of position, including both laser intensity variation from disorder in the optics and intensity attenuation into the sample.

- *Point spread function* $P(\mathbf{x}; \mathbf{x}')$ – the image of a point particle due to diffraction of light, including effects from index mismatch and finite pinhole diameter.

plus three minor additional fit model components:

- *Image Background* c, $B(\mathbf{x})$ – the overall exposure of the image c and the background values corresponding to a blank image without dye, B.

- *Rastering Step Size* z_{scale} – the displacement distance of the lens as it rasters along the optical axis.

- *Sensor noise* σ – the noise due to shot noise from finite light intensity reaching the sensor or electronic noise at the sensor.

These components are combined with a fitted scale factor c to form the image through convolution

$$\mathcal{M}(\mathbf{x}) = B(\mathbf{x}) + \int \mathrm{d}^3 x' \ [I(\mathbf{x}')(1 - \Pi(\mathbf{x}')) + c\Pi(\mathbf{x}')]P(\mathbf{x} - \mathbf{x}'; \mathbf{x}) \qquad (2.5)$$

which is then sampled at discrete pixel locations to give the final image $M_i = \mathcal{M}(\mathbf{x}_i)$.

Here, we describe each part of our model in detail along with our explanations and motivations behind any simplifications. In subsequent sections we will also discuss other aspects of image formation which may result in other model choices and why we omit them from the final form of the model.

Platonic image

The Platonic image must accurately represent the continuous distribution of fluorescent dye in the sample on the finite, pixelated image domain. The colloidal sample consists of a collection of spherical particles embedded in the solvent, with either only the particles or only the solvent dyed. Our Platonic image should then consist of the union of images of individual spherical particles, with their corresponding radii and positions. Thus, if we have a method to accurately represent one colloidal sphere, we can easily construct the Platonic image in our generative model.

A naïve way to generate the Platonic image of one sphere would be simply to sample the dye distributions at the different pixel locations, with each pixel being either 0 (if it is outside the sphere) or 1 (if it is inside the sphere) with no aliasing. This method will not work, since a pixel value in the Platonic image can only change when a sphere's position or radii has shifted by one pixel. This method of Platonic image formation would produce a generative model that does not adequately distinguish between particle locations separated by less than 1 pixel or 100 nm! Simply multiplying the resolution and corresponding coarse-graining of the boolean cut by a factor of N in each dimension increases the resolution of this

method to $1/N$ pixels. However, calculating these high resolution platonic spheres is computationally expensive, requiring 10^9 operations to draw spheres capable of determining positions within 0.01 px. While we will use these images as ground truth to verify our approximate platonic formation, we consider an alternative image formation path to arrive at a fast sphere generation method.

To find the correct representation of a Platonic sphere, we examine the mechanism of image formation in Eq. 2.5. The final image results from a convolution of the Platonic image with the point-spread function $P(\mathbf{x} - \mathbf{x}'; \mathbf{x})$. Thus, we need a representation of a sphere that will produce the correct image after being convolved with the point-spread function. To do this, we recall that a convolution is a multiplication in Fourier space. However, creating the image in Fourier space is problematic since, due to the truncation from the finite number of pixels in the image there will be undesirable ringing (i.e. Gibbs phenomenon) in the Platonic image. Moreover, each update of one particle requires updating all the pixels in the image, which is exceedingly slow for large images.

Instead, we look for a functional form in real space that approximates the numerically exact truncated Fourier series given by $\tilde{\Pi}(\mathbf{k}; \mathbf{p}, a) = 4\pi a^3 (j_1(k)/k)e^{i\mathbf{k}\cdot\mathbf{p}}$. We can view the truncation operation as a multiplication in Fourier space by a boxcar $H(1 - |q_x|)H(1 - |q_y|)H(1 - |q_z|)$, where $\mathbf{q}$ is the variable inverse to position, measured in px^{-1}. By the convolution theorem, this truncation corresponds to a convolution in real space with $\text{sinc}(x)\,\text{sinc}(y)\,\text{sinc}(z)$, using the inverse Fourier transform of the boxcar as the sinc function. However, the convolution with the sinc function is analytically intractable. To circumvent this, we approximate the sinc function by a Gaussian. This gives a representation of the correctly-aliased

Platonic image $\Pi(\mathbf{x}; \mathbf{a})$ of a sphere of radius a as

$$\Pi(\mathbf{x}) = \mathcal{S}(\mathbf{x}) * \left[\left(2\pi\sigma_x^2\sigma_y^2\sigma_z^2 \right)^{-1/2} e^{-x^2/2\sigma_x^2} e^{-y^2/2\sigma_y^2} e^{-z^2/2\sigma_z^2} \right] \qquad (2.6)$$

where $\mathcal{S}(\mathbf{x}; \mathbf{p}, a) = H(|\mathbf{x} - \mathbf{p}| - a)$ where $H(x)$ is the Heaviside step function, which is either 0 or 1 depending on whether $|\mathbf{x} - \mathbf{p}| > a$ or $< a$, and $*$ denotes convolution. The Gaussian widths σ should be approximately 1 px; however for $z_{\text{scale}} \neq 1$ σ_z will not be the same as σ_x and σ_y.

While Eq. 2.6 does not generally admit a simple solution, there is a closed-form functional form for the symmetric case $\sigma_x = \sigma_y = \sigma_z$. In the symmetric case ($z_{\text{scale}} = 1$) Eq. 2.6 takes the form

$$\Pi(\mathbf{x}) = \frac{1}{2} \left[\text{erf}\left(\frac{a-r}{\sigma\sqrt{2}} \right) + \text{erf}\left(\frac{a+r}{\sigma\sqrt{2}} \right) \right] - \frac{1}{\sqrt{2\pi}} \frac{\sigma}{r} \left[e^{-(r-a)^2/2\sigma^2} - e^{-(r+a)^2/2\sigma^2} \right]$$
$$(2.7)$$

where r is the distance from the particle's center. The first bracketed group of terms corresponds to treating the sphere as a flat surface, and the second bracketed group corresponds to the effects the sphere's curvature on the integral. In each sub-grouping, the first term that depends on $r - a$ reflects the contribution due to the particle's nearer edge, and the second term that depends on $r + a$ reflects the contribution due to the particle's farther edge. We then fit σ in Eq. 2.7 to best match the exact Fourier space image of a sphere, giving a value $\sigma \approx 0.276$.

Although Eq. 2.6 does not admit a simple solution for $z_{\text{scale}} \neq 1$, we can use the exact form for $z_{\text{scale}} = 1$ to construct an approximate solution. Since both $\text{erf}(x)$ and e^{-x^2} approach their asymptotic values extremely rapidly, and since at the best fit $\sigma \approx 0.276$ $(a + r)/\sigma \gg 1$ for even moderately small radii, the terms $\text{erf}((a + r)/\sigma\sqrt{2}) \approx 0.5$ and $\exp(-(r+a)^2/2\sigma^2) \approx 0$ to an excellent accuracy. Next, we write the position vector in terms of its direction $\hat{\mathbf{x}}$ and a vector $\boldsymbol{\delta}\mathbf{x}$ as $\mathbf{x} \equiv a\hat{\mathbf{x}} + \boldsymbol{\delta}\mathbf{x}$, and replace $(a-r)/\sigma$ in Equation (2.7) by $\sqrt{(\delta x/\sigma_x)^2 + (\delta y/\sigma_y)^2 + (\delta z/\sigma_z)^2}$. Note that

this approximation is exact in the limit of infinite sphere radii. Empirically, we find that this approximation works quite well, giving differences in the Platonic image of a few percent from a numerical solution to Eq. 2.6 as well as high resolution boolean cut real-space spheres (see Fig. 2.4).

While this implementation of the Platonic image correctly captures most of the effects of finite-pixel size, there are still some minor details that need to be fixed to give unbiased images. By construction, Eq. (2.7) conserves volume – its integral over all space is $4/3\pi a^3$ since the Gaussian kernel is normalized. However, when $\Pi(\mathbf{x})$ is sampled on a pixelated grid, its sum is not exactly $4/3\pi a^3$ but is slightly different, depending on the position of the particle's center relative to a voxel's center. The slight change in volume is important for two reasons. First, the convolution with the PSF in our image generation (see next subsection) suppresses high-frequency portions of the image, but it does not affect the $\mathbf{q} = \mathbf{0}$ component, *i.e.* the image sum or the particle volume. Since we aim to create a Platonic image that accurately represents the final image, we need the $\mathbf{q} = \mathbf{0}$ component of the Platonic image to be correct. Secondly, as discussed in section 2.5.4 the real microscope image is actually an integral over a finite pixel area. As such, the image recorded on the detector preserves the particle's volume or $\mathbf{q} = \mathbf{0}$ component of the image. To circumvent this issue of incorrect particle volume, instead of drawing the particle at its actual radius we draw it with a slightly different radius that preserves the particle's volume, which we accomplish with an iterative scheme. The results of this iterative scheme are shown in Fig. 2.4 along with the errors it introduces. Incidentally, the effects of image pixelation on image moments higher than $\langle 1 \rangle$, *e.g.* $\langle \mathbf{x} \rangle$ and its effects on the particle positions, are much smaller than the noise floor in our data at a moderate SNR (see section 2.5.4).
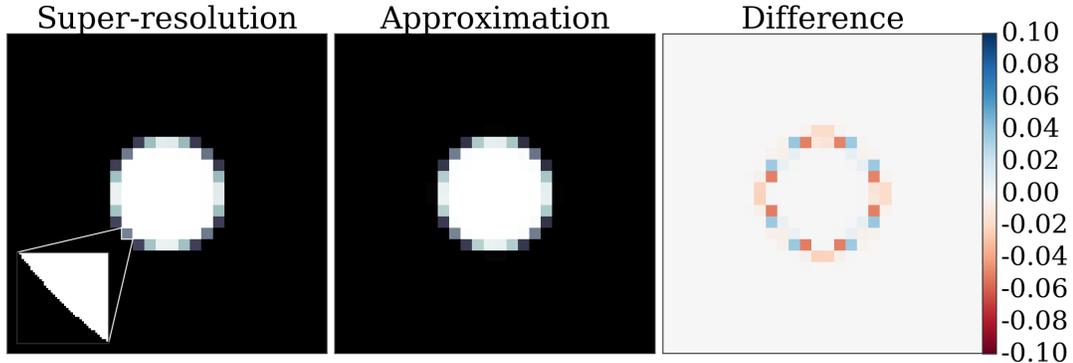
Figure 2.4: **Platonic sphere generation**. A comparison of our approximate pla-
tonic sphere generation method to a sphere created by performing a
boolean cut $\Pi(\mathbf{x}) = \int_{\text{pixel}} d\mathbf{x}' H(|\mathbf{x} - \mathbf{x}' - \mathbf{p}| - a)$ on a lattice 100×
higher in resolution in each dimension compared to the final image.
On the left we show the super resolution sphere with fractional volume
error $\delta V/V = 10^{-6}$ and an inset displaying the jagged edges caused
by discrete jumps in distance. This is in contrast to the iterative ap-
proximate platonic sphere with volume error $\delta V/V = 10^{-16}$ drawn at
an effective radius with error $\delta a/a = 5 \times 10^{-3}$. The errors of individ-
ual pixels along the center of the sphere are shown in the right panel
showing a high frequency structure with a maximal relative value 0.08.
These high frequency features are dramatically reduced later in the im-
age formation process through the convolution with the point spread
function.

The Platonic image needs to represent accurately all objects in the image, not

just the spheres. In particular, when the solvent is dyed there is almost always

a dark coverslip visible in the image. We model this dark coverslip as a slab

occupying a half-space. The slab is characterized by a z-position and by a unit

normal $\hat{\mathbf{n}}$ denoting the perpendicular to the plane. To capture accurately sub-pixel

displacements of the slab, we use the image of a slab convolved with a Gaussian

as above for a sphere; for the slab this gives a simple error (erf) function.

(3, 3, 2) (7, 5, 2) (11, 7, 2) (17, 11, 2) (35, 19, 2)
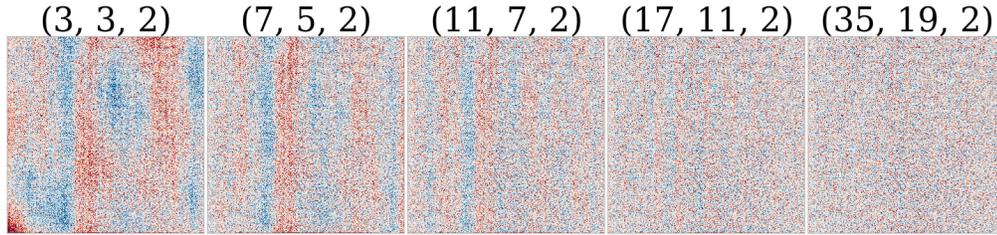
Figure 2.5: **ILM residuals**. Fits to a blank confocal image using a Legendre polynomial given by Eq. 2.9 of the order specified above the subplot where the tuple represents the number of terms in the summations (i, j, k). We see non-uniform structures in low order fits using $(3, 3, 2)$ terms continuing until there are clear stripes present at $(11, 7, 2)$ due to the line-scan nature of our confocal. Finally, most structure is gone by the high order fits to ILMs of order $(17, 11, 2)$ and $(35, 19, 2)$.

Illumination field

In order to illuminate the sample, confocal microscopes scan a laser over the field of view using several distinct patterns including point, line, and disc scanning. This illumination laser travels through the optics train and interacts with fluorescent dye in the suspension causing it to emit light in a second wavelength which is then detected. The intensity of this illumination pattern depends on the aberrations in the optics as well as dirt in the optical train which creates systematic fluctuations in illumination across the field of view. Accounting for these variations is important as they can account for as much as $\sim 20\%$ of the intensity variation in the generated image. In the case of our line scanning confocal microscope, these patterns manifest themselves as stripe patterns perpendicular to the scan direction overlaid on spherical aberrations which cause the corners of the image to dim.

We describe the variation in the illumination across the sample by decomposing

the underlying field into various basis functions. Using polynomials P_i as the basis, the illumination field can be written as

$$I(\mathbf{x}) = \sum_i \sum_j \sum_k c_{ijk} P_i(x_x) P_j(x_y) P_k(x_z) \, ; \qquad (2.8)$$

empirically Legendre polynomials work well for our optimization scheme. Since the variation of illumination in our line-scan confocal is mostly a series of stripes propagated along the scan direction, we can alternatively use a one-dimensional Barnes interpolant as part of the basis functions $P_i(x_x)$. The Barnes interpolation is a method of interpolating between unstructured data using a given weight kernel [24], similar to inverse distance weighting, using a truncated Gaussian kernel so as to allow for strictly local updates to the high frequency illumination structure.

This overall polynomial form can be simplified because the variation of light intensity is a separate function of x,y and z – the disorder present in the optics affects every layer of the image the same with an overall modulation in the z direction. We model this variation by a product of two polynomials

$$I(\mathbf{x}) = \left(\sum_i \sum_j c_{ij} P_i(x_x) P_j(x_y) \right) \left(\sum_k d_k P_k(x_z) \right) \qquad (2.9)$$

In particular, in the case of line-scanning confocal microscopes, it is necessary to use a high order polynomial in x and y to account for correlated illumination heterogeneities due to disorder being dragged across the field of view. The variation of the illumination field $I(z)$ accounts for the intensity variation usually associated with the point spread function (next section). For simplicity, we group all intensity effects into $I(\mathbf{x})$ in this work.

How well do these functional forms fit to experimental data for a line-scanning confocal microscope? We acquire blank images of a water-glycerol mixture as a function of depth and fit this data with Legendre polynomials in the form of Eq. 2.9.

As a function of the expansion order, we look at the magnitude and patterns of the residuals. In Fig. 2.5, we see large scale structure in the ILM residuals, suggesting that higher powers are necessary. Increasing the ILM order, we find stripes in x emerge due to the line-scan nature of our machine. Finally, at higher order we are able to adequately capture all illumination variation independent of depth into the sample. In practice, we do not typically use such a high order of Legendre polynomial but instead augment $P_i(x_x)$ with a Barnes interpolant to give

$$I(\mathbf{x}) = \left[\left(\sum_i \sum_j c_{ij} P_i(x_x) P_j(x_y)\right) + B(x_x)\right]\left(\sum_k d_k P_k(x_z)\right) \qquad (2.10)$$

Fitting the ILM correctly is essential for finding the correct particle positions and radii. Fig. 2.6 demonstrates the effect of featuring a real confocal image with an illumination field of insufficient order. In the left panel is an image featured with a high-degree polynomial illumination of 9th order in the x-direction and of 5th order in the y- and z- directions. While these polynomials are high-order, they are not high enough to capture all of the structure in the light illumination. There is a clear bias in the featured radii, with particle radii being systematically larger on the edge of the image and smaller in the middle. These biases arise from large stripes in the confocal illumination due to the line-scanning nature of our confocal. Using a higher-order 25th degree polynomial in the x-direction (upper right panel) eliminates the effect of these stripes, as visible in the featured particle radii plotted as a function of x in the bottom panel. Note that the particle radii may be biased by as much as 1 px or 100 nm due to effects of the spatially varying illumination field.
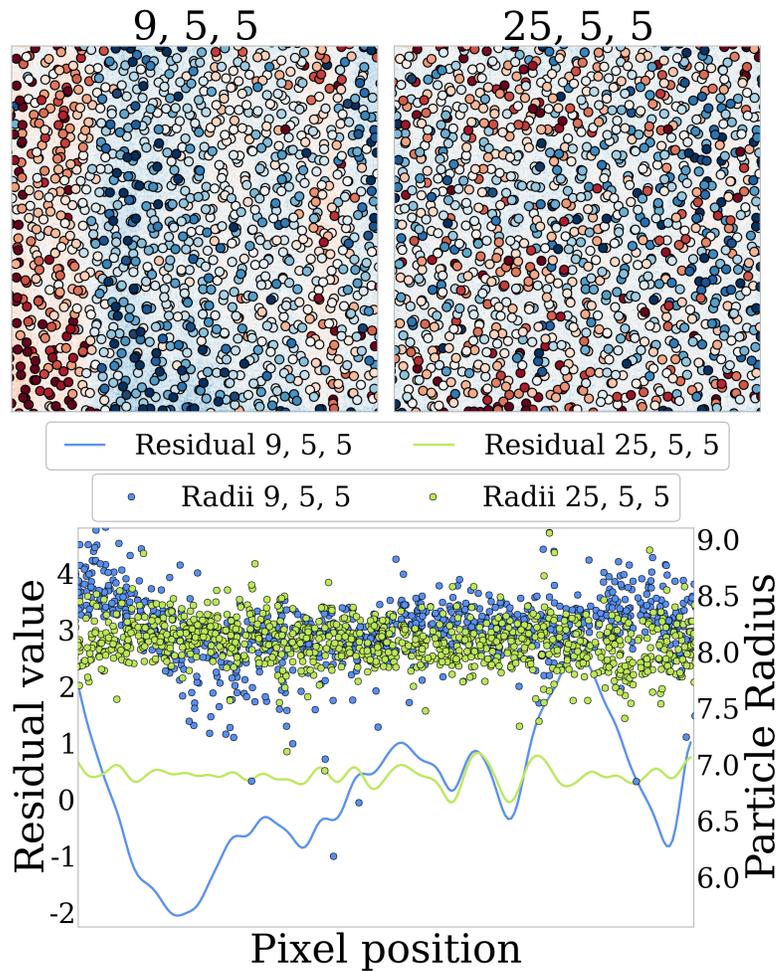
Figure 2.6: **ILM generated biases**. Using an incorrect illumination field results in significant biases. The upper left panel shows an image featured with a $(9, 5, 5)$ order polynomial in (x, y, z). In the foreground are the featured particle radii, color-coded according to their difference from the mean. In the background is the residuals of the featured image. Clear stripes are visible in both the featured radii and the residuals. The particles are systematically much larger on the left side of the image, before decreasing in size in the middle and increasing again in a small stripe on the image's right side. In contrast, when the image is featured with a higher-order $(25, 5, 5)$ degree polynomial, shown in the upper right, these systematic residuals disappear. The bottom panel shows the particle radii and image residuals for the two illumination fields as a function of the image x direction.

Point spread function

Due to diffraction, the illuminating laser light focused from the microscope's lens and the detected fluorescent light collected from the sample are not focused to a single point. Instead, the light is focused to finite-sized diffraction-limited blur. To reconstruct an image correctly we need to account for the effects of diffraction in image formation.

A confocal microscope first illuminates the sample with light focused through the microscope lens. The lens then collects the light emitted from fluorophores distributed in the sample. As a result, the final image of a point source on the detector results from two separate terms: an illumination point-spread function P_{ilm} that describes the focusing of the incoming laser light, and a detection point spread function P_{det} that describes the focused fluorescent light collected from the emitted fluorophores. Since a fluorophore is only imaged if it is both excited by the laser illumination and detected by the camera, the resulting point-spread function for a confocal with an infinitesimal pinhole is the product of the illumination and detection point-spread functions: $P(\mathbf{x}) = P_{\mathrm{ilm}}(\mathbf{x})P_{\mathrm{det}}(\mathbf{x})$. For a confocal with a finite-sized pinhole, this product becomes an convolution over the pinhole area. The two separate point-spread functions (PSFs) P_{ilm} and P_{det} can be calculated from solutions to Maxwell's equations in the lens train [106, 258, 271, 181]. The PSFs can be written as integrals over wavefronts of the propagating light.

An additional complication arises from the presence of an optical interface. Most microscope lenses are essentially "perfect" lenses, creating a perfect focus in the geometric optics limit. However, refraction through the optical interface destroys this perfect focus and creates an image with spherical aberration. In addition, the refracted rays shift the point of least confusion of the lens from its
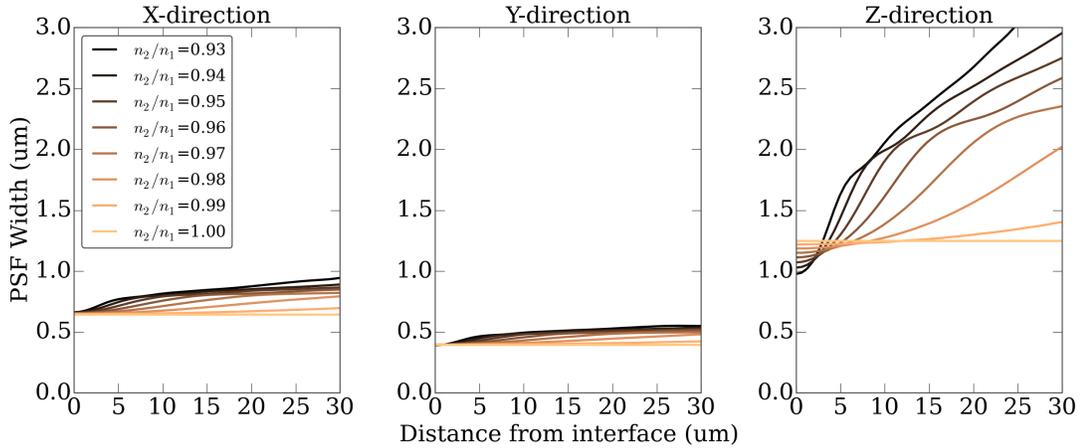
Figure 2.7: **PSF widths vs depth**. The x (left panel), y (center panel), and z (right panel) widths of the PSF as a function of distance from the interface, for various refractive index mismatches. The width of the point-spread function generally increases with depth and with index mismatch due to increased spherical aberrations. The width is broadest in the z (axial) direction, and is narrower in the y direction than along the x direction of the line illumination.

original geometric focus. For a confocal geometry, this spherical aberration and focal shift depend on the distance of the nominal focal point from the optical interface z_{int}.

All of these effects have been calculated in detail by many previous researchers [106, 258, 271, 181]. The PSFs depend on several parameters: the wave vectors of the incoming and outgoing light k_{in} and k_{out}, the ratio of the indices of refraction $n_{\mathrm{sample}}/n_{\mathrm{lens}}$ of the optical train design and the sample, the numerical aperture of the lens or its acceptance angle α, and the distance focused into the sample z_{int}. For completeness, we repeat the key results here, as we will be using this exact PSF to generate fake images and to study how to best approximate the class of PSFs for confocal microscopes. In polar coordinates, the illumination PSF $P_{\mathrm{ilm}}(\rho, \phi, z)$ for illuminating light with wave vector k_{in} traveling through a lens focused to a

depth z_{int} from the interface is [106]

$$P_{\text{ilm}}(\mathbf{x}) = |K_1|^2 + |K_2|^2 + \frac{1}{2}|K_3|^2 + \cos 2\phi \left[K_1 K_2^* + K_2 K_1^* + \frac{1}{2}|K_3|^2 \right] \quad , \text{ where}$$

$$\begin{pmatrix} K_1 \\ K_2 \\ K_3 \end{pmatrix} = \int_0^\alpha \sqrt{\cos\theta'} \sin\theta' e^{-ik_{\text{in}}f(z,\theta')} \begin{pmatrix} \frac{1}{2}(\tau_s(\theta') + \tau_p(\theta')\cos\theta_2)J_0(k_{\text{in}}\rho\sin\theta') \\ \frac{1}{2}(\tau_s(\theta') - \tau_p(\theta')\cos\theta_2)J_2(k_{\text{in}}\rho\sin\theta') \\ J_1(k_{\text{in}}\rho\sin\theta')\tau_p(\theta')\frac{n_1}{n_2}\sin\theta' \end{pmatrix} d\theta'$$

$$f(\theta) = z_{\text{int}}\cos\theta - \frac{n_2}{n_1}(z_{\text{int}} - z)\sqrt{1 - \left(\frac{n_1}{n_2}\right)^2 \sin^2\theta}$$

$$(2.11)$$

Here $\tau_s(\theta')$ and $\tau_p(\theta)$ are the Fresnel reflectivity coefficients for s and p polarized light, J_n is the Bessel function of order n, and θ_2 is the angle of the refracted ray entering at an angle θ' ($n_2 \sin\theta_2 = n_1 \sin\theta'$). To derive this equation from equation (12) in Ref. [107], we used the additional assumption that all distance scales in the image (including z_{int}) are small compared to the focal length of the lens. The corresponding detection PSF P_{det} is identical to P_{ilm} except for the removal of the $\sqrt{\cos\theta}$ and the replacement of k_{in} by the wave vector of the fluorescent light k_{out}. For an infinitesimal pinhole, the complete PSF is the product of these two point spread functions, $P(\mathbf{x}; z_{\text{int}}) = P_{\text{ilm}}(\mathbf{x}; z_{\text{int}})P_{\text{det}}(\mathbf{x}; z_{\text{int}})$.

In order to extract the PSF from experimental data, we must devise a simple parameterized description or effective model for this PSF. Based on calculations of the exact PSF, $\approx 90\%$ of the function can be described by a Gaussian [271]. We verified this for PSFs calculated from Eq. 2.11, and found that although the presence of aberrations from the interface worsens the Gaussian approximation, generally a Gaussian accounts for $\approx 90\%$ of the PSF except for in the most aberrated cases (large index mismatch imaging deep into the sample).

In light of this, our simplest approximation of the PSF is as an anisotropic

Gaussian with different widths in x, y, and z, with the widths changing with distance from the interface. We therefore parameterize the Gaussian widths as a function of depth,

$$P(\mathbf{x}; z) = \prod_i \frac{e^{-x_i^2/2\sigma_i^2(z)}}{\sqrt{2\pi}\sigma_i(z)} \tag{2.12}$$

where each width $\sigma_i(z)$ can be described as a general function of z such as a polynomial, Legendre, Hermite or otherwise.

How much does it matter if we don't take into account the depth dependence of the point spread function? In section 2.5.3, we show that there is a strong dependence on the fit Gaussian parameters, but how much does this affect the quality of the inferred parameters? We generate fake confocal images using a simulated, exact PSF with random distribution of particles up to a depth of 30 μm. Featuring this data using a 3D anisotropic Gaussian, we find a strong depth-dependent bias in the featured z position and radii measurements. Using a low order z-dependent Gaussian PSF decreases this bias only slightly. Therefore, we employ the full line-scan PSF calculation into our model.

Background

Due to background, the detector CCD pixels always reads a non-zero value even when there is no light incident on them. We incorporate this into our generative model by fitting a nonzero background level to the images. Ideally, this background would be constant at every pixel location. Empirically, however, we find from blank images that this background varies with pixel location in the detector (see Fig.2.9). For our confocal microscope, we find the background is slowly-varying in the optical plane, perhaps due to different dwell times for different regions of the line scan, and does not vary in z. As a result, the background is well-modeled by
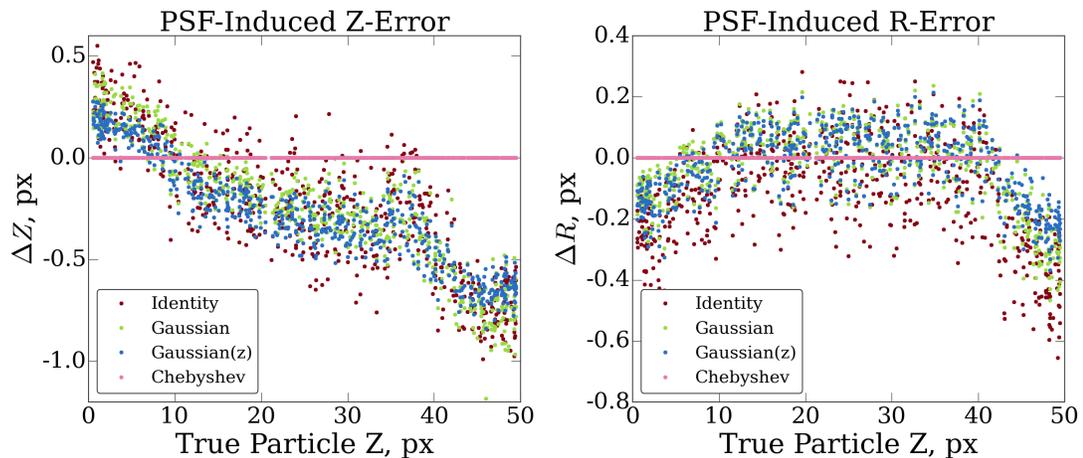
Figure 2.8: **PSF generated biases**. Using an incorrect point-spread function results in significant biases, as PSF leakage affects neighboring particle fits. Moreover, since the PSF gets significantly broader with depth, using a spatially constant PSF, there are systematic biases with depth in both the z positions (left panel) and a characteristic drift in the fitted radii errors with depth (right panel), as shown for the delta-function (identity), an (x, y, z) anisotropic Gaussian, and a depth-varying Gaussian point-spread function. In contrast, using the correct Chebyshev PSF eliminates the errors in both the radii and z positions.

a low-order polynomial in x and y.

However, due to the long-tails of the PSF, the coverslip slab of the image affects the image in a much larger z region than that of a typical particle. Rather than dealing with this by using a large point-spread function, we use a small point spread function to capture the effects of the PSF's moderate tails on the particles and slab, and fit a polynomial in z to capture the residual slab correction. This residual correction is mathematically the same as a background level in the detector. As a result, while the "true" background in the image is $P(x, y)$, our model uses a background $P(x, y) + P_{slab}(z)$.
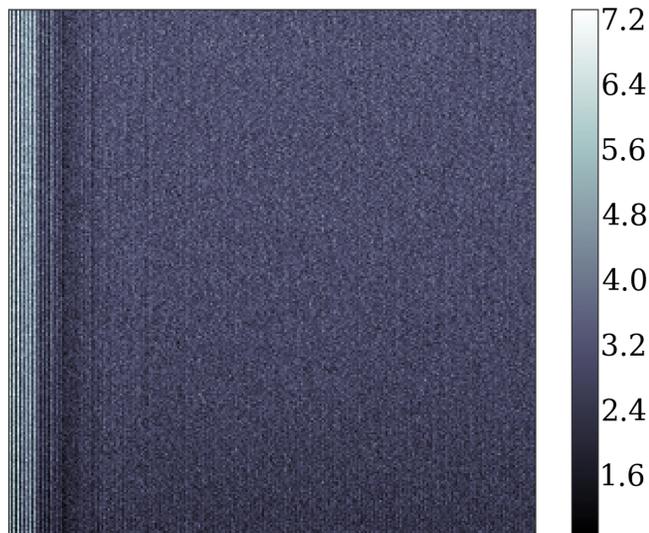
Figure 2.9: **Experimental background image**. The measured background from our line-scan confocal microscope captured by adjusting the exposure to a full brightness image, removing the sample, and capturing a set of images with no illumination including room lights. Note that the range of values is from 1 to 7 out of a maximum 255 given by the 8-bit resolution of the CCD. While only a variation of 3%, we have seen in the illumination field section that this can create a bias that significantly alters our inference as a function of the position in the field of view. To remove this bias we fit the background field to a low order polynomial and add it to our model image.

Sensor noise

The last feature of the generative model is our understanding of the unrecoverable parts of the image: noise. To study the intrinsic noise spectrum of the confocal microscope, we subtract the long wavelength behavior from the blank image of Fig. 2.5. After removing the background we find that the noise appears white and is well approximated by a Gaussian distribution (see Fig. 2.10). There are, however, some highly localized non-Gaussian parts to the noise spectrum, arising due to the specific nature of our confocal. For instance, at high scan speeds slight intensity fluctuations in the laser's power couple to the dwell time on each stripe of line-scanned pixels. This produces periodic stripes across the image with a

wavevector mostly parallel to the scan direction, but with a random noisy phase. How can we handle these sources of correlated noise and do they affect the quality of our reconstruction?

In principle, these correlated noise sources can be represented in the Bayesian model by introducing a full noise covariance matrix. That is, instead of writing that log-likelihood as the product of all pixel values, we can write

$$\log \mathcal{L}(\mathbf{M} \mid \mathbf{d}) = -\frac{1}{2}(M_i - d_i)\, \Lambda_{ij}^{-1}\, (M_j - d_j) \tag{2.13}$$

where Λ_{ij}^{-1} is the covariance matrix between each pixel residual in the entire image. In our optimization, we would form a low dimensional representation for this covariance matrix and allow it to vary until we find a maximum. In doing so, we would reconstruct the image and the correlated noise simultaneously. In practice, this introduces a large computational overhead due to the need for a full image convolution during each update as well as many new free parameters that need to be optimized.

Therefore, we address the effect of correlated noise by working in reverse – we identify the several intense Fourier peaks in the confocal noise spectrum and remove them from the raw data before the fitting process. An example of this noise pole removal is given in Fig. 2.10. There, we can see that removing only 5 distinct poles (Fig. 2.10(d)) removes almost all visible correlated noise structure while changing the overall noise magnitude by a negligible amount. This small shift in estimated noise magnitude only affects the estimate of the errors associated with parameters such as positions and radii in a proportional way. Since these correlated errors are very small and do not bias our inferred parameters, we often ignore them in our analysis entirely.
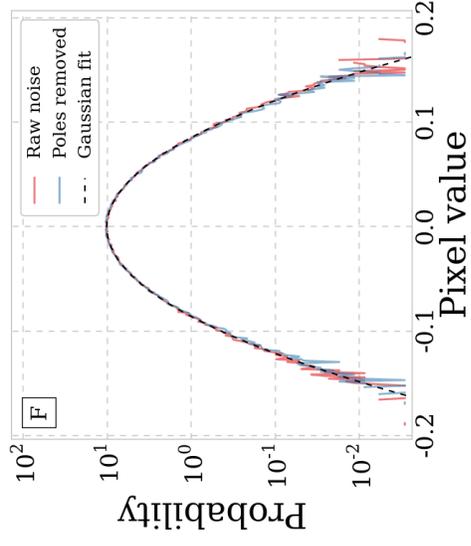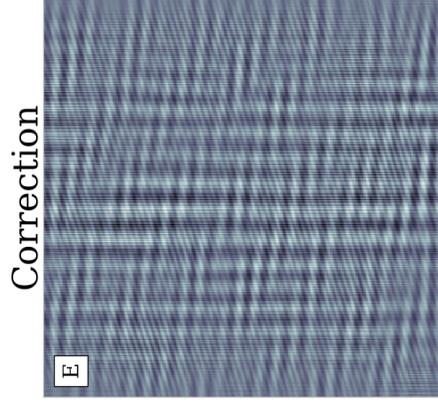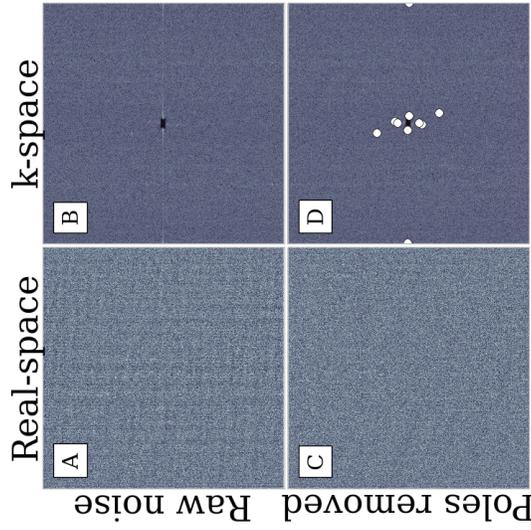
Figure 2.10: **Noise spectrum** (a) Real-space plot of residuals representing the intrinsic noises generated in line-scanning confocal microscopy. This noise spectrum was generated by subtracting the background from a blank sample as in Fig.2.5. Notice that while most of the signal appears to be white noise, there is a systematic modulation along the x coordinate and high frequency features in the y scan direction. (b) Fourier power of the noise spectrum given in (a). The high frequency modulation can now be seen as two small 'poles' in the Fourier power spectrum. Note the dark box in the center of the spectrum is created by subtracting the high order polynomial background from the blank image. In (c) and (d) we present the real and Fourier space noise after removing several discrete peaks in the Fourier intensity that represent correlated noise sources. The removed signal can be seen in (e) showing the stripes created by the scanning nature of the confocal microscope. In (f) we show the histogram of residuals from (a) and (c). In solid red we plot the data and in dashed black lines we plot a Gaussian fit to the residuals with a width $\sigma = 0.0398$, showing that the noise spectrum is well approximated by a Gaussian distribution after taking into account long wavelength background features.

38

2.5.4   Model considerations

Here, we investigate several complexities of image formation in confocal micro-scopes and systematically analyze whether or not it is necessary to include them in our generative model. In particular, we will first analyze how much complexity we must introduce into the model elements listed in the previous section, including the platonic image, illumination field, and point spread function. We will also look at elements of image formation which we have not explicitly included in our model. For example, confocal microscopes build a 3D image by scanning in various 2D patterns as discussed in Section 2.5.3 as well as in the z-direction. There is noise in this scanning procedure which may affect the image formation process. The final image that comes from this scan is a cropped view of a much larger sample where the edges of this image are influenced by the excluded exterior particles. Additionally, this final image is a pixelated version of a continuous electric field, potentially modifying our inferred parameters. Finally, while the exposure is made by the camera, particles undergo diffusional motion, blurring their apparent location. In this next section we address each of these image formation complexities.

We would like to systematically investigate at what level omitting a detail of the image formation from the model affects the fitted parameters. We can understand this quantitatively by examining the optimization procedure. Let us assume that the true image formation is completely described by a set of N parameters Θ. Then, near its maximum, the log-likelihood is approximately quadratic: $\log \mathcal{L} = \frac{1}{2} \sum_{ij} H_{ij} \Theta_i \Theta_j$, where the true value of the parameters is arbitrarily set to $\Theta = 0$. Empirically, we find that with the starting parameter values provided by our initial featuring, the log-likelihood is extremely well-approximated by a quadratic.

If our model were complete, then the maximum of $\log \mathcal{L}$ would be exactly at the
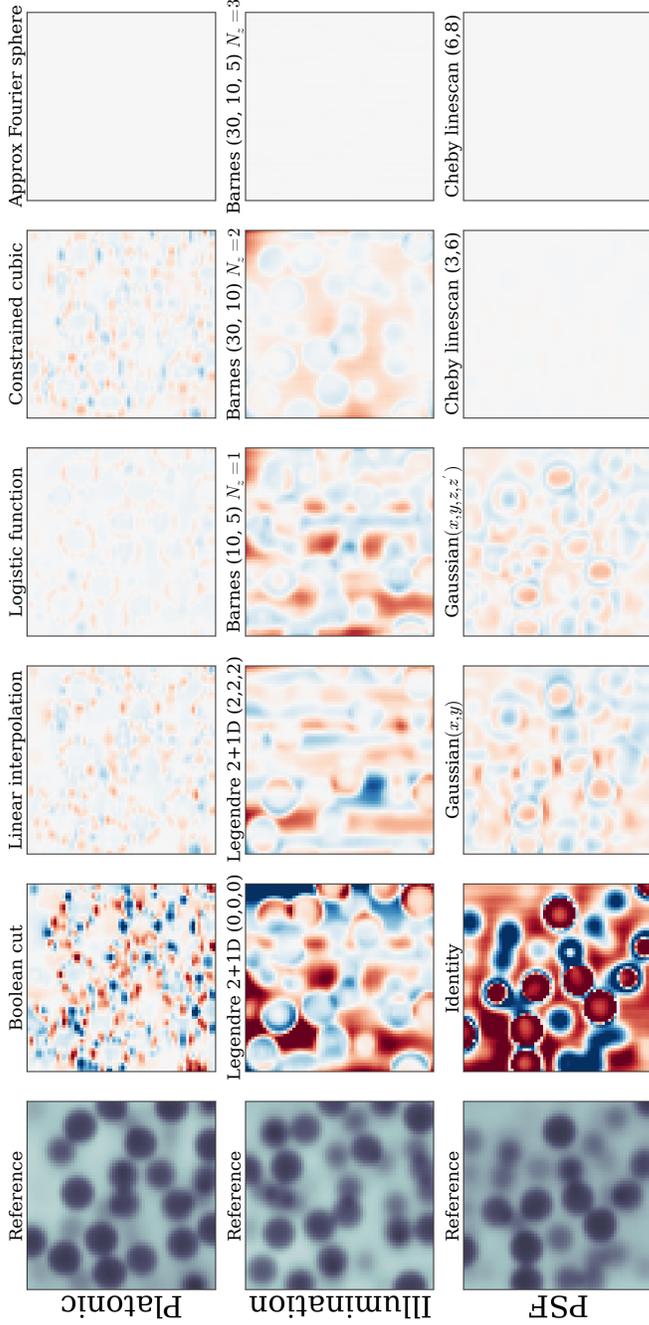
Figure 2.11: **Component complexity residuals.** Here we visually demonstrate the results of choosing between different forms of model components as well as different parameterizations of a single component. We generate simulated microscope images using the model components which we employ when fitting experimental data (left column) and fit them with different choices of platonic image (top row), illumination field (middle row), and point spread function (bottom row). Each choice is labeled above its panel showing the residuals and each row is on a common color scale. In the case of the platonic forms, the boolean cut, linear interpolation, and constrained cubic display higher order multipole errors while the logistic function's first correction is to the monopole (volume) term (as shown by the presence of rings). In the illumination field, stripes are present in the residuals until we use a Barnes interpolant with 30 control points. Past that, the ability to capture intensity as a function of depth is the remaining term which we are able to fit with a single extra Legendre polynomial in the z-direction. Finally, in the case of the PSF, we see hard boundaries transitioning to softer boundaries using a Gaussian PSF in both 3D and 3+1D. The residuals all but disappear when the image is fit with our exact line-scan confocal PSF model (Eq. 2.11) approximated by a Chebyshev polynomial in 3+1D.

40

true parameter values $\Theta = 0$. However, our model is incomplete. This means that, instead of fitting all N parameters Θ, we only fit the first (say) M parameters, which for convenience we denote as θ. Thus we can write the log-likelihood as three separate terms:

$$\log \mathcal{L} = \frac{1}{2} \sum_{i,j=1}^{M} H_{ij}\theta_i\theta_j + \sum_{i=M+1}^{N} \sum_{j=1}^{M} H_{ij}\Theta_i\theta_j + \frac{1}{2} \sum_{i,j=M+1}^{N} H_{ij}\Theta_i\Theta_j \quad . \qquad (2.14)$$

The first term, containing only the parameters θ that we are fitting, is the quadratic in the reduced space, with a maximum at the true parameter values. The unimportant third term reflects the separate contribution to $\log \mathcal{L}$ of the unknown or ignored portions of the model, and is constant in the θ space. However, the second term mixes both the fitted parameters θ and the unknown parameters Θ_j. This mixing results in a linear shift of $\log \mathcal{L}$ in the θ space away from the true parameters, and causes a systematic bias due to an incomplete model. Minimizing $\log \mathcal{L}$ with respect to θ gives the fitted values of the parameters gives an equation for the best-fit incomplete model parameters θ:

$$\theta_j = \sum_{k=1}^{M} \bar{H}_{jk}^{-1} \sum_{i=M+1}^{N} H_{ik}\Theta_i \qquad (2.15)$$

where $\bar{H}^{-1}$ is the inverse of the sub-block $\bar{H}$ of the Hessian matrix H that corresponds to the fitted parameters θ.

We can use equation 2.15 to estimate the effect on one of the estimated parameters θ_j, if we ignore one aspect of the generative model Θ_k. Ignoring the off-diagonal terms in H^{-1} to capture the scaling gives $\theta_j \approx H_{kj}\Theta_k/H_{jj}$. Thus, the error in the fitted parameter θ_j is proportional to both the coupling H_{kj} between that parameter and the ignored aspect of the generative model, and the magnitude of the error of the generative model Θ_k.

Component complexities

There are several choices one can make concerning the form and complexity of each of the components of our model image. As discussed in the Section 2.5.3, we have implemented many forms of the platonic image, illumination field, and point spread function and each one of these forms has a varying number of parameters with which to fit. How do we decide which form to use and at which complexity (number of parameters) to stop? To decide on a per-image basis, we could employ Occam's factor, which is a measure of the evidence that a model is correct given the data [162]. In practice, however, we are mainly concerned with how these models influence the underlying observables which we are attempting to extract. That is, we wish to use knowledge of the physical system to check which model best predicts the particle locations and sizes. To do so (as mentioned in the main manuscript), we often turn to particle sizes versus time as well as particle overlaps, both physical statements that assert almost no assumptions on our system.

We can also get a sense of the magnitude of the effect these choices have on inferred positions and radii by creating synthetic data and fitting it using a simpler model. In Fig. 2.11 we show the residuals of such fits for various simplifications made to the platonic form, illumination field, and point spread function. In the left columns of the figure we see the reference image formed using the most complex image model available and in each row the residuals for each choice with a description of that choice above the panel. For all but the last column, in which we fit the image with the exact model once again, we can see systematic errors in the fit. We compute how much these residuals influence the extracted positions and radii and report these errors in Table 2.1. In particular, most choices of platonic image aside from the naive boolean cut do not influence particle featuring below

an SNR of 30. However, the complexity of the illumination field always matters until all long wavelength structure is removed from the image. Finally, the choice of PSF is crucial, requiring the use of a calculated confocal PSF to even approach the CRB.

Scan jitter

Confocal microscopes operate by taking an image with the lens at a fixed z position to create one layer of the three-dimensional image, then moving the lens up a fixed amount to take the next layer. In our generative model, we assume that these steps of the lens (and the resultant image slices) are perfectly equally spaced by an amount which is fitted internally. However, a real confocal microscope will have some error in the vertical positioning of the lens. As a result, the actual image taken will not be sampled at exactly evenly spaced slices in z, but at slices that are slightly shifted by a random amount.

To test the effect of this z-scan jitter on our parameter estimation, we simulate images taken by a confocal microscope with imperfect z-positioning. Instead of sampling the image at a deterministic z position, we instead sampled the image at a z position shifted from the ideal position by an uncorrelated Gaussian amount of varying standard deviation. A representative image of a 5 px radius particle with a step positioning error of 10% is shown in Fig. 2.12(a). There is very little difference between this image with z jitter and the perfectly-sampled image, as shown by the difference image in panel b. We then fit an ensemble of these images at varying image SNR levels, over a random sampling of image noise, z-jitter noise,

	Fitting model type	Position error (px)	Radius error (px)
Platonic form	Boolean cut	0.03376	0.01577
	Linear interpolation	0.00778	0.00386
	Logistic function	0.00411	0.00352
	Constrained cubic	0.00674	0.00249
	Approx Fourier sphere	0.00000	0.00000
Illumination	Legendre 2+1D (0,0,0)	0.18051	0.13011
	Legendre 2+1D (2,2,2)	0.06653	0.03048
	Barnes (10, 5) $N_z = 1$	0.13056	0.06997
	Barnes (30, 10) $N_z = 2$	0.04256	0.02230
	Barnes (30, 10, 5) $N_z = 3$	0.00074	0.00022
PSF	Identity	0.54427	0.57199
	Gaussian(x, y)	0.47371	0.14463
	Gaussian(x, y, z, z')	0.34448	0.04327
	Cheby linescan (3,6)	0.03081	0.00729
	Cheby linescan (6,8)	0.00000	0.00000

Table 2.1: **Position and radii errors by model complexity.** Here we tabulate the position and radius errors associated with the model component choices made in Fig. 2.11. Note that while the components with the largest impact on determining underlying parameters are the ILM and PSF, the choice of platonic image cannot be ignored in order to reach the theoretical maximum resolution. Interestingly, in the case of PSF selection, Gaussian(x, y, z, z') (3+1D) is almost no better at extracting particle positions than Gaussian(x, y) (2D). However, its ability to extract particle sizes increases by 3 since it takes into account the variation of the PSF in space. Additionally, in the case of the ILM, capturing the stripes in the illumination using a 30 control point Barnes increases the resolution by 3 whereas capturing the illumination's dependence in depth causes the resolution to increase 10 fold.
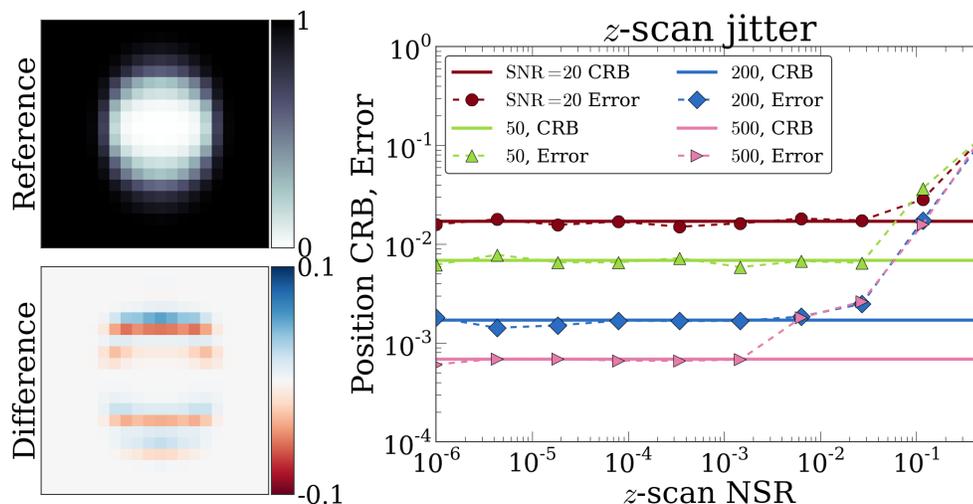
Figure 2.12: **Lens Positioning Jitter** (a) The xz cross-section of a simulated image of a 5 px radius colloidal particle taken with a 10% error in the lens positioning. (b) The difference between the image with positioning error and a reference image with zero positioning error. The differences between the images are both random and small, for this image no more than 7% of the perfect image intensity. (c) The effect of lens positioning error on featured particle positions, at signal-to-noise ratios of 20, 50, 200, and 500. The solid symbols and dashed lines show the position error for images with imperfect lens positioning, while the solid lines denote the Cramer-Rao bound for an image with no positioning error. At lens positioning errors of $\approx 10\%$ or larger, the error in featured positions from the z-slice jitter dominates that from the simple image noise, even for an SNR of 20. However, the featuring error due to a z jitter of $\approx 1\%$ is less than the error due to image noise, for any noise level than can be captured by an 8-bit camera.

and random shifts of particle positions by a fraction of a pixel.

The results of these fits are shown in Fig. 2.12c, showing the actual error in the featured positions versus the size of the z-positioning noise. For our confocal which is equipped with a hyper-fine z-positioning piezo, we expect the z positioning error to be a few nm, or a few percent of a pixel. For a 3% error in positioning, the signal-to-noise ratio must be ≈ 100 for the effects of z-positioning jitter to be comparable to the theoretical minimum effect from the image noise. This small

effect of the error is partially due to the large size of our particle. If each z slice of the image is randomly displaced with standard deviation σ, then we expect roughly a $\sigma/\sqrt{N}$ scaling for the final error in the particle's z-position, where N is the number of z slices the particle appears in. A 5 px diameter particle with a 4 px axial point-spread function occupies ≈ 18 difference slices, decreasing the effect of scan noise by a factor of ≈ 4 and putting it below the CRB for our data.

As the error in z-positioning increases, however, the effect on the featured particle positions increases correspondingly. The error due to a $\approx 10\%$ z jitter is comparable to the CRB for image noises of $SNR = 20$. For exceptionally large z-jitters of 40% the error due to the lens positioning dominates all other sources of error. However, even with this large error in lens positioning, the error in featured positions is still only 10% of a pixel, or about 10 nm in physical units.

Missing and Edge particles

The point spread function delocalizes the particle's image over a region larger than the particle's size. As a result, if two particles are close enough together, their images can overlap. This overlapping is a significant problem for heuristics such as centroid fitting, as the true particle centers do not coincide with the fitted centroid. In contrast, PERI's accuracy is negligibly affected by the presence of a second, close particle, since PERI correctly incorporates close particles in its generative model. The CRB of two touching, 5 px diameter particles increases by only $\approx 3\%$, and PERI finds particles to the same accuracy when close.

However, large systematic errors can affect PERI when one of these particles is missing in the generative model. This situation is illustrated in its simplest form in Fig. 2.13. If one of the two touching particles is missing from the generative model,
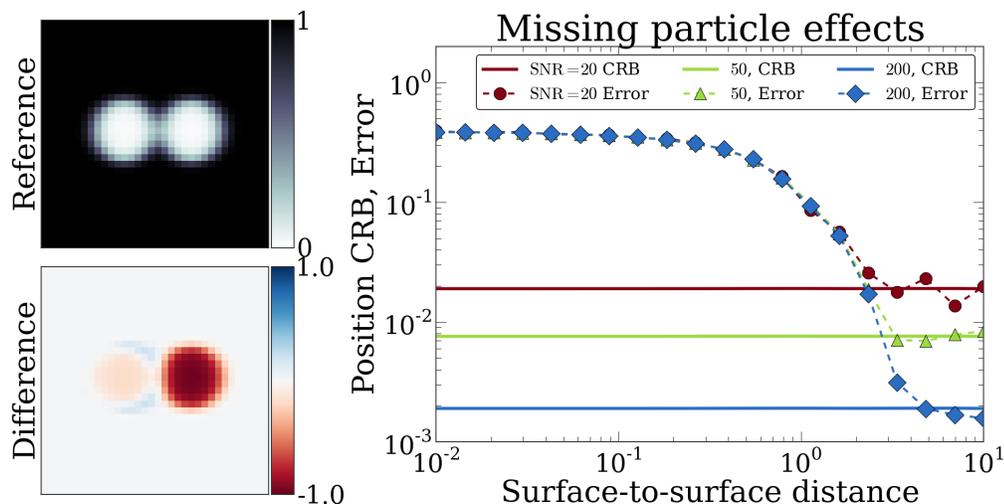
Figure 2.13: **Effect of missing particles**. (a) The xz-cross section of an image of two 5 px radius particles placed in contact. (b) The difference image for a bad generative model that includes only the particle on the left. To minimize the effect of the missing right particle, the left particle is drawn to the right and expanded in radius. This effect is visible as the red and blue ring on the right border of the left particle. (c) The error in position along the separation axis, as a function of true surface-to-surface distance, for a model with a missing particle. When the particles are separated by ≈ 10 px the featured particle is located correctly. However, as the particles get closer than ≈ 2 px significant biases start to appear. These biases saturate at a separation of ≈ 0.1 px, corresponding to a featuring error of ≈ 0.4 px.

then the second particle will be enlarged and drawn into the first particle's void to compensate, as shown in panel b. As a result, the missing second particle will severely bias the fitted positions and radii of the first particle. Figure 2.13c shows the magnitude of this effect. For particles separated by 1 px or less, significant biases on the order of 0.4 px appear in the identified particle's featured position. These biases matter at essentially all values of the SNR, only being comparable to the CRB for SNR < 1. As a result, it is essential for PERI to identify all the particles in the image to return accurate results. For this reason, we take extra precaution and thoroughly search the image for missing particles before fitting, as detailed in section 2.5.5.

The biases caused by missing particles appear whether or not the missing particle is located inside or outside the image. As a result, accurately locating edge particles requires identifying all their nearby particles, even ones that are outside the image! We attempt to solve this problem by padding the Platonic and model images and the ILM by a significant portion, and including this padded extra-image region in both the add/remove and relaxation portions of the PERI algorithm. Nevertheless, it is extremely difficult to locate all the particles outside the image, for obvious reasons. As such, there is the possibility for moderate systematic errors to enter for particles located at or near the edge.

Nevertheless, if the exterior particle is identified, PERI correctly locates the interior particle, as shown in Fig. 2.14. To demonstrate this, we create simulated images of two particles near the boundary of an image. One particle is placed at $z = a$ so that its edge just touches the boundary while the other is placed at $z = -(a + \delta)$ on the other side of the border. We plot the CRB of the interior particle and the measurement errors of both PERI and trackpy as a function of the exterior particle's coordinate in Fig. 2.14. While the CRB only changes by a factor of 2 as the particles come within contact, the featuring errors grow drastically for traditional featuring methods due to biases introduced by the exterior particle. For this same data set, PERI featuring errors follow the CRB allowing precise unbiased featuring of particles at the edge of images.

However, this apparent conundrum of edge particles presents an interesting positive side-effect. Missing edge particles affect the fits because they contribute a significant amount to the image. As such, we might expect that a particle outside the field of view can still be located very precisely. This prediction is borne out by a calculation of the Cramér-Rao bound, as shown in Fig. 2.15.
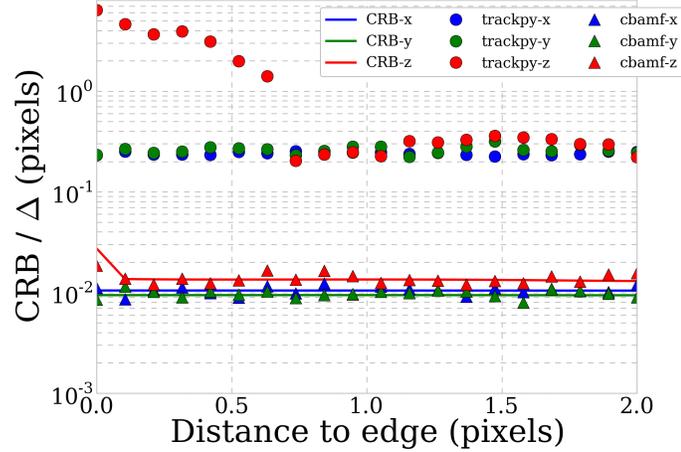
Figure 2.14: **Influence of particles outside of the image**. Here we study how particles external to the image can influence the perceived positions of other interior edge particles. Here we place one particle at $x = a$ and a second particle at $x = -(a+\delta)$ so that one is completely inside the image and the other outside. We plot the CRB for the x, y, and z positions and radius a of the interior particle as well as measured errors for PERI in triangles and trackpy in circles as a function of the position of the second particle. When the exterior particle is further than a pixel outside the image we see that the measurements of the interior particle are constant. However, as the PSF of the exterior particle begins to overlap the interior particle the CRB and all measured errors increase dramatically. While PERI's measured error continues to follow the CRB, trackpy's error increases beyond pixel resolution. Note that pixel separations at the edge are generic in colloidal images especially in dense suspensions.

Until the particle and PSF fall off the edge of the image (distance $> 1R$), the CRB remains constant for all particle parameters. When the particle is centered on the image edge (distance of 0), the CRB is twice that of the bulk, intuitively corresponding to a loss of half of the information about the particle. As the volume of the particle leaves the image, the CRB decreases as $1/\delta^2$ until the particle is no longer part of the image. Interestingly, Fig. 2.15 shows that the PSF constrains the particle position to within 0.1 px even when the particle is entirely out of the image! If correctly seeded with a moderate guess for the particle position outside
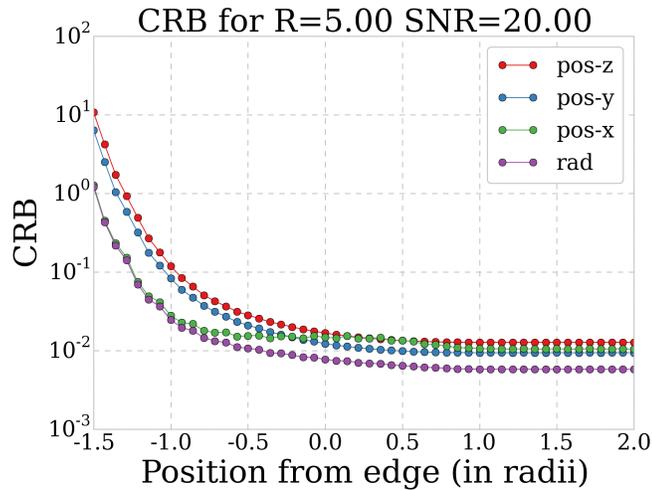
Figure 2.15: **CRB of edge particles**. Here we calculate the Cramér-Rao bound of the x, y, and z positions as well as radius (in red, blue, green, purple respectively) for an isolated particle as a function of its distance to the edge of the image. For positive displacement (inside the image) we see very little change with position as expected. As parts of the PSF leak out of the image (displacements close to zero, positive) we see that the expected error increases slightly since information is lost. Finally, as the particle itself leaves the image, information is lost more dramatically as indicated by a sharp rise in the CRB. However, note that even at a displacement of one radius a, the PSF allows us to locate the particle outside of the image to within a pixel. While in practice it is difficult to identify these particles systematically, their presence can greatly influence the measured positions of other edge particles.

the image, PERI will locate the particle to a precision of the Cramér-Rao bound. However, in practice it is very difficult to seed these particles into PERI, as a slight change of the intensity at the image edge could be either a missing particle outside the image or a slight variation in the ILM near the image edge. Nevertheless, PERI is very good at locating particles that are partially outside the image.
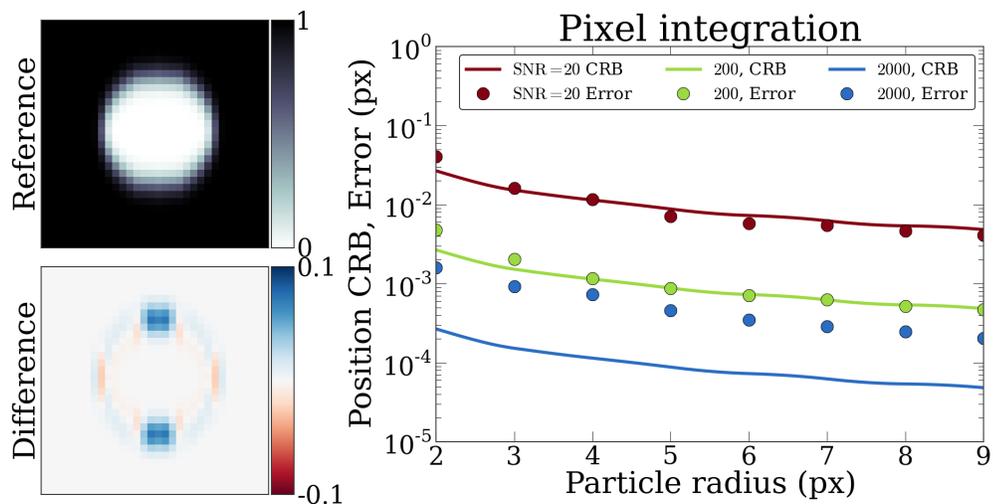
Figure 2.16: **Pixel Integration** (a) The xz cross-section of a simulated image of a 5 px radius colloidal particle, where each pixel contains the light intensity integrated over its area instead of sampled at its center. (b) The difference between the pixel-integrated image and a reference image sampled at the center of the pixels. The differences between the images are small (10%) and centered in a ring which has mean 0 and is positioned at the particle's edge. (c) The effect of pixel integration on featured particle positions as a function of particle radius, at signal-to-noise ratios of 20, 200, and 2000. The solid symbols and dashed lines show the position error for images generated with pixel integration and fit without, while the solid lines denote the Cramer-Rao bound for the images (without pixel integration). Integrating over a pixel area has no effect on the featured positions for any SNR compatible with an 8-bit depth camera. The effect of pixel integration only starts to matter for an SNR ≥ 400 (not shown).

Pixel intensity integration

Our generative model considers the image formed on the camera as if the camera pixels had an infinitesimal size. In reality, the camera pixels have a finite extent. As a result, the image at each pixel on the camera is not a discrete sampling of the light intensity, as in our generative model, but is instead an integration in the detector plane over the pixel's size.

To check whether the effect of pixel integration matters, we generated images

that were up-sampled by a factor of 8 in the xy-plane. We then numerically integrated these images over the size of each pixel. A representative image is shown in Fig. 2.16a. There is very little difference between the xy-integrated image and the generative model, as visible in panel b. We then fitted an ensemble of these xy-integrated pixel images, both over an ensemble of noise samples and over an ensemble of particle positions shifted by a random fraction of a pixel. The results are shown in Fig. 2.16c. We find that there is no discernible effect of pixel integration at a SNR of 200 or less. The error due to neglecting pixel integration becomes comparable to that due to noise only for $SNR \geq 400$, which is significantly higher than the maximum allowed by an ordinary 8-bit camera. Thus, the effect of integrating over a pixel size for a colloidal particle essentially always has a negligible effect on the fitted parameters.

Diffusional motion

A typical colloidal particle is not fixed in its location, but diffuses about due to Brownian motion. For an isolated colloidal particle, this Brownian motion results in a random walk with mean displacement $\langle \mathbf{x} \rangle = \mathbf{0}$ and a mean-square displacement $\langle x^2 \rangle = 6Dt$ that is linear in time, with a diffusion constant $D = kT/6\pi\eta R$ where η is the solvent viscosity and R the particle radius. As a result, the microscope takes an image not of a colloidal particle at a single position, but of an integrated image of the colloidal particle over the trajectory that it has diffused.

First, at what length- and time- scales is a colloidal particle de-localized due to Brownian motion by a scale that is larger than the resolution? For a 1 μm diameter particle in water to diffuse the 1 nm resolution provided by PERI takes a fantastically small time of $t = 1$ nm^2/D $\approx 10\mu$s. Even for our relatively viscous

samples of $\approx 80\%$ glycerol and 20% water this time slows down to only $\approx 600\mu s$. These times are orders of magnitude faster $\approx$ 5ms required by our confocal to take a 3D image of the particle, corresponding to a 8 nm displacement. Thus, a freely diffusing particle has always diffused much more than the featuring errors than the uncertainty intrinsic to PERI.

However, this does not mean that the precision past 8 nm is empty. The particle's positions are Gaussian distributed about its mean value during the exposure time. While the extent of the distribution is much larger than the PERI featuring errors, the particle's mean position during the exposure time is well-defined. Moreover, the actual image on the camera from the diffusing particle is a convolution of the particle's trajectory with a single particle image. Since this convolution is like an averaging, we might expect that the small Brownian excursions are averaged out in the image formation, and that the image allows for accurate featuring of the particle's mean position.

We can use the formalism of Eq. 2.15 to show that Brownian motion does not affect our featuring accuracies. Let the particle's mean position be $\bar{\mathbf{x}}_0$, and its Brownian trajectory be $\mathbf{x}_0(t)$. Then the actual image $I(\mathbf{x})$ on the detector is

$$I(\mathbf{x}) = \frac{1}{t_{\text{exp}}} \int_0^{t_{\text{exp}}} I_0(\mathbf{x}_0(t))\, dt = I_0(\bar{\mathbf{x}}_0) + \frac{1}{t_{\text{exp}}} \int_0^{t_{\text{exp}}} I_0(\mathbf{x}_0(t)) - I_0(\bar{\mathbf{x}}_0)\, dt \quad (2.16)$$

where $I_0(\mathbf{x})$ is the image of one particle at position $\mathbf{x}$ and t_{exp} is the camera exposure time. As before, we view the actual image as $I(\mathbf{x}) = I_0(\bar{\mathbf{x}}_0; \theta) + (1 - \Theta)\Delta I$, in terms of a group of fitted parameters $\boldsymbol{\theta}$ and an additional parameter Θ describing the effects of Brownian motion ΔI. For the true image $\Theta = 0$ but for our model image $\Theta = 1$. Then equation 2.15 says the error will be $\theta_j \approx H_{kj}/H_{jj}$, where $H_{\Theta j} = \partial_\Theta \partial_{\theta_j} I = \partial_{\theta_j} \Delta I$. However, for small displacements the effect of Brownian
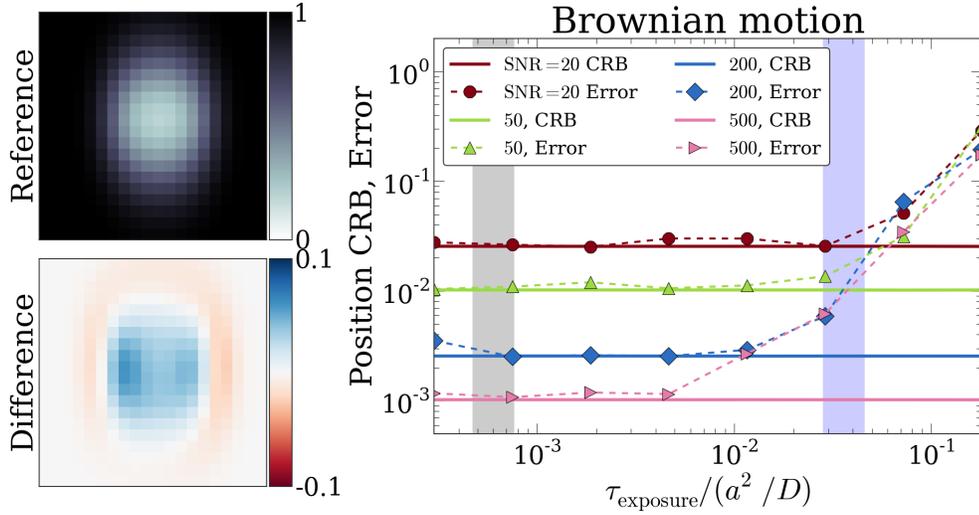
Figure 2.17: **Brownian Motion** (a) The xz cross-section of a simulated image of a 5 px radius colloidal particle undergoing strong Brownian motion $\tau_{\text{exposure}}/(R^2 D) = 0.01$ during the image formation. (b) The difference between the diffusing-particle image and a reference image without diffusion. The differences between the images are small (10%) and are mostly in a ring with mean 0 at the particle's edge. (c) The effect of Brownian motion on featured particle positions as a function of exposure time, at signal-to-noise ratios of 20, 50, 200, and 500. The image exposure time for our camera is located in the shaded grey band for 20/80 water/glycerol and blue band for pure water. The solid symbols and dashed lines show the error between the fitted positions and the mean position in the particle's trajectory, while the solid lines denote the Cramer-Rao bound for the generated images. At our exposure times and SNR of 20, the effects of Brownian motion are small compared to those from noise in the image. Interestingly, for higher SNR or slower exposure times, Brownian motion starts to have a noticeable effect and must be incorporated into the image generation model.

motion on the image is

$$\Delta I = \frac{1}{t_{\text{exp}}} \int_0^{t_{\text{exp}}} \frac{\partial I(\bar{\mathbf{x}}_0)}{\partial x_i}(\mathbf{x} - \bar{\mathbf{x}}_0)\, dt = 0$$

since $\partial I(\bar{\mathbf{x}}_0)/\partial x_i$ does not depend on time. As a result, $\partial_{\theta_k}\partial_\Theta \Delta I = 0$ and there is no affect of Brownian motion on the image to first order in the displacements, *i.e.* when the particle displacement is moderately small compared to the radius.

Finally, in Fig. 2.17 we show empirically that the effect of Brownian motion is negligible for our exposure times. To create an image of a diffusing particle captured by a slow camera, we simulated a 200 point Brownian trajectory of a $R = 5$ px radius particle, generating an image for each point in the particle's trajectory. We then took the average of these images as the noise-free image captured by the microscope. One such image is shown in Fig. 2.17a. Once again, there is a slight difference (10%, as shown in panel b) between the slow image of a diffusing particle and the reference image taken of a particle at a single location. We then fitted an ensemble of these images, over a variety of both Brownian trajectories and noise samples. Figure 2.17c shows the results of these fits as a function of the mean displacement during the collection $\tau_{\mathrm{exposure}}/(R^2 D)$, where τ_{exposure} is the exposure time of the camera and D the particle's diffusion constant. Brownian motion has a negligible effect on the featured positions for our experimental images of freely-diffusing particles (camera exposure time of 100 ms and $D = 0.007$ $\mu\mathrm{m}^2/\mathrm{s}$ corresponding to a 1 μm particle in 80:20 glycerol:water, corresponding to $\tau_{\mathrm{exposure}}/(R^2 D) \approx 10^{-3}$). Interestingly, however, to achieve a higher localization accuracy at a higher SNR of ≈ 200, Brownian motion must be correctly taken into account in the image formation. Incorporating Brownian motion at these high signal-to-noise ratios would allow the teasing out of information about the particle's trajectory from a single image.

2.5.5 Implementation

A typical confocal image is roughly $512 \times 512 \times 100$ pixels in size and contains 10^4 particles meaning that the number of degrees of freedom in our fit is roughly 10^7 described by 10^5 parameters, a daunting space to optimize. On modern hardware

using the highly optimized FFTW, the typical time for an FFT the size of a single image is ~ 1 sec. Given this time, a single sweep through all parameters would take an entire week while a full optimization would consume a year of computer time. However, since particles have finite size, we are able to optimize most of these parameters locally with a small coupling to global parameters (ILM, PSF). Additionally, the finite intensity resolution of microscope sensors, typically 8 or 16 bits, allows us to make further simplifications to our model. Here we describe the practical algorithmic optimizations that we have made as well as the optimization schedule that we have devised to quickly reach the best fit model.

Partial image updates

First and foremost, we optimize our fitting procedure by working in image updates and only updating parts of the image that are required at any one time. In order to modify the position of one particle by a small amount, the number of pixels that are affected is simply $(2a + w)^3$ where a is the particle radius and w is the PSF width, both in pixels. For a typical particle, the ratio of this volume to the entire image volume is typically 10^{-4} which represents a speed up of the same factor due to the roughly linear scaling of FFT performance with problem size ($N \log N$). The practical scale of the width w is set by the bit resolution of the microscope camera used to image the sample. For a bit depth b, values less than 2^{-b} are not unique, corresponding to a change in signal value of $1/255 \approx 0.004$ for an 8-bit camera. For confocal microscope PSFs, the distance scale associated with this signal change is only several pixels. Therefore, we employ a technique common applied to inter-atomic potentials in molecular dynamic simulation – we simply cutoff the PSF at this distance scale allowing for exact partial updates. By cutting off the PSF, we are able to incrementally apply image updates in an

exact procedure (up to floating point errors). For example, when moving a single particle from $\mathbf{x}_0$ to $\mathbf{x}_1$, we must simply calculate the local image change given by

$$\Delta\mathcal{M}(\mathbf{x}) = \int \mathrm{d}^3\mathbf{y}\, cP(\mathbf{x} - \mathbf{y}; \mathbf{y})(\Pi(\mathbf{x}_1) - \Pi(\mathbf{x}_0)) \qquad (2.17)$$

then calculate $\mathcal{M} + \Delta\mathcal{M}$ only in a small local region around the particle being updated. We are able to use similar update rules for all variables except for those effecting the entire image such as the PSF, offset, z_{scale}, and estimate of the SNR.

Additionally, in our code, we generously employ the principle of "space-time trade-off" in which we cache intermediate results of all model components and reuse them later in the computation. In particular, we maintain a full platonic image and illumination field, which we update along with the model image. We also cache all calculated PSF values so that we may utilize previous results until the PSF is sampled. In doing so, we limited in our current implementation by the speed of the FFT, which takes 70% of the total runtime.

Optimization of parameters and sampling for error

Once an approximate initial guess is obtained by more traditional featuring methods [63], we optimize the parameters by fitting using a modified Levenberg-Marquardt routine. Our Levenberg-Marquardt algorithm uses previously-reported optimization strategies designed for large parameter spaces [248]. However, a Levenberg-Marquardt minimization requires the matrix $J_{i\alpha} \equiv \partial m(x_i)/\partial\theta_\alpha$, which is the gradient of each pixel in the model with respect to all the parameters. For the $\approx 10^5$ parameters and 10^7 pixels in our image, this matrix would be many thousand times too large to store in memory. Instead, we construct a random approximation to $J_{i\alpha}$ by using a random sub-section of pixels x_i in the image to

compute J. This approach works well for the global parameters (PSF, ILM, etc) but fails for the particles, which appear in a relatively small number of pixels. For the particles, we instead fit small groups of adjacent particles using the full $J_{i\alpha}$ for the affected pixels. As the global parameters and particle parameters are coupled, we iterate by optimizing first the globals, then the particles, and repeating until the optimization has converged.

Once the model is optimized, we use Monte Carlo sampling to estimate parameter errors. Our Monte Carlo sampler sweeps over each parameter and updates the particle position, accepting or rejecting based on the change in the log-likelihood of the model. We use slice sampling to produce highly uncorrelated samples, allowing an excellent error estimate from only a few sweeps. Our error sampling doubles as a check for convergence. If, after sampling, the log-likelihood increases by a significant amount, then the optimization has not converged and either more Monte Carlo sampling or better optimization is needed. In practice, we check with $\approx 5 - 10$ Monte Carlo sweeps, and check if the log-likelihood has changed by a few times $\sqrt{N}$, where N is the number of parameters in the model.

2.5.6   Benchmarks of featuring algorithms

We check our algorithm by benchmarking it against physically realistic image models, as shown in Fig. 2.18. To generate these images, we employ the exact PSF calculation as described in Sec. 2.5.3, the ILM extracted from real confocal data, as well as complex correlated noise created in line-scanning confocal microscopes. We create particle distributions using hard-sphere Brownian dynamics to ensure that the configurations tested are statistically similar to those found in experiments. For comparison, we also show the relative performance of the well established method
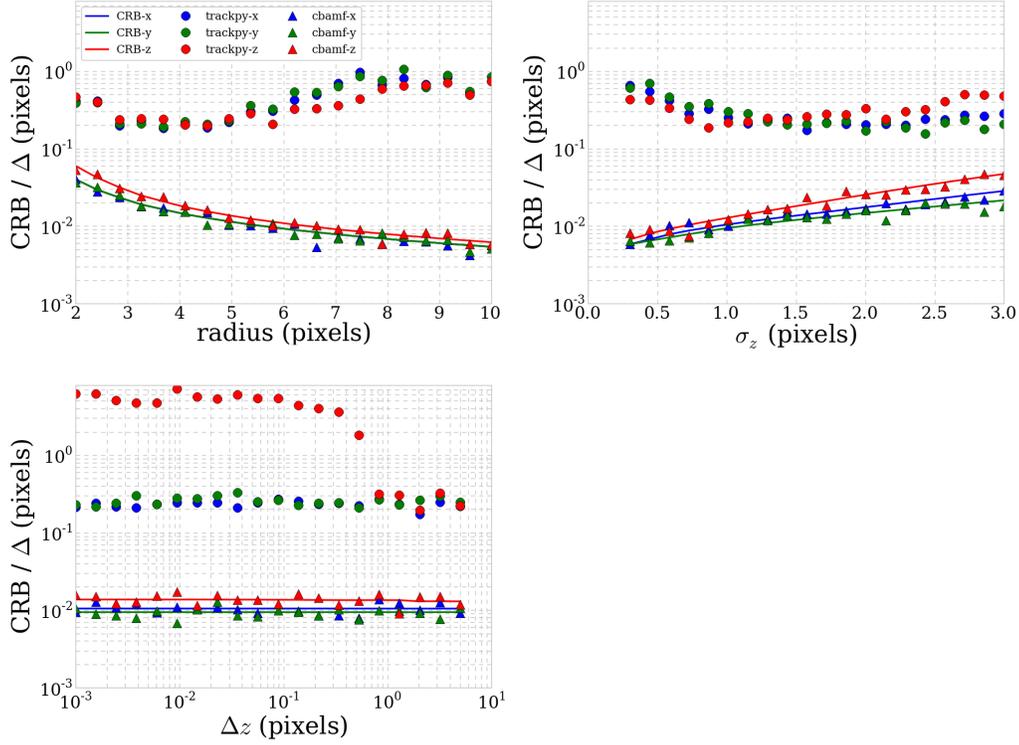
Figure 2.18: **Accuracy benchmark** – We compare the performance of PERI with traditional featuring methods as well as the calculated Cramér-Rao bound. In the upper left panel, we plot the CRB for x, y, and z positions of a single particle as a function of radius for a constant ILM and exact PSF. For 1.3 μm particles, the CRB crosses 1 nm. In circles we plot the featuring results from Crocker-Grier (CG) which are mostly independent of radius. In triangles, we see that PERI errors follow the CRB for all particles sizes. In the upper right panel, we plot the same quantities but for a particle of radius $a = 5$ px as a function of PSF scale. In the lower left we plot the same quantities for an image containing 2 particles as a function of the separation Δz. While PERI closely follows the CRB independently of separation, CG becomes biased as the particle PSFs overlap.

of Crocker-Grier [63], the standard in sphere extraction algorithms in the colloidal community.

First, we compare the methods' ability to locate a single, isolated particle as a function of radius as in Fig. 2.18(a). In thick lines we can see that the CRB for the x position of an isolated particle scales as $1/a$. Overlaid, we see the average error of positions over many different positions and noise instantiations so as to remove effects of pixel bias and the peculiarities of a single noise instance. We see that PERI closely follows the expected error of $0.6 - 3$ nm, following the CRB for all radii. Both centroid methods are affected by pixel bias leading to an average error of roughly 0.3 px or 40 nm regardless of particle size.

Next, we check the dependence on PSF size, as determined by the effects of spherical aberrations when imaging deeply into an index-mismatched sample. As a function of the PSF size in z (directly related to the distance from the interface), we find once again that PERI lies along the CRB while the others have an error of at least 0.3 px, increasing with z_{int}. We next study the effect of a second, nearby particle on the fits, by measuring the average error as a function of separation between two identical particles. When the particles are close enough, delocalization from the PSF causes the centroid methods to incorrectly place the particles too close together. In contrast, PERI is not affected by the interparticle separation.

Finally, we check how the complexity of our synthetic data affects the accuracy of standard featuring methods. In Table 2.2 we see, surprisingly, that there is a non-monotonic relationship between positional error and image complexity, becoming optimal when there is significant striping in the image but little variation in depth. However, the rate of missing particles decreases significantly with simpler models and rising to as much as 40% for our most complex model images. The effective

resolution of CG is never much smaller than a single pixel in these synthetic tests, most likely due to pixel edge biases.

2.5.7 Experimental Details

The microscope is a Zeiss LSM 5 Live inverted confocal microscope, used in conjunction with an infinity-corrected 100x immersion oil lens (Zeiss Plan-Apochromat, 1.4 NA, immersion oil with index $n = 1.518$). The LSM 5 Live confocals operate by line-scanning. Rather than rastering a single point at a time to form the image, a line-scanning confocal images an entire line at once. An image of a line is focused onto the sample, and the sample fluorescence is detected on a line CCD. Rastering this line allows images to be collected much faster. However, the different line-scanning optics worsen the point-spread function compared to a point-scanning confocal and cause illumination imperfections such as dirt to be smeared out over one direction in the image. Importantly, our confocal is outfitted with a hyper-fine piezo scanner which gives precise z-positioning of the lens. This precise z-positioning is important for accurate reconstruction of images – with the less-precise standard positioning our image reconstruction and results suffer considerably. The data shown in Fig. blah were taken at 108.1 in-plane frames per second, with a time between successive image stacks of either 10 or 1.25 seconds.

Our experimental images consist of ≈ 1.3 μm silica particles (MicroPearl) suspended in a mixture of glycerol and water. The glycerol/water mixture is tuned

Polydispersity	Illumination field	Point spread function	Position error	% Identified
0.0	Legendre 2+1D (0,0,0)	Identity	1.458	0.81
0.0	Legendre 2+1D (2,2,2)	Gaussian(x, y)	1.218	0.84
0.01	Barnes (10, 5), $N_z = 1$	Gaussian(x, y, z, z')	1.015	0.75
0.05	Barnes (30, 10)), $N_z = 2$	Cheby linescan (3,6)	0.819	0.64
0.10	Barnes (30, 10, 5), $N_z = 3$	Cheby linescan (6,8)	1.085	0.64

Table 2.2: **Crocker-Grier featuring errors.** We show the effect of image complexity on position error and miss rate for the CG featuring method using synthetic data. Surprisingly, there is a non-monotonic behavior of error with complexity, hitting a maximum for highly striped images that don't vary strongly with depth. However, the featuring miss rate steadily decreases with complexity, reaching only 60% identified particles with our most realistic images.

to match the refractive index of the particles by minimizing the sample scattering. For these particles we find the optimal refractive index is $n \approx 1.437$ corresponding to $\approx 80\%$ glycerol and 20% water. We match the index of refraction of the spheres and the suspending fluid to within a few parts per thousand, resulting in practically zero scattering by the spheres of the laser or fluorescent light. The glycerol has the additional advantage of creating a very viscous suspension, slowing down the Brownian motion of the particles. We add fluorescein sodium salt to dye the suspending fluid. The fluorescein diffuses rapidly compared to the particles, and is effectively uniformly distributed throughout the regions occupied by the fluid. By using a considerable amount of dye and a low laser power, we minimize photobleaching during our experiments.

CHAPTER 3

SCALING THEORY OF CONTINUUM DISLOCATION DYNAMICS IN THREE DIMENSIONS: SELF-ORGANIZED FRACTAL PATTERN FORMATION[1]

3.1 Introduction

Dislocations in plastically deformed crystals, driven by their long-range interactions, collectively evolve into complex heterogeneous structures where dislocation-rich cell walls or boundaries surround dislocation-depleted cell interiors. These have been observed both in single crystals [128, 174, 224] and polycrystals [252] using transmission electron microscopy (TEM). The mesoscopic cellular structures have been recognized as scale-free patterns through fractal analysis of TEM micrographs [88, 87, 99, 270]. The complex collective behavior of dislocations has been a challenge for understanding the underlying physical mechanisms responsible for the development of emergent dislocation morphologies.

Complex dislocation microstructures, as an emergent mesoscale phenomenon, have been previously modeled using various theoretical and numerical approaches. Discrete dislocation dynamics (DDD) models have provided insights into the dislocation pattern formations: parallel edge dislocations in a two-dimensional system evolve into 'matrix structures' during single slip [21], and 'fractal and cell structures' during multiple slip [22, 23]; random dislocations in a three-dimensional system self-organize themselves into microstructures through junction formation,

---

cross-slip, and short-range interactions [163, 92]. However, DDD simulations are limited by the computational challenges on the relevant scales of length and strain. Beyond these micro-scale descriptions, CDD has also been used to study complex dislocation structures. Simplified reaction-diffusion models have described persistent slip bands [259], dislocation cellular structures during multiple slip [98], and dislocation vein structures [217]. Stochasticity in CDD models [21, 95, 99] or in the splittings and rotations of the macroscopic cells [191, 192, 229] have been suggested as an explanation for the formation of organized dislocation structures. The source of the noise in these stochastic theories is derived from either extrinsic disorder or short-length-scale fluctuations.

In a recent manuscript [52], we analyzed the behavior of a grossly simplified continuum dislocation model for plasticity [8, 212, 155, 52] – a physicist's 'spherical cow' approximation designed to explore the minimal ingredients necessary to explain key features of the dynamics of deformation. Our simplified model ignores many features known to be important for cell boundary morphology and evolution, including slip systems and crystalline anisotropy, dislocation nucleation, lock formation and entanglement, line tension, geometrically unnecessary forest dislocations, etc. However, our model does encompass a realistic order parameter field (the Nye dislocation density tensor [187] embodying the GNDs), which allows detailed comparisons of local rotations and deformations, stress, and strain. It is not a realistic model of a real material, but it is a model material with a physically sensible evolution law. Given these simplifications, our model exhibited a surprisingly realistic evolution of cellular structures. We analyzed these structures in two-dimensional simulations (full three-dimensional rotations and deformations, but uniform along the z-axis) using both the fractal box counting method [88, 87, 99, 270] and the single-length-scale scaling meth-

ods [122, 119, 121, 168] used in previous theoretical analyses of experimental data. Our model qualitatively reproduced the self-similar, fractal patterns found in the former, and the scaling behavior of the cell sizes and misorientations under strain found in the latter (power-law refinement of the cell sizes, power-law increases in misorientations, and scaling collapses of the distributions).

There are many features of real materials which are not explained by our model. We do not observe distinctions between 'geometrically necessary' and 'incidental' boundaries, which appear experimentally to scale in different ways. The fractal scaling observed in our model may well be cut off or modified by entanglement, slip-system physics, quantization of Burger's vector [135] or anisotropy – we cannot predict that real materials should have fractal cellular structures; we only observe that our model material does so naturally. Our spherically symmetric model obviously cannot reproduce the dependence of morphological evolution on the axis of applied strain (and hence the number of activated slip systems); indeed, the fractal patterns observed in some experiments [99, 270] could be associated with the high-symmetry geometry they studied [101, 261]. While many realistic features of materials that we ignore may be important for cell-structure formation and evolution, our model gives clear evidence that these features are not essential to the formation of cellular structures when crystals undergo plastic deformation.

In this longer manuscript, we provide an in-depth analysis of three plasticity models. We show how they (and more traditional models) can be derived from the structures of the broken symmetries and order parameters. We extend our simulations to 3D, where the behavior is qualitatively similar with a few important changes. Here we focus our attention on relaxation (rather than strain), and on correlation functions (rather than fractal box counting or cell sizes and misorien-

tations).

Our model exhibits fractal cell structures directly upon relaxation from randomly deformed initial conditions (Sec. 3.3.2). This is not the case for realistic materials, where the dislocation evolution cannot be postponed to start after the plastic deformation is imposed. Indeed, cellular structures in real materials emerge only after significant deformation; presumably this feature is missing in our model because our model has no impediment to cross-slip or multiple slip, and no entanglement of dislocations. This initial relaxation should not be viewed as annealing or dislocation creep. A proper description of annealing must include dislocation line tension effects, since the driving force for annealing is the reduction in total dislocation density – our dislocations annihilate when their Nye Burger's vector density cancels under evolution, not because of the dislocation core energies. Creep involves dislocation climb, which (for two of our three models) is forbidden. Instead, we view this initial relaxation as the evolution under an instantaneous external plastic deformation – the dislocations produced by random, rapid hammer blows (Sec. 3.3.2). The resulting cellular structures are qualitatively very similar to those we observe under external strain [52, 51], except that they are statistically isotropic. Indeed, we believe that the relaxation evolution we study here mimics almost precisely what we would observe under an imposed time-increasing random plastic deformation from an initially uniform state – slow hammer blows producing similar patterns to relaxation after rapid ones (see Sec. 3.3.2).

We focus here on correlation functions, rather than the methods used in previous analyses of experiments. Correlation functions have a long, dignified history in the study of systems exhibiting emergent scale invariance – materials at continuous thermodynamic phase transitions [46], fully developed turbulence [161, 55],

and crackling noise and self-organized criticality [226]. We study not only numerical simulations of these correlations, but provide also extensive analysis of the relations between the correlation functions for different physical quantities and their (possibly universal) power-law exponents. The decomposition of the system into cells (needed for the cell-size and misorientation distribution analyses [122, 119, 121, 168]) demands the introduction of an artificial cutoff misorientation angle, and demands either laborious human work or rather sophisticated numerical algorithms [1]. These sections of the current manuscript may be viewed both as a full characterization of the behavior of our simple model, and as an illustration of how one can use correlation functions to analyze the complex morphologies in more realistic models and in experiments providing 2D or 3D real-space data. We believe that analyses that explicitly decompose structures into cells remain important for systems with single changing length-scale: grain boundary coarsening should be studied both with correlation functions and with explicit studies of grain shape and geometry evolution, and the same should apply to cell-structure models and experiments that are not fractal. But our model, without such an intermediate length-scale, is best analyzed using correlation functions.

Our earlier work [52] focused on 2D. How different are our predictions in 3D? In this paper, we explore three different CDDs that display similar dislocation fractal formation in 3D and confirm analytically that correlation functions of the GND density, the plastic distortion, and the crystalline orientation, all share a single underlying critical exponent, up to exponent relations, dependent only on the type of dynamics. Unlike our 2D simulations, where forbidding climb led to rather distinct critical exponents, all three dynamics in 3D share quite similar scaling behaviors.

We begin our discussion in Sec. 3.2.1 by defining the various dislocation, distortion, and orientation fields. In Sec. 3.2.2, we derive standard local dynamical evolution laws using traditional condensed matter approaches, starting from both the non-conserved plastic distortion and the conserved GND densities as order parameters. Here, we also explain why these resulting dynamical laws are inappropriate at the mesoscale. In Sec. 3.2.3, we show how to extend this approach by defining appropriate constitutive laws for the dislocation flow velocity to build novel dynamics [140]. There are three different dynamics we study: i) isotropic climb-and-glide dynamics (CGD) [8, 9, 10, 212, 155], ii) isotropic glide-only dynamics, where we define the part of the local dislocation density that participates in the local *mobile dislocation population*, keeping the local volume conserved at all times (GOD-MDP) [52], iii) isotropic glide-only dynamics, where glide is enforced by a *local vacancy pressure* due to a co-existing background of vacancies that have an infinite energy cost (GOD-LVP) [11]. All three types of dynamics present physically valid alternative approaches for deriving a coarse-grained continuum model for GNDs. In Sec. 3.3, we discuss the details of numerical simulations in both two and three dimensions, and characterize the self-organized critical complex patterns in terms of correlation functions of the order parameter fields. In Sec. 3.4, we provide a scaling theory, and derive relations among the critical exponents of these related correlation functions, and conclude in Sec. 3.5.

3.2 Continuum models

3.2.1 Order parameter fields

Conserved order parameter field

A dislocation is the topological defect of a crystal lattice. In a continuum theory, it can be described by a coarse-grained variable, the GND density [2], (also called the net dislocation density or the Nye dislocation density), which can be defined by the GND density tensor

$$\rho(\mathbf{x}) = \sum_\alpha (\hat{\mathbf{t}}^\alpha \cdot \hat{\mathbf{n}}) \hat{\mathbf{n}} \otimes \mathbf{b}^\alpha \delta(\mathbf{x} - \boldsymbol{\xi}^\alpha), \tag{3.1}$$

so

$$\rho_{km}(\mathbf{x}) = \sum_\alpha \hat{\mathbf{t}}_k^\alpha b_m^\alpha \delta(\mathbf{x} - \boldsymbol{\xi}^\alpha), \tag{3.2}$$

measuring the sum of the net flux of dislocations α located at $\boldsymbol{\xi}$, tangent to $\hat{\mathbf{t}}$, with Burgers vector $\mathbf{b}$, in the neighborhood of $\mathbf{x}$, through an infinitesimal plane with the normal direction along $\hat{\mathbf{n}}$, seen in Fig. 3.1. In the continuum, the discrete sum of line singularities in Eqs. (3.1) and (3.2) is smeared into a continuous (nine-component) field, just as the continuum density of a liquid is at root a sum of point contributions from atomic nuclei.

Since the normal unit pseudo-vector $\hat{\mathbf{n}}$ is equivalent to an antisymmetric unit bivetor $\hat{\mathbf{E}}$, $\hat{\mathbf{E}}_{ij} = \varepsilon_{ijk}\hat{\mathbf{n}}_k$, we can reformulate the GND density as a three-index tensor

$$\varrho(\mathbf{x}) = \sum_\alpha (\hat{\mathbf{t}}^\alpha \cdot \hat{\mathbf{n}}) \hat{\mathbf{E}} \otimes \mathbf{b}^\alpha \delta(\mathbf{x} - \boldsymbol{\xi}^\alpha), \tag{3.3}$$

so

$$\varrho_{ijm}(\mathbf{x}) = \sum_\alpha (\hat{\mathbf{t}}^\alpha \cdot \hat{\mathbf{n}}) \hat{\mathbf{E}}_{ij} b_m^\alpha \delta(\mathbf{x} - \boldsymbol{\xi}^\alpha), \tag{3.4}$$

measuring the same sum of the net flux of dislocations in the neighborhood of $\mathbf{x}$, through the infinitesimal plane indicated by the unit bivetor $\hat{\mathbf{E}}$. This three-index variant will be useful in Sec. 3.2.3, where we adapt the equations of Refs. 212 and 155 to forbid dislocation climb (GOD-MDP).

According to the definition of $\hat{\mathbf{E}}$, we can find the relation between ρ and ϱ

$$\varrho_{ijm}(\mathbf{x}) = \sum_{\alpha}(\hat{\mathbf{t}}_l^{\alpha}\hat{\mathbf{n}}_l)\varepsilon_{ijk}\hat{\mathbf{n}}_k b_m^{\alpha}\delta(\mathbf{x} - \boldsymbol{\xi}^{\alpha}) = \varepsilon_{ijk}\rho_{km}(\mathbf{x}). \tag{3.5}$$

It should be noted here that dislocations cannot terminate within the crystal, implying that

$$\partial_i\rho_{ij}(\mathbf{x}) = 0, \tag{3.6}$$

or

$$\varepsilon_{ijk}\partial_k\varrho_{ijl}(\mathbf{x}) = 0. \tag{3.7}$$

Within plastic theories, the gradient of the total displacement field $\mathbf{u}$ represents the compatible total distortion field [134] $\beta_{ij} = \partial_i u_j$, which is the sum of the elastic and the plastic distortion fields [134], $\beta = \beta^{\mathrm{p}} + \beta^{\mathrm{e}}$. Due to the presence of dislocation lines, both β^{p} and β^{e} are incompatible, characterized by the GND density ρ

$$\rho_{ij} = \epsilon_{ilm}\partial_l\beta_{mj}^{\mathrm{e}}, \tag{3.8}$$

$$= -\epsilon_{ilm}\partial_l\beta_{mj}^{\mathrm{p}}. \tag{3.9}$$

The elastic distortion field β^{e} is the sum of its symmetric strain and antisymmetric rotation fields,

$$\beta^{\mathrm{e}} = \epsilon^{\mathrm{e}} + \omega^{\mathrm{e}}, \tag{3.10}$$

where we assume linear elasticity, ignoring the 'geometric nonlinearity' in these tensors. Substituting the sum of two tensor fields into the incompatibility relation

71
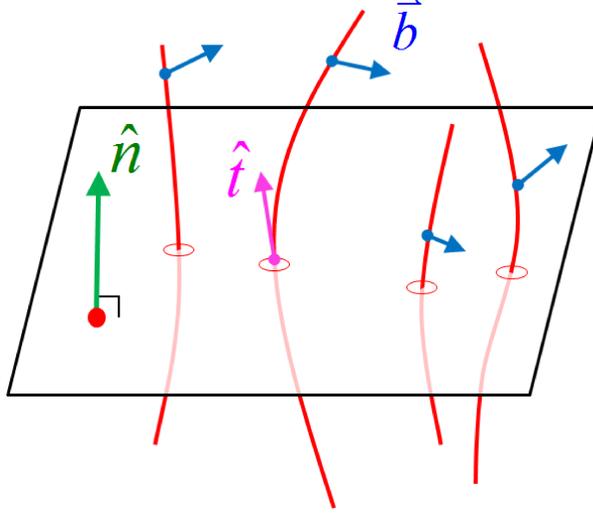
Figure 3.1: **Representation of the crystalline line defect — dislocation**. Each curved line represents a dislocation line with the tangent direction $\hat{\mathbf{t}}$, and the Burgers vector $\mathbf{b}$ which characterizes the magnitude and direction of the distortion to the lattice. The two-index GND density ρ_{km} (Eqs. 3.1 and 3.2) is the net flux of the Burgers vector density $\mathbf{b}$ along $\hat{\mathbf{e}}^{(m)}$ through an infinitesimal piece of a plane with normal direction $\hat{\mathbf{n}}$ along $\hat{\mathbf{e}}^{(k)}$. The three-index version ϱ_{ijm} (Eqs. 3.3 and 3.4) is the flux density through the plane along the axes $\hat{\mathbf{e}}^{(i)}$ and $\hat{\mathbf{e}}^{(j)}$, with the unit bivetor $\hat{\mathbf{E}} = \hat{\mathbf{e}}^{(i)} \wedge \hat{\mathbf{e}}^{(j)}$.

Eq. (3.8) gives

$$\rho_{ij} = \varepsilon_{ikl}\partial_k \omega_{lj}^{e} + \varepsilon_{ikl}\partial_k \epsilon_{lj}^{e}. \tag{3.11}$$

The elastic rotation tensor ω^{e} can be rewritten as an axial vector, the crystalline orientation vector $\mathbf{\Lambda}$

$$\Lambda_k = \frac{1}{2}\varepsilon_{ijk}\omega_{ij}^{e}, \tag{3.12}$$

or

$$\omega_{ij}^{e} = \varepsilon_{ijk}\Lambda_k. \tag{3.13}$$

Thus we can substitute Eq. (3.13) into Eq. (3.11)

$$\rho_{ij} = (\delta_{ij}\partial_k \Lambda_k - \partial_j \Lambda_i) + \varepsilon_{ikl}\partial_k \epsilon_{lj}^{e}. \tag{3.14}$$

For a system without residual elastic stress, the GND density thus depends only on the varying crystalline orientation [156].

Dynamically, the time evolution law of the GND density emerges from the conservation of the Burgers vector

$$\frac{\partial}{\partial t}\rho_{ik} = -\varepsilon_{ijq}\partial_j J_{qk}, \qquad (3.15)$$

or

$$\frac{\partial}{\partial t}\varrho_{ijk} = -\varepsilon_{ijm}\varepsilon_{mpq}\partial_p J_{qk} = -g_{ijpq}\partial_p J_{qk}, \qquad (3.16)$$

where J represents the Burgers vector flux, and the symbol g_{ijpq} indicates $\varepsilon_{ijm}\varepsilon_{mpq} = \delta_{ip}\delta_{jq} - \delta_{iq}\delta_{jp}$.

Non-conserved order parameter field

The natural physicist's order parameter field ϱ, characterizing the incompatibility, can be written in terms of the plastic distortion field β^{p}

$$\varrho_{ijk} = \varepsilon_{ijm}\rho_{mk} = -g_{ijls}\partial_l \beta^{\mathrm{p}}_{sk}. \qquad (3.17)$$

In the linear approximation, the alternative order parameter field β^{p} fully specifies the local deformation $\boldsymbol{u}$ of the material, the elastic distortion β^{e}, the internal long-range stress field σ^{int} and the crystalline orientation (the Rodrigues vector $\boldsymbol{\Lambda}$ giving the axis and angle of rotation).

According to Eq. (3.9) and Eq. (3.15), the flux J of the Burgers vector can be expressed in terms of the dynamics of the plastic distortion tensor β^{p}

$$\frac{\partial \beta^{\mathrm{p}}_{ij}}{\partial t} = J_{ij}. \qquad (3.18)$$

The plastic distortion β^{p} can be specified by the GND density ρ and a strain-history deformation field $\boldsymbol{\psi}$ in Fourier space [3],

$$
\begin{aligned}
\widetilde{\beta}_{ij}^{\mathrm{p}}(\mathbf{k}) &= -i\varepsilon_{ilm}\frac{k_l}{k^2}\widetilde{\rho}_{mj}(\mathbf{k}) + ik_i\widetilde{\psi}_j(\mathbf{k}) \\
&\equiv \widetilde{\beta}_{ij}^{\mathrm{p,I}}(\mathbf{k}) + \widetilde{\beta}_{ij}^{\mathrm{p,H}}(\mathbf{k}),
\end{aligned} \tag{3.19}
$$

hence decomposing $\widetilde{\beta}^{\mathrm{p}}$ into two parts. $\beta^{\mathrm{p,I}}$ is the intrinsic field specified by the GND density. Similar to ρ, $\beta^{\mathrm{p,I}}$ is also divergence free: $\partial_i\beta_{ij}^{\mathrm{p,I}} = 0$, i.e., $k_i\widetilde{\beta}_{ij}^{\mathrm{p,I}} = 0$. $\beta^{\mathrm{p,H}}$ is a (curl-free) gradient of $\boldsymbol{\psi}$, depending upon the strain history and contributing nothing to the GND density. This decomposition will become important to us in Sec. 3.3.3, where the correlation functions of $\beta^{\mathrm{p,I}}$ and $\beta^{\mathrm{p,H}}$ will scale differently with distance.

In the presence of external loading, we can express the appropriate free energy $\mathcal{F}$ as the sum of two terms: the elastic interaction energy of GNDs, and the energy of interaction with the applied stress field. The free energy functional is

$$
\mathcal{F} = \int d^3\mathbf{x}\left(\frac{1}{2}\sigma_{ij}^{\mathrm{int}}\epsilon_{ij}^{\mathrm{e}} - \sigma_{ij}^{\mathrm{ext}}\epsilon_{ij}^{\mathrm{p}}\right). \tag{3.20}
$$

Alternatively, it can be rewritten in Fourier space

$$
\begin{aligned}
\mathcal{F} = &-\int \frac{d^3\mathbf{k}}{(2\pi)^3}\left(\frac{1}{2}M_{ijmn}(\mathbf{k})\widetilde{\beta}_{ij}^{\mathrm{p}}(\mathbf{k})\widetilde{\beta}_{mn}^{\mathrm{p}}(-\mathbf{k})\right. \\
&\left. +\widetilde{\sigma}_{ij}^{\mathrm{ext}}(\mathbf{k})\widetilde{\beta}_{ij}^{\mathrm{p}}(-\mathbf{k})\right).
\end{aligned} \tag{3.21}
$$

3.2.2 Traditional dissipative continuum dynamics

There are well known approaches for deriving continuum equations of motion for dissipative systems, which in this case produce a traditional von Mises-style theory [208], useful at longer scales. We begin by reproducing these standard equations.

For the sake of simplicity, we ignore external stress (σ_{ij} simplified to σ_{ij}^{int}) in the following three subsections. We start by using the standard methods applied to the non-conserved order parameter β^{p}, and then turn to the conserved order parameter ϱ.

Dissipative dynamics built from the non-conserved order parameter field β^{p}

The plastic distortion β^{p} is a non-conserved order parameter field, which is utilized by the engineering community to study texture evolution and plasticity of mechanically deformed structural materials. The simplest dissipative dynamics in terms of β^{p} minimizes the free energy by steepest descents

$$\frac{\partial}{\partial t}\beta_{ij}^{\text{p}} = -\Gamma\frac{\delta\mathcal{F}}{\delta\beta_{ij}^{\text{p}}}, \tag{3.22}$$

where Γ is a positive material-dependent constant. We may rewrite it in Fourier space, giving

$$\frac{\partial}{\partial t}\widetilde{\beta}_{ij}^{\text{p}}(\mathbf{k}) = -\Gamma\frac{\delta\mathcal{F}}{\delta\widetilde{\beta}_{ij}^{\text{p}}(-\mathbf{k})}. \tag{3.23}$$

The functional derivative $\delta\mathcal{F}/\delta\widetilde{\beta}_{ij}^{\text{p}}(-\mathbf{k})$ is the negative of the long-range stress

$$\frac{\delta\mathcal{F}}{\delta\widetilde{\beta}_{ij}^{\text{p}}(-\mathbf{k})} = -M_{ijmn}(\mathbf{k})\widetilde{\beta}_{mn}^{\text{p}}(\mathbf{k}) \equiv -\widetilde{\sigma}_{ij}(\mathbf{k}). \tag{3.24}$$

This dynamics implies a simplified version of von Mises plasticity

$$\frac{\partial}{\partial t}\widetilde{\beta}_{ij}^{\text{p}}(\mathbf{k}) = \Gamma\widetilde{\sigma}_{ij}(\mathbf{k}). \tag{3.25}$$

Dissipative dynamics built from the conserved order parameter field ϱ

We can also derive an equation of motion starting from the GND density ϱ, as was done by Rickman and Viñals [208]. For this dissipative dynamics Eq. (3.16), the

simplest expression for J is

$$J_{qk} = -\Gamma'_{ablq}\partial_l\frac{\delta\mathcal{F}}{\delta\varrho_{abk}}, \tag{3.26}$$

where the material-dependent constant tensor Γ' must be chosen to guarantee a decrease of the free energy with time.

The infinitesimal change of $\mathcal{F}$ with respect to the GND density ϱ is

$$\delta\mathcal{F}[\varrho] = \int d^3\mathbf{x}\,\frac{\delta\mathcal{F}}{\delta\varrho_{ijk}}\delta\varrho_{ijk}. \tag{3.27}$$

The free energy dissipation rate is thus $\delta\mathcal{F}/\delta t$ for $\delta\varrho = \frac{\partial\varrho}{\partial t}\delta t$, hence

$$\frac{\partial}{\partial t}\mathcal{F}[\varrho] = \int d^3\mathbf{x}\,\frac{\delta\mathcal{F}}{\delta\varrho_{ijk}}\frac{\partial\varrho_{ijk}}{\partial t}. \tag{3.28}$$

Substituting Eq. (3.16) into Eq. (3.28) and integrating by parts gives

$$\frac{\partial}{\partial t}\mathcal{F}[\varrho] = \int d^3\mathbf{x}\left(g_{ijpq}\partial_p\frac{\delta\mathcal{F}}{\delta\varrho_{ijk}}\right)J_{qk}. \tag{3.29}$$

Substituting Eq. (3.26) into Eq. (3.29) gives

$$\frac{\partial}{\partial t}\mathcal{F}[\varrho] = -\int d^3\mathbf{x}\left(g_{ijpq}\partial_p\frac{\delta\mathcal{F}}{\delta\varrho_{ijk}}\right)\left(\Gamma'_{ablq}\partial_l\frac{\delta\mathcal{F}}{\delta\varrho_{abk}}\right). \tag{3.30}$$

Now, to guarantee that energy never increases, we choose $\Gamma'_{ablq} = \Gamma g_{ablq}$, ($\Gamma$ is a positive material-dependent constant), which yields the rate of change of energy as a negative of a perfect square

$$\frac{\partial}{\partial t}\mathcal{F}[\varrho] = -\int d^3\mathbf{x}\,\Gamma\sum_{q,k}\left(g_{ablq}\partial_l\frac{\delta\mathcal{F}}{\delta\varrho_{abk}}\right)^2. \tag{3.31}$$

Using Eqs. (3.16) and (3.26), we can write the dynamics in terms of ϱ

$$\frac{\partial}{\partial t}\varrho_{ijk} = \Gamma g_{ijpq}g_{ablq}\partial_p\partial_l\frac{\delta\mathcal{F}}{\delta\varrho_{abk}}. \tag{3.32}$$

Substituting the functional derivative $\delta\mathcal{F}/\delta\varrho_{abk}$, Eq. (3.102), into Eq. (3.32) and comparing to Eq. (3.16) tells us

$$\frac{\partial}{\partial t}\varrho_{ijk}(\mathbf{x}) = -\Gamma g_{ijpq}\partial_p\sigma_{qk}(\mathbf{x}) = -g_{ijpq}\partial_p J_{qk}(\mathbf{x}), \tag{3.33}$$

where

$$J_{qk} = \Gamma \sigma_{qk} \tag{3.34}$$

duplicating the von Mises law (Eq. 3.25) of the previous subsection. The simplest dissipative dynamics of either non-conserved or conserved order parameter fields thus turns out to be the traditional linear dynamics, a simplified von Mises law.

The problem with this law for us is that it allows for plastic deformation in the absence of dislocations, i.e., the Burgers vector flux can be induced through the elastic loading on the boundaries, even in a defect-free medium. This is appropriate on engineering length scales above or around a micron, where statistically stored dislocation (SSD) dominates the plastic deformation. (Methods to incorporate their effects into a theory like ours have been provided by Acharya *et al.* [11, 213] and Varadhan *et al.* [254])

By ignoring the SSDs, our theory assumes that there is an intermediate coarse-graining length scale, large compared to the distance between dislocations and small compared to the distance where the cancelling of dislocations with different Burger's vectors dominates the dynamics. We believe this latter length scale is given by the distance between cell walls (as discussed in Sec. 3.3.2). The cell wall misorientations are geometrically necessary. On the one hand, it is known [137, 120] that neighboring cell walls often have misorientations of alternating signs, so that on coarse-grained length scales just above the cell wall separation one would expect explicit treatment of the SSDs would be necessary. On the other hand, the density of dislocations in cell walls is high, so that a coarse-graining length much smaller than the interesting structures (and hence where we believe SSDs are unimportant) should be possible. (Our cell structures are fractal, with no characteristic 'cell size'; this coarse-graining length sets the minimum cutoff scale of the fractal, and

the grain size or inhomogeneity length will set the maximum scale.) With this assumption, to treat the formation of cellular structures, we turn to theories of the form given in Eq. (3.15), defined in terms of dislocation currents J that depend directly on the local GND density.

3.2.3 Our CDD model

The microscopic motion of a dislocation under external strain depends upon temperature. In general, it moves quickly along the glide direction, and slowly (or not at all) along the climb direction where vacancy diffusion must carry away the atoms. The glide speed can be limited by phonon drag at higher temperatures, or can accelerate to nearly the speed of sound at low temperatures [113]. It is traditional to assume that the dislocation velocity is over-damped, and proportional to the component of the force per unit dislocation length in the glide plane.

To coarse-grain this microscopics, for reasons described above, we choose a CDD model whose dislocation currents vanish when the GND density vanishes. Limkumnerd and Sethna [155] derived a dislocation current J for this case using a closure approximation of the underlying microscopics. Their work reproduced (in the case of both glide and climb) an earlier dynamical model proposed by Acharya and collaborators [8, 212] assuming a single velocity field for the dislocations.

In our CGD and GOD-LVP dynamics (Sections 3.2.3 and 3.2.3 below), we also assume that all dislocations in the infinitesimal volume at $\mathbf{x}$ are moving with a common velocity $\boldsymbol{v}(\mathbf{x})$. This common velocity ansatz was first mentioned by Mura [176], but as a warning that it is an incorrect assumption. Microscopically, different dislocations in a region experience Peach-Koehler forces in different di-

rections, and will not move in tandem. (In real materials the dislocation dynamics is intermittent, as dislocations bow out or depin from junctions and disorder, and engage in complex dislocation avalanches. Our model has no pinning and hence no metastability: the single velocity approximation is an additional assumption.) Indeed, it is the difference in velocities for dislocations on different slip systems that leads to entanglement. It is also not justified in a coarse-grained theory (unlike fluid mechanics, where momentum conservation and Galilean invariance leads to an emergent collective local velocity for systems in local equilibrium). However, we shall see that variants of the local velocity ansatz allow one to construct physically sensible 'model materials' – perhaps not the correct theory for a particular material, but a sensible framework to generate theories of plastic deformation. This ansatz has been supplemented by constitutive laws for the velocity field by Acharya [8] and collaborators [9, 10, 212, 11, 254] to generate CDD theories. We follow their argument in Sec. 3.2.3 to derive the dynamics allowing both glide and climb, and then modify it to remove climb in Sec. 3.2.3. We also derive a second variant version of glide-only dynamics in Sec. 3.2.3 by approaching the limit of infinite vacancy energy, which reproduces a model proposed by Acharya and Roy [11].

Climb-glide dynamics (CGD)

We start with a model presuming (perhaps unphysically) that vacancy diffusion is so fast that dislocations climb and glide with equal mobility. The elastic Peach-Koehler force due to the stress $\sigma(\mathbf{x})$ on the local GND density is given by $f_u^{PK} = \sigma_{mk}\varrho_{umk}$. We assume that the velocity $\boldsymbol{v} \propto \boldsymbol{f}^{PK}$, giving a local constitutive relation

$$v_u \propto \sigma_{mk}\varrho_{umk}. \tag{3.35}$$
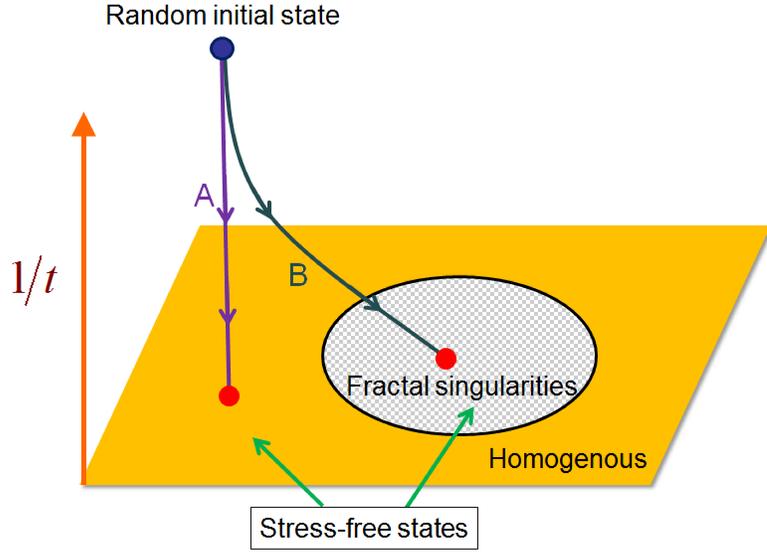
Figure 3.2: **Relaxation of various CDD models**. The blue dot represents the initial random plastically-deformed state; the red dots indicate the equilibrated stress-free states driven by different dynamics. Curve A: steepest decent dynamics leads to the trivial homogeneous equilibrated state, discussed in Sec. 3.2.2. Curve B: our CDD models settle the system into non-trivial stress-free states with wall-like singularities of the GND density, discussed in Sec. 3.2.3.

How should we determine the proportionality constant between velocity and force? In experimental systems, this is complicated by dislocation entanglement and short-range forces between dislocations. Ignoring these features, the velocity of each dislocation should depend only on the stress induced by the other dislocations, not the local density of dislocations [4]. We can incorporate this in an approximate way by making the proportionality factor in Eq. (3.35) inversely proportional to the GND density. We measure the latter by summing the square of all components of ϱ, hence $|\varrho| = \sqrt{\varrho_{ijk}\varrho_{ijk}/2}$ and $v_u = \frac{D}{|\varrho|}\sigma_{mk}\varrho_{umk}$, where D is a positive material-dependent constant. This choice has the additional important feature that the evolution of a sharp domain wall whose width is limited by the lattice cutoff is unchanged when the lattice cutoff is reduced.

The flux J of the Burgers vector is thus

$$J_{ij} = v_u \varrho_{uij} = \frac{D}{|\varrho|} \sigma_{mk} \varrho_{umk} \varrho_{uij}. \tag{3.36}$$

Notice that this dynamics satisfies our criterion that $J = 0$ when there are no GNDs (i.e., $\varrho = 0$).

Substituting this flux J (Eq. 3.36) into the free energy dissipation rate (Eq. 3.108) gives

$$\frac{\partial \mathcal{F}}{\partial t} = -\int d^3\mathbf{x} \; \sigma_{ij} J_{ij} = -\int d^3\mathbf{x} \; \frac{|\varrho|}{D} v^2 \leq 0. \tag{3.37}$$

Glide-only dynamics: mobile dislocation population (GOD-MDP)

When the temperature is low enough, dislocation climb is negligible, i.e., dislocations can only move in their glide planes. Fundamentally, dislocation glide conserves the total number of atoms, which leads to an unchanged local volume. Since the local volume change in time is represented by the trace J_{ii} of the flux of the Burgers vector, conservative motion of GNDs demands $J_{ii} = 0$. Limkumnerd and Sethna [155] derived the equation of motion for dislocation glide only, by removing the trace of J from Eq. (3.36). However, their dynamics fails to guarantee that the free energy monotonically decreases. Here we present an alternative approach.

We can remove the trace of J by modifying the first equality in Eq. (3.36),

$$J'_{ij} = v'_u \left(\varrho_{uij} - \frac{1}{3} \delta_{ij} \varrho_{ukk} \right), \tag{3.38}$$

where $\varrho'_{uij} = \varrho_{uij} - \frac{1}{3} \delta_{ij} \varrho_{ukk}$ can be viewed as a subset of 'mobile' dislocations moving with velocity $\boldsymbol{v}'$.

Substituting the current (Eq. 3.38) into the free energy dissipation rate

81

(Eq. 3.108) gives

$$\frac{\partial \mathcal{F}}{\partial t} = -\int d^3\mathbf{x}\; \sigma_{ij}\left(v_u' \varrho_{uij}'\right). \tag{3.39}$$

If we choose the velocity $v_u' \propto \sigma_{ij}\varrho_{uij}'$, the appropriate free energy monotonically decreases in time. We thus express $v_u' = \frac{D}{|\varrho|}\varrho_{uij}'\sigma_{ij}$, where D is a positive material-dependent constant, and the prefactor $1/|\varrho|$ is added for the same reasons, as discussed in the second paragraph of Sec. 3.2.3.

The current J' of the Burgers vector is thus written [52]

$$
\begin{aligned}
J_{ij}' &= v_u' \varrho_{uij}' \\
&= \frac{D}{|\varrho|}\sigma_{mn}\left(\varrho_{umn} - \frac{1}{3}\delta_{mn}\varrho_{ull}\right)\left(\varrho_{uij} - \frac{1}{3}\delta_{ij}\varrho_{ukk}\right).
\end{aligned}
\tag{3.40}
$$

This natural evolution law becomes much less self-evident when expressed in terms of the traditional two-index version ρ (Eqs. 3.1&3.2)

$$
\begin{aligned}
J_{ij}' &= \frac{D}{|\varrho|}\bigg(\sigma_{in}\rho_{mn}\rho_{mj} - \sigma_{mn}\rho_{in}\rho_{mj} - \frac{1}{3}\sigma_{mm}\rho_{ni}\rho_{nj} \\
&\quad + \frac{1}{3}\sigma_{mm}\rho_{in}\rho_{nj} - \frac{\delta_{ij}}{3}\Big(\sigma_{kn}\rho_{mn}\rho_{mk} - \sigma_{mn}\rho_{kn}\rho_{mk} \\
&\quad - \frac{1}{3}\sigma_{mm}\rho_{nk}\rho_{nk} + \frac{1}{3}\sigma_{mm}\rho_{kn}\rho_{nk}\Big)\bigg),
\end{aligned}
\tag{3.41}
$$

(which is why we introduce the three-index variant ϱ).

This current J' makes the free energy dissipation rate the negative of a perfect square in Eq. (3.110).

Glide-only dynamics: local vacancy-induced pressure (GOD-LVP)

At high temperature, the fast vacancy diffusion leads to dislocation climb out of the glide direction. As the temperature decreases, vacancies are frozen out so that

82

dislocations only slip in the glide planes. In Sec. 3.8.1, we present a dynamical model coupling the vacancy diffusion to our CDD model. Here we consider the limit of frozen-out vacancies with infinite energy costs, which leads to another version of glide-only dynamics.

According to the coupling dynamics Eq. (3.118), we write down the general form of dislocation current

$$J''_{ij} = \frac{D}{|\varrho|}\left(\sigma_{mn} - \delta_{mn}p\right)\varrho_{umn}\varrho_{uij},$$ (3.42)

where p is the local pressure due to vacancies.

The limit of infinitely costly vacancies ($\alpha \to \infty$ in Sec. 3.8.1) leads to the traceless current, $J''_{ii} = 0$. Solving this equation gives a critical local pressure p^c

$$p^c = \frac{\sigma_{pq}\varrho_{spq}\varrho_{skk}}{\varrho_{uaa}\varrho_{ubb}}.$$ (3.43)

The corresponding current J'' of the Burgers vector in this limit is thus written

$$J''_{ij} = \frac{D}{|\varrho|}\left(\sigma_{mn} - \frac{\sigma_{pq}\varrho_{spq}\varrho_{skk}}{\varrho_{uaa}\varrho_{ubb}}\delta_{mn}\right)\varrho_{umn}\varrho_{uij},$$ (3.44)

reproducing the glide-only dynamics proposed by Acharya and Roy [11].

Substituting the current (Eq. 3.44) into the free energy dissipation rate (Eq. 3.108) gives

$$\frac{\partial \mathcal{F}}{\partial t} = -\int d^3\mathbf{x}\frac{D}{|\varrho|}\left[f_i^{PK}f_i^{PK} - \left(\frac{d_i f_i^{PK}}{|\mathbf{d}|}\right)^2\right] \leq 0,$$ (3.45)

where $f_i^{PK} = \sigma_{mn}\varrho_{imn}$ and $d_i = \varrho_{ikk}$. The equality emerges when the force $\mathbf{f}^{PK}$ is along the same direction as $\mathbf{d}$.

Unlike the traditional linear dissipative models, our CDD model, coarse grained from microscopic interactions, drives the random plastic distortion to non-trivial

stress-free states with dislocation wall singularities, as schematically illustrated in Fig. 3.2.

Our minimal CDD model, consisting of GNDs evolving under the long-range interaction, provides a framework for understanding dislocation morphologies at the mesoscale. Eventually, it can be extended to include vacancies by coupling them to the dislocation current (as discussed in Sec. 3.8.1, or extended to include disorder, dislocation pinning, and entanglement by adding appropriate interactions to the free energy functional and refining the effective stress field. It has already been extended to include SSDs incorporating traditional crystal plasticity theories [11, 213, 254].

3.3   Results

3.3.1   Two and three dimensional simulations

We perform simulations in 2D and 3D for the dislocation dynamics of Eq. (3.15) and Eq. (3.18), with dynamical currents defined by CGD (Eq. 3.36), GOD-MDP (Eq. 3.40), and GOD-LVP (Eq. 3.44). We numerically observe that simulations of Eqs. (3.15), (3.18) lead to the same results statistically (i.e., the numerical time step approximations leave the physics invariant). We therefore focus our presentation on the results of Eq. (3.18), where the evolving field variable β^{p} is unconstrained. Our CGD and GOD-MDP models have been quite extensively simulated in one and two dimensions and relevant results can be found in Refs. 52, 155, and 154. In this paper, we concentrate on periodic grids of spatial extent L in both two [52] and three dimensions. The numerical approach we

use is a second-order central upwind scheme designed for Hamilton-Jacobi equations [138] using finite differences. This method is quite efficient in capturing $\delta-$shock singular structures [55], even though it is flexible enough to allow for the use of approximate solvers near the singularities.

Our numerical simulations show a close analogy to those of turbulent flows [55]. As in three-dimensional turbulence, defect structures lead to intermittent transfer of morphology to short length scales. As conjectured [203, 202] for the Euler equations or the inviscid limit of Navier-Stokes equations, our simulations develop singularities in finite time [52, 155]. Here these singularities are δ-shocks representing grain-boundary-like structures emerging from the mutual interactions among mobile dislocations [56]. In analogy with turbulence, where the viscosity serves to smooth out the vortex-stretching singularities of the Euler equations, we have explored the effects of adding an artificial viscosity term to our equations of motion [55]. In the presence of artificial viscosity, our simulations exhibit nice numerical convergence in all dimensions [56]. However, in the limit of vanishing viscosity, the solutions of our dynamics continue to depend on the lattice cutoff in higher dimensions, (our simulations only exhibit numerical convergence in one dimension). Actually, the fact that the physical system is cut off by the atomic scale leads to the conjecture that our equations are in some sense non-renormalizable in the ultraviolet. These issues are discussed in detail in Refs. 55 and 56.

In the vanishing viscosity limit, our simulations exhibit fractal structure down to the smallest scales. When varying the system size continuously, the solutions of our dynamics exhibit a convergent set of correlation functions of the various order parameter fields, which are used to characterize the emergent self-similarity.

In both two and three dimensional simulations, we relax the deformed sys-
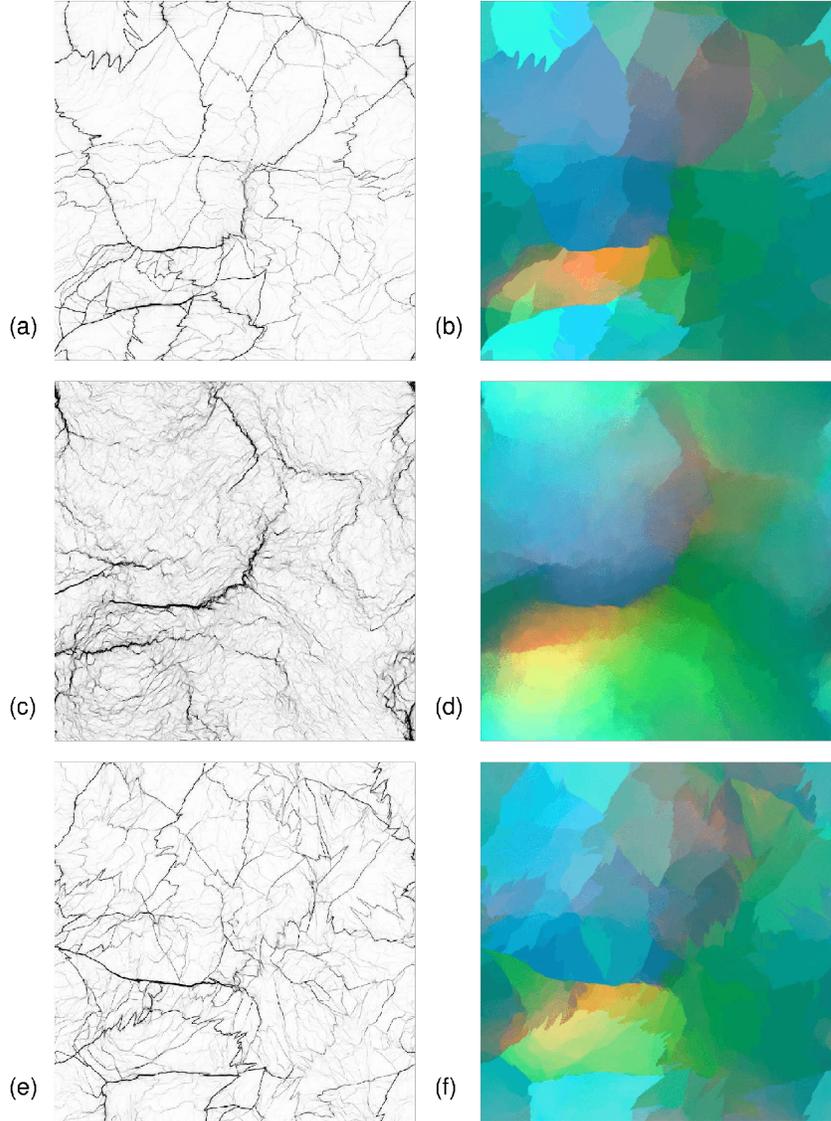
Figure 3.3: **Complex dislocation structures in two dimensions** (1024^2) for the relaxed states of an initially random distortion. *Top:* Dislocation climb is allowed; *Middle:* Glide only using a mobile dislocation population; *Bottom:* Glide only using a local vacancy pressure. *Left:* Net GND density $|\varrho|$ plotted linearly in density with dark regions a factor $\sim 10^4$ more dense than the lightest visible regions. (a) When climb is allowed, the resulting morphologies are sharp, regular, and close to the system scale. (c) When climb is forbidden using a mobile dislocation population, there is a hierarchy of walls on a variety of length scales, getting weaker on finer length scales. (e) When climb is removed using a local vacancy pressure, the resulting morphologies are as sharp as those (a) allowing climb. *Right:* Corresponding local crystalline orientation maps, with the three components of the orientation vector **Λ** linearly mapped onto a vector of RGB values. Notice the fuzzier cell walls (c) and (d) suggests a larger fractal dimension.
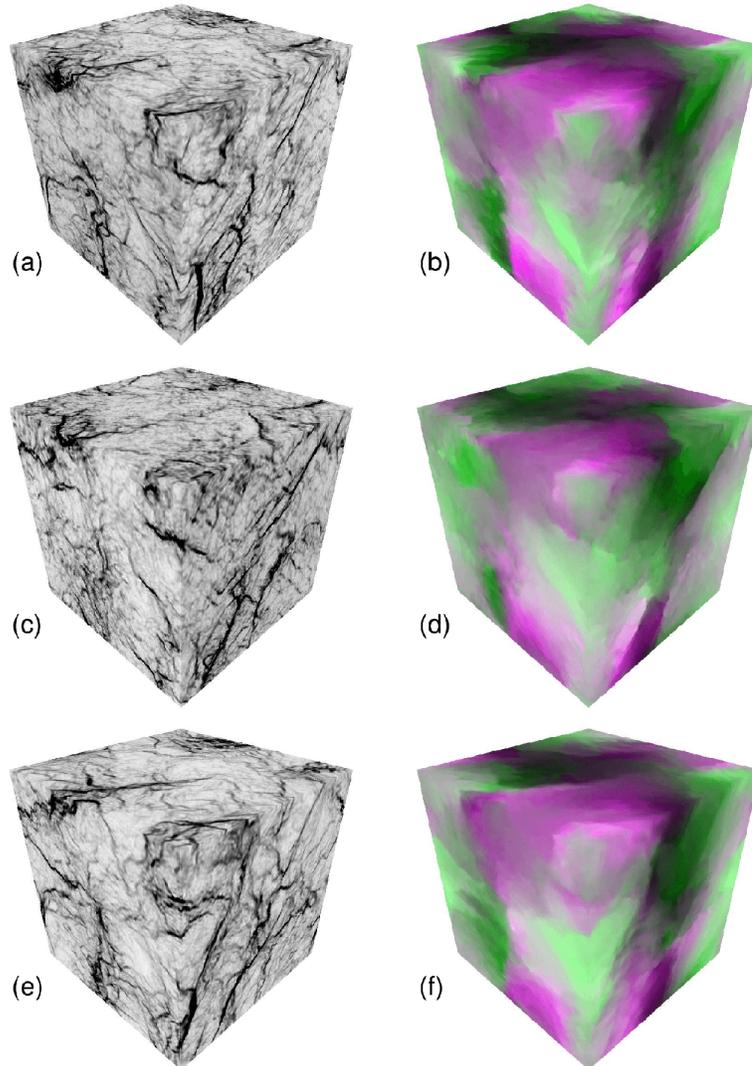
Figure 3.4: **Complex dislocation structures in three dimensions** (128^3) for the relaxed states of an initially random distortion. Notice these textured views on the surface of simulation cubes. *Top:* Dislocation climb is allowed; *Middle:* Glide only using a mobile dislocation population; *Bottom:* Glide only using a local vacancy pressure. *Left:* Net GND density $|\varrho|$ plotted linearly in density with dark regions a factor $\sim 10^3$ more dense than the lightest visible regions. The cellular structures in (a), (c), and (e) seem similarly fuzzy; our theory in three dimensions generates fractal cell walls. *Right:* Corresponding local crystalline maps, with the three components of the orientation vector $\boldsymbol{\Lambda}$ linearly mapped onto a vector of RGB values.
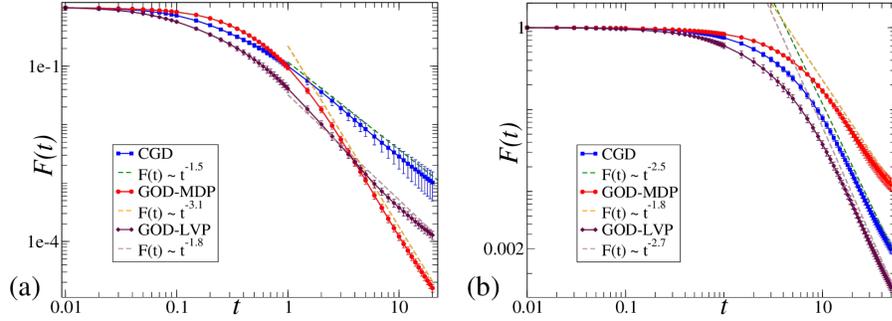
Figure 3.5: **The elastic free energy decreases to zero as a power law in time in both two and three dimensions**. In both (a) and (b), we show that the free energy $\mathcal{F}$ decays monotonically in time, and goes to zero as a power law for CGD, GOD-MDP, and GOD-LVP simulations, as the system relaxes in the absence of external strain.

tem with and without dislocation climb in the absence of external loading. Here, the initial plastic distortion field β^{p} is still a Gaussian random field with correlation length scale $\sqrt{2}L/5 \sim 0.28L$ and initial amplitude $\beta_0 = 1$. (In our earlier work [52], we described this length as $L/5$, using a non-standard definition of correlation length scale.) In 2D, Figure 3.3 shows that CGD and GOD-LVP simulations (top and bottom) exhibit much sharper, flatter boundaries than GOD-MDP (middle). This difference is quantitatively described by the large shift in the static critical exponent η in 2D for both CGD and GOD-LVP. In our earlier work [52], we announced this difference as providing a sharp distinction between high-temperature, non-fractal grain boundaries (for CGD), and low-temperature, fractal cell wall structures (for GOD-MDP). This appealing message did not survive the transition to 3D; Figure 3.4 shows basically indistinguishable complex cellular structures, for all three types of dynamics. Indeed, Table 3.1 shows only a small change in critical exponents, among CGD, GOD-MDP, and GOD-LVP. During both two and three dimensional relaxations, their appropriate free energies monotonically decay to zero as shown in Fig. 3.5.
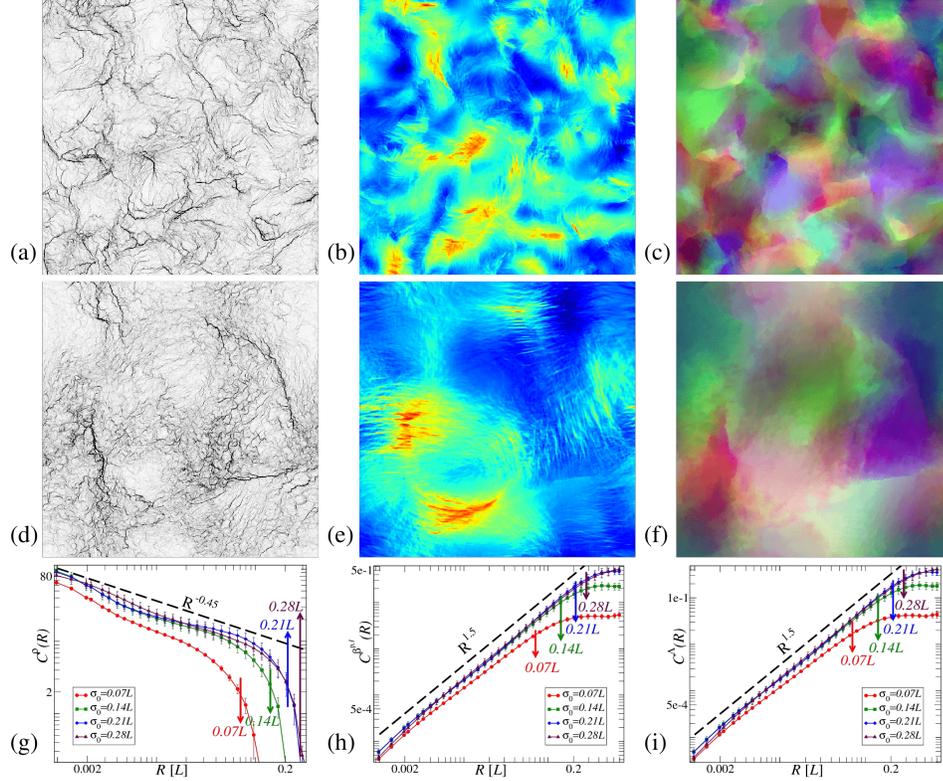
Figure 3.6: **Relaxation with various initial length scales in two dimensions**. GNDs are not allowed to climb due to the constraint of a mobile dislocation population in these simulations. (a), (b), and (c) are the net GND density map $|\varrho|$, the net plastic distortion $|\beta^{\mathrm{p}}|$ (the warmer color indicating the larger distortion), and the crystalline orientation map in a fully-relaxed state evolved from an initial random plastic distortion with correlated length scale $0.07L$. They are compared to the same sequence of plots, (d), (e), and (f), which are in the relaxed state with the initial length scale $0.21L$ three times as long. Notice the features with the longest wave length reflecting the initial distortion length scales. (g), (h), and (i) are the scalar forms (discussed in Sec. 3.3.3) of correlation functions of the GND density ρ, the intrinsic plastic distortion $\beta^{\mathrm{p,I}}$, and the crystalline orientation $\mathbf{\Lambda}$ for well-relaxed states with initial length scales varying from $0.07L$ to $0.28L$. They exhibit power laws independent of the initial length scales, with cutoffs set by the initial lengths. (The scaling relation among their critical exponents will be discussed in Sec. 3.4.)

89
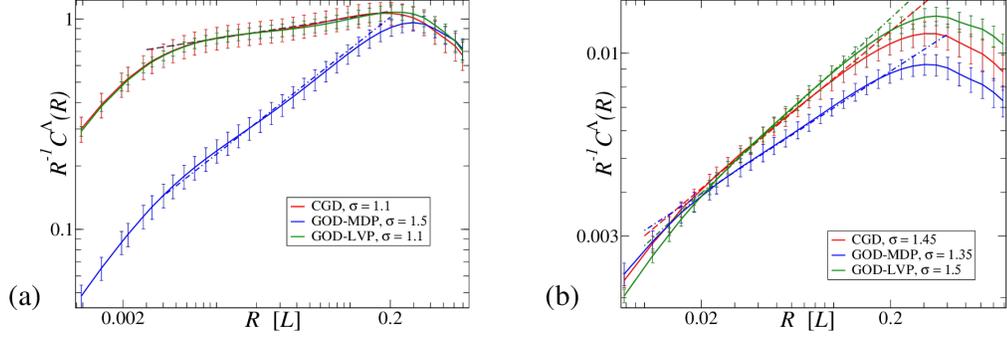
Figure 3.7: **Correlation functions of Λ in both two and three dimensions**. In (a) and (b), red, blue, and green lines indicate CGD, GOD-MDP, and GOD-LVP simulations, respectively. *Left:* Correlation functions of **Λ** are measured in relaxed, unstrained 1024^2 systems; *Right:* These correlation functions are measured in relaxed, unstrained 128^3 systems. All dashed lines show estimated power laws quoted in Table 3.1.
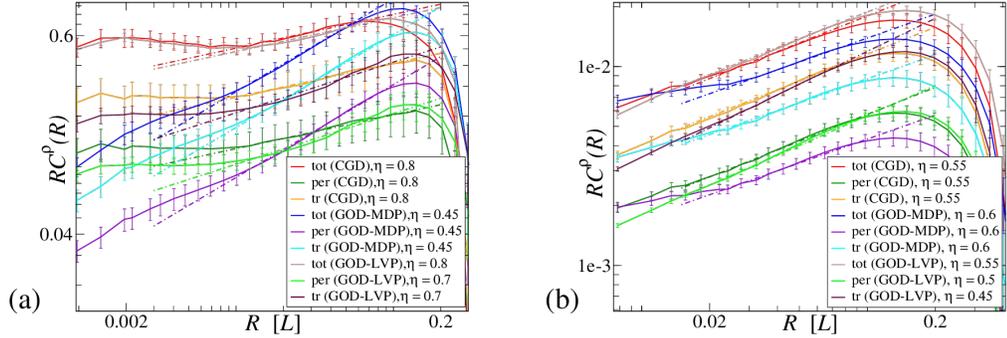
Figure 3.8: **Correlation functions of ϱ in both two and three dimensions**. *Left:* (a) is measured in relaxed, unstrained 1024^2 systems; *Right:* (b) is measured in relaxed, unstrained 128^3 systems. All dashed lines show estimated power laws quoted in Table 3.1. Notice all three scalar forms of the correlation functions of GND density share the same power law.

3.3.2 Self-similarity and initial conditions

Self-similar structures, as emergent collective phenomena, have been studied in mesoscale crystals [52], human-scale social network [239], and the astronomical-scale universe [255]. In some models [255], the self-similarity comes from *scale-free* initial conditions with a power-law spectrum [194, 60]. In our CDD model, our simulations start from a random plastic distortion with a Gaussian distribution

characterized by a single length scale. The scale-free dislocation structure sponta-
neously emerges as a result of the deterministic dynamics.

Our Gaussian random initial condition is analogous to hitting a bulk material
randomly with a hammer. The hammer head (the dent size scale) corresponds to
the correlated length. We need to generate inhomogeneous deformations like ran-
dom dents, because our theory is deterministic and hence uniform initial conditions
under uniform loading will not develop patterns.

We have considered alternatives to our imposition of Gaussian random defor-
mation fields as initial conditions. (a) As an alternative to random initial defor-
mations, we could have imposed a more regular (albeit nonuniform) deformation –
starting with our material bent into a sinusoidal arc, and then letting it relax. Such
simulations produce more symmetric versions of the fractal patterns we see; indeed,
our Gaussian random initial deformations have correlation lengths 'hammer size'
comparable to the system size, so our starting deformations are almost sinusoidal
(although different components have different phases). (b) To explore the effects
of multiple uncorrelated random domains (multiple small dents), we reduce the
Gaussian correlation length as shown in Fig. 3.6. We find that the initial-scale de-
formation determines the maximal cutoff for the fractal correlations in our model.
In other systems (such as two-dimensional turbulence) one can observe an 'inverse
cascade' with fractal structures propagating to long length scales; we observe no
evidence of these here. (c) As an alternative to imposing an initial plastic defor-
mation field and then relaxing, we have explored deforming the material slowly
and continuously in time. Our preliminary 'slow hammering' explorations turn
the Gaussian initial conditions $\beta^{\mathrm{p}0}$ into a source term, modifying Eq. 3.18 with an
additional term to give $\partial_t \beta^{\mathrm{p}}_{ij} = J_{ij} + \beta^{\mathrm{p}0}{}_{ij}/\tau$. Our early explorations suggest that

slow hammering simulations will be qualitatively compatible with the relaxation of an initial rapid hammering. In this paper, to avoid the introduction of the hammering time scale τ, we focus on the (admittedly less physically motivated) relaxation behavior.

In real materials, initial grain boundaries, impurities, or sample sizes, can be viewed as analogies to our initial dents — explaining the observation of dislocation cellular structures both in single crystals and polycrystalline materials.

Figure 3.6 shows relaxation without dislocation climb (due to the constraint of a mobile dislocation population) at various initial length scales in 2D. From Fig. 3.6(a) to (f), the net GND density, the net plastic distortion, and the crystalline orientation map, measured at two well-relaxed states evolved from different random distortions, all show fuzzy fractal structures, distinguished only by their longest-length-scale features that originate from the initial conditions. In Fig. 3.6(g), (h), and (i), the correlation functions of the GND density ρ, the intrinsic plastic distortion $\beta^{\mathrm{p,I}}$, and the crystalline orientation $\mathbf{\Lambda}$ are applied to characterize the emergent self-similarity, as discussed in the following section 3.3.3. They all exhibit the same power law, albeit with different cutoffs due to the initial conditions.

3.3.3   Correlation functions

Hierarchical dislocation structures have been observed both experimentally [128, 174, 224, 252] and in our simulations [52]. Early work analyzed experimental cellular structures using the fractal box counting method [99] or by separating the systems into cells and analyzing their sizes and misorientations [122, 119, 121,

168]. In our previous publication, we analyzed our simulated dislocation patterns using these two methods, and showed broad agreement with these experimental analyses [52]. In fact, lack of the measurements of physical order parameters leads to incomplete characterization of the emergent self-similarity [5]. We will not pursue these methods here.

In our view, the emergent self-similarity should best be exhibited by the correlation functions of the order parameter fields, such as the GND density ρ, the plastic distortion β^{p}, and the crystalline orientation vector $\boldsymbol{\Lambda}$. Here we focus on scalar invariants of the various tensor correlation functions.

For the vector correlation function $\mathcal{C}_{ij}^{\boldsymbol{\Lambda}}(\mathbf{x})$ (Eq. 3.46), only the sum $\mathcal{C}_{ii}^{\boldsymbol{\Lambda}}(\mathbf{x})$ is a scalar invariant under three dimensional rotations. For the tensor fields ρ and β^{p}, their two-point correlation functions are measured in terms of a complete set of three independent scalar invariants, which are indicated by 'tot' (total), 'per' (permutation), and 'tr' (trace). In searching for the explanation of the lack of scaling [52] for β^{p} (see Sec. 3.3.3), we checked whether these independent invariants might scale independently. In fact, most of them share a single underlying critical exponent, except for the trace-type scalar invariant of the correlation function of $\beta^{\mathrm{p,I}}$, which go to a constant in well-relaxed states, as discussed in Sec. 3.4.1.

Correlation function of crystalline orientation field

As dislocations self-organize themselves into complex structures, the relative differences of the crystalline orientations are correlated over a long length scale.

For a vector field, like the crystalline orientation $\boldsymbol{\Lambda}$, the natural two-point

correlation function is

$$
\begin{aligned}
\mathcal{C}_{ij}^{\boldsymbol{\Lambda}}(\mathbf{x}) &= \langle (\Lambda_i(\mathbf{x}) - \Lambda_i(0))(\Lambda_j(\mathbf{x}) - \Lambda_j(0)) \rangle \\
&= 2\langle \Lambda_i \Lambda_j \rangle - 2\langle \Lambda_i(\mathbf{x})\Lambda_j(0) \rangle.
\end{aligned}
\tag{3.46}
$$

Note that we correlate changes in $\boldsymbol{\Lambda}$ between two points. Just as for the height-height correlation function in surface growth [46], adding a constant to $\boldsymbol{\Lambda}(\mathbf{x})$ (rotating the sample) leads to an equivalent configuration, so only differences in rotations can be meaningfully correlated.

It can be also described in Fourier space

$$
\widetilde{\mathcal{C}}_{ij}^{\boldsymbol{\Lambda}}(\mathbf{k}) = 2\langle \Lambda_i \Lambda_j \rangle (2\pi)^3 \delta(\mathbf{k}) - \frac{2}{V}\widetilde{\Lambda}_i(\mathbf{k})\widetilde{\Lambda}_j(-\mathbf{k}).
\tag{3.47}
$$

In an isotropic medium, we study the scalar invariant formed from $\mathcal{C}_{ij}^{\boldsymbol{\Lambda}}$

$$
\mathcal{C}^{\boldsymbol{\Lambda}}(\mathbf{x}) = \mathcal{C}_{ii}^{\boldsymbol{\Lambda}}(\mathbf{x}) = 2\langle \Lambda^2 \rangle - 2\langle \Lambda_i(\mathbf{x})\Lambda_i(0) \rangle.
\tag{3.48}
$$

Figure 3.7 shows the correlation functions of crystalline orientations in both 1024^2 and 128^3 simulations. The large shift in critical exponents seen in 2D (Fig. 3.7(a)) for both CGD and GOD-LVP is not observed in the fully three dimensional simulations (Fig. 3.7(b)).

Correlation function of GND density field

As GNDs evolve into δ-shock singularities, the critical fluctuations of the GND density can be measured by the two-point correlation function $\mathcal{C}^\rho(\mathbf{x})$ of the GND density, which decays as the separating distance between two sites increases. The complete set of rotational invariants of the correlation function of ρ includes three

scalar forms

$$\mathcal{C}^{\rho}_{tot}(\mathbf{x}) \;=\; \langle \rho_{ij}(\mathbf{x})\rho_{ij}(0)\rangle, \tag{3.49}$$

$$\mathcal{C}^{\rho}_{per}(\mathbf{x}) \;=\; \langle \rho_{ij}(\mathbf{x})\rho_{ji}(0)\rangle, \tag{3.50}$$

$$\mathcal{C}^{\rho}_{tr}(\mathbf{x}) \;=\; \langle \rho_{ii}(\mathbf{x})\rho_{jj}(0)\rangle. \tag{3.51}$$

Figure 3.8 shows all the correlation functions of GND density in both 1024^2 and 128^3 simulations. These three scalar forms of the correlation functions of ρ exhibit the same critical exponent η, as listed in Table 3.1. Similar to the measurements of $\mathcal{C}^{\Lambda}$, the large shift in critical exponents seen in 2D (Fig. 3.8(a)) for both CGD and GOD-LVP is not observed in the fully three dimensional simulations (Fig. 3.8(b)).
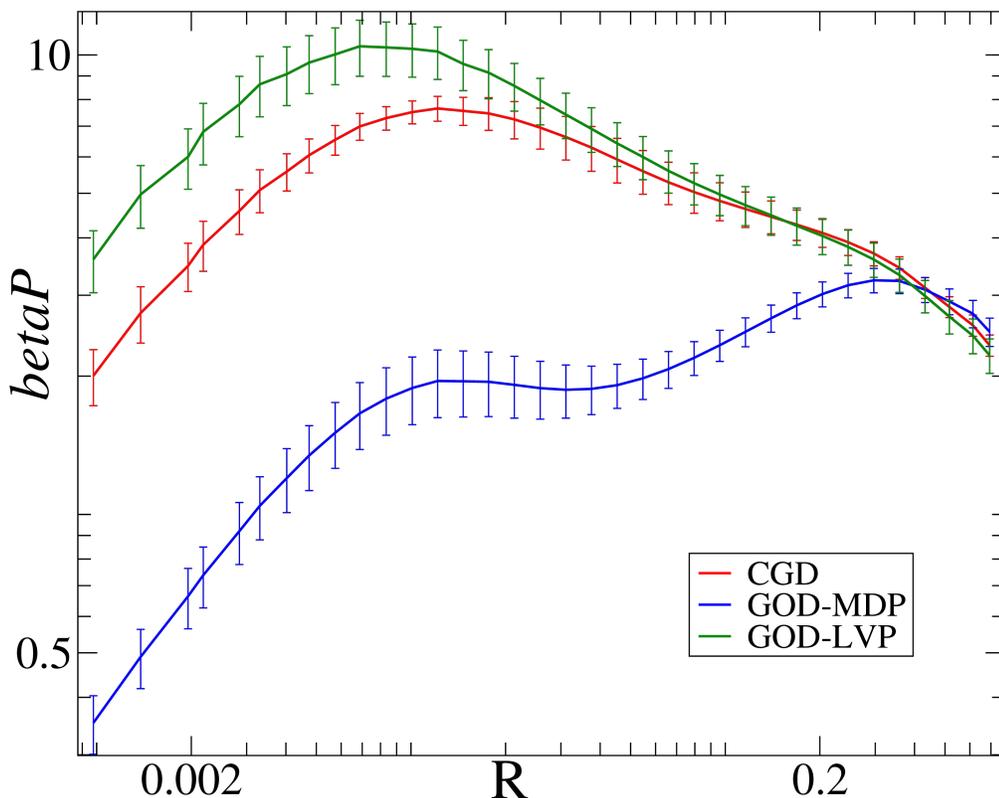
Figure 3.9: **Correlation functions of β^{p} in two dimensions**. Red, blue, and green lines indicate CGD, GOD-MDP, and GOD-LVP simulations, respectively. None of these curves shows a convincing power law.
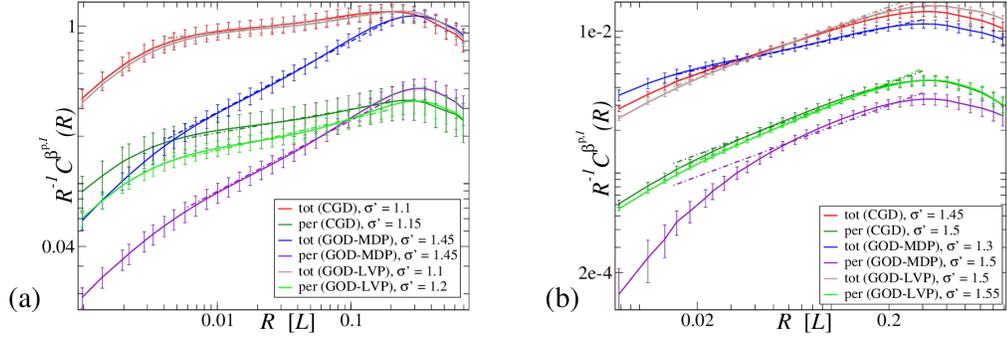
Figure 3.10: **Correlation functions of $\beta^{\mathrm{p,I}}$ in both two and three dimensions**. In (a) and (b), the correlation functions of the intrinsic part of plastic distortion field are shown. *Left:* (a) is measured in relaxed, unstrained 1024^2 systems; *Right:* (b) is measured in in relaxed, unstrained 128^3 systems. All dashed lines show estimated power laws quoted in Table 3.1. Notice that we omit the correlation functions of $\mathcal{C}_{tr}^{\beta^{\mathrm{p,I}}}$, which are independent of distance, and unrelated to the emergent self-similarity, as shown in Sec. 3.4.1.

Correlation function of plastic distortion field

The plastic distortion β^{p} is a mixture of both the divergence-free $\beta^{\mathrm{p,I}}$ and the curl-free $\beta^{\mathrm{p,H}}$. Figure 3.9 shows that β^{p} does not appear to be scale invariant, as observed in our earlier work [52]. It is crucial to study the correlations of the two physical fields, $\beta^{\mathrm{p,I}}$ and $\beta^{\mathrm{p,H}}$, separately.

Similarly to the crystalline orientation $\mathbf{\Lambda}$, we correlate the differences between $\beta^{\mathrm{p,I}}$ at neighboring points. The complete set of scalar invariants of correlation

functions of $\beta^{\mathrm{p,I}}$ thus includes the three scalar forms

$$
\begin{aligned}
\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) &= \langle(\beta_{ij}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{ij}^{\mathrm{p,I}}(0))(\beta_{ij}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{ij}^{\mathrm{p,I}}(0))\rangle \\
&= 2\langle\beta_{ij}^{\mathrm{p,I}}\beta_{ij}^{\mathrm{p,I}}\rangle - 2\langle\beta_{ij}^{\mathrm{p,I}}(\mathbf{x})\beta_{ij}^{\mathrm{p,I}}(0)\rangle; & (3.52)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{C}_{per}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) &= -\langle(\beta_{ij}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{ij}^{\mathrm{p,I}}(0))(\beta_{ji}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{ji}^{\mathrm{p,I}}(0))\rangle \\
&= -2\langle\beta_{ij}^{\mathrm{p,I}}\beta_{ji}^{\mathrm{p,I}}\rangle + 2\langle\beta_{ij}^{\mathrm{p,I}}(\mathbf{x})\beta_{ji}^{\mathrm{p,I}}(0)\rangle; & (3.53)
\end{aligned}
$$

$$
\begin{aligned}
\mathcal{C}_{tr}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) &= \langle(\beta_{ii}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{ii}^{\mathrm{p,I}}(0))(\beta_{jj}^{\mathrm{p,I}}(\mathbf{x}) - \beta_{jj}^{\mathrm{p,I}}(0))\rangle \\
&= 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle - 2\langle\beta_{ii}^{\mathrm{p,I}}(\mathbf{x})\beta_{jj}^{\mathrm{p,I}}(0)\rangle; & (3.54)
\end{aligned}
$$

where an overall minus sign is added to $\mathcal{C}_{per}^{\beta^{\mathrm{p,I}}}$ so as to yield a positive measure.

In Fig. 3.10, the correlation functions of the intrinsic plastic distortion $\beta^{\mathrm{p,I}}$ in both 1024^2 and 128^3 simulations exhibit a critical exponent σ'. These measured critical exponents are shown in Table 3.1.

3.4 Scaling theory

The emergent self-similar dislocation morphologies are characterized by the rotational invariants of correlation functions of physical observables, such as the GND density ρ, the crystalline orientation $\mathbf{\Lambda}$, and the intrinsic plastic distortion $\beta^{\mathrm{p,I}}$. Here we derive the relations expected between these correlation functions, and show that their critical exponents collapse into a single underlying one through a generic scaling theory.

In our model, the initial elastic stresses are relaxed via dislocation motion, leading to the formation of cellular structures. In the limit of slow imposed deformations, the elastic stress goes to zero in our model. We will use the absence of external stress to simplify our correlation function relations. (Some relations can

be valid regardless of the existence of residual stress.) Those relations that hold only in stress-free states will be labeled 'sf'; they will be applicable in analyzing experiments only insofar as residual stresses are small.

3.4.1 Relations between correlation functions

$\mathcal{C}^\rho$ and $\mathcal{C}^\Lambda$

For a stress-free state, we thus ignore the elastic strain term in Eq. (3.14) and write in Fourier space

$$\widetilde{\rho}_{ij}(\mathbf{k}) \overset{\text{sf}}{=} -ik_j\widetilde{\Lambda}_i(\mathbf{k}) + i\delta_{ij}k_k\widetilde{\Lambda}_k(\mathbf{k}). \tag{3.55}$$

First, we can substitute Eq. (3.55) into the Fourier-transformed form of the correlation function Eq. (3.49)

$$
\begin{aligned}
\widetilde{\mathcal{C}}^\rho_{tot}(\mathbf{k}) \quad &\overset{\text{sf}}{=} \quad \frac{1}{V}\left(-ik_j\widetilde{\Lambda}_i(\mathbf{k}) + i\delta_{ij}k_k\widetilde{\Lambda}_k(\mathbf{k})\right) \\
&\qquad \times \left(ik_j\widetilde{\Lambda}_i(-\mathbf{k}) - i\delta_{ij}k_m\widetilde{\Lambda}_m(-\mathbf{k})\right) \\
&\overset{\text{sf}}{=} \quad \frac{1}{V}(\delta_{ij}k^2 + k_ik_j)\widetilde{\Lambda}_i(\mathbf{k})\widetilde{\Lambda}_j(-\mathbf{k}).
\end{aligned}
\tag{3.56}
$$

Multiplying both sides of Eq. (3.47) by $(\delta_{ij}k^2 + k_ik_j)$ gives

$$(\delta_{ij}k^2 + k_ik_j)\widetilde{\mathcal{C}}^\Lambda_{ij}(\mathbf{k}) \overset{\text{sf}}{=} -\frac{2}{V}(\delta_{ij}k^2 + k_ik_j)\widetilde{\Lambda}_i(\mathbf{k})\widetilde{\Lambda}_j(-\mathbf{k}). \tag{3.57}$$

Comparing Eq. (3.57) and Eq. (3.56), we may write $\widetilde{\mathcal{C}}^\rho_{tot}$ in terms of $\widetilde{\mathcal{C}}^\Lambda_{ij}$ as

$$\widetilde{\mathcal{C}}^\rho_{tot}(\mathbf{k}) \overset{\text{sf}}{=} -\frac{1}{2}(\delta_{ij}k^2 + k_ik_j)\widetilde{\mathcal{C}}^\Lambda_{ij}(\mathbf{k}). \tag{3.58}$$

Second, we can substitute Eq. (3.55) into the Fourier-transformed form of the correlation function Eq. (3.50)

$$\widetilde{\mathcal{C}}^\rho_{per}(\mathbf{k}) \overset{\text{sf}}{=} \frac{2}{V}k_ik_j\widetilde{\Lambda}_i(\mathbf{k})\widetilde{\Lambda}_j(-\mathbf{k}). \tag{3.59}$$

Multiplying both sides of Eq. (3.47) by $k_i k_j$ and comparing with Eq. (3.59) gives

$$\widetilde{\mathcal{C}}_{per}^{\rho}(\mathbf{k}) \overset{\text{sf}}{=} -k_i k_j \widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}). \tag{3.60}$$

Finally, we substitute Eq. (3.55) into the Fourier-transformed form of the correlation function Eq. (3.51)

$$\widetilde{\mathcal{C}}_{tr}^{\rho}(\mathbf{k}) \overset{\text{sf}}{=} \frac{4}{V} k_i k_j \widetilde{\Lambda}_i(\mathbf{k}) \widetilde{\Lambda}_j(-\mathbf{k}). \tag{3.61}$$

Repeating the same procedure of deriving $\widetilde{\mathcal{C}}_{per}^{\rho}$, we write $\widetilde{\mathcal{C}}_{tr}^{\rho}$ in terms of $\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}$ as

$$\widetilde{\mathcal{C}}_{tr}^{\rho}(\mathbf{k}) \overset{\text{sf}}{=} -2k_i k_j \widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}). \tag{3.62}$$

Through an inverse Fourier transform, we convert Eq. (3.58), Eq. (3.60), and Eq. (3.62) back to real space to find

$$\mathcal{C}_{tot}^{\rho}(\mathbf{x}) \overset{\text{sf}}{=} \frac{1}{2}\partial^2 \mathcal{C}^{\mathbf{\Lambda}}(\mathbf{x}) + \frac{1}{2}\partial_i \partial_j \mathcal{C}_{ij}^{\mathbf{\Lambda}}(\mathbf{x}), \tag{3.63}$$

$$\mathcal{C}_{per}^{\rho}(\mathbf{x}) \overset{\text{sf}}{=} \partial_i \partial_j \mathcal{C}_{ij}^{\mathbf{\Lambda}}(\mathbf{x}), \tag{3.64}$$

$$\mathcal{C}_{tr}^{\rho}(\mathbf{x}) \overset{\text{sf}}{=} 2\partial_i \partial_j \mathcal{C}_{ij}^{\mathbf{\Lambda}}(\mathbf{x}). \tag{3.65}$$

$\mathcal{C}^{\beta^{\mathrm{p,I}}}$ and $\mathcal{C}^{\mathbf{\Lambda}}$

The intrinsic part of the plastic distortion field is directly related to the GND density field. In stress-free states, the crystalline orientation vector can fully describe the GND density. We thus can connect $\mathcal{C}^{\beta^{\mathrm{p,I}}}$ to $\mathcal{C}^{\mathbf{\Lambda}}$.

First, substituting $\widetilde{\beta}_{ij}^{\mathrm{p,I}} = -i\varepsilon_{ilm} k_l \widetilde{\rho}_{mj}/k^2$ into the Fourier-transformed form of

Eq. (3.52) gives

$$
\begin{aligned}
\widetilde{\mathcal{C}}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) &= 2\langle \beta_{ij}^{\mathrm{p,I}} \beta_{ij}^{\mathrm{p,I}} \rangle (2\pi)^3 \delta(\mathbf{k}) - \frac{2}{V} \left(-i\epsilon_{ilm} \frac{k_l}{k^2} \widetilde{\rho}_{mj}(\mathbf{k}) \right) \\
&\quad \times \left(i\epsilon_{ist} \frac{k_s}{k^2} \widetilde{\rho}_{tj}(-\mathbf{k}) \right) \\
&= 2\langle \beta_{ij}^{\mathrm{p,I}} \beta_{ij}^{\mathrm{p,I}} \rangle (2\pi)^3 \delta(\mathbf{k}) - \frac{2}{k^2} \left(\frac{1}{V} \widetilde{\rho}_{mj}(\mathbf{k}) \widetilde{\rho}_{mj}(-\mathbf{k}) \right).
\end{aligned}
\tag{3.66}
$$

During this derivation, some terms vanish due to the geometrical constraint on ρ, Eq. (3.6). Multiplying $-k^2/2$ on both sides of Eq. (3.66) and applying the Fourier-transformed form of Eq. (3.49) gives

$$
-\frac{k^2}{2} \widetilde{\mathcal{C}}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) = \widetilde{\mathcal{C}}_{tot}^{\rho}(\mathbf{k}).
\tag{3.67}
$$

In stress-free states, we can substitute Eq. (3.58) into Eq. (3.67)

$$
-\frac{k^2}{2} \widetilde{\mathcal{C}}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) \overset{\mathrm{sf}}{=} \widetilde{\mathcal{C}}_{tot}^{\rho,sf}(\mathbf{k}) = -\frac{1}{2} \left(\delta_{ij} k^2 + k_i k_j \right) \widetilde{\mathcal{C}}_{ij}^{\Lambda}(\mathbf{k}),
\tag{3.68}
$$

which is rewritten after multiplying $-2/k^2$ on both sides

$$
\widetilde{\mathcal{C}}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) \overset{\mathrm{sf}}{=} \widetilde{\mathcal{C}}^{\Lambda}(\mathbf{k}) + \frac{k_i k_j}{k^2} \widetilde{\mathcal{C}}_{ij}^{\Lambda}(\mathbf{k}).
\tag{3.69}
$$

Second, substituting $\widetilde{\beta}_{ij}^{\mathrm{p,I}} = -i\varepsilon_{ilm} k_l \widetilde{\rho}_{mj}/k^2$ into the Fourier-transformed form of Eq. (3.53) gives

$$
\begin{aligned}
\widetilde{\mathcal{C}}_{per}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) &= -2\langle \beta_{ij}^{\mathrm{p,I}} \beta_{ji}^{\mathrm{p,I}} \rangle (2\pi)^3 \delta(\mathbf{k}) + \frac{2}{V} \left(-i\epsilon_{ilm} \frac{k_l}{k^2} \widetilde{\rho}_{mj}(\mathbf{k}) \right) \\
&\quad \times \left(i\epsilon_{jst} \frac{k_s}{k^2} \widetilde{\rho}_{ti}(-\mathbf{k}) \right) \\
&= -2\langle \beta_{ij}^{\mathrm{p,I}} \beta_{ji}^{\mathrm{p,I}} \rangle (2\pi)^3 \delta(\mathbf{k}) + \frac{2}{k^2} \widetilde{\mathcal{C}}_{tot}^{\rho}(\mathbf{k}) \\
&\quad - \frac{2}{k^2} \widetilde{\mathcal{C}}_{tr}^{\rho}(\mathbf{k}) - \frac{2}{V k^4} k_i k_j \widetilde{\rho}_{mj}(\mathbf{k}) \widetilde{\rho}_{mi}(-\mathbf{k}),
\end{aligned}
\tag{3.70}
$$

where we skip straightforward but tedious expansions and the geometrical constraint on ρ, Eq. (3.6). Notice that this relation is correct even in the presence of stress.

In stress-free states, we substitute Eqs. (3.55), (3.58), (3.62) into Eq. (3.70), and ignore the constant zero wavelength term

$$
\begin{aligned}
\widetilde{\mathcal{C}}_{per}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) \;\overset{\mathrm{sf}}{=}\; & -\frac{1}{k^2}(k^2\delta_{ij}+k_ik_j)\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}) + \frac{4}{k^2}k_ik_j\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}) \\
& -\frac{2k_ik_j}{Vk^4}\left(-ik_j\widetilde{\Lambda}_m(\mathbf{k}) + i\delta_{mj}k_k\widetilde{\Lambda}_k(\mathbf{k})\right) \\
& \times\left(ik_i\widetilde{\Lambda}_m(-\mathbf{k}) - i\delta_{mi}k_n\widetilde{\Lambda}_n(-\mathbf{k})\right) \\
\overset{\mathrm{sf}}{=}\; & 2\frac{k_ik_j}{k^2}\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}).
\end{aligned}
\tag{3.71}
$$

Finally, substituting $\widetilde{\beta}_{ij}^{\mathrm{p,I}} = -i\varepsilon_{ilm}k_l\widetilde{\rho}_{mj}/k^2$ into the Fourier-transformed form of Eq. (3.54) gives

$$
\begin{aligned}
\widetilde{\mathcal{C}}_{tr}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) \;=\; & 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle(2\pi)^3\delta(\mathbf{k}) - \frac{2}{V}\left(-i\epsilon_{ilm}\frac{k_l}{k^2}\widetilde{\rho}_{mi}(\mathbf{k})\right) \\
& \times\left(i\epsilon_{jst}\frac{k_s}{k^2}\widetilde{\rho}_{tj}(-\mathbf{k})\right) \\
=\; & 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle(2\pi)^3\delta(\mathbf{k}) - \frac{2}{k^2}\widetilde{\mathcal{C}}_{tot}^{\rho}(\mathbf{k}) + \frac{2}{k^2}\widetilde{\mathcal{C}}_{per}^{\rho}(\mathbf{k}) \\
& +\frac{2}{Vk^4}k_ik_j\widetilde{\rho}_{mi}(\mathbf{k})\widetilde{\rho}_{mj}(-\mathbf{k}),
\end{aligned}
\tag{3.72}
$$

valid in the presence of stress. Here we repeat a similar procedure as was used to derive in Eq. (3.70).

In stress-free states, we substitute Eqs. (3.55), (3.58), (3.60) into Eq. (3.72)

$$
\begin{aligned}
\widetilde{\mathcal{C}}_{tr}^{\beta^{\mathrm{p,I}}}(\mathbf{k}) \;\overset{\mathrm{sf}}{=}\; & 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle(2\pi)^3\delta(\mathbf{k}) \\
& +\frac{1}{k^2}(k^2\delta_{ij}+k_ik_j)\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}) - \frac{2}{k^2}k_ik_j\widetilde{\mathcal{C}}_{ij}^{\mathbf{\Lambda}}(\mathbf{k}) \\
& +\frac{2k_ik_j}{Vk^4}\left(-ik_i\widetilde{\Lambda}_m(\mathbf{k}) + i\delta_{mi}k_k\widetilde{\Lambda}_k(\mathbf{k})\right) \\
& \times\left(ik_j\widetilde{\Lambda}_m(-\mathbf{k}) - i\delta_{mj}k_n\widetilde{\Lambda}_n(-\mathbf{k})\right) \\
\overset{\mathrm{sf}}{=}\; & 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle(2\pi)^3\delta(\mathbf{k}),
\end{aligned}
\tag{3.73}
$$

which is a trivial constant in space.

Through an inverse Fourier transform, Eqs. (3.69), (3.71), and (3.73) can be converted back to real space, giving

$$\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) \stackrel{\mathrm{sf}}{=} \mathcal{C}^{\mathbf{\Lambda}}(\mathbf{x}) + \frac{1}{4\pi}\int d^3\mathbf{x}'\left(\frac{\delta_{ij}}{R^3} - 3\frac{R_iR_j}{R^5}\right)\mathcal{C}_{ij}^{\mathbf{\Lambda}}(\mathbf{x}'), \tag{3.74}$$

$$\mathcal{C}_{per}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) \stackrel{\mathrm{sf}}{=} \frac{1}{2\pi}\int d^3\mathbf{x}'\left(\frac{\delta_{ij}}{R^3} - 3\frac{R_iR_j}{R^5}\right)\mathcal{C}_{ij}^{\mathbf{\Lambda}}(\mathbf{x}'), \tag{3.75}$$

$$\mathcal{C}_{tr}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) \stackrel{\mathrm{sf}}{=} 2\int d^3\mathbf{x}'\beta_{ii}^{\mathrm{p,I}}(\mathbf{x}')\beta_{jj}^{\mathrm{p,I}}(\mathbf{x}') = 2\langle\beta_{ii}^{\mathrm{p,I}}\beta_{jj}^{\mathrm{p,I}}\rangle, \tag{3.76}$$

where $\mathbf{R} = \mathbf{x}' - \mathbf{x}$. According to Eqs. (3.69) and (3.71), we can extract a relation

$$\mathcal{C}_{per}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) - 2\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) + 2\mathcal{C}^{\mathbf{\Lambda}}(\mathbf{x}) \stackrel{\mathrm{sf}}{=} const. \tag{3.77}$$

We can convert Eq. (3.67) through an inverse Fourier transform

$$\mathcal{C}_{tot}^{\rho}(\mathbf{x}) = \frac{1}{2}\partial^2\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}), \tag{3.78}$$

or

$$\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}) = -\frac{1}{2\pi}\int d^3\mathbf{x}'\frac{\mathcal{C}_{tot}^{\rho}(\mathbf{x}')}{R}, \tag{3.79}$$

valid in the presence of residual stress.

3.4.2 Critical exponent relations

When the self-similar dislocation structures emerge, the correlation functions of all physical quantities are expected to exhibit scale-free power laws. We consider the simplest possible scenario, where single variable scaling is present to reveal the minimal number of underlying critical exponents.

First, we define the critical exponent η as the power law describing the asymptotic decay of $\mathcal{C}_{tot}^{\rho}(\mathbf{x}) \sim |\mathbf{x}|^{-\eta}$, one of the correlation functions for the GND density

Correlation functions	Scaling theory	Simulations					
		Climb&Glide		Glide Only (MDP)		LVP Glide Only (LVP)	
		2D(1024^2)	3D(128^3)	2D(1024^2)	3D(128^3)	2D(1024^2)	3D(128^3)
C_{tot}^{ρ}	η	0.80 ± 0.30	0.55 ± 0.05	0.45 ± 0.25	0.60 ± 0.20	0.80 ± 0.30	0.55 ± 0.05
C_{per}^{ρ}	η	0.80 ± 0.20	0.55 ± 0.05	0.45 ± 0.20	0.60 ± 0.20	0.70 ± 0.30	0.50 ± 0.05
C_{tr}^{ρ}	η	0.80 ± 0.20	0.55 ± 0.05	0.45 ± 0.20	0.60 ± 0.10	0.70 ± 0.30	0.45 ± 0.05
C^{Λ}	$2 - \eta$	1.10 ± 0.65	1.45 ± 0.25	1.50 ± 0.30	1.35 ± 0.25	1.10 ± 0.65	1.50 ± 0.25
$C_{tot}^{\beta\mathrm{p,I}}$	$2 - \eta$	1.10 ± 0.60	1.45 ± 0.15	1.45 ± 0.25	1.30 ± 0.20	1.10 ± 0.60	1.50 ± 0.20
$C_{per}^{\beta\mathrm{p,I}}$	$2 - \eta$	1.15 ± 0.45	1.50 ± 0.25	1.45 ± 0.25	1.50 ± 0.50	1.20 ± 0.45	1.55 ± 0.25

Table 3.1: **Critical exponents for correlation functions at stress-free states.**

tensor (summed over components). If we rescale the spatial variable $\mathbf{x}$ by a factor b, the correlation function $\mathcal{C}^\rho$ is rescaled by the power law as

$$\mathcal{C}^\rho_{tot}(b\mathbf{x}) = b^{-\eta}\mathcal{C}^\rho_{tot}(\mathbf{x}). \tag{3.80}$$

Similarly, the correlation function of the crystalline orientation field $\boldsymbol{\Lambda}$ is described by a power law, $\mathcal{C}^{\boldsymbol{\Lambda}}(\mathbf{x}) \sim |\mathbf{x}|^\sigma$, where σ is its critical exponent. We repeat the rescaling by the same factor b

$$\mathcal{C}^{\boldsymbol{\Lambda}}(b\mathbf{x}) = b^\sigma \mathcal{C}^{\boldsymbol{\Lambda}}(\mathbf{x}). \tag{3.81}$$

Since $\mathcal{C}^\rho_{tot}$ can be written in terms of $\mathcal{C}^{\boldsymbol{\Lambda}}$, Eq. (3.63), we rescale this relation by the same factor b

$$\mathcal{C}^\rho_{tot}(b\mathbf{x}) \overset{\text{sf}}{=} \frac{1}{2}\left[\frac{\partial}{b}\right]^2 \mathcal{C}^{\boldsymbol{\Lambda}}(b\mathbf{x}) + \frac{1}{2}\left[\frac{\partial_i}{b}\right]\left[\frac{\partial_j}{b}\right]\mathcal{C}^{\boldsymbol{\Lambda}}_{ij}(b\mathbf{x}). \tag{3.82}$$

Substituting Eq. (3.81) into Eq. (3.82) gives

$$\begin{aligned}
\mathcal{C}^\rho_{tot}(b\mathbf{x}) &\overset{\text{sf}}{=} b^{\sigma-2}\left[\frac{1}{2}\partial^2\mathcal{C}^{\boldsymbol{\Lambda}}(\mathbf{x}) + \frac{1}{2}\partial_i\partial_j\mathcal{C}^{\boldsymbol{\Lambda}}_{ij}(\mathbf{x})\right] \\
&\overset{\text{sf}}{=} b^{\sigma-2}\mathcal{C}^\rho_{tot}(\mathbf{x}).
\end{aligned} \tag{3.83}$$

Comparing with Eq. (3.80) gives a relation between σ and η

$$\sigma = 2 - \eta. \tag{3.84}$$

We can repeat the same renormalization group procedure to analyze the critical exponents of the other two scalar forms of the correlation functions of the GND density field. Clearly, $\mathcal{C}^\rho_{per}$ and $\mathcal{C}^\rho_{tr}$ share the same critical exponent η with $\mathcal{C}^\rho_{tot}$.

Also, we can define the critical exponent σ' as the power law describing the asymptotic growth of $\mathcal{C}^{\beta^{\mathrm{p},\mathrm{I}}}_{tot}(\mathbf{x}) \sim |\mathbf{x}|^{\sigma'}$, one of the correlation functions for the

intrinsic part of the plastic distortion field. We can rescale the correlation function $\mathcal{C}^{\beta^{\mathrm{p,I}}}$

$$\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(b\mathbf{x}) = b^{\sigma'}\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x}). \tag{3.85}$$

We rescale the relation Eq. (3.78) by the same factor b, and substitute Eq. (3.85) into it

$$
\begin{aligned}
\mathcal{C}_{tot}^{\rho}(b\mathbf{x}) &= \frac{1}{2}\left[\frac{\partial}{b}\right]^2 \mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(b\mathbf{x}) = b^{\sigma'-2}\left[\frac{1}{2}\partial^2 \mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}(\mathbf{x})\right] \\
&= b^{\sigma'-2}\mathcal{C}_{tot}^{\rho}(\mathbf{x}).
\end{aligned}
\tag{3.86}
$$

Comparing with Eq. (3.80) also gives a relation between σ' and η

$$\sigma' = 2 - \eta. \tag{3.87}$$

Since both $\mathcal{C}_{tot}^{\beta^{\mathrm{p,I}}}$ and $\mathcal{C}^{\mathbf{\Lambda}}$ share the same critical exponent $2 - \eta$, it is clear that $\mathcal{C}_{per}^{\beta^{\mathrm{p,I}}}$, the other scalar form of the correlation functions of the intrinsic plastic distortion field, also shares this critical exponent, according to Eq. (3.77).

Thus the correlation functions of three physical quantities (the GND density ρ, the crystalline orientation $\mathbf{\Lambda}$, and the intrinsic plastic distortion $\beta^{\mathrm{p,I}}$) all share the same underlying universal critical exponent η for self-similar morphologies, in the case of zero residual stress, and still hold in the limit of slow imposed deformation. Table 3.1 verifies the existence of single underlying critical exponent in both two and three dimensional simulations for each type of dynamics. Imposed strain, studied in Ref. 52, could in principle change η, but the scaling relations derived here should still apply. The strain, of course, breaks the isotropic symmetry, allowing even more allowed correlation functions to be measured.

3.5 Conclusion

In our earlier works [155, 52, 55], we have proposed a flexible framework of CDD to study complex mesoscale phenomena of collective dislocation motion. Traditionally, deterministic CDDs have missed the experimentally ubiquitous feature of cellular pattern formation. Our CDD models have made progress in that respect. In the beginning, we focused our efforts on describing coarse-grained dislocations that naturally develop dislocation cellular structures in ways that are consistent with experimental observations of scale invariance and fractality, a target achieved in Ref. 52. However, that paper studied only 2D, instead of the more realistic 3D.

In this manuscript, we go further in many aspects of the theory extending the results of our previous work:

We provide a derivation of our theory that explains the differences with traditional theories of plasticity. In addition to our previously studied climb-glide (CGD) and glide-only (GOD-MDP) models, we extend our construction in order to incorporate vacancies, and re-derive [11] a different glide-only dynamics (GOD-LVP) which we show exhibits very similar behavior in 2D to our CGD model. It is worth mentioning that in this way, the GOD-LVP and the CGD dynamics become statistically similar in 2D, while the previously studied, less physical, GOD-MDP model provides rather different behavior in 2D [52].

We present 3D simulation results here for the first time, showing qualitatively different behavior from that of 2D. In 3D, all three types of dynamics – CGD, GOD-MDP and GOD-LVP – show similar non-trivial fractal patterns and scaling dimensions. Thus our 3D analysis shows that the flatter 'grain boundaries' we observe in the 2D simulations are not intrinsic to our dynamics, but are an artifact

of the artificial z-independent initial conditions. Experimentally, grain boundaries are indeed flatter and cleaner than cell walls, and our theory no longer provides a new explanation for this distinction. We expect that the dislocation core energies left out of our model would flatten the walls, and that adding disorder or entanglement would prevent the low-temperature glide-only dynamics from flattening as much.

We also fully describe, in a statistical sense, multiple correlation functions – the local orientation, the plastic distortion, the GND density – their symmetries and their mutual scaling relations. Correlation functions of important physical quantities are categorized and analytically shown to share one stress-free exponent. The anomaly in the correlation functions of β^{p}, which was left as a question in our previous publication [52], has been discussed and explained. All of these correlation functions and properties are verified with the numerical results of the dynamics that we extensively discussed.

As discussed in Sec. 3.1, our model is an immensely simplified caricature of the deformation of real materials. How does it connect to reality?

First, we show that a model for which elastic strain energy minimization determines the dynamics produces realistic cell wall structures even while ignoring slip systems, crystalline anisotropy [119], pinning, junction formation, and statistically stored dislocations. The fact that low-energy dislocation structures (LEDS) provides natural explanations for many properties of these structures has long been emphasized by Kuhlmann-Wilsdorf [136]. Intermittent flow, forest interactions, and pinning will in general impede access to low energy states. These real-world features, our model suggests, can be important for the morphology of the cell wall structures but are not the root cause of their formation nor of their evolution under

stress (discussed in previous work [52]).

One must note, however, that strain energy minimization does not provide the explanation for wall structures in our model material. Indeed, there is an immense space of dislocation densities which make the strain energy zero [156], including many continuous densities. Our dynamics relaxes into a small subset of these allowed structures – it is the dynamics that leads to cell structure formation here, not purely the energy. In discrete dislocation simulations and real materials, the quantization of the Burger's vector leads to a weak logarithmic energetic preference for sharp walls. This $-\mu\mathbf{b}/(4\pi(1-\nu))\theta\log\theta$ energy of low-angle grain boundaries yields a $\log 2$ preference for one wall of angle θ rather than two walls of angle $\theta/2$. This leads to a 'zipping' together of low angle grain boundaries. Since $\mathbf{b} \to 0$ in a continuum theory, this preference is missing from our model. Yet, we still find cell wall formation suggesting that such mechanisms are not central to cell wall formation.

Second, how should we connect our fractal cell wall structures with those (fractal or non-fractal) seen in experiments? Many qualitatively different kinds of cellular structures are seen in experiments – variously termed cell block structures, mosaic structures, ordinary cellular structures, ... Hansen et al. [101] recently categorized these structures into three types, and argue that the orientation of the stress with respect to the crystalline axes largely determines which morphology is exhibited. The cellular structures in our model, which ignores crystalline anisotropy, likely are the theoretical progenitors of all of these morphologies. In particular, Hansen's type 1 and type 3 structures incorporate both 'geometrically necessary' and 'incidental dislocation' boundaries (GNBs and IDBs), while type 2 structures incorporate only the latter. Our simulations cannot distinguish between

these two types, and indeed qualitatively look similar to Hansen's type 2 structures. One should note that the names of these boundaries are misleading – the 'incidental' boundaries do mediate geometrical rotations, with the type 2 boundaries at a given strain having similar average misorientations to the geometrically necessary boundaries of type 1 structures (Ref. 31, Fig. 8). It is commonly asserted that the IDBs are formed by statistical trapping of stored dislocations; our model suggests that stochasticity is not necessary for their formation.

Third, how is our model compatible with traditional plasticity, which focuses on the total density of dislocation lines? Our model evolves the net dislocation density, ignoring the *geometrically unnecessary* or *statistically stored* dislocations with canceling Burger's vectors. These latter dislocations are important for yield stress and work hardening on macroscales, but are invisible to our theory (since they do not generate stress). Insofar as the cancellation of Burger's vectors on the macroscale is due to cell walls of opposing misorientations on the mesoscale, there needs to be no conflict here. Also our model remains agnostic about whether cell boundaries include significant components of geometrically unnecessary dislocations. However, our model does assume that the driving force for cell boundary formation is the motion of geometrically necessary dislocations, as opposed to (for example) inhomogeneous flows of statistically stored dislocations.

There still remain many fascinating mesoscale experiments, such as dislocation avalanches [167, 71], size-dependent hardness (smaller is stronger) [251], and complex anisotropic loading [223, 160], that we hope to emulate. We intend in the future to include several relevant additional ingredients to our dynamics, such as vacancies (Sec. 3.8.1), impurities (Sec. 3.8.2), immobile dislocations/SSDs and slip systems, to reflect real materials.

3.6 Physical quantities in terms of the plastic distortion tensor

In an isotropic infinitely large medium, the local deformation $\boldsymbol{u}$, the elastic distortion β^{e} and the internal long-range stress σ^{int} can be expressed [177, 155] in terms of the plastic distortion field β^{p} in Fourier space:

$$
\begin{aligned}
\widetilde{\boldsymbol{u}}_i(\mathbf{k}) &= N_{ikl}(\mathbf{k})\widetilde{\beta}_{kl}^{\mathrm{p}}(\mathbf{k}), \\
N_{ikl}(\mathbf{k}) &= -\frac{i}{k^2}(k_k\delta_{il} + k_l\delta_{ik}) - i\frac{\nu k_i\delta_{kl}}{(1-\nu)k^2} \\
&\quad + i\frac{k_i k_k k_l}{(1-\nu)k^4}; \quad (3.88)\\
\widetilde{\beta}_{ij}^{\mathrm{e}}(\mathbf{k}) &= T_{ijkl}(\mathbf{k})\widetilde{\beta}_{kl}^{\mathrm{p}}(\mathbf{k}), \\
T_{ijkl}(\mathbf{k}) &= \frac{1}{k^2}(k_i k_k\delta_{jl} + k_i k_l\delta_{jk} - k^2\delta_{ik}\delta_{jl}) \\
&\quad + \frac{k_i k_j}{(1-\nu)k^4}(\nu k^2\delta_{kl} - k_k k_l); \quad (3.89)\\
\widetilde{\sigma}_{ij}^{\mathrm{int}}(\mathbf{k}) &= M_{ijmn}(\mathbf{k})\widetilde{\beta}_{mn}^{\mathrm{p}}(\mathbf{k}), \\
M_{ijmn}(\mathbf{k}) &= \frac{2u\nu}{1-\nu}\left(\frac{k_m k_n\delta_{ij} + k_i k_j\delta_{mn}}{k^2} - \delta_{ij}\delta_{mn}\right) \\
&\quad + u\left(\frac{k_i k_m}{k^2}\delta_{jn} + \frac{k_j k_n}{k^2}\delta_{im} - \delta_{im}\delta_{jn}\right) \\
&\quad + u\left(\frac{k_i k_n}{k^2}\delta_{jm} + \frac{k_j k_m}{k^2}\delta_{in} - \delta_{in}\delta_{jm}\right) \\
&\quad - \frac{2u}{1-\nu}\frac{k_i k_j k_m k_n}{k^4}. \quad (3.90)
\end{aligned}
$$

All these expressions are valid for systems with periodic boundary conditions.

According to the definition Eq. (3.12) of the crystalline orientation $\boldsymbol{\Lambda}$, we can replace ω^{e} with β^{e} and ϵ^{e} by using the elastic distortion tensor decomposition Eq. (3.10)

$$
\Lambda_i = \frac{1}{2}\varepsilon_{ijk}(\beta_{jk}^{\mathrm{e}} - \epsilon_{jk}^{\mathrm{e}}). \quad (3.91)
$$

Here the permutation factor acting on the symmetric elastic strain tensor gives

zero. Hence we can express the crystalline orientation vector $\boldsymbol{\Lambda}$ in terms of β^{p} by using Eq. (3.89)

$$
\begin{aligned}
\widetilde{\Lambda}_i(\mathbf{k}) &= \frac{1}{2}\varepsilon_{ijk}\left\{\frac{1}{k^2}(k_jk_s\delta_{kt} + k_jk_t\delta_{ks} - k^2\delta_{js}\delta_{kt})\right. \\
&\quad \left. +\frac{k_jk_k}{(1-\nu)k^4}(\nu k^2\delta_{st} - k_sk_t)\right\}\widetilde{\beta}^{\mathrm{p}}_{st}(\mathbf{k}) \\
&= \frac{1}{2k^2}(\varepsilon_{ijt}k_jk_s + \varepsilon_{ijs}k_jk_t - k^2\varepsilon_{ist})\widetilde{\beta}^{\mathrm{p}}_{st}(\mathbf{k}).
\end{aligned}
$$

(3.92)

3.7 Energy dissipation rate

3.7.1 Free energy in Fourier space

In the absence of external stress, the free energy $\mathcal{F}$ is the elastic energy caused by the internal long-range stress

$$
\mathcal{F} = \int d^3\mathbf{x}\, \frac{1}{2}\sigma^{\mathrm{int}}_{ij}\epsilon^{\mathrm{e}}_{ij} = \int d^3\mathbf{x}\, \frac{1}{2}C_{ijmn}\epsilon^{\mathrm{e}}_{ij}\epsilon^{\mathrm{e}}_{mn},
$$

(3.93)

where the stress is $\sigma^{\mathrm{int}}_{ij} = C_{ijmn}\epsilon^{\mathrm{e}}_{mn}$, with C_{ijmn} the stiffness tensor.

Using the symmetry of C_{ijmn} and ignoring large rotations, $\epsilon^{\mathrm{e}}_{ij} = (\beta^{\mathrm{e}}_{ij} + \beta^{\mathrm{e}}_{ji})/2$, we can rewrite the elastic energy $\mathcal{F}$ in terms of β^{e}

$$
\mathcal{F} = \int d^3\mathbf{x}\, \frac{1}{2}C_{ijmn}\beta^{\mathrm{e}}_{ij}\beta^{\mathrm{e}}_{mn}.
$$

(3.94)

Performing a Fourier transform on both β^{p}_{ij} and β^{p}_{mn} simultaneously gives

$$
\begin{aligned}
\mathcal{F} &= \int d^3\mathbf{x}\int \frac{d^3\mathbf{k}}{(2\pi)^3}\int \frac{d^3\mathbf{k}'}{(2\pi)^3}\, e^{i(\mathbf{k}+\mathbf{k}')\mathbf{x}} \\
&\quad \times \left(\frac{1}{2}C_{ijmn}\widetilde{\beta}^{\mathrm{e}}_{ij}(\mathbf{k})\widetilde{\beta}^{\mathrm{e}}_{mn}(\mathbf{k}')\right).
\end{aligned}
$$

(3.95)

111

Integrating out the spatial variable $\mathbf{x}$ leaves a δ−function $\delta(\mathbf{k} + \mathbf{k}')$ in Eq. (3.95). We hence integrate out the k-space variable $\mathbf{k}'$

$$\mathcal{F} = \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{1}{2} C_{ijmn} \widetilde{\beta}_{ij}^{\mathrm{e}}(\mathbf{k}) \widetilde{\beta}_{mn}^{\mathrm{e}}(-\mathbf{k}).$$ (3.96)

Substituting Eq. (3.89) into Eq. (3.96) gives

$$
\begin{aligned}
\mathcal{F} &= \int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{1}{2} \big(C_{ijmn} T_{ijpq}(\mathbf{k}) T_{mnst}(-\mathbf{k})\big) \widetilde{\beta}_{pq}^{\mathrm{p}}(\mathbf{k}) \widetilde{\beta}_{st}^{\mathrm{p}}(-\mathbf{k}) \\
&= -\int \frac{d^3\mathbf{k}}{(2\pi)^3} \frac{1}{2} M_{pqst}(\mathbf{k}) \widetilde{\beta}_{pq}^{\mathrm{p}}(\mathbf{k}) \widetilde{\beta}_{st}^{\mathrm{p}}(-\mathbf{k}),
\end{aligned}
$$ (3.97)

where we skip straightforward but tedious simplifications.

When turning on the external stress, we repeat the same procedure used in Eq. (3.95), yielding

$$\mathcal{F}^{\mathrm{ext}} = -\int d^3\mathbf{x}\, \sigma_{ij}^{\mathrm{ext}} \beta_{ij}^{\mathrm{p}} = -\int \frac{d^3\mathbf{k}}{(2\pi)^3} \, \widetilde{\sigma}_{ij}^{\mathrm{ext}}(\mathbf{k}) \widetilde{\beta}_{ij}^{\mathrm{p}}(-\mathbf{k}).$$ (3.98)

3.7.2 Calculation of energy functional derivative with respect to the GND density ϱ

According to Eq. (3.17), the infinitesimal change of the variable $\delta\varrho$ is given in terms of $\delta\beta^{\mathrm{p}}$

$$\delta\varrho_{ijk} = -g_{ijls} \partial_l \big(\delta\beta_{sk}^{\mathrm{p}}\big).$$ (3.99)

Substituting Eq. (3.99) into Eq. (3.27) and applying integration by parts, the infinitesimal change of $\mathcal{F}$ is hence rewritten in terms of β^{p}

$$\delta\mathcal{F}[\beta^{\mathrm{p}}] = \int d^3\mathbf{x}\, g_{ijls} \partial_l \left(\frac{\delta\mathcal{F}}{\delta\varrho_{ijk}}\right) \delta\beta_{sk}^{\mathrm{p}}.$$ (3.100)

112

According to Eq. (3.24), it suggests

$$\delta \mathcal{F}[\beta^{\mathrm{p}}] = \int d^3\mathbf{x} \, \frac{\delta \mathcal{F}}{\delta \beta^{\mathrm{p}}_{sk}} \delta \beta^{\mathrm{p}}_{sk} = \int d^3\mathbf{x} \, (-\sigma_{sk}) \delta \beta^{\mathrm{p}}_{sk}. \tag{3.101}$$

Comparing Eq. (3.100) and Eq. (3.101) implies

$$g_{ijls} \partial_l \left(\frac{\delta \mathcal{F}}{\delta \varrho_{ijk}} \right) = -\sigma_{sk}, \tag{3.102}$$

up to a total derivative which we ignore due to the use of periodic boundary conditions.

3.7.3 Derivation of energy dissipation rate

We can apply variational methods to calculate the dissipation rate of the free energy. As is well known, the general elastic energy $\mathcal{E}$ in a crystal can be expressed as $\mathcal{E} = \frac{1}{2} \int d^3\mathbf{x} \, \sigma_{ij} \epsilon^{\mathrm{e}}_{ij}$, with $\epsilon^{\mathrm{e}}_{ij}$ the elastic strain. An infinitesimal change of $\mathcal{E}$ is:

$$\delta \mathcal{E} = \frac{1}{2} \int d^3\mathbf{x} \, \sigma_{ij} \delta \epsilon^{\mathrm{e}}_{ij} + \frac{1}{2} \int d^3\mathbf{x} \, \delta \sigma_{ij} \epsilon^{\mathrm{e}}_{ij} = \int d^3\mathbf{x} \, \sigma_{ij} \delta \epsilon^{\mathrm{e}}_{ij}, \tag{3.103}$$

where we use $\sigma_{ij} \delta \epsilon^{\mathrm{e}}_{ij} = C_{ijkl} \epsilon^{\mathrm{e}}_{kl} \delta \epsilon^{\mathrm{e}}_{ij} = \delta \sigma_{ij} \epsilon^{\mathrm{e}}_{ij}$.

So the infinitesimal change of the free energy Eq. (3.20) is

$$\delta \mathcal{F} = \int d^3\mathbf{x} \left(\sigma^{\mathrm{int}}_{ij} \delta \epsilon^{\mathrm{e}}_{ij} - \sigma^{\mathrm{ext}}_{ij} \delta \epsilon^{\mathrm{p}}_{ij} \right). \tag{3.104}$$

We apply the relation $\epsilon^{\mathrm{e}} = \epsilon - \epsilon^{\mathrm{p}}$, where ϵ^{p} is the plastic strain and ϵ is the total strain:

$$\delta \mathcal{F} = \int d^3\mathbf{x} \left(\sigma^{\mathrm{int}}_{ij} \delta \epsilon_{ij} - \sigma^{\mathrm{int}}_{ij} \delta \epsilon^{\mathrm{p}}_{ij} - \sigma^{\mathrm{ext}}_{ij} \delta \epsilon^{\mathrm{p}} \right). \tag{3.105}$$

Using the symmetry of σ_{ij} and ignoring large rotations, $\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$, we can rewrite the first term of Eq. (3.105) as $\int d^3\mathbf{x} \, \sigma^{\mathrm{int}}_{ij} \delta(\partial_i u_j)$. Integrating by parts

yields $\int d^3\mathbf{x}\left(\partial_i(\delta u_j\sigma^{\mathrm{int}}_{ij}) - \delta u_j\partial_i\sigma^{\mathrm{int}}_{ij}\right)$. We can convert the first volume integral to a surface integral, which vanishes for an infinitely large system. Hence

$$\delta\mathcal{F} = \int d^3\mathbf{x}\left(\partial_i\sigma^{\mathrm{int}}_{ij}\delta u_j - (\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})\delta\epsilon^{\mathrm{p}}_{ij}\right). \tag{3.106}$$

The first term of Eq. (3.106) is zero assuming instantaneous elastic relaxation due to the local force equilibrium condition,

$$\delta\mathcal{F} = -\int d^3\mathbf{x}\,(\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})\delta\beta^{\mathrm{p}}_{ij}, \tag{3.107}$$

using the symmetry of σ_{ij} and $\epsilon^{\mathrm{p}}_{ij} = \frac{1}{2}(\beta^{\mathrm{p}}_{ij} + \beta^{\mathrm{p}}_{ji})$.

The free energy dissipation rate is thus $\delta\mathcal{F}/\delta t$ for $\delta\beta^{\mathrm{p}}_{ij} = \frac{\partial\beta^{\mathrm{p}}}{\partial t}\delta t$, hence

$$\begin{aligned} \frac{\partial\mathcal{F}}{\partial t} &= -\int d^3\mathbf{x}\,(\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})\frac{\partial\beta^{\mathrm{p}}_{ij}}{\partial t} \\ &= -\int d^3\mathbf{x}\,(\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})J_{ij}. \end{aligned} \tag{3.108}$$

When dislocations are allowed to climb, substituting the CGD current Eq. (3.36) into Eq. (3.108) implies that the free energy dissipation rate is strictly negative

$$\begin{aligned} \frac{\partial\mathcal{F}}{\partial t} &= -\int d^3\mathbf{x}\,(\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})\left[v_l\varrho_{lij}\right] \\ &= -\int d^3\mathbf{x}\,\frac{|\varrho|}{D}v^2 \leq 0. \end{aligned} \tag{3.109}$$

When removing dislocation climb by considering the mobile dislocation population, we substitute Eq. (3.40) into Eq. (3.108) to guarantee that the rate of the change of the free energy density is also the negative of a perfect square

$$\begin{aligned} \frac{\partial\mathcal{F}}{\partial t} &= -\int d^3\mathbf{x}(\sigma^{\mathrm{int}}_{ij} + \sigma^{\mathrm{ext}}_{ij})\left[v'_l\left(\varrho_{lij} - \frac{1}{3}\delta_{ij}\varrho_{lkk}\right)\right] \\ &= -\int d^3\mathbf{x}\,\frac{|\varrho|}{D}v'^2 \leq 0. \end{aligned} \tag{3.110}$$

114

3.8 Model Extensions: Adding vacancies and disorder to CDD

3.8.1 Coupling vacancy diffusion to CDD

In plastically deformed crystals at low temperature, dislocations usually move only in the glide plane because vacancy diffusion is almost frozen out. When temperature increases, vacancy diffusion leads to dislocation climb out of the glide plane. At intermediate temperatures, slow vacancy diffusion can enable local creep. The resulting dynamics should couple the vacancy and dislocation fields in non-trivial ways. Here we couple the vacancy diffusion to the dislocation motion in our CDD model.

We introduce an order parameter field $c(\mathbf{x})$, indicating the vacancy concentration density at the point $\mathbf{x}$. The free energy $\mathcal{F}$ is thus expressed

$$\mathcal{F} = \mathcal{F}^{Dis} + \mathcal{F}^{Vac} = \int d^3\mathbf{x} \left(\frac{1}{2}\sigma_{ij}\epsilon_{ij}^{\mathrm{e}} + \frac{1}{2}\alpha(c - c_0)^2 \right), \tag{3.111}$$

where α is a positive material parameter related to the vacancy creation energy, and c_0 is the overall equilibrium vacancy concentration density.

Assuming that GNDs share the velocity $\boldsymbol{v}$ in an infinitesimal volume, we write the current J for GNDs

$$J_{ij} = v_u \varrho_{uij}. \tag{3.112}$$

The current trace J_{ii} describes the rate of volume change, which acts as a source and sink of vacancies. The coupling dynamics for vacancies is thus given as

$$\partial_t c = \gamma \nabla^2 c + J_{ii}, \tag{3.113}$$

where γ is a positive vacancy diffusion constant.

The infinitesimal change of the free energy $\mathcal{F}$ (Eq. 3.111) is

$$\delta\mathcal{F} = \int d^3\mathbf{x}\left(\frac{\delta\mathcal{F}^{Dis}}{\delta\beta^{\mathrm{p}}_{ij}}\delta\beta^{\mathrm{p}}_{ij} + \frac{\delta\mathcal{F}^{Vac}}{\delta c}\delta c\right). \tag{3.114}$$

We apply Eq. (3.107) and $\delta\mathcal{F}^{Vac}/\delta c = \alpha(c - c_0)$

$$\delta\mathcal{F} = \int d^3\mathbf{x}\left(-\sigma_{ij}\delta\beta^{\mathrm{p}}_{ij} + \alpha(c - c_0)\delta c\right). \tag{3.115}$$

The free energy dissipation rate is thus $\delta\mathcal{F}/\delta t$ for $\delta\beta^{\mathrm{p}}_{ij} = \frac{\partial\beta^{\mathrm{p}}}{\partial t}\delta t$ and $\delta c = \frac{\partial c}{\partial t}\delta c$, hence

$$\frac{\partial\mathcal{F}}{\partial t} = -\int d^3\mathbf{x}\left(\sigma_{ij}\frac{\partial\beta^{\mathrm{p}}_{ij}}{\partial t} - \alpha(c - c_0)\frac{\partial c}{\partial t}\right). \tag{3.116}$$

Substituting the current J (Eq. 3.112) and Eq. (3.113) into Eq. (3.116) gives

$$\begin{aligned}
\frac{\partial\mathcal{F}}{\partial t} &= -\int d^3\mathbf{x}\left(\sigma_{ij}(v_u\varrho_{uij}) - \alpha(c - c_0)(\gamma\nabla^2 c + v_u\varrho_{uii})\right)\\
&= -\int d^3\mathbf{x}\left((\sigma_{ij} - \alpha(c - c_0)\delta_{ij})\varrho_{uij}\right)v_u\\
&\quad -\int d^3\mathbf{x}\,\alpha\gamma(\nabla c)^2, \tag{3.117}
\end{aligned}$$

where we integrate by parts by assuming an infinitely large system.

If we choose the velocity $v_u = \frac{D}{|\varrho|}\left(\sigma_{ij} - \alpha(c - c_0)\delta_{ij}\right)\varrho_{uij}$, ($D$ is a positive material dependent constant and $1/|\varrho|$ is added for the same reasons as discussed in Sec. 3.2.3), the free energy is guaranteed to decrease monotonically. The coupling dynamics for both GNDs and vacancies is thus

$$\begin{cases}
\partial_t\beta^{\mathrm{p}}_{ij} = \frac{D}{|\varrho|}\left(\sigma_{mn} - \alpha(c - c_0)\delta_{mn}\right)\varrho_{umn}\varrho_{uij},\\
\partial_t c = \gamma\nabla^2 c + \frac{D}{|\varrho|}\left(\sigma_{mn} - \alpha(c - c_0)\delta_{mn}\right)\varrho_{umn}\varrho_{ukk}.
\end{cases} \tag{3.118}$$

This dynamics gives us a clear picture of the underlying physical mechanism: the vacancies contribute an extra hydrostatic pressure $p = -\alpha(c - c_0)$.

3.8.2 Coupling disorder to CDD

In real crystals, the presence of precipitates or impurities results in a force pinning nearby dislocations. We can mimic this effect by incorporating a spatially varying random potential field $V(\mathbf{x})$.

In our CDD model, we can add the interaction energy between GNDs and random disorder into the free energy $\mathcal{F}$ (Eq. 3.20)

$$\mathcal{F} = \mathcal{F}_E + \mathcal{F}_I = \int d^3\mathbf{x}\left(\frac{1}{2}\sigma^{\text{int}}_{ij}\epsilon^{\text{e}}_{ij} - \sigma^{\text{ext}}_{ij}\epsilon^{\text{P}}_{ij} + V(\mathbf{x})|\varrho|\right), \quad (3.119)$$

where $\mathcal{F}_E$ indicates the elastic free energy corresponding to the integral of the first two terms, and $\mathcal{F}_I$ indicates the interaction energy, the integral of the last term.

An infinitesimal change of the free energy is written

$$\delta\mathcal{F} = \delta\mathcal{F}_E + \delta\mathcal{F}_I = \int d^3\mathbf{x}\left(\frac{\delta\mathcal{F}_E}{\delta\beta^{\text{P}}_{ij}}\delta\beta^{\text{P}}_{ij} + \frac{\delta\mathcal{F}_I}{\delta\beta^{\text{P}}_{sk}}\delta\beta^{\text{P}}_{sk}\right). \quad (3.120)$$

In an infinitely large system, Eq. (3.107) gives

$$\frac{\delta\mathcal{F}_E}{\delta\beta^{\text{P}}_{ij}} = -(\sigma^{\text{int}}_{ij} + \sigma^{\text{ext}}_{ij}), \quad (3.121)$$

and Eq. (3.100) implies

$$\begin{aligned}
\delta\mathcal{F}_I &= \int d^3\mathbf{x}\, g_{ijls}\partial_l\left(\frac{\delta\mathcal{F}_I}{\delta\varrho_{ijk}}\right)\delta\beta^{\text{P}}_{sk} \\
&= \int d^3\mathbf{x}\, g_{ijls}\partial_l\left(V(\mathbf{x})\frac{\varrho_{ijk}}{|\varrho|}\right)\delta\beta^{\text{P}}_{sk}.
\end{aligned} \quad (3.122)$$

Substituting Eq. (3.121) and Eq. (3.122) into Eq. (3.120) gives

$$\begin{aligned}
\delta\mathcal{F} &= -\int d^3\mathbf{x}\left(\sigma^{\text{int}}_{ij} + \sigma^{\text{ext}}_{ij} - g_{mnli}\partial_l\left(V(\mathbf{x})\frac{\varrho_{mnj}}{|\varrho|}\right)\right)\delta\beta^{\text{P}}_{ij} \\
&= -\int d^3\mathbf{x}\, \sigma^{\text{eff}}_{ij}\delta\beta^{\text{P}}_{ij}.
\end{aligned} \quad (3.123)$$

where the effective stress field is $\sigma_{ij}^{\text{eff}} = \sigma_{ij}^{\text{int}} + \sigma_{ij}^{\text{ext}} - g_{mnli}\partial_l\left(V(\mathbf{x})\frac{\varrho_{mnj}}{|\varrho|}\right)$.

By replacing σ_{ij} with σ_{ij}^{eff} in the equation of motion of either allowing climb (Eq. 3.36) or removing climb (Eqs. 3.40 and 3.44), we achieve the new CDD model that models GNDs interacting with disorder.

CHAPTER 4

VISUALIZATION, COARSENING AND FLOW DYNAMICS OF
FOCAL CONIC DOMAINS IN SIMULATED SMECTIC-A LIQUID
CRYSTALS[1]

4.1 Introduction

Translational order is *frail*. Most broken symmetry states respond elastically until deformations are large. In contrast, crystals fracture or plastically yield at strains of a few parts per thousand. In equilibrium, they form grain boundaries – expelling rotation gradients into walls – when subject to *atomic-scale* boundary displacements. An analogous expulsion occurs in smectics, which expel deviations from equal-layer spacing in a manner that can be mapped onto the Meissner/Higgs effect [67]. Instead of grain boundaries, this expulsion of strain in smectics results in a remarkable patterns of singular ellipses, hyperbolas, and parabolas known as focal conic domains (FCDs, Fig. 4.1), which are the signature of the smectic one-dimensional layered structure. Smectics provide a window into deep properties of translational order, lending insight into crystalline behavior.

FCDs have appealed to theorists and experimentalists since the early days [81], partially because of their unique geometric origin. In its minimum energy state, a smectic has lamellar layers spaced at equal distances. Equal layer spacing implies a singularity at the centers of curvature of the surfaces. This constraint of equal layer spacing, surprisingly, determines the allowed shapes of the smectic's lamella. The lamella choose surfaces whose centers of curvature trace out curves rather than

costly two-dimensional internal boundaries. These surfaces are called cyclides of Dupin [112]; their centers of curvature trace out one-dimensional conic sections, typically confocal ellipses and hyperbolas. The resulting structures in smectics are known as FCDs.

(a)

(b)

(c)

(d)
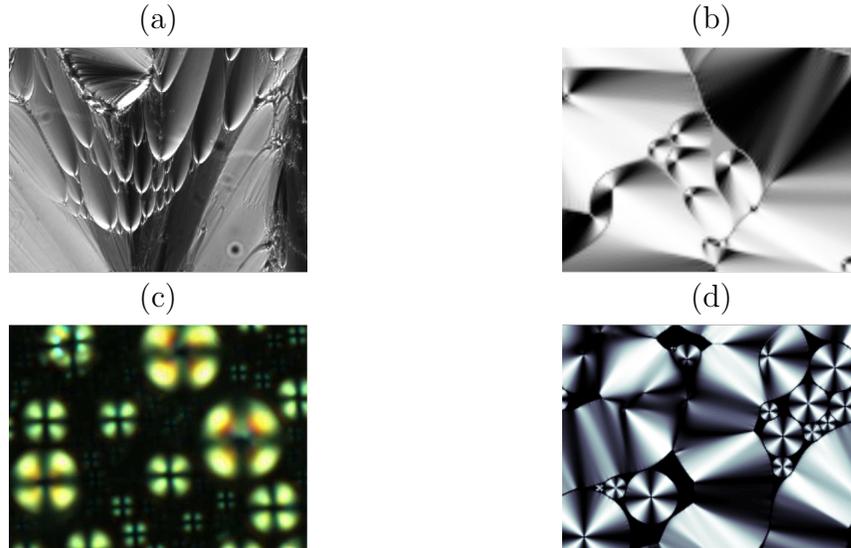
Figure 4.1: **Example experimental and simulation images of smectic-A**. Experimental (a and c) and simulation (b and d) results for polarizer microscopy images of a section of smectic-A slab for planar (a and b) and homeotropic anchoring (c and d).

On a practical level, an understanding of focal conic dynamics is necessary for the description of a variety of liquid-crystalline states, such as smectic-A [68, 132], smectic-C and C* [36, 180], lyotropic lamellar [34, 35], twist-bend [164], and even metallotropic liquid crystals [165]. We focus our attention on smectic-A's, which are the simplest case. Our current understanding of focal conic structures in smectics at rest includes the study of geometrical and energetic properties [228, 240, 130], the effects of anchoring for several substrates [233, 115, 116], the role played by dislocations [166, 133], and beautiful insights extracted from a hidden symmetry of the Poincaré group [14]. When a smectic is driven by external dilatative stresses, experiments on initially planar-aligned samples show a sequence

of elastic and plastic strain patterns that ultimately lead to a polygonal array of parabolic focal conic lines [211]. More recently, experiments on smectic samples with antagonistic anchoring conditions subjected to shear flow report on the emergence of satellite defects [47]. Further recent developments on the smectic rheology have been reported in [84, 85]. However, progress in simulating smectic dynamics has been slow, perhaps because of the challenge of incorporating defect dynamics into Ericksen-Leslie-Parodi theory. Simulations of smectics are often based on atomistic and molecular dynamics approaches [90, 15, 238]. Numerical solutions of the Ericksen-Leslie equations and Monte Carlo methods using the Frank free energy have been reported for nematics (see e.g. [64, 97]). As far as the authors know, there has been no report of the observation of focal conic domains in smectic simulations.

In this paper we present results of our simulations of an effective theory of smectic-A liquid crystals. Our dynamics is an extension of Ericksen-Leslie-Parodi dynamics and the Oseen-Frank free energy [190, 272], in that we allow focal conic singularites by allowing the order parameter to change magnitude, but we continue to forbid dislocations. The use of modern GPU computing makes these simulations feasible. Our simulations naturally form FCDs upon relaxation of random initial conditions and allow us to study these fascinating defects both during formation and under mechanical loading. We find good comparisons with experiments performed under similar situations. Our approach allows us to investigate focal conic structures in great detail through simulations, and provides us with an invaluable tool to understand their several aspects, ranging from energetics, topology and geometry to anchoring and mechanical strain effects, nicely complementing current experimental approaches [235].

4.2 Equations of Motion

Our description of the smectic starts from its elastic free energy

$$\Psi = \int d\boldsymbol{r} \left[F(N_\mu, \partial_\mu N_\nu) + \boldsymbol{\lambda} \cdot \nabla \times \boldsymbol{N} \right] \quad , \tag{4.1}$$

which is a functional of the layer-normal field and its derivatives. The layer normal field $\boldsymbol{N}$ can be written in terms of the scalar displacement field u as $\boldsymbol{N} = \boldsymbol{N}_0 - \nabla u$ [68, 45], where $\boldsymbol{N}_0$ is the undeformed layer normal. The free-energy density F is given by:

$$\begin{aligned} F = {} & \frac{B}{4}(1 - N^4)^2 + K\, N^2\, (\nabla \cdot \boldsymbol{N})^2 \\ & + \frac{1}{2} K_{24}\, N^2\, \nabla \cdot \left[(\boldsymbol{N} \cdot \nabla)\, \boldsymbol{N} - \boldsymbol{N}\, (\nabla \cdot \boldsymbol{N}) \right]. \end{aligned} \tag{4.2}$$

Here, the first term penalizes compression or extension of the layers away from $N = 1$. The second and third terms are related to splay and saddle-splay distortions, which are inherited from the Oseen-Frank elastic free energy [190, 272, 79]. Notice that the order parameter $\boldsymbol{N}$ plays a dual role, and is very close to a unit vector field away from the focal singularities because of the small de Gennes' length (we will elaborate on this choice of dynamics in the next few paragraphs). The Lagrange multiplier $\boldsymbol{\lambda}$ forbids dislocations by ensuring that the layer-normal field is curl-free, since the vector $\nabla \times \boldsymbol{N}$ is the density of dislocations (the Burger's vector in units of the average layer spacing is given by the contour and area integrals $\oint \boldsymbol{N} \cdot d\boldsymbol{\ell} = \int_\Gamma \nabla \times \boldsymbol{N} \cdot d\boldsymbol{s}$). We will treat the effects and dynamics of dislocations in a separate paper [2]. Note that there is no term in the free energy to account for anchoring at the boundaries. We instead enforce strict anchoring, by doubling the simulation volume and using suitably symmetrized initial conditions, to enforce the homeotropic or planar boundary conditions (see section 4.3 for more details).

Note that a more general smectic free energy should depend on two order

---

[2]D. B. Liarte, A. Acharya, M. Bierbaum, and J. P. Sethna, *in preparation.*

parameters, such as the displacement field and the tensorial liquid-crystalline order parameter. For uniaxial order, it is possible to write the elastic free energy as a functional of the Frank director $\boldsymbol{n}$ and layer normal $\boldsymbol{N}$ vector fields. Assuming these vectors are parallel (they should be in smectics-A), and that their sizes are nearly constant (they will be constant except near singularities), it is possible to minimize the free energy with respect to one of the fields, yielding a relationship between N^2 and n^2, and derive a (complicated) effective free energy in terms of a single field. For the sake of simplicity, we bypassed this analysis, and started with a single order parameter. The unusual amplitude dependence ($\sim N^2$) multiplying the K and K_{24} elastic terms is motivated by gradient distortions of the form $(\nabla Q)^2$, which are proportional to N^2 for nematic uniaxial ordering [264], where $Q = ((Q_{ij}))$ is the Maier-Saupe tensorial order parameter. Since $\boldsymbol{n}$ and $\boldsymbol{N}$ are parallel, we use N^4 in the first term of Eq. (4.2) because the lowest order invariant in a Landau-de Gennes theory ($\text{tr}Q^2$) is proportional to N^4. Strictly speaking, neglecting an effective coupling between $\boldsymbol{n}$ and $\boldsymbol{N}$, the compression term should be proportional to $(1 - N^2)^2$, as in the first term of the r.h.s. of Eq. (4.4). Later on we will get back to this choice for the smectic dynamics (see Eqs. (4.4) and (4.5)).

To arrive at the smectic's dynamical equations of motion, we evolve the layer normal field in the simplest possible form, assuming $\boldsymbol{N}$ relaxes directly towards equilibrium. These dynamics give a partial differential equation for the gradient-descent evolution of $\boldsymbol{N}$:

$$\gamma \, \dot{\boldsymbol{N}} = -\left(\frac{\delta \Psi}{\delta \boldsymbol{N}} - \left\langle \frac{\delta \Psi}{\delta \boldsymbol{N}} \right\rangle \right), \tag{4.3}$$

where the angle brackets denote a spatial average and γ is a viscosity constant; γ can be written in terms of Leslie coefficients as $\gamma = \alpha_3 - \alpha_2$ [68]. The second term of (4.3) ensures that the net number of layers in the cell given by $\boldsymbol{N}_0$ does not

change during the gradient descent step. Equations (4.1-4.3) differ from Ericksen-Leslie-Parodi (ELP) dynamics in a few aspects. We relax the constraint of equal layer spacing $|\mathbf{N}| = 1$, which is ensured in ELP theory by means of a Lagrange multiplier, and we consider amplitude-dependent elastic constants. Apart from the dependence on $\mathbf{N}$, our dynamics is a particular case of ELP theory in the limit of infinite fluid viscosity. As a result, our centers of mass move affinely with the external shear and only the orientation of the molecules change.

We have also considered two other choices for the energy-gradient dynamics, which are not completely described by Eqs. (4.1-4.3). For future reference, we label the dynamics described in the last paragraph as *dynamics I*. For our second choice (*dynamics II*), we keep equations (4.1) and (4.3), but replace the free energy density by:

$$F_{II} = \frac{B}{4}\left(1 - N^2\right)^2 + K\left(\nabla \cdot \mathbf{N}\right)^2 . \tag{4.4}$$

Note that this choice of dynamics does not make contact with the tensorial order parameter Q. Since there is no amplitude dependence, the saddle-splay term is a surface term that vanishes in a system with periodic boundary conditions. On the one hand, the absence of a saddle-splay term limits the morphology of the allowed focal conic domains, since this term is associated with the Gaussian curvature energy of the layers [132] (the splay term is associated with the mean curvature). On the other hand, the equations of motion are simpler for *dynamics II*, so that we can implement simulations in a more efficient way, and study the numerical effects of varying grid sizes and de Genne's lengths (see Appendix 4.8.1). Finally, for our third choice (*dynamics III*) we consider the free energy as a functional of the displacement field and its derivatives, and replace Eq. (4.2) by:

$$F_{III} = \frac{B}{4}\left[1 - (\nabla u)^2\right]^2 + K\left(\nabla \cdot \nabla u\right)^2 , \tag{4.5}$$

which is equivalent to (4.4), and Eq. (4.3) by

$$\frac{\gamma}{\lambda^2}\dot{u} = -\frac{\delta\Psi}{\delta u}, \tag{4.6}$$

where γ is a viscosity constant, and λ is a length scale that we take to be the grid spacing a for convenience. This roughly corresponds to a smectic where the motion of layers is the dynamical bottleneck, rather than the reorientation of molecules (and hence the layer normals). The numerical evolution is slower for this choice of dynamics, probably due to derivatives of higher order in the equations of motion. Fig. 4.2 shows polarizer microscopy images of a simulated smectic-A planar section, starting from the same initial condition (which has been used in Fig. 4.1(b)) and evolved using *dynamics I* (a), *II* (b) and *III* (c). For (a) and (b), we evolved the initial configuration for a period of about $t = 2000\tau$, where $\tau \equiv \gamma/B$. (c) was obtained using *dynamics III* for a longer time ($\sim 10000\tau$). The morphology in (c) resembles the FCD pattern shown in Fig. 2d of [165] for metallotropic liquid crystals.

(a) (b) (c)
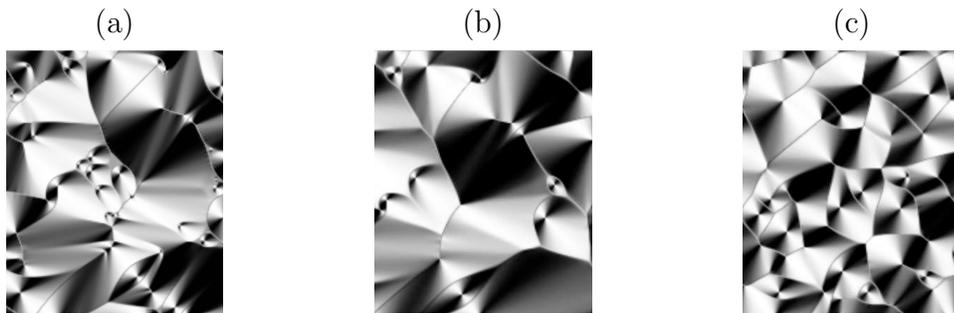
Figure 4.2: **Simulation results for a planar section of smectic-A** as viewed through polarizer microscopy images, simulated using *dynamics I* (a), *II* (b), and *III* (c).

To impose the external shear and extensional flows, we assume the layers are dragged with a displacement field determined by the flow. For simple shear, the layers are dragged in the x direction according to the displacement field

$$u_x^{\mathrm{s}}(x, y, z; t) = \frac{A}{l_z}(z - l_z)\sin(\omega t), \tag{4.7}$$

where l_z is the system size in the z direction, A is the amplitude, and ω is the frequency of oscillation; our simulations are done at a fixed Ericksen number $\gamma \omega l_z^2 / K \approx 129$. Extensional dynamics are implemented by stretching the smectic in the z direction while it contracts in the orthogonal x and y directions, as described by the set of equations

$$l_z(t) = l_z(0) f(t), \quad l_{x,y}(t) = \frac{l_{x,y}(0)}{\sqrt{f(t)}}$$

$$f(0) = 1, \quad f(t) > 0, \, \forall \, t \in [0, \infty) \quad ,$$

(4.8)

where l_x, l_y, and l_z are the grid sizes along the x, y, and z directions. To incorporate shear and dilatational dynamics simultaneously with the director relaxation, we employ an operator splitting method, alternatively applying gradient-descent motion from Eq. 4.3 and one of the loading dynamics from Eqs. (4.7) & (4.8).

4.3 Experimental and simulation setup

We perform analogous experiments on 8CB in the SmA phase, using a custom-built shear cell that allows precise control of the plate separation for gaps as small as 2-5 μm while keeping the plates parallel to < 1 part in 10^3 [157], allowing us to explore a large range of strain amplitudes and Ericksen numbers. The shear cell is outfitted with two parallel glass plates, which we use as the sample boundaries, and imaged with cross-polarized microscopy. We treat the glass slides with cetyl-trimethylammonium bromide for homeotropic anchoring and with a poly-imide treating for planar anchoring.

At the beginning of our simulations, we generate normally distributed random grids for each spatial component of the layer-normal field. We then enforce anchoring constraints, and use a Gaussian filter to smooth the field on short length scales.

To implement boundary conditions, we double the grid size in the z direction, and require that

$$N_x(l_z + z) = N_x(l_z - z), \; N_y(l_z + z) = N_y(l_z - z),$$
$$N_z(l_z + z) = -N_z(l_z - z), \tag{4.9}$$

for planar anchoring, and

$$N_x(l_z + z) = -N_x(l_z - z), \; N_y(l_z + z) = -N_y(l_z - z),$$
$$N_z(l_z + z) = N_z(l_z - z), \tag{4.10}$$

for homeotropic anchoring, with $0 \leq z \leq l_z - 1$. Mixed homeotropic and planar boundary conditions can be enforced in a similar way by quadrupling the thickness of the simulation grid. In order to remove the curl component of the field, we use a Helmholtz decomposition in Fourier space. The resulting components are divided by the mean length of the director field so that the field has average unit norm. We use an Euler integrator with adaptive step size [200] in order to integrate our partial differential equations. The driving code is written in Python. Each step of the integration is evaluated using parallel computing on a GPU using CUDA. Spatial derivatives are evaluated with Fourier methods (FFTs). In this letter we present results for fixed values for the ratio $K_{24}/K = -1.5$, and for deGenne's length scale $\xi = \sqrt{K/B} = 0.2a$, where a is the finite-difference grid spacing. (Larger ξ produces similar results with blurred features; see Appendix 4.8.1). Except in the study of dilatational flow, we have presented results for fixed values of $l_z = 320\,\xi$. A systematic study of the dependence of the dynamical behavior on sample thickness is beyond the scope of the present work.

4.4   Visualization

From these random initial conditions, our gradient-descent dynamics forms FCDs which closely resemble those seen in experiments as shown in Fig. 4.1. We visualize the focal conics domains in our simulations through several techniques we have developed. We render polarizer microscopy images, as shown in Fig. 4.1, by ray-tracing light using the Fresnel equations for anisotropic dielectrics [132]. Figure 4.3 shows a plot of the free energy density F, overlaid with cross-sections of the layer surfaces (contours of constant $\boldsymbol{N}_0 \cdot \boldsymbol{r} - u$). The free energy is high at the focal lines, where the layer contours form cusps.
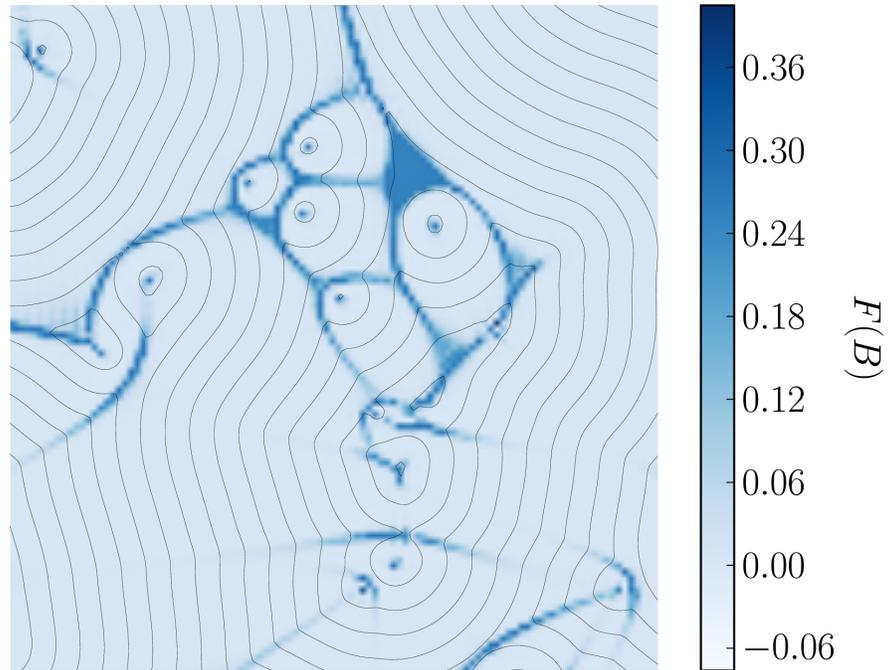
Figure 4.3: **Simulated energy density** (white-blue density plot) and some sections of the layer surfaces (black lines) at the top section $z = l_z$ of the system with planar anchoring (Fig 4.1b).

Three visualizations of the three-dimensional smectic structure are shown in

figure 4.4. Figure 4.4a is a volume rendering visualization of the free energy density where each voxel is given a color and degree of transparency that is associated with its free energy density. The high energy regions (red) have organized into the characteristic focal conic structure of smectics, forming multiple ellipses, each with a hyperbola coming out of its focus. The focal conic character of the smectic structures is reinforced by the loci of the principal centers of curvature of the smectic layers, shown in Fig. 4.4b, which coincide with the regions of high energy density in 4.4a [132]. To calculate the radii of curvature, we project each layer's second fundamental form tensor $M_{\mu\nu} = \partial_\mu \partial_\nu u$ into the layer-surface tangent plane. The principal radii of curvature are equal to the inverse of the principal curvatures, which are the eigenvalues of the projected $M_{\mu\nu}$. The shared surface normals intersect at the centers of curvature for the layers, which form the confocal conics [228]. Finally, Fig. 4.4c shows three-dimensional level surfaces of the mass-density field.

(a)

(b)

(c)
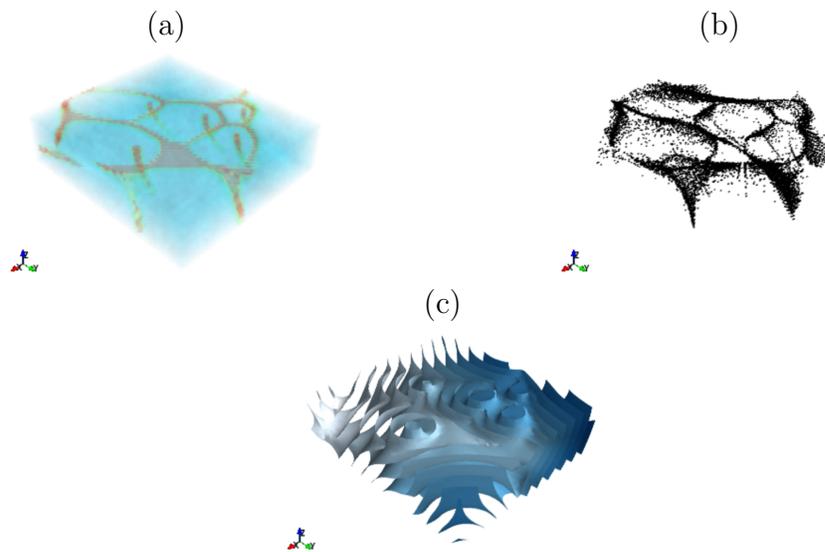
Figure 4.4: **3D visualizers of a simulation of smectic-A liquid crystals with planar anchoring**. (a) Volume rendering visualization of the energy density; (b) loci of the centers of curvature of the layer surfaces; (c) layer surfaces.

4.5 Coarsening

To study the coarsening dynamics of focal conics we simulate with no anchoring at the boundary, since the boundary constraint introduces a length scale for the distribution of the layer-surface radii of curvature. As the system evolves, our dynamics seem to energetically favor ellipses with large linear eccentricity c. The layers around singular ellipses become flatter with increasing c, and converge to planes when $c \to \infty$. This is the dominant coarsening mechanism in our simulations. The coarsening of focal conics becomes slower with increasing time, but it does not stop until a uniform flat configuration is reached. (Our computational defect structures can be stabilized with simulated 'dust' particles on the glass slides, by introducing spatially-dependendent energetic anchoring on the boundaries.) To quantify the coarsening, we investigate the probability distribution of the principal radii of curvature R, which define a characteristic distance to the focal conics, and are distributed according to a function $P(R, t)$ which also depends on time. The scaling assumption states that the morphology at late times statistically scales with a single length scale R^*, so in particular $P(R, t) \propto \Pi(R/R^*(t))$ for some (possibly universal) function $\Pi(X)$. In Appendix 4.8.2, we propose two possible arguments yielding the cutoff radius of curvature $R^* \sim t^{1/4}$, and $R^* \sim [t \ln(t/\tau_c)]^{1/4}$. Fig 4.5 shows scaling collapse plots that are consistent with both of these possibilities, (see Appendix 4.8.2 for more details, and for a discussion of the decay in the energy density with time). Using the first scaling form ($\sim t^\alpha$), we observe that the exponent $\alpha = 0.5$ gives a better numerical collapse of the data. We do not show the collapse plot for this exponent because it does not have theoretical motivation. Incidentally, the inclusion of logarithmic corrections (due to the singularities near the conic sections) makes the collapse worse for $\alpha = 0.5$, but improves the collapse

using $\alpha = 0.25$. These results do not change if we use *dynamics II* [3]. The approach to equilibrium by *increasing* eccentricity to minimize bending energy is an interesting contrast to the typical approach to equilibrium of decreasing eccentricity to minimize a surface energy.

4.6   Flow dynamics

Our simulations and experiments also provide a window to understand dynamics of focal conic domains under shear. From our simulations, simple shear oscillations parallel to the glass slides primarily act to accelerate the focal conics' coarsening. When we shear stabilized focal conic structures, our simulations show that the focal conics are not significantly altered by the planar shear, in qualitative agreement with our experiments with strong homeotropic anchoring. In addition, our simulations allow us to tune the smectic's anchoring at the boundary. As a result, our simulations promise to discern the effects of anchoring imperfections, such as weak or spatially-modulated anchoring, on the rich structures that can be produced in experiments (see supplemental animations [151]).

Under dilative strain (stretching the layer spacing), homeotropic smectic-A liquid crystals are known to release free energy by forming undulations [69, 59, 58], and focal conic domains [211, 48]. In Fig. 4.6a we show simulation results for the total free energy as a function of time for a dilative dynamics with $f(t) = 1 + A(1 - \cos \omega t)$, where $A = 0.25$, and $\omega = 2\pi/1000 \, (\tau^{-1})$. The first sharp peak at about $t_1 = 100 \, \tau$ marks the onset of an undulation pattern, which is depicted

---

[3]Previously, using a different amplitude dependence of the elastic constants ($\sim N^4$), we were motivated by an apparent self-similar structure to propose a third scaling form ($\sim t^\alpha$, with $\alpha \approx 0.29$). The interested reader should refer to the second version of this manuscript in the arXiv repository [152].

in the layers contour plot in the inset of Fig. 4.6a. Linear stability analysis using the methods of [69, 59, 58] leads to a critical strain threshold ϵ_c that is given by the solution of the equation (see Appendix 4.8.3):

$$\epsilon_c = \frac{\pi \xi}{l_z} \sqrt{\frac{1 - 6\epsilon_c + 6\epsilon_c^2}{(1 - 3\epsilon_c + 2\epsilon_c^2)^2}} \approx \frac{\pi \xi}{l_z}, \tag{4.11}$$

since ϵ_c is small. This analysis results in a buckling wavelength of $\approx 9a \approx 0.04 l_x$, which is consistent with our simulations (see inset of Fig. 4.6a), as is the onset strain of the instability (the first peak in Fig. 4.6a) is later than the instability onset by approximately a factor of two, because of the growth time of the undulation pattern). The second peak of the free energy signals the onset of a configuration which evolves towards a complex pattern of focal conic domains. Fig. 4.6b and 4.6c show crossed-polarizer images obtained from simulations and experiments at maximum strain, respectively. We found compatible results using *dynamics II*.

Note the fascinating fact that the critical change in length $\epsilon l_z \approx \pi \xi$ is a microscopic length. Except near a critical point, one expects ξ to be of the order of a molecular size; the instability threshold [4] for a bulk material happens when one stretches it by one molecular length [143]! A simple calculation for crystals shows an analogous result for grain boundaries: a bent crystal's ground state has dislocations once the net displacements become of the order of the lattice constant (up to a logarithm of the crystal size over the atomic size). Unlike crystals which are metastable, smectics are unstable under long-wavelength deformations with atomic-scale displacements – the lower-energy defective state has no associated nucleation barrier. Thus the equilibrium continuum elastic theory of materials with broken translation invariance is frail [40, 41]– it is only valid in general for *microscale* net displacement differences over *macroscale* distances.

---

[4]This instability to defect structures is *not* true for two-dimensional smectics with undulating [74, 59] boundary conditions [117, Ch. 4].

4.7 Conclusions

To conclude, we have presented results from numerical simulations and experiments of smectic-A liquid crystals. In our simulations, focal conic domains spontaneously emerge out of random initial configurations. The numerical reproduction of the experimental morphologies is striking, for both planar and homeotropic boundary conditions. Our several visualization tools comprise the description of the most important structural features of FCDs. We have also characterized the coarsening of FCDs, by studying the scaling behaviour associated with the distribution of the layer surface radii of curvature, which is the length scale related to the size of the focal conics. Finally, we have studied the system response to strain, which includes a numerical and experimental investigation of the classical dilatational instability, correctly predicting the instability threshold, and the onset of a state populated with parabolic focal conic domains.

4.8 Appendix

4.8.1 Convergence tests

In this section we present some results of a test to analyze the effects of small grid sizes and small de Genne's length ($\xi = 0.2a$) in our simulations. We start with a smoothened random initial field of linear size $16a$, and evolve it for a very short time ($\approx 0.5\tau$), using *dynamics II* with $\xi = 0.2a$ [5] and no anchoring at the boundaries. We then duplicate the resulting configuration into larger lattices, with linear size

---

[5]We do not use *dynamics I* in these calculations because it is harder to control numerical instabilities for larger de Genne's length.

32, 64 and 128. To generate a smooth interpolation between lattice points of the larger lattices, we copy the Fourier components of small wave number, and leave the coefficients associated with short wavelengths equal zero. To be consistent with the periodic boundary conditions, we divide the smaller cubic grid into eight equal pieces, and copy the configuration of each piece into the corresponding corner of the larger grid in Fourier space. To estimate the finite-size error, we evaluate

$$\sigma_b = \max_{\mu \in \{x,y,z\}} \max_{i,j,k} \left| N_\mu(b\,i, b\,j, b\,k) - N'_\mu(i, j, k) \right| \tag{4.12}$$

where $\boldsymbol{N}$ and $\boldsymbol{N'}$ are the large and small lattices, with linear sizes l and $l' = b\,l$, respectively. The indices i, j, and k are grid coordinates of $\boldsymbol{N'}$. The second row of Table 4.1 shows the error comparison for this initial configuration. Since the configuration is copied (with a smooth interpolation) from the smaller to the larger lattices, this error is of order 10^{-15}. We then evolve this initial state for each grid for a period of time of about 200τ, keeping $\xi = 0.2a$ for $l = 16$, and using *dynamics II*. In order to have comparable simulations, we multiply K and K_{24} by four (thus increasing ξ by a factor of two) each time we double the grid size, since the wave vectors are divided by two, and the splay and saddle-splay terms contribute with two gradient terms. The results for $\xi_{L=16} = 0.2a$ and $t = 200\tau$ are shown in the third row of Table 4.1. Note that the difference between the 32^3 lattice (with $\xi = 0.4a$) and the 128^3 lattice (with $\xi = 1.6a$) is of just about two percent. For the sake of completeness, we started with the same initial state, and evolved each lattice using *dynamics II* with $\xi_{L=16} = 0.4a$ (so that $\xi = 0.8a, 1.6a$, and $3.2a$ for $L = 32, 64$ and 128, respectively). The results for this case are shown in the fourth row. Notice that $\sigma_2(64, 128)$ for $\xi_{L=16} = 0.2a$ is comparable to $\sigma_4(32, 128)$ for $\xi_{L=16} = 0.4a$, because the 64^3 lattice in the third row and the 32^3 lattice in the fourth row are simulated using the same de Genne's length. The same comparison holds between $\sigma_2(32, 128)$ for $\xi_{L=16} = 0.2a$ and $\sigma_4(16, 128)$ for $\xi_{L=16} = 0.4a$. Note

that this test analyzes convergence upon increasing both resolution and de Genne's length, and that $\xi = 0.4a$ gives sensible results even for small lattices. We recall that most of the results that have been presented in this paper were obtained using *dynamics I* with $\xi = 0.2a$ for large grids ($256^2 \times 64$). We keep the results for *dynamics I*, even though it is harder to control numerical instabilities in this case, because that is our only choice with a Gaussian curvature energy, which is associated with the saddle-splay term. We emphasize that we do not observe a significant change for our results for coarsening and dilative dynamics when we use *dynamics II* with $\xi = 0.4$.

	σ_8 $(16, 128)$	σ_4 $(32, 128)$	σ_2 $(64, 128)$
Initial state	10^{-15}	10^{-15}	10^{-15}
$\xi_{L=16} = 0.2a$ at $t = 200\tau$	0.55	0.03	0.0007
$\xi_{L=16} = 0.4a$ at $t = 200\tau$	0.02	0.0009	0.0007

Table 4.1: **Convergence of simulation results with grid size**. Comparison of the errors of evolved simulation configurations for grids with increasing resolution and de Genne's length.

4.8.2 Scaling exponent for the coarsening of focal conics

The principal radii of curvature R define a characteristic distance to the focal conics: equally spaced layers develop singularities at their centers of curvature. These radii have a time-dependent probability distribution $P(R|t)$. Scaling suggests that all correlation functions should scale with a single length scale $R^*(t)$ that diverges at late times, hence $P(R|t) \approx \Pi(R/R^*)/R^*$ for some perhaps universal function $\Pi(X)$. (Here the last factor preserves normalization: $\int P(R)dR = \int \Pi(X)dX = 1$.) In coarsening problems, it is often possible to use simple energetic arguments to derive the power law divergence $R^*(t) \propto t^\alpha$; for example, phase separation in systems without hydrodynamic flow has $\alpha = 1/3$ for

conserved order parameters and $\alpha = 1/2$ for non-conserved order parameters. Here we give two possible scaling forms, of increasing sophistication. The first mimics the standard energetic arguments and the second provides a refined argument including a logarithmic correction due to defect cores (Fig. 4.5).

Away from the defect cores, where $|\boldsymbol{N}| \approx 1$, the free energy density scales as R^{-2}. So the average energy density is

$$\mathcal{E}(R^*) = \int P(R)/R^2 dR. \tag{4.13}$$

In traditional coarsening, one assumes that the integral for $\mathcal{E}(R^*)$ converges at zero, so $\mathcal{E}(R^*) \sim 1/(R^*)^2$. This leads to a force (tension) $T = \delta\mathcal{E}/\delta R^* \sim 1/(R^*)^3$. Since the order parameter is non-conserved [6],

$$\dot{R} = -\gamma T, \tag{4.14}$$

where γ is an effective inverse viscosity (see section 11.4 of [225]). This can be solved giving $R^* \sim t^{1/4}$ (Fig. 4.5a), and hence $\mathcal{E}(R^*(t)) \sim 1/t^{1/2}$ (Fig. 4.7).

How does this change if we consider the defect cores, where $|\boldsymbol{N}| < 1$? The energy in the cones, near the focal conic line singularities, scales as the length of the conics times $\ln(R/\xi)$, where ξ is de Gennes' length scale. Within a focal domain of size R^*, near the singular ellipse and hyperbola $R \to 0$, the volume fraction $P(R) \sim R$, so that $\Pi(X) \sim X$ for small X. This leads to a divergence in the integrated energy near the focal conic singularities, which is cut off by ξ,

$$\begin{aligned}
\mathcal{E}(R^*) &= \int_\xi^\infty (1/R^2)\Pi(R/R^*)/R^* dR \\
&= \int_{\xi/R^*}^\infty \Pi(X)/(XR^*)^2)dX \\
&\sim (1/R^*)^2 \int_{\xi/R^*}^1 X/X^2 dX \\
&= \log(\xi/R^*)/(R^*)^2,
\end{aligned} \tag{4.15}$$

---

[6]*Dynamics III* could perhaps correspond to conserved dynamics.

(see section 10.5 of [132]). After some calculation, Eq. (4.14) implies

$$R^* \sim [t \ln(t/\tau_c)]^{1/4} \,, \tag{4.16}$$

in the limit of large R or t. So if the focal domains are all of the same length scale R^*, and the relaxation of the core singularites dominates the coarsening, we expect a $t^{1/4}$ scaling with a logarithmic correction, as in Fig. 4.5(b). There is a large range of values for τ_c which collapse our data. Fig. 4.5b shows a scaling collapse plot with the logarithmic corrections for $\tau_c = 100\tau$. Note that Eq. (4.16) only applies for times times t larger than τ_c, hence the range of times used in the collapse plot of Fig. 4.5(b). Our data for $t < 200\tau$ do not fit well in the collapse plot, even when we consider lower τ_c so that t is still greater than τ_c; we surmise that τ_c is associated with the time needed to form line singularities. Unfortunately, we have not been able to verify this hypothesis, since there is no surface anchoring in this case and the three-dimensional visualizers are not useful at early stages of the dynamic evolution.

Both of these scaling forms are compatible with the data, given the limited scaling regime (less than a decade in length, corresponding to less than three decades in 'size'); $P(R, t)$ is clearly still evolving in shape from its non-universal initial form.

4.8.3 Linear stability analysis for SmA under dilative strain

We consider a situation where a thin slab of homeotropic smectic-A is subject to dilative stress [69, 59, 58, 68, 45, 132]. In this case, the smectic layers are parallel

to the glass slides, so that the stretching of the gap promotes an increase of the interlayer spacing. Planar-layer configurations store a considerable amount of bulk energy as strain is increased, which is released with the formation of an undulation pattern after a critical strain is reached. Here we use the methods of [69, 59, 58] to study the formation of undulation instabilities on smectic-A liquid crystals.

The displacement field associated with an undulation pattern of a smectic-A can be written as:

$$u(\boldsymbol{r}) = \epsilon z + u_0 \cos(qx) \sin(kz), \tag{4.17}$$

where we take $k = \pi/l_z$ to enforce strict homeotropic anchoring. Our elastic free energy density is given by:

$$f = \frac{B}{4}(1 - N^4)^2 + K (\nabla \cdot \boldsymbol{N})^2. \tag{4.18}$$

Notice that we have not included a saddle-splay term, nor have we considered amplitude dependence of the elastic constant K, since their effects are negligible. Roughly speaking, the amplitude dependence of K_{24} gives rise to higher-order terms for the undulating solution. Hence we can approximate N^2 multiplying K and K_{24} by one, and the saddle-splay term becomes a surface term that vanishes for periodic boundary conditions. Also, we do not need include a Lagrange multiplier, since $\boldsymbol{N} = \boldsymbol{z} - \nabla u$ is curl-less if u is given by Eq. (4.17). The free energy density (4.18) can be written in terms of the displacement field as

$$
\begin{aligned}
f &= \frac{B}{4}\left\{1 - \left[1 + \left(\frac{\partial u}{\partial x}\right)^2 - 2\frac{\partial u}{\partial z}\right]^2\right\}^2 \\
&\quad + K \left(\frac{\partial^2 u}{\partial x^2}\right)^2.
\end{aligned}
\tag{4.19}
$$

We can combine Eqs. (4.19) and (4.17) in order to write

$$
\begin{aligned}
\frac{f}{B} = {} & \xi^2 q^4 u_0{}^2 \sin^2(kz)\cos^2(qx) \\
& + \frac{1}{4}\left\{\left[q^2 u_0{}^2 \sin^2(kz)\sin^2(qx)\right.\right. \\
& \left.\left. - 2(ku_0\cos(kz)\cos(qx) + \epsilon) + 1\right]^2 - 1\right\}^2,
\end{aligned}
\tag{4.20}
$$

where $\xi = \sqrt{K/B}$ is de Gennes' length scale. To find the stability threshold we integrate out the free energy density over one period in the x-direction, and from 0 to l_z in the z-direction:

$$
F = \int_0^{\frac{2\pi}{q}} dx \int_0^{l_z} dz\, f(x, z).
\tag{4.21}
$$

The stability threshold is given by the solution of the equation:

$$
\frac{\partial^2 F}{\partial u_0{}^2} = 0,
\tag{4.22}
$$

or,

$$
\begin{aligned}
& 4k^2\left(6\epsilon^2 - 6\epsilon + 1\right) \\
& + q^2\left(\xi^2 q^2 - 8\epsilon^3 + 12\epsilon^2 - 4\epsilon\right) = 0.
\end{aligned}
\tag{4.23}
$$

For given ξ and l_z, this equation defines a curve in the $\epsilon \times q$ plane. Fig 4.8 shows the critical strain as a function of q for $\xi = 0.2a$ and $l_z = 64a$, corresponding to our simulation parameters, where a is the finite-difference grid spacing. The strain is minimal for

$$
q = \frac{\sqrt{4\epsilon^3 - 6\epsilon^2 + 2\epsilon}}{\xi}.
\tag{4.24}
$$

Eq. (4.24) can be plugged back into Eq. (4.23), so that,

$$
\epsilon_c = \frac{\pi\xi}{l_z}\sqrt{\frac{1 - 6\epsilon_c + 6\epsilon_c{}^2}{(1 - 3\epsilon_c + 2\epsilon_c{}^2)^2}} \approx \frac{\pi\xi}{l_z},
\tag{4.25}
$$

where the approximate solution on the r.h.s. of (4.25) is valid since ϵ is small. Notice that our approximate critical strain $(\pi\xi/l_z)$ corresponds to half of the value

139

obtained in [69, 59, 58], because we use a slightly different form for the free energy density. Also, it is interesting to point out that the critical change in length $\epsilon_c l_z \approx \pi \xi$ is a microscopic length (see main text).

Fig 4.6 shows simulation results for smectic-A liquid crystals under dilative stress with $\xi = 0.2a$ and $l_z = 64a$. We simulate $f(t) = 1 + A(1 - \cos(\omega t))$ (see Eq. (6) of main text), with $A = 0.25$, and $\omega = 2\pi/1000 \, \tau^{-1}$. Eq (4.25) predicts a strain threshold of $\epsilon_c \approx 0.01$, and a critical wavelength of $q \approx 9a \approx 0.04 l_x$, which is consistent with our simulations (Fig 4.6b). However, since $\epsilon = 1 - f^{-1}$, we can rearrange $f(t)$ in order to write

$$t_c = \frac{1}{\omega} \cos^{-1}\left(1 - \frac{\epsilon_c}{A(1 - \epsilon_c)}\right) \approx 45\,\tau, \tag{4.26}$$

which is about half of the time threshold associated with the first peak of Fig 4.6a. We suggest that the time scale associated with the growth of the undulation pattern accounts for the discrepancy between the simulation threshold and the analytical estimate. We tested our stability analysis directly by adding a small perturbation $\delta \boldsymbol{N}$ to $\boldsymbol{N}_0$. Under a gradient descent infinitesimal evolution of $\boldsymbol{N}$, we expect that $\mathcal{F}[\delta N_z(t + \delta t)] = \exp(\lambda_k \delta t)\mathcal{F}[\delta N_z(t)]$, where $\mathcal{F}$ denotes a Fourier transform operator. An exponent λ_k is less than zero for stable planar configurations, and reaches zero at the critical strain for some wavenumber q. Careful numerical calculations for λ_k lead to $t_c \approx 45\tau$, and $q \approx 9a$, in agreement with our analytical estimate. There is no significant change (apart from shifts of numerical values) in the analysis and numerical results using *dynamics II*.

(a)

(b)
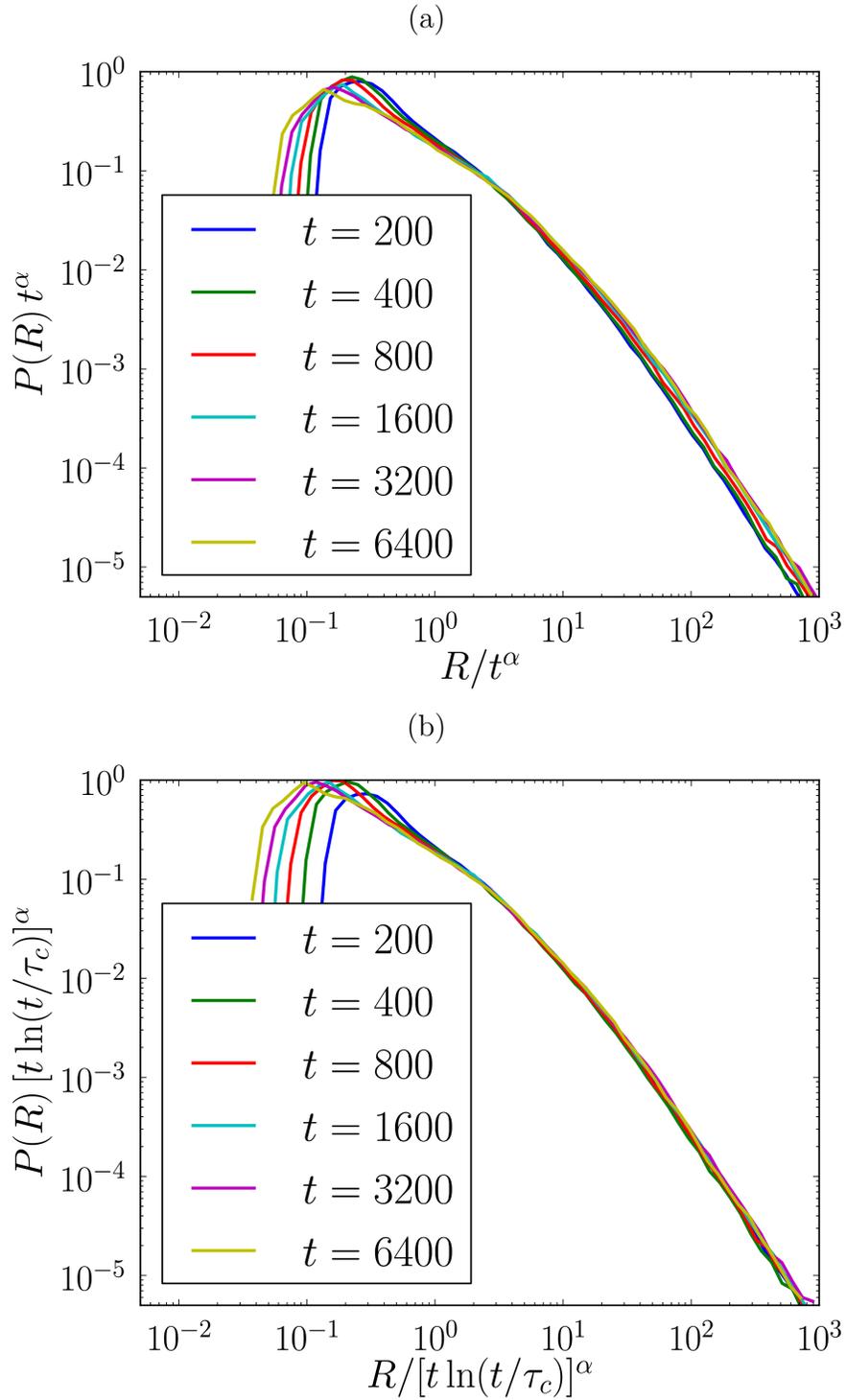
Figure 4.5: **Scaling collapses of the distribution layer-surface radii of curvature** on a logarithmic scale. (a) Naive power-law scaling, $R^*(t) \sim t^{1/4}$. (b) Incorporation of logarithmic corrections to scaling, $R^*(t) \sim [t \log(t/\tau_c)]^{1/4}$, with $\tau_c = 100\tau$.
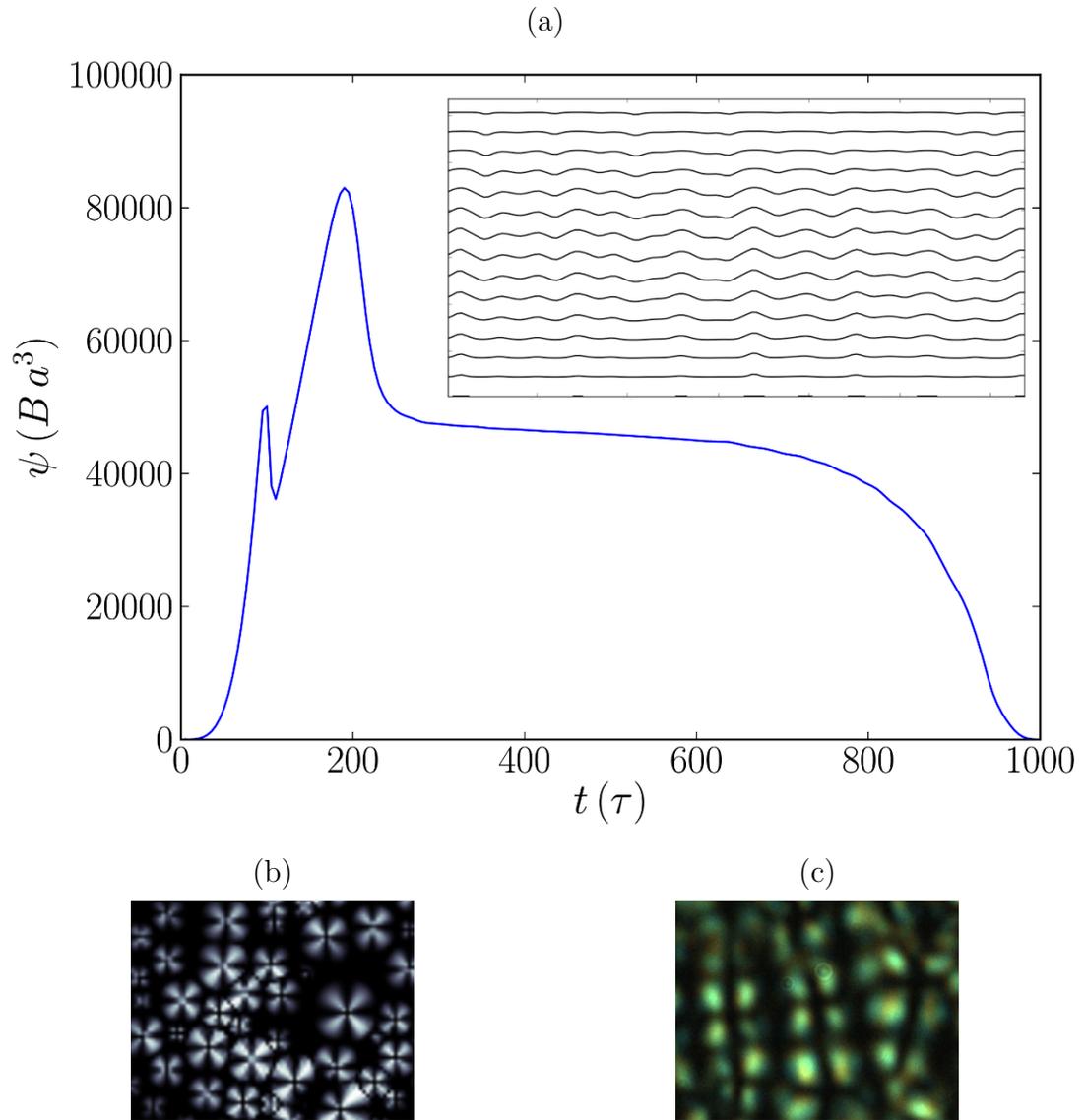
Figure 4.6: **Simulation and experimental results for SmA under dilative stress**. (a) Total free energy as a function of time. In the inset we show layers in the x-z plane showing undulation pattern at $t = 120\tau$. (b) Simulation and (c) experimental results for crossed-polarizer images showing a pattern of focal conic domains, at strain amplitudes of 0.33 and 0.13, respectively.
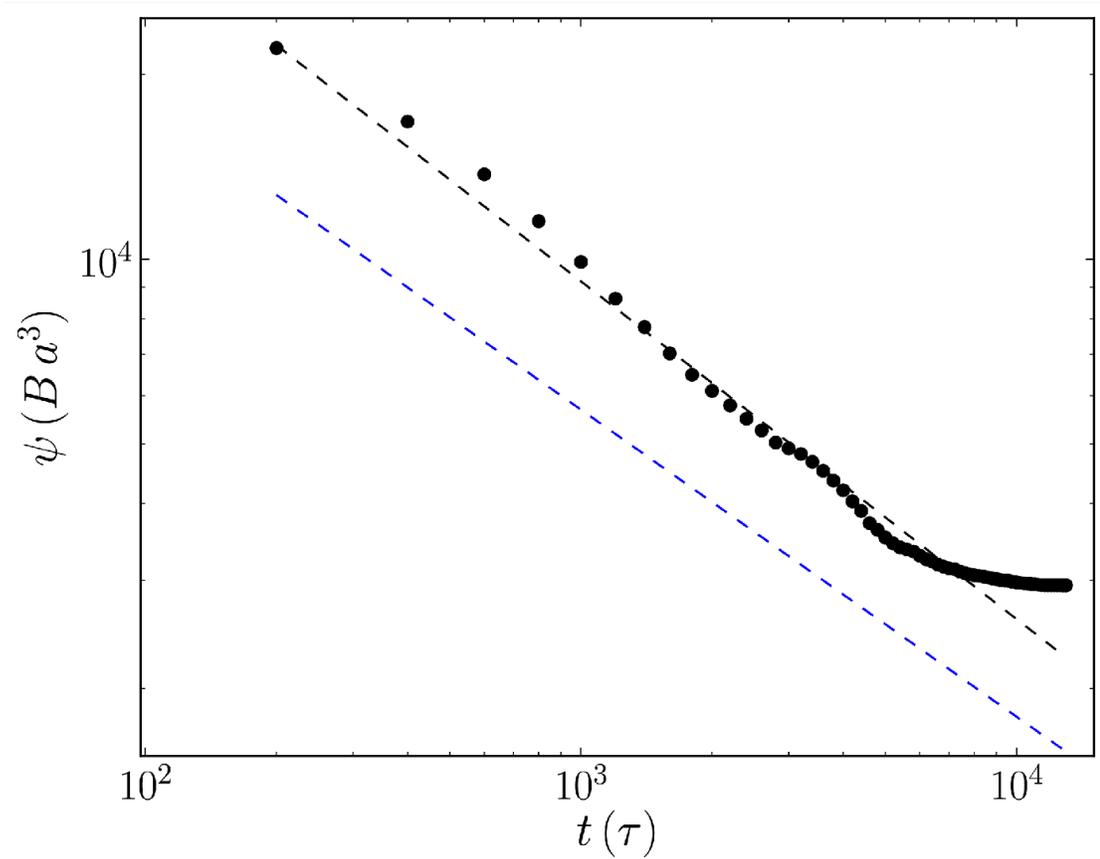
Figure 4.7: **Energy minimization through gradient descent**. Total free energy ψ (symbols) as a function of time in log-log scale. The blue dashed line correspond to the behavior predicted from the naive argument $R^* \sim t^{1/4}$. The black dashed line is a best fit ($\sim t^{-0.55}$).
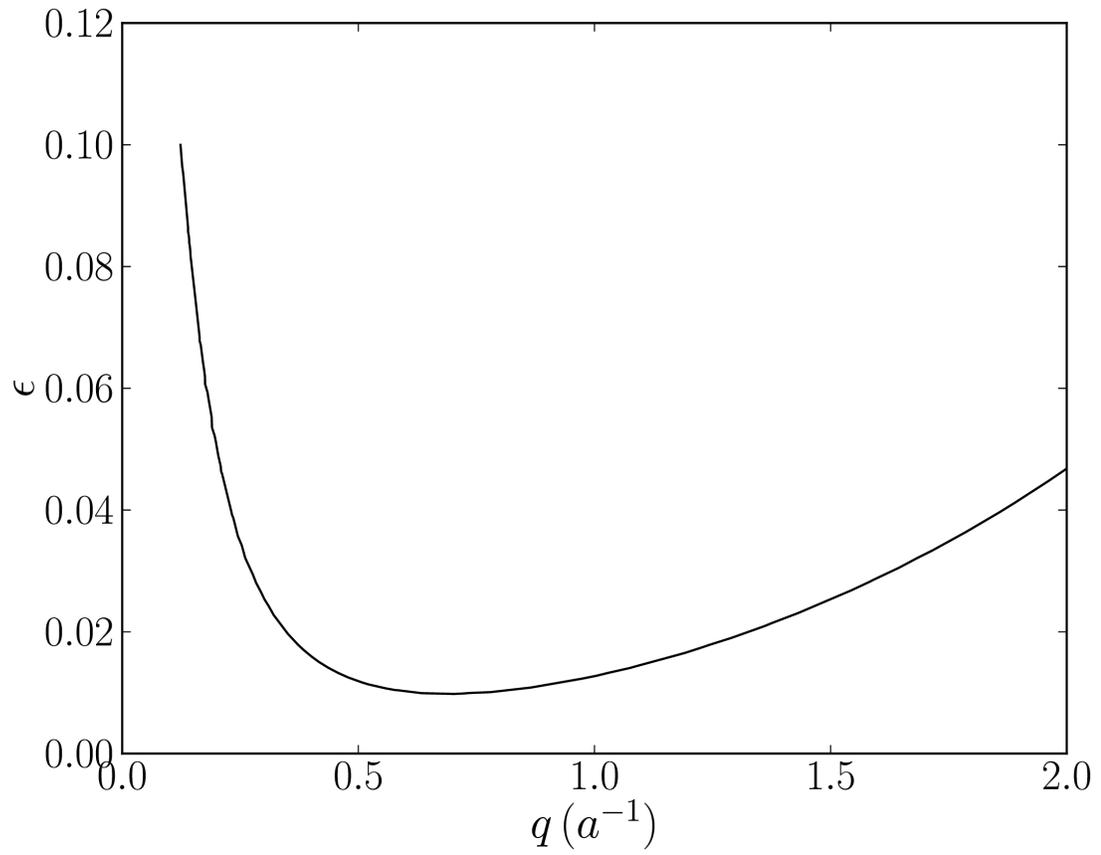
Figure 4.8: **Critical strain at which undulations form**. Critical strain as a function of q for $\xi = 0.2a$ and $l_z = 64a$.

CHAPTER 5

MEASURING NONLINEAR STRESSES GENERATED BY
DEFECTS IN 3D COLLOIDAL CRYSTALS[1]

5.1 Local stress measurements

Bulk measurements of the nonlinear materials response have shown that fascinating mechanical behaviors emerge when crystals are plastically deformed [123]. Such measurements however, average over the rich spatial heterogeneity in structure and stress distributions. This averaging makes it difficult to determine how microscopic mechanisms collude to determine a crystal's bulk behavior. Pioneering measurements of local crystalline strains have done much to elucidate the heterogeneity in the linear stress response of crystals [118, 219, 25, 125, 131, 150]. Despite these advances however, applying such techniques to measure the nonlinear stress distributions in crystals with defects has remained prohibitive since it is impossible to *a priori* determine how the nonlinear modulus varies with strain or even define a strain when the structure is highly distorted. Consequently, it has been difficult to experimentally determine even the qualitative interactions between defects that give rise to these fascinating mechanical behaviors under large deformations.

Here, building on the technological advances offered by high-speed confocal microscopy, we use Stress Assessment from Local Structural Anisotropy (SALSA) to directly measure the complete stress tensor down to the single particle-scale in a 3D colloidal crystal. Hard-sphere colloidal crystals have been widely employed as a model system to investigate many fundamental and important processes including defect nucleation [219], crystal melting [16, 195], and crystal growth [253].

---

[1]Neil Y.C. Lin, Matthew Bierbaum, James P. Sethna, & Itai Cohen
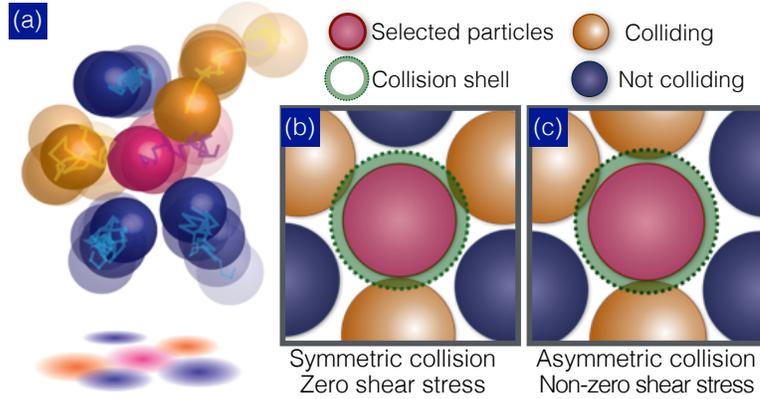
Figure 5.1: **Particle-level stress measurements (SALSA)**. (a) Particles exhibit Brownian motions (trajectories segments) and exert stresses on the selected particle (red sphere) when the neighboring particles collide with it. The energy density (stress) per collision is k_BT/Ω^α. (b) Schematic illustrating the SALSA algorithm for hard spheres. A thin shell (Δ=106 nm) is constructed to identify colliding particles (yellow spheres), which lie within distance $2a + \Delta$ from the selected particle. The shear stress is zero when the colliding particles' configuration is symmetric. When the collisions are asymmetric, the shear stress is non-zero. The schematics here are two dimensional, but all presented calculations are fully three dimensional.

In Brownian hard-sphere systems, the force with which particles collide can be related to the thermal energy k_BT. Therefore, using a time series of featured particle positions [72, 63], we determine the thermal collision probability, and directly report the stress arising from these Brownian collisions. Our derivation (see Supplemental Information) shows the stress tensor $\sigma_{ij}^\alpha = \sigma_{ij}(\mathbf{x}^\alpha)$ at particle α can be approximated by

$$\sigma_{ij}^\alpha = \frac{k_BT}{\Omega^\alpha}\left(\frac{a}{\Delta}\right)\langle\psi_{ij}^\alpha(\Delta)\rangle \tag{5.1}$$

where k_BT is the thermal energy, Ω^α is the volume occupied by the particle, a is the particle radius, and Δ is the cutoff distance from contact (SI and SI video). Here $\langle\psi_{ij}^\alpha(\Delta)\rangle$ is the time-averaged *local structural anisotropy* or fabric tensor for the particle, $\langle\psi_{ij}^\alpha(\Delta)\rangle = \langle\sum_{\beta\in nn}\hat{\mathbf{r}}_i^{\alpha\beta}\hat{\mathbf{r}}_j^{\alpha\beta}\rangle$, where nn is the set of particles that lie within a distance $2a + \Delta$ from particle α, ij are spatial indices, and $\hat{\mathbf{r}}^{\alpha\beta}$ is the
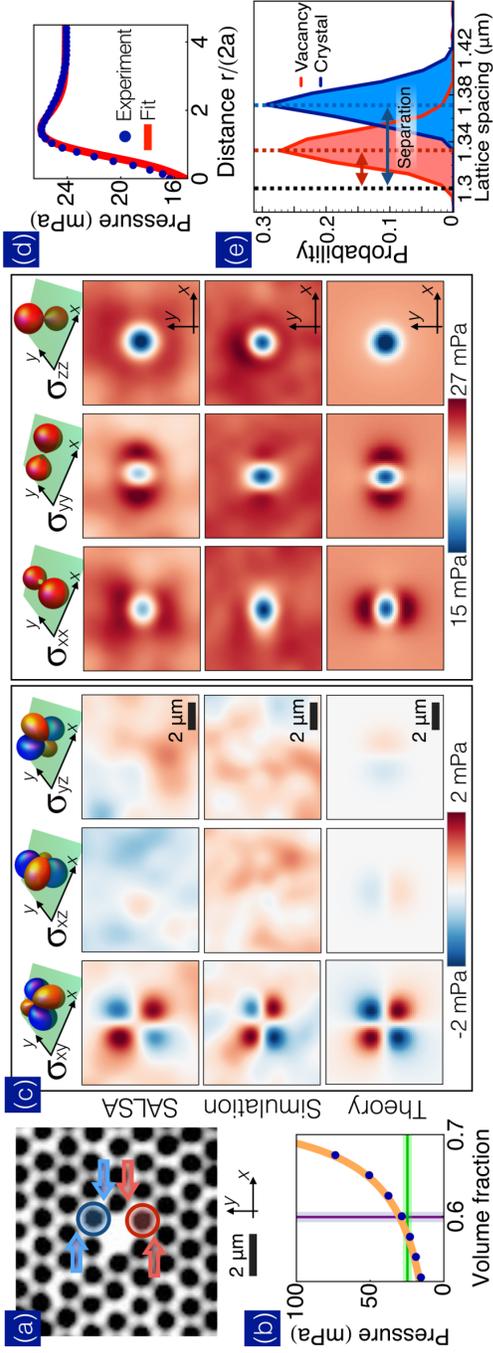
Figure 5.2: **Stress around a vacancy.** (a) Confocal image of a crystal with an isolated vacancy defect with no other defects within five lattice spacings in the plane, or in the adjacent layers along the (111) direction. The hard-sphere interparticle potential, large particle size, and high volume fraction slow down vacancy diffusion, which allows for time averaging. To account for particle polydispersity (SI) and the background stress we imaged 20 independent defects. Each vacancy is imaged for 20 seconds at a scan rate of 2 stacks per second, yielding a total of 800 snapshots for averaging. The absence of a particle at the defect core results in asymmetric collisions on the particles surrounding the vacancy. The upper right (blue) and lower right (red) particles are under positive and negative shear σ_{xy}. (b) SALSA accurately reports the mean pressure of the tested colloidal crystal with volume fraction 0.59 (purple line). Green line is the SALSA value. The blue dots are our hard sphere simulation results, and orange curve is the prediction from the literature. The shades of green and purple lines are the standard deviations. (c) All stress components around a vacancy determined using SALSA (upper row), simulation (middle row), and nonlinear elasticity (lower row). In contrast to the significant feature in σ_{xy}, we find that the small fluctuations in σ_{xz} and σ_{yz} are less than 20% of the variation in σ_{xy}. These measurements are consistent with the fact that both σ_{xz} and σ_{yz} exhibit nodes along the x-y plane. Additionally, we find that the normal stresses σ_{xx} and σ_{yy} demonstrate elastic dipoles that align horizontally and vertically, respectively.(d) Pressure is plotted as a function of distance d from the vacancy core. Both experimental (Blue) and theory (Red) results show clear stress enhancements near (red) and far away from (blue) the vacancy. The observed 50% change in surface-surface spacing would correspond to a $\sim 8.3\%$ local volume change in a defect free crystal.

147

unit vector between particle α and particle β. In the *local structural anisotropy* calculation, the trace $\hat{\mathbf{r}}_i^{\alpha\beta}\hat{\mathbf{r}}_i^{\alpha\beta}$ is the total number of neighbors while the remaining components captures the anisotropy of the collisions [32]. The time averaged fabric tensor of each particle accurately captures the probability of thermally induced collisions arising from the spatial distribution of its neighbors (Fig. 5.1). Scaling the probability by the energy density per collision $k_B T/\Omega^\alpha$, we then determine the Cauchy stress at the selected particle's position. This capability enables us to measure the local stress distributions surrounding crystalline defects such as vacancies (0D), dislocations (1D), and grain boundaries (2D).

5.1.1   Vacancy

Vacancies dominate mass transport in crystals by playing key roles in electromigration growth of voids in integrated circuit interconnects, impurity diffusion, and dislocation creep and climb. These processes are governed by the vacancy interaction arising from the stress field. Whether the stress field surrounding the core is linear or nonlinear directly determines the qualitative interaction between vacancies and influences our understandings of those processes. To measure the stress field using SALSA, we create a crystal of $2a = 1.3$ μm diameter silica particles via sedimentation in an index matched water-glycerol mixture. We image the 3D microstructure of isolated vacancies (Fig. 5.2 (a)) and determine their stress fields.

The mean pressure of our crystal sample is $\sim$ 24 mPa (green line in Fig 5.2((b))), which is consistent with previous numerical predictions (orange curve) [13] and our Brownian dynamics simulations (blue dots) for hard-sphere crystals at $\phi \sim 0.59$ (purple line). The top row of images in Fig 5.2(c) show the vacancy 3D stress isosurfaces predicted by linear elasticity. The six independent

stress components determined by SALSA are shown in the next row of Fig 5.2(c). For simplicity we show 2D cuts of each stress component along the (111) or x–y plane (green planes) centered at the vacancy core in the upper images. We also conduct Brownian dynamics simulations (see SI) and directly calculate particle stresses (second to last row of Fig 5.2(c)). The simulation results give quantitatively similar features for all stress components. For example, as shown in the first column of Fig 5.2(c), σ_{xy} exhibits a quadrupole distribution, which arises from the asymmetric collisions due to the absence of a particle at the vacancy core (blue and red arrows in Fig 5.2(a)).

The vacancy stresses also show non-trivial trends in the radial pressure distribution as shown in Fig 5.2(d) that are not captured by isotropic linear elasticity. In particular, while linear elasticity predicts a constant pressure outside the vacancy core, here we observe a pressure bump at $r \sim 3a$ that results from a $\approx 50\%$ reduction in particle surface separation near the core (double arrows in Fig 5.2(e)). In hard sphere systems this reduced separation, hence increased local collision rate, leads to an enhancement of the local modulus.

To account for this changing modulus, we develop an isotropic elastic model including all terms up to third-order with finite strain. Using the volume change ($\Delta V = 8.4\%$) estimated in experiments and literature values of the bulk ($K = 93\,\text{mPa}$) and shear ($\mu = 92\,\text{mPa}$) moduli for our system's volume fraction [201], we fit the pressure distribution by adjusting the three third-order isotropic elastic constants. We find the predicted stress distributions quantitatively reproduce all stress components (last row of Fig 5.2(c)) as well as the radial pressure distribution (red line Fig. 5.2(e)). Furthermore, the local modulus at the pressure ring region can be determined from the fitting. We find that the bulk modulus at that region

149

more than doubles to 213 mPa. This drastically increased modulus is consistent with the value from numerical studies of bulk hard spheres [201, 13] at the local interparticle spacing of the pressure ring region. Overall, the strongly enhanced local modulus indicates a significant hardening near the defect core.

While linear isotropic theory predicts no interaction between vacancies, our findings indicate vacancies attract within the length scale associated with the pressure bump, as was predicted by numerical studies [28, 66, 145]. This attraction can be understood by noting that the volume change ΔV due to one vacancy is negative and so the $P\Delta V$ term in the elastic energy leads to a force that attracts that vacancy to the pressure ring of the other (see SI). Therefore, we estimate the elastic energy of the attraction $\sim 2.6\ k_BT$ at $r \sim 4a$. Since this attraction is several times larger than the thermal energy, it will significantly accelerate the aggregation of vacancies. In an atomic crystal, this large vacancy aggregate will form a void. For hard sphere crystals without attractive interactions, void formation is inhibited by large configurational entropies found at very low equilibrium defect density. At the vacancy densities in many experimental systems, however, voids form in equilibrium [28] and neighboring particles surrounding a void will 'evaporate' into the void, filling it with liquid-state particles in local equilibrium with the surrounding crystal.

5.1.2   Dislocation

Dislocations are one-dimensional topological defects whose collective interactions determine macroscale plasticity including work hardening, yield stress, and fatigue. At the high defect densities involved in such processes however, interactions are significantly altered by nonlinear stress fields surrounding these defects. One
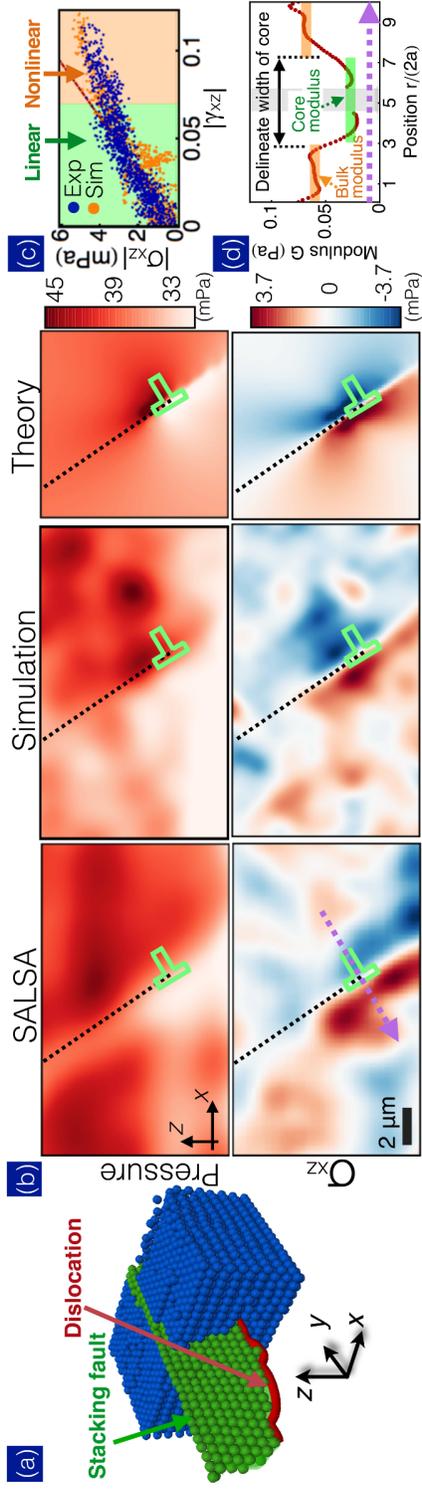
Figure 5.3: **Experimental dislocation stress.** (a) 3D reconstruction of a partial dislocation (red line) and the associated stacking fault plane (blue) analyzed using the Dislocation Extraction Algorithm. (b) Pressure and shear stress, σ_{xz}, around a dislocation determined using SALSA (left column), simulation (middle column), and linear elasticity (right column). The experimental and simulation data are depth-averaged. In simulation, the experimental particle positions are used to determine appropriate initial and boundary conditions. The system is relaxed prior to recording the stresses to avoid particle overlaps due to featuring uncertainties. For the theory calculation, we use the observed Burgers vector and orientation of the partial dislocation to calculate the corresponding stress fields. (c) Both experimental and simulation stress-strain relations show softening behaviors at strains $|\gamma| \geq 0.08$ (orange region). (d) Shear modulus versus position for fields within $2\mu m$ of the purple line in (b). The modulus decreases by $\sim 50\%$ at the defect core, which is approximately four particles wide. Since the modulus value fluctuates in the gray area due to the sign changes in stress and strain, the corresponding points are removed for clarity.

151

critical conjecture that has been widely employed in the dislocation simulation literature is that the modulus softens at the dislocation core [42, 93]. This conjecture however, has never been validated.

To study the dislocation stress field using SALSA, we grow a crystal on a patterned template with a lattice spacing 1.5% larger than the equilibrium crystal lattice. A 3D reconstruction of the particle configuration is shown in Fig. 5.3(a). The dislocation (red) delineates the lower bound of a stacking fault (green) embedded in a crystalline region (blue) which has been clipped for visual clarity. The dislocation is slightly curved (variation$\sim 2a$) and aligned along the y-axis corresponding to the $(1\bar{1}0)$ direction of the fcc lattice. The dislocation core is highlighted with a $(\perp)$ and has a Burgers vector $1/6(\bar{1}\bar{1}2)$, which corresponds to a Shockley partial, the most prominent dislocation in fcc metals.

Using SALSA we measure the stresses near the dislocation and show the pressure (upper row) and shear stress, σ_{xz} (lower row) in Fig. 5.3. The stress field is averaged along the dislocation line to eliminate the effects of polydispersity. To confirm SALSA accurately extracts the stress features in this more complicated defect structure, we compare to stresses calculated by direct Brownian dynamics simulations that are seeded by the experiment data (middle column Fig. 5.3) (SI). Both experimental and simulation results show comparable features. Overall, we observe a pressure gradient across the stacking fault, and a shear stress dipole centered at the defect core. These general trends are consistent with predictions of linear isotropic elastic theory (right column Fig. 5.3) indicating that dislocation curvature does not qualitatively alter the stress distribution. However, both SALSA (blue) and the simulation (orange) results show a non-linear strain softening in highly strained regions near the defect core (Fig. 5.3 (c)). This local modulus

drop allows us to visualize the precise location and size of the dislocation core. To do so, we focus on the cross-section region denoted by the dashed line in Fig. 5.3(b), and plot nonlinear shear modulus $(d\sigma_{xz}/d\gamma_{xz})$ versus position $(r/2a)$ in Fig. 5.3(d) (SI). The modulus decreases by $\sim 50\%$ on both sides of the dislocation core, which is about four particles in width. Overall, our measured modulus profile clearly shows the softening and provides the first experimental evidence supporting the non-singular continuum assumption widely employed in dislocation theories and simulations [42, 93], in which the divergence in the stress at the dislocation core is cut off. Moreover, this modulus softening regularizes the interactions between dislocations and dramatically influences the dislocation creep behavior in crystals.

5.1.3 Grain boundaries

Grain boundaries are 2D structures important for crystal growth [82], melting kinetics [16, 195], transport properties [91], and can substantially harden materials through internal stress variation [209, 131, 150]. While X-ray microbeam experiments have been used to reveal *strain* fluctuations at the scale of 100 nm [150], measuring *stress* remains challenging at these scales, especially at the grain boundaries where particles are highly disordered.

To visualize such stresses using SALSA, we grow polycrystals using the same method described in the vacancy section (see Fig. 5.4 (a) for a confocal image). We plot the measured pressure and shear stress σ_{xy} in the left column of Fig. 5.4(b). Just as for the dislocation simulation, we employ the featured particle positions as initial configurations, and simulate stresses in the polycrystal. The simulation results (right column in Fig. 5.4(b)) show similar features to the SALSA stress distributions in both pressure and shear components.
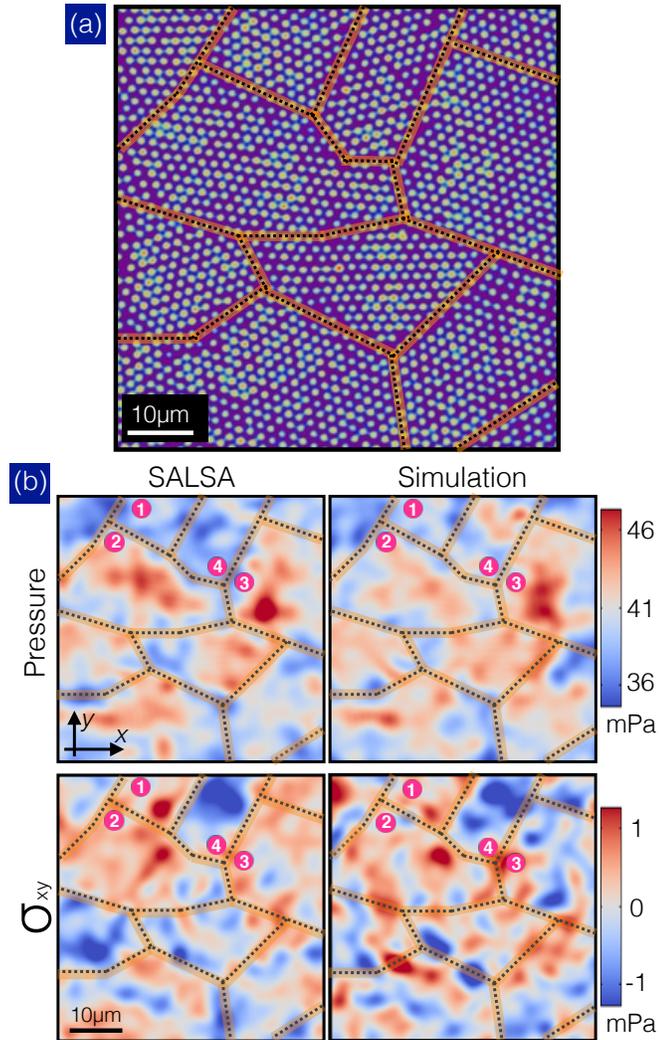
Figure 5.4: **Stress near grain boundaries**. (a) One confocal image slice of a time series consisting of fifty 3D stacks. (b) Pressure (upper) and σ_{xy} (lower) fields of the polycrystal. Both experimental (left) and simulation (right) results show qualitatively similar stress distribution features. The grain boundaries are indicated with dashed lines.

The spatial fluctuations in both pressure and shear stress seen in Fig. 5.4 (b) are significant compared to relevant stress scales. The standard deviation in pressure (≈ 6 mPa) is about 15% of the mean pressure whereas the shear stress fluctuation (≈ 0.7 mPa). To provide intuition, this stress level is about 30% of the stress magnitude one lattice constant from a dislocation core, the principle component of a tilt grain boundary. Moreover, we find that both pressure and shear stresses fluctuate between and within grains. For example, the mean pressure difference between grain 1 and 2 is 5 mPa (25%) whereas grain 3 shows an *intragrain* fluctuation of $\approx 10\%$ the mean pressure. Similar trends can be seen in the shear stress difference between grains 1 and 4, and the fluctuations within grain 3.

Overall, our observation of the stress fluctuations in the polycrystal is consistent with previous simulations [222], and X-ray microbeam measurements [131, 150], where neighboring grains consisting of millions of atoms were found to have substantially different strains. The SALSA measurements indicate such stress fluctuations also arise *within* grains consisting of only hundreds of particles. These small crystallites are reminiscent of the nano-scale grains in atomic crystals. Previous atomistic simulations have predicted the stress fluctuations in a strained nanocrystal are predominately localized to the grain boundaries [222]. In our colloidal crystal grains however, the stress fluctuations are spread roughly evenly throughout the grains (See SI for direct comparison). Our sample however has not been subject to shear. We conjecture that condensation of stress under plastic strain arises from trapping of dislocations at grain boundaries, grain boundary slip [222], or an as of yet unidentified mechanism.

In conclusion, we measure, for the first time, the microscale stress fields of crystalline defect cores that determine fundamental mechanisms governing pro-

cesses ranging from local defect interactions to macroscale yielding. We illustrate the specific significance of this microscale measurement in three canonical defects. The measured pressure enhancement around the vacancy core settles the controversy between theory and simulations, and provides critical insights into the origin of attraction between vacancies. The observed softening at the dislocation core validates the decades-long conjecture of the non-singular stresses in numerous dislocation simulations. Finally, the evenly-spread stress fluctuation in the polycrystal predicts hardening of grains when the crystal undergoes plastic deformation. Such stress measurements will be even more valuable when applied to systems driven further out of equilibrium by applied strains since it will directly measure the stress precursors that generate material failure.

5.2 Supplemental information

5.2.1 Derivation of the SALSA method

Derivation

For a hard-sphere colloidal system, its bulk pairwise Brownian stress can be calculated by an integral over the three-dimensional pair correlation function $g(\mathbf{r})$ [37, 100, 142, 78].

$$\sigma_{ij}^{B} = nk_BTa \int_{S_2} \hat{\mathbf{r}}_i\hat{\mathbf{r}}_j ng(\mathbf{r})dS_2 + nk_BT\delta_{ij} \tag{5.2}$$

Here, n is the number density of particles, k_BT is the thermal energy, a is the particle radius, and i, j are the indices of the stress tensor. The second term is the ideal gas stress contribution arising from the kinetic energy. The off-diagonal

156

components of the first term, which usually dominate at high volume fractions, simply capture the anisotropy of $g(\mathbf{r})$ at contact surface S_2. To determine the stresses at the particle-level, we write the pair correlation function, $g(\mathbf{r})$, as an ensemble average of delta functions.

$$
\begin{aligned}
\sigma_{ij}^B &= \left\langle \frac{k_B T a}{\Omega} \int_{S_2} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_j \frac{N}{V} \frac{V}{N^2} \sum_\alpha \sum_{\beta \neq \alpha} \delta(\mathbf{r} - \mathbf{r}^{\alpha\beta}) dS_2 \right\rangle + n k_B T \delta_{ij} \\
&= \left\langle \frac{1}{N} \sum_\alpha \frac{k_B T a}{\Omega} \int_{S_2} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_j \sum_{\beta \neq \alpha} \delta(\mathbf{r} - \mathbf{r}^{\alpha\beta}) dS_2 \right\rangle + n k_B T \delta_{ij} \quad (5.3)
\end{aligned}
$$

where V, N, Ω are system volume, particle number, and particle volume, respectively. The bracket $\langle ... \rangle$ denotes an average over configurations and can be replaced with a time average in our system. In Eq. 5.3 we identify the outer sum as being an average over the particles in the sample, so we obtain an individual stress tensor for each,

$$
\begin{aligned}
\sigma_{ij}^\alpha &= \frac{k_B T a}{\Omega^\alpha} \left\langle \int_{S_2} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_j \sum_{\beta \neq \alpha} \delta(\mathbf{r} - \mathbf{r}^{\alpha\beta}) dS_2 \right\rangle + n k_B T \delta_{ij} \\
&= \frac{k_B T a}{\Omega^\alpha} \bar{\Psi}_{ij}^\alpha + n k_B T \delta_{ij} \quad (5.4)
\end{aligned}
$$

The elements of the sum $\bar{\Psi}_{ij}^\alpha$ can be identified as the fabric tensor linear density of particle α as the units work out to be $[1/L]$. This fabric tensor density directly reports the angular distribution of neighbors in contact with a particle while the magnitude of its trace is related to the total number of neighbors. To calculate $\bar{\Psi}_{ij}^\alpha$ in simulation and experiment, it is necessary to perform an average over a narrow interval $\Delta \ll 1$

$$
\begin{aligned}
\bar{\Psi}_{ij}^\alpha &= \left\langle \int_{S_2} \sum_{\beta \neq \alpha} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_j \delta(\mathbf{r} - \mathbf{r}^{\alpha\beta}) dS_2 \right\rangle \\
&\approx \left\langle \frac{1}{\Delta} \int_{S_2} \int_a^{a+\Delta} \sum_{\beta \neq \alpha} \hat{\mathbf{r}}_i \hat{\mathbf{r}}_j \delta(\mathbf{r} - \mathbf{r}^{\alpha\beta}) dS_2 dr \right\rangle \\
&= \frac{1}{\Delta} \left\langle \sum_{\beta \in \Delta} \hat{\mathbf{r}}_i^{\alpha\beta} \hat{\mathbf{r}}_j^{\alpha\beta} \right\rangle \quad (5.5)
\end{aligned}
$$

where Δ is the thickness of the measurement shell and $\hat{\mathbf{r}}^{\alpha\beta}$ is the unit vector connecting the centers of particles α and β. Using this particular form, the SALSA formula reads

$$
\begin{aligned}
\sigma_{ij}^{\alpha} &= \frac{k_B T}{\Omega^{\alpha}} \left(\frac{a}{\Delta}\right) \left\langle \sum_{\beta \in \Delta} \hat{r}_i^{\alpha\beta} \hat{r}_j^{\alpha\beta} \right\rangle + n k_B T \delta_{ij} \\
&= \frac{k_B T}{\Omega^{\alpha}} \left(\frac{a}{\Delta}\right) \Psi_{ij}^{\alpha}(\Delta) + n k_B T \delta_{ij}
\end{aligned}
\tag{5.6}
$$

where Ψ_{ij}^{α} can now be identified as the fabric tensor of particle α. In general, the shell thickness $\Delta \ll 1$ should be small enough that the measurement result σ_{ij}^{α} becomes independent of the particular choice of Δ. In the calibration section, we perform experiments and simulations to confirm this independence. The details concerning the particle volume Ω^{α} and converting the pointwise stresses to continuum fields are discussed in later sections.

Finally, it is possible to extend our particle-level stress calculations to other finite potentials including the depletion force [198, 210, 249], paramagnetic [269] and electrostatic [265] interactions, and soft particles [184, 50]. While similar calculations have been performed at the bulk scale [269], extending our method to these systems would allow us to further investigate their heterogeneous elastic properties.

Local structural anisotropy tensor

The local structural anisotropy tensor reports the instantaneous arrangement of colliding neighbors. The trace of this tensor, $\hat{r}_i \hat{r}_i$, is the total number of neighbors while the remaining terms capture the anisotropy of the collisions. As an illustration, we consider a two-dimensional case. Assuming one particle is at the origin and in contact with one other particle at an angle θ with respect to the x-axis then

the fabric tensor can be written as

$$
r_i r_j = \begin{bmatrix} \hat{\mathbf{r}}_x \hat{\mathbf{r}}_x & \hat{\mathbf{r}}_x \hat{\mathbf{r}}_y \\ \hat{\mathbf{r}}_y \hat{\mathbf{r}}_x & \hat{\mathbf{r}}_y \hat{\mathbf{r}}_y \end{bmatrix} = \begin{bmatrix} \cos^2 \theta & \sin \theta \cos \theta \\ \sin \theta \cos \theta & \sin^2 \theta \end{bmatrix}.
$$

Here, the trace $\sin^2 \theta + \cos^2 \theta$ is always 1 and a maximum shear is achieved at four locations, $\theta = \pm\, 45°$ and $\pm\, 135°$. These directions correspond to the maximal compression and extension axes. For instance, if θ is 30°, then the trace is $(\sin^2 30°) + (\cos^2 30°) = 1$, and the shear component is $\hat{\mathbf{x}}\hat{\mathbf{y}} = \hat{\mathbf{y}}\hat{\mathbf{x}} = (\sin 30°)(\cos 30°) = \sqrt{3}/4$. Similarly, if θ is 90°, while the trace remains unity, the shear component becomes $\hat{\mathbf{x}}\hat{\mathbf{y}} = \hat{\mathbf{y}}\hat{\mathbf{x}} = (\sin 90°)(\cos 90°) = 0$.

From pointwise functions to continuum fields

In the literature of molecular dynamics simulation, the virial

$$
s_{ij}^\alpha = \frac{1}{2} \sum_{\beta \neq \alpha} \frac{\partial V(r)}{\partial r} \frac{r_i r_j}{r} + m v_i v_j \tag{5.7}
$$

has been widely used to report stresses at the atomic level [70, 148, 222]. In particular, this quantity has been used to measure many interesting phenomena including the stress correlation length in liquid metals [148] and stress fluctuations near grain boundaries [222]. While the sum of the virials s_{ij} divided by system volume V yields the bulk stress of the system, there are multiple choices that can be made as to how to incorporate local volume variation and how to smooth the pointwise stress into a continuum field. Similar to the Irving-Kirkwood-Noll procedure, which constructs continuum fields from the underlying discrete distribution with phase averaging, we perform a spatial average to obtain a macroscopic measurement at the particle-level [243]. The final continuum stress field $\sigma_{ij}^{cont}(\mathbf{x}; t)$ is

159
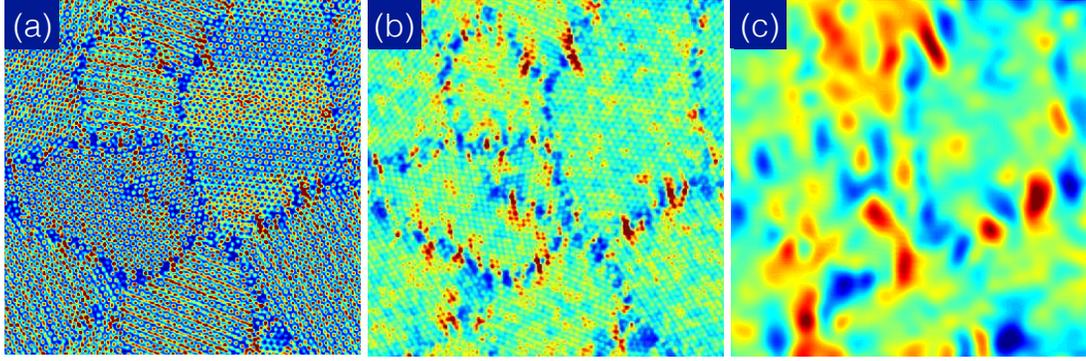
Figure 5.5: **Spatially smoothed stress fields of a simulated polycrystal**. We plot the smoothed stressed field of a polycrystal as a function of filter size $r_w/2a$ at (a) 0.2 (b) 0.3 and (c) 1. For small $r_w/2a$, only the points containing particle centers with assigned stresses have nonzero values. As the value of $r_w/2a$ increases, the stress field becomes smoother and more continuous. Note that since the weighting Gaussian function is normalized, the overall stress of the system is the same for all different $r_w/2a$. The color scale is adjusted for each sub figure to emphasize the stress variation features. In all calculations, we use the value $r_w/2a = 1$, where the features of individual particles are no longer distinguishable.

given by

$$\sigma_{ij}^{cont}(\mathbf{x};t) = \int w(\mathbf{y} - \mathbf{x})\sigma_{ij}^{pt}(\mathbf{x};t)d\mathbf{y} \tag{5.8}$$

where $w(\mathbf{r})$ is a weighting kernel function, and $\sigma_{ij}^{pt}(\mathbf{x};t)$ is the pointwise function. Here, the weighting function w must be normalized so that the total energy is conserved during spatial averaging. We use a normalized Gaussian function that weights values closer to $\mathbf{x}$ more strongly than other points that are further away:

$$w(r) = \pi^{-3/2}r_w^{-3}e^{-\frac{r^2}{2r_w^2}} \tag{5.9}$$

where $r = \|\mathbf{r}\|$, and the filter size r_w is chosen to emphasize the length scale of the continuum fields of interest [243]. Similar smoothing algorithms have been implemented in previous literature [179, 178, 102, 27].

In our experiments and simulations, we set $r_w/2a = 1$ so that we remove stress features on length scales smaller than a particle. The pointwise stress is constructed by first assigning calculated particle virial, $\sigma_{ij}^\alpha \Omega^\alpha$ to their corresponding grid boxes, then dividing their values by the box volume. This pointwise function shows singularities at particle centers and zero elsewhere. Finally, we smooth the field using the Gaussian kernel, Eq. 5.9. We show the spatial averaged continuum fields of a simulated polycrystal for three different filter sizes r_w in Fig. 5.5.

In this spatial averaging procedure, the influence of local volume variation on the stress distribution is already incorporated. In particular, as the occupied volume of a particle increases, the stress is proportionally reduced due to the increased effective averaging volume. Furthermore, the final fields are nearly independent of the grid size and the mean stresses are constant at all values of r_w.

To explore how the filter size r_w affects our final SALSA stress fields, we show the experimental vacancy stress field, σ_{xy} for five different values of r_w in Fig. 5.6(a). We find that the feature of the stress quadrupole[2] remains discernible up to $r_w = 2.7(2a)$. Note in Fig. 5.6(a), while the quadrupole distribution theoretically diverges as $1/r^3$, it is cut off by the lattice and smoothed over the distance r_w, squelching the stress features near the vacancy core. Similarly in Fig. 5.6(b),

---

[2]There is a nomenclature problem when describing the stress and strain around a point defect. Some describe them as quadrupolar fields, reflecting the similar $\cos 2\theta$ angular dependence with electrical quadrupoles. This angular dependence is special, however, to isotropic elastic theory; also, the displacement field has a 'monopolar' isotropic component $\Delta V/(4\pi r^2)$ with no angular dependence. (However, as noted in Fig. 5.7, the off-diagonal stresses for an isotropic vacancy still look like quadrupoles.) Some describe them as elastic dipoles, reflecting both the common $1/r^3$ force law with electric dipoles and their origins as 'force dipoles' (two opposite forces acting on nearby points in the elastic medium.) Finally, others (particularly studying elasticity in amorphous systems) name the field after Eshelby – thus bypassing the choice by using an uninformative label. Here, we use the term quadrupole, mostly because the 'strength' of a point defect is quantified by a 3x3 symmetric matrix (either the strain quadrupole Q_{ij} described here, or the corresponding stress quadrupole / force dipole.) We must note, however, that in the present case of vacancies in a cubic environment, $Q_{ij} = \Delta V \delta_{ij}$ has no correspondence to electric quadrupoles, which are traceless.
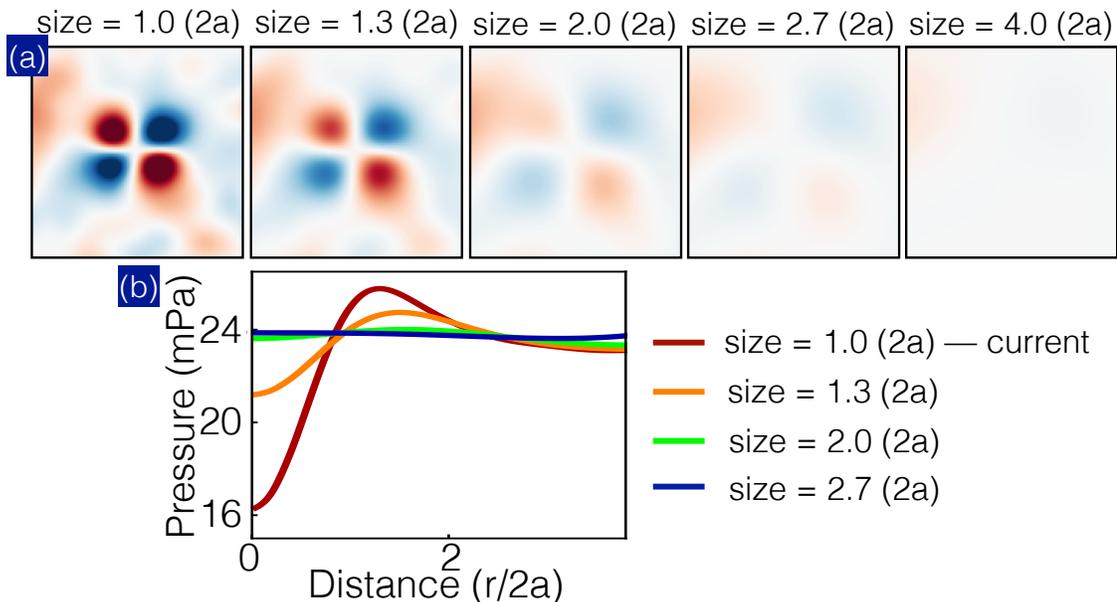
Figure 5.6: **Spatially smoothed stress fields of a vacancy**. (a)The values of smoothing length scale r_w from left to right are 1.0×2a, 1.3×2a, 2.0×2a, 2.7×2a and 4.0×2a, respectively. (b) The radial distribution of pressure for different r_w.

we find that the near field of the pressure is strongly affected by the filter, but the long-range nonlinear pressure ring at $r_w = 1(2a)$ and $1.3(2a)$ match each other. Indeed, the long-range stresses from all defects have correspondingly slow variations, and hence will be invariant to the choice of r_w. The filter size choice is a balance – hiding noise at lengths $r < r_w$ to enhance features of size $r > r_w$.

In terms of elastic theory, the filter size can be thought of a regularization for the theory, renormalizing higher order terms in the elastic free energy. To see how smoothing affects these terms, consider how smoothing changes a linear elastic free energy

$$\mathcal{F}_0 = \int d^3r \, S_{ijkl} \, \sigma_{ij} \, \sigma_{kl} \tag{5.10}$$

where S_{ijkl} is the elastic compliance tensor and σ_{ij} is the stress tensor. If we smooth the stress tensor using a Gaussian kernel of size r_w this linear elastic theory changes

to

$$\mathcal{F}' = \int d^3r\, d^3k\, d^3k'\ S_{ijkl}[\sigma_{ij}(\mathbf{k})e^{-k^2r_w^2/2}e^{i\mathbf{k}\cdot\mathbf{r}}][\sigma_{kl}(\mathbf{k}')e^{-k'^2r_w^2/2}e^{i\mathbf{k}'\cdot\mathbf{r}}]$$

$$= \int d^3k\ S_{ijkl}\, \sigma_{ij}(\mathbf{k})\, \sigma_{kl}(-\mathbf{k})\, e^{-k^2r_w^2}$$

$$\approx S_{ijkl} \int d^3k\ (1 - k^2r_w^2)\, \sigma_{ij}(\mathbf{k})\, \sigma_{kl}(-\mathbf{k})$$

$$= \int d^3k\ S_{ijkl}\, \sigma_{ij}(\mathbf{k})\, \sigma_{kl}(-\mathbf{k}) - \int d^3k\ S_{ijkl}\, k^2r_w^2\sigma_{ij}(\mathbf{k})\sigma_{kl}(-\mathbf{k})$$

$$= \mathcal{F}_0 - r_w^2 \int d^3k\ S_{ijkl}(k_m\sigma_{ij}(\mathbf{k}))(k_m\sigma_{kl}(-\mathbf{k}))$$

What new terms must we add to $\mathcal{F}'$ to cancel the second term? Since each factor of k introduces a gradient, we can check that a stress gradient term $S_{ijkl}\partial_m\partial_m\sigma_{ij}\sigma_{kl}$ has the correct form.

$$S_{ijkl} \int d^3r\ \partial_m\partial_m(\sigma_{ij}\, \sigma_{kl}) = S_{ijkl}\, \partial_m\partial_m \int d^3r\, d^3k\, d^3k'\ e^{-i\mathbf{k}\cdot\mathbf{r}}\sigma_{ij}(\mathbf{k})\, e^{-i\mathbf{k}'\cdot\mathbf{r}}\sigma_{kl}(\mathbf{k}')$$

$$= -2S_{ijkl} \int d^3r\, d^3k\, d^3k'\ k_m k'_m e^{-i\mathbf{k}\cdot\mathbf{r}}\sigma_{ij}(\mathbf{k})\, e^{-i\mathbf{k}'\cdot\mathbf{r}}\sigma_{kl}(\mathbf{k}')$$

$$= -2S_{ijkl} \int d^3k\ k_m k_m \sigma_{ij}(\mathbf{k})\sigma_{kl}(-\mathbf{k})$$

Indeed, the energy regularized by smoothing by r_w is the original free energy plus a filter size-dependent term times a stress gradient energy:

$$\mathcal{F}' \approx \mathcal{F}_0 + \frac{r_w^2}{2} S_{ijkl} \int d^3r\ \partial_m\partial_m(\sigma_{ij}\, \sigma_{kl}) \tag{5.11}$$

Similarly, nonlinear and other terms in the energy will produce regularization-dependent counter terms. We will now demonstrate that these gradient terms contribute less to the free energy than other nonlinear terms in the case of hard sphere crystals.

Let us consider the long and short wavelength behavior of the first nonlinear and gradient terms of the elastic free energy to see which dominates the behavior

163

both close and far from a defect. For isotropic elasticity, the allowed terms that arise in the free energy density are

$$\mathcal{F} = C_{ijkl}\epsilon_{ij}\epsilon_{kl} + D_{ijklmn}\epsilon_{ij}\epsilon_{kl}\epsilon_{mn} + E_{ijklmn}\partial_i\partial_j\epsilon_{kl}\epsilon_{mn} \tag{5.12}$$

where the allowed elements of the elastic constant tensors C, D, and E are determined by the symmetries of the system being studied. In the case of isotropic materials, these tensors must be built using terms that are formed by δ_{ij} and Δ_{ijkl} and $\mathcal{D}_{ijklmn}$, the Kronecker delta and the four and six index equivalents of the Kronecker delta. That is, the parts of the free energy may be written

$$\mathcal{F}_C = c_0\,\epsilon_{ii}\epsilon_{jj} + c_1\,\epsilon_{ij}\epsilon_{ij}$$

$$\mathcal{F}_D = d_0\,\epsilon_{ii}\epsilon_{jj}\epsilon_{kk} + d_1\,\epsilon_{ij}\epsilon_{ij}\epsilon_{kk} + d_2\,\epsilon_{ij}\epsilon_{jk}\epsilon_{ki}$$

$$\mathcal{F}_E = e_0\,\partial_i\partial_i\epsilon_{jj}\epsilon_{kk} + e_1\,\partial_i\partial_j\epsilon_{ij}\epsilon_{kk} + \cdots$$

Which terms should we keep to describe the elastic fields? We consider their contributions in the case of the elastic field of a vacancy defect – since the displacement field for linear theory goes as $u(r) \sim \Delta V/r^2$, the strain field goes as $\epsilon \sim \Delta V/r^3$. The quadratic term then has energy density which scales as r^{-6}, while the nonlinear cubic term (ϵ^3) scales as r^{-9} and the gradient term as $r^{-1}r^{-1}r^{-3}r^{-3} = r^{-8}$. For short range behavior, the nonlinear contributions are nearly equal, with the cubic term contributing more for $r \ll 1$. However, we also need to consider the magnitude and scaling of the coefficients for each term. In the case of the cubic elastic constants, we know from the equation of state that the pressure diverges at maximal packing as $P \sim (\phi_c - \phi)^{-1}$, implying that e.g. the bulk modulus diverges as $K = \phi\frac{\partial P}{\partial \phi} \sim (\phi_c - \phi)^{-2}$ [196]. If we write the scaling in terms of the two length scales in the problem, the lattice constant a and the surface to surface distance Δ, we find that the cubic term in the free energy goes

as $(a/\Delta)^{-2}r^{-9}$ and the first gradient term as a^2r^{-8}. In the case $(a/\Delta) \ll 1$, we find that the cubic terms dominates the gradient contributions. It is for this reason as well as the fact that gradient terms are not unique in our smoothing scheme that we only consider nonlinear elasticity without gradient terms.

5.2.2   Calibration of the SALSA method

To validate our SALSA measurements, it is crucial to calibrate the method and evaluate its performance and dependence on input parameters. We divide the calibration section into three parts: A) comparison with simulation stresses, B) contact criteria dependence, and C) force balance in a smoothed field. We show that the SALSA method accurately captures the stress fields as calculated by Brownian dynamics very well. The stress field determined by using SALSA is not significantly sensitive to the particle contact criteria e.g. the shell thickness Δ. Finally, we discuss a systematic residual force within the vacancy core in the continuum stress field due to the smoothing process and compare the elastic fields to similarly smoothed continuum calculations.

Calibration: SALSA versus actual stress

To evaluate how well SALSA is able to report the correct stress field of a complex system, we use Brownian dynamics (BD) to generate a vacancy in a simulated crystal, whose orientation is matched to that found in our experiments. We use SALSA to determine all six independent stress components and compare their values with the BD stress calculation using the same set of position data (see Fig. 5.7). The continuum stress field from BD is constructed by spatially averaging

pointwise virials $F_i x_j$. We see that the BD and SALSA stresses exhibit a good match. We also find that SALSA stresses become more quantitatively similar to the actual stress fields as a longer time-average is performed (not shown).
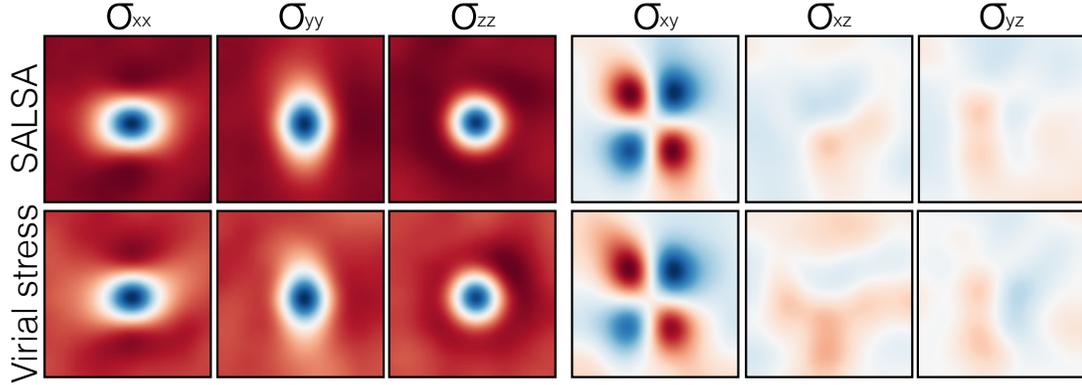
Figure 5.7: **SALSA and actual stresses comparison in simulation**. Stresses determined through the SALSA method (top row) are compared with the stresses directly calculated in simulation (bottom row) using the same dataset.

Calibration: Contact criteria dependence

There is another parameter in the SALSA method, which is the shell thickness Δ used to identify particles in contact. This shell thickness directly determines the number of particles that are included in the stress calculation, larger Δ allowing for shorter time averages. However, at large Δ, the radial distribution of particles $g(r)$ will vary through out the thickness of the shell, leading to systematic errors in stress. We test for the optimum by calculating the pressure of a system versus $\Delta/2a$. Fig. 5.8 shows the SALSA pressure $P(\Delta)/P(\Delta_g)$ versus $\Delta/2a$ for both experimental and simulation data. Here, $\Delta_g/2a$ (gray line in Fig. 5.8) is the cutoff thickness used throughout our analysis of the experiments, where $2a + \Delta_g$ roughly coincides with the first peak of $g(r)$. We find that the measured pressure has negligible dependence on $\Delta/2a$ for $\Delta/2a \leq 0.12$. In the experiment the

Figure 5.8: **SALSA pressure versus Δ in experiment and simulation**. Pressure measured using different shell thickness Δ is plotted as a function of Δ. The pressures values P(Δ) are normalized by P(Δ_g) where $\Delta_g + 2a$ is the p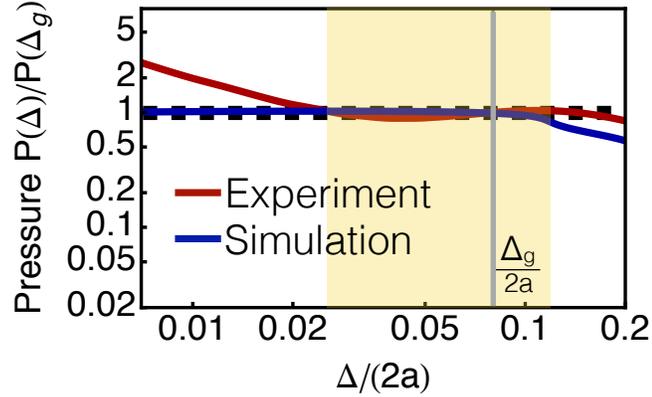osition of the $g(r)$ first peak, which is denoted by the gray line. The vacancy experimental data (red curve) shows that SALSA method generates quantitatively consistent results in a wide range of shell thickness $0.03(2a) \leq \Delta \leq 0.15(2a)$ (shaded area). The corresponding length scale of this thickness range is 40 nm $\leq \Delta \leq 230$ nm. The simulation data (blue curve) demonstrate an even wider pressure plateau that extends beyond $\Delta \sim 10^{-4}(2a)$. The diverging trend of the experimental pressure at small Δ occurs due to slight particle overlaps arising from featuring uncertainties.

normalized pressure deviates from 1 when Δ is smaller than $\sim 3\%$ of the particle size, corresponding to $\approx 1/4$ of a pixel (~ 35 nm). This trend arises due to particle overlaps from featuring uncertainties. Overall, as shown in Fig. 5.8, both the experimental and simulation results indicate an insignificant correlation between the SALSA pressure and shell thickness Δ in the range of 35 nm $\leq \Delta \leq 230$ nm (yellow shaded area).

Finally, we also investigate how the shell thickness Δ affects the spatial distribution of stresses. We show the experimental pressure and σ_{xy} fields near a vacancy for four different Δ in Fig. 5.9. Again, we find that the stress fields for all Δ demonstrate qualitatively similar trends, where the pressure exhibits an enhancement surrounding the defect core and σ_{xy} shows a quadrupole distribution. This
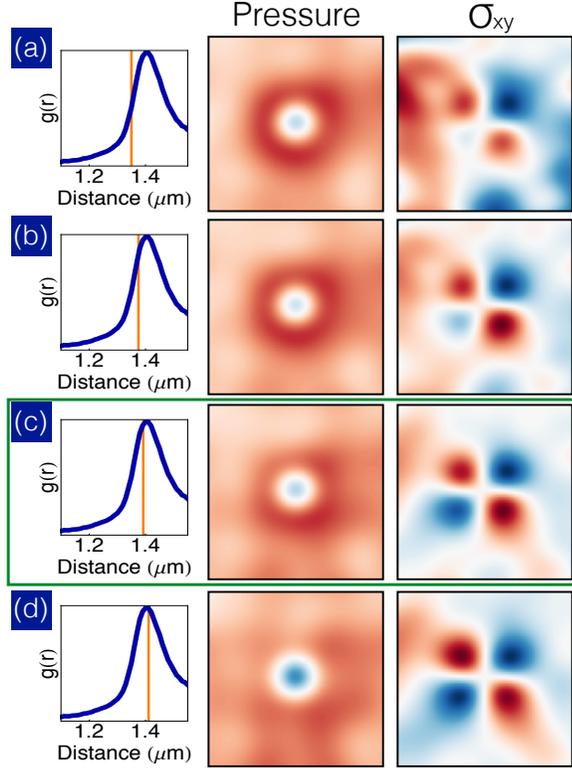
Figure 5.9: **Pressure and σ_{xy} fields for different Δ.** Pressure (middle row) and shear stress σ_{xy} (right row) distributions for four different values of shell thicknesses. (c) The contact criteria used throughout our experiments. The pair correlation functions $g(r)$ (left row) are plotted to illustrate the differences between the shell thickness Δ choices (orange lines).

weak Δ dependence is consistent with the previous experimental studies [53, 158] where the authors have found that different contact criterion consistently generate similar bulk Brownian stresses. Here, we provide a similar calibration but at the particle level. As shown in Fig. 5.9, it is remarkable that SALSA is able to produce consistent results with a wide range of shell thickness 35 nm $\leq \Delta \leq$ 150 nm. This wide window of Δ choice promises robust stress measurements in the typical colloidal experiments with 3D imaging, where the particle positions can be precisely determined with a sub-pixel resolution ($\leq$ 50 nm) using standard featuring algorithms.
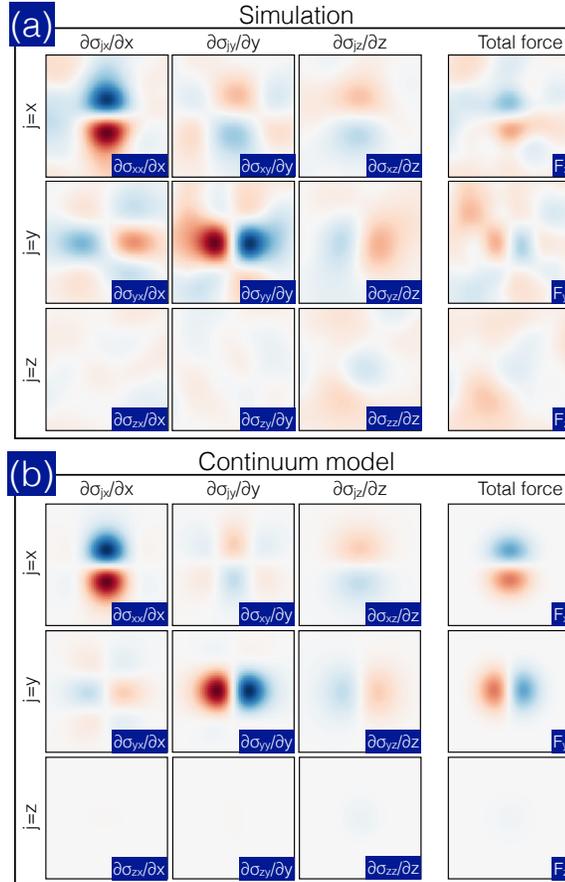
Figure 5.10: **Force balance of vacancy**. The right-most frames show the divergence of the stress fields from simulation (a) and continuum theory (b). In the group of nine panels in each of these subfigures, we show the components of the divergence of the stress. Summing along each row, we find the total force as the sum of the gradients. Both the BD simulation and continuum elastic fields show systematic force dipoles in their center indicating that the small residual forces are produced not from the SALSA method but from the choices made in creating a continuum field.

Calibration: Mechanical equilibrium of a smoothed field

In principal, it is possible to further determine the continuous force field by calculating the divergence of the smoothed stress distribution. In our experiments, where all the studied regions are stationary, the calculated force field should be zero implying a mechanical equilibrium. It has recently been shown that this me-

chanical equilibrium of the microscopic stress closely depends on the details of the stress definition at particle-level [247]; it is important to check it for SALSA. To investigate this issue, we calculate the divergences of the vacancy SALSA stresses in simulation (Fig. 5.10 (a)). We find that the force fields are consistent with zero everywhere except the region of the defect core, where the force shows a dipole distribution in all components. Since there is no particle in the region that violates mechanical equilibrium, it is unclear whether this force imbalance leads to a particle drift or not. Nevertheless, to confirm that this is only a result of the smoothing procedure, we perform a similar analysis on calculated continuum elastic fields. In doing so, we mimic the SALSA measurements by introducing a pressure hole in the center, and smooth the stress fields with the same kernel used in the simulation. As shown in Fig. 5.10 (b), the force fields from continuum theory also display force dipoles consistent with SALSA and BD stress fields. This consistency clearly indicates that the force imbalance mostly arises from the pressure drop and smoothing algorithm rather than the SALSA calculation.

To further characterize the magnitude of the force imbalance, we perform a similar mechanical equilibrium calibration in a simulated polycrystal. In particular, we construct a three dimensional box enclosing a grain boundary. Then we determine the net force acting on this box by calculating the tractions from stresses. We find that the forces correspond to the shear and normal tractions approximately cancel each other, indicating a good mechanical equilibrium. The residual force can then be related to a drift arising from this force imbalance given that the system is over-damped. Finally, we find that the drift, which is independent of the box size, is less than 5% of the particle diameter over the entire simulation.

5.2.3 Vacancy Stress Fields

Experimental details

We create a colloidal crystal consisting of 1.3 μm diameter silica particles via sedimentation in a sealed sample cell. The particles are suspended in a water-glycerol mixture with a refractive index matching the silica particles. This matched refractive index allows us to acquire confocal images of the sample. Vacancy defects form spontaneously during sedimentation (along with stacking faults and grain boundaries), and are imaged directly using a high-speed confocal microscope. In the measurements, we select isolated vacancies that contain no other defects within five lattice spacings in the plane or in either of the adjacent layers perpendicular to the plane.

Since the system is thermal, it is important to perform a time average to correctly determine the equilibrium stress field. Therefore, we record, analyze, and average the stress fields over 60 snapshots (20 seconds). We further average the stress field over 20 vacancies to reduce the effects of polydispersity and the local vacancy environment. In Fig. 5.11, we show the confocal images of all measured vacancies in the experiment (the horizontal slices of full 3D confocal image stacks) to illustrate the vacancy morphology. As shown in Fig. 5.11, the independent vacancies have random orientations with respect to the microscope and must be aligned before averaging. With all images aligned in the same orientation, we then calculate the stress field of each sample and average over 20 seconds. This time interval is sufficiently long for the colloids to explore the local phase space as the time required to diffuse one particle separation (100 nm) in the absence of obstructing neighbors is about 0.35 s. Finally, we average the per-vacancy stress

field over all 20 samples.
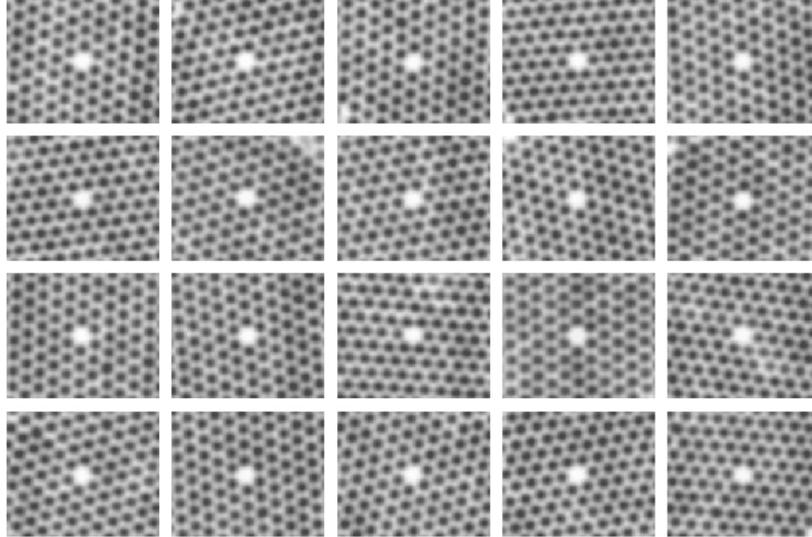
Figure 5.11: **Confocal images of 20 isolated vacancies**. We display the raw confocal data before rotation and alignment.

Since SALSA solely relies on the particle positions to determine the stress field, the correctness of final measurements directly depends on the accuracy of particle featuring and noise in the experiment. We employ both time and sample averages to remove noise and improve such measurement accuracies. Uncorrelated noise such as the current noise in the electronics, can be effectively reduced by using a time average. On the other hand, there is correlated noise that result in a persistent featuring bias in time, including spherical aberration in the optics [232, 231, 246], the microscope point-spread-function [30, 159, 230], and particle polydispersity [199]. In the vacancy experiment, we minimize the effects of spherical aberration and point-spread-function by index-matching the sample within 0.1% and confining the imaging field to a few particles away from the coverslip. Polydispersity also affects our ability to determine which particles are in contact. For instance, when two larger particles are touching, the SALSA method may identify them as not in contact because their center-to-center distance is larger

than $2\langle a \rangle + \Delta$, the mean distance between particles plus the shell thickness. The polydispersity (≈ 35 nm for our silica colloids) is about 30% of the shell thickness $\Delta = 106$ nm used in our experiment. In the vacancy experiment, by averaging the stress field over 20 different samples, the collision uncertainty due to polydispersity is further reduced by a factor of $\sqrt{20} = 4.5$.
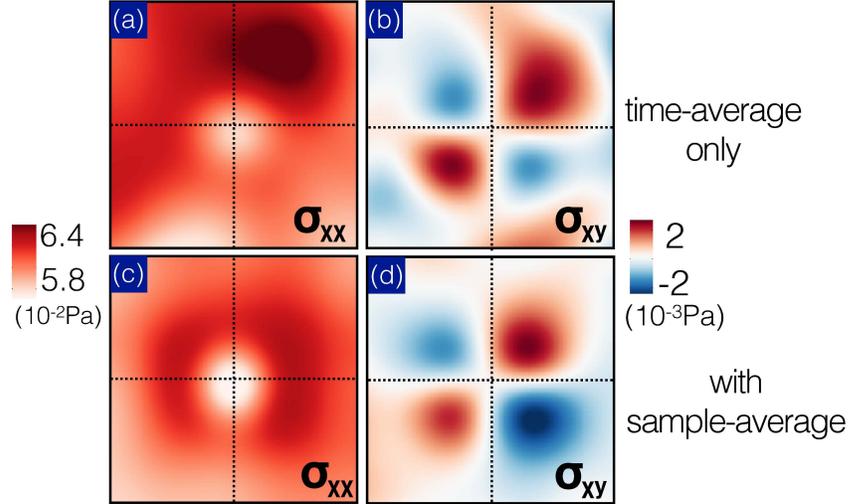
Figure 5.12: **Effects of sample average in addition to time average**. Comparison of time average and sample average on a normal component σ_{xx} (a,c) and shear component σ_{xy} (b,d). We see that sample average improves the measurement of the normal component more than the shear components due to the effects of polydispersity.

Furthermore, we find that the polydispersity has different effects on the pressure and shear stress measurements. We show two representative stress components σ_{xx} and σ_{xy}, first time averaged, Fig. 5.12 (a) and (b), and then with an additional sample average, Fig. 5.12 (c) and (d). We find that the pressure does not fully capture the enhanced stress ring around the vacancy defect if only a time average is applied. However, the shear component of the same time-averaged data already shows a clear quadrupole structure that is very similar to the one with additional sample averaging. This finding implies that while the pressure measurement may

rely on a more precise identification of touching particles, the shear measurement is relatively robust and insensitive to the noise. In contrast to the normal stresses that are strongly associated with the collision (touching) probability of surrounding particles, the shear components are more related to the angular anisotropy of the neighboring particle configurations. Finally, it is also possible to avoid the effects of polydispersity by determining the individual particle size and taking the size variation into account [139].

Simulated vacancy stresses

We confirm the experimental findings by simulating the stress of a vacancy in a colloidal crystal using nearly hard-sphere Brownian dynamics. Here, we numerically simulate the Langevin dynamics of N particles interacting through a very sharp radial potential $V(r)$. We do so using cell neighbor lists calculated on an NVIDIA GPU integrating with the velocity Verlet algorithm. The virials of each particle, calculated through Eq. 5.7, are used to compare directly to the stresses calculated with the SALSA method.

Based on previous literature [158], we have tried several interparticle potentials including the Yukawa potential, pure power law, and smoothed power law. Here we use the smoothed power law to ensure continuity in derivatives:

$$V(r) = \mathcal{E} \left(\frac{r}{r_0} - 1 \right)^2 \left(\frac{r_0}{r} \right)^{24} \tag{5.13}$$

To simulate the vacancy, we begin with a periodic cell of a fcc crystal with the same physical parameters as the experimental setup. We create a periodic cell containing $2^{14} = 16384$ particles at a packing fraction of $\phi = 0.59$, temperature $T = 300 \, k_B T$ and viscosity $\eta = 10^3$. We remove the center particle and simulate for

2500 snapshots where each snapshot is separated by $t = 10\tau$ diffusion times. The stress of these particle configurations is calculated using both the simulation virial and SALSA and averaged over the entire simulation time. A direct comparison of these stress fields can be found in Fig. 5.7. We do find some quantitative differences between the experiment and simulation. For example, the pressure of the simulated vacancy is plotted as a function of distance from the defect core in Fig. 5.13. While the morphology in each individual stress component is very close to that seen in experiments, we do not find as strong of a pressure enhancement around the simulated vacancy. This softening is most likely due to the softened core of the potential we use. Overall, however, all the qualitative features are reproduced in each of experiment, simulation, and theory.

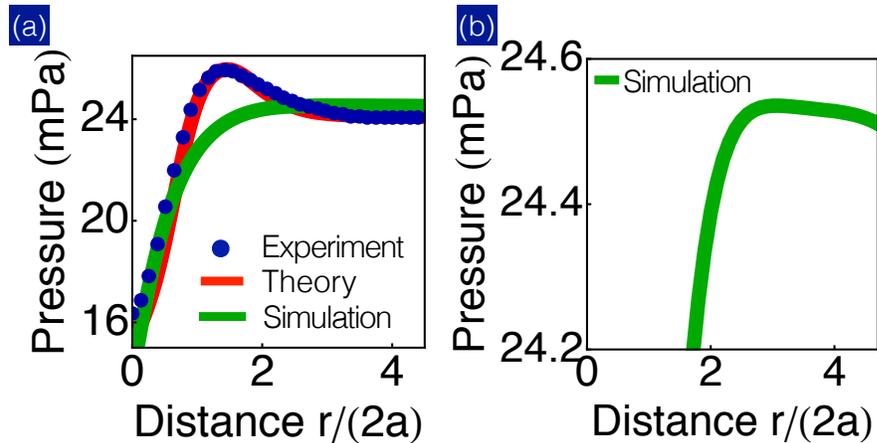Figure 5.13: **Vacancy pressure distribution**. Comparison of experimental (blue dots in (a)), theoretical (red line in (a)), and simulation (green line in (b)) pressure. Despite its relatively insignificant feature, the simulated vacancy pressure also exhibits a pressure enhancement at $r \sim 4a$. All individual stress components in experiment, theory, and simulation show qualitatively similar results.

Continuum elastic theory

Linear elasticity To further measure the performance of SALSA, we also compare the stress fields calculated in simulation and experiment to the corresponding continuum elastic theory. In linear isotropic elastic theory, a vacancy's displacement field can be described by a radial function determined entirely by the local volume change $u_i(r) = \Delta V / r^2 \, \hat{\mathbf{r}}_i$. This can be seen by looking at the elastic free energy which can be written

$$\mathcal{F}_{\text{linear}} = \frac{1}{2} K \epsilon_{ii}^2 + \mu (\epsilon_{ij} - 1/3 \, \delta_{ij} \epsilon_{ll})^2 \tag{5.14}$$

where K is the bulk modulus, μ is the shear modulus and ϵ_{ij} the strain tensor, the symmetrized Jacobian of the displacement field u, $\epsilon_{ij} = 1/2(\partial_i u_j + \partial_j u_i)$. Since the material is isotropic, we make u a radial function such that $u_i(\mathbf{r}) = u(r)\hat{\mathbf{r}}_i$. We then minimize the free energy with respect to this displacement field through the Euler-Langrange equations $\frac{\partial F}{\partial u_i} - \frac{\partial}{\partial x_i} \frac{\partial F}{\partial \partial u_i} = 0$, giving a differential equation for the displacement $r^2 u'' + 2ru' - 2u = 0$ whose solution is given by

$$u(r) = \frac{\Delta V}{4\pi r^2} + \frac{P_\infty}{3K} r \tag{5.15}$$

Here, ΔV is the local volume change, P_∞ is the pressure at long length scales due to boundary conditions and K is again the bulk modulus. This displacement field leads to a strain and stress field in linear elasticity that has the form

$$
\begin{aligned}
\epsilon_{ij} &= \frac{1}{2}(\partial_i u_j + \partial_j u_i) \\
&= \frac{u}{r}\left(\delta_{ij} - \frac{r_i r_j}{r^2}\right) + u'\frac{r_i r_j}{r^2} \\
\sigma_{ij} &= K\epsilon_{ll}\delta_{ij} + 2\mu(\epsilon_{ij} - 1/3\delta_{ij}\epsilon_{kk}) \\
&= K\left(2\frac{u}{r} + u'\right)\delta_{ij} + 2\mu\left(\frac{u}{r} - u'\right)\left(\frac{1}{3}\delta_{ij} - \frac{r_i r_j}{r^2}\right)
\end{aligned}
$$

Using this stress field, we find the pressure field is a constant

$$P = \sigma_{ii}/3 = K\left(2\frac{u}{r} + u'\right) = P_\infty \tag{5.16}$$

Geometric nonlinearity However, both the simulation and experimental data show a pressure ring that suggests we need to move to higher order elasticity to accurately describe the stress field of the colloidal vacancy. The first natural attempt to capture this ring can be done by including the geometric nonlinearity (also known as finite strain), an extra term in the strain field that makes it rotationally invariant but is higher order in displacement. Doing so, we find that

$$\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i + \partial_i u_k \partial_j u_k)$$
$$= \frac{u}{r}\left(\delta_{ij} - \frac{r_i r_j}{r^2}\right) + u'\frac{r_i r_j}{r^2} + \frac{1}{2}\left[\left(\frac{u}{r}\right)^2\left(\delta_{ij} - \frac{r_i r_j}{r^2}\right) + u'^2\frac{r_i r_j}{r^2}\right]$$
$$= \bar{U}\left(\delta_{ij} - \frac{r_i r_j}{r^2}\right) + \bar{U}'\frac{r_i r_j}{r^2}$$

where we can define $\bar{U} = u/r + 1/2(u/r)^2$ and $\bar{U}' = u' + 1/2u'^2$. We then calculate the stress which is linear in the strain, arriving at the same answer as previously, except with these variables substituted. Lastly, we find the differential equation for the radial displacement field via the Euler-Lagrange equations, yielding a long nonlinear ODE, which we omit for brevity. Fitting the experimental data using this form yields a small pressure enhancement which cannot be tuned to quantitatively match the experimental data without resorting to unphysical values for the bulk and shear modulus.

Nonlinear elasticity Motivated by arguments made in Section II, we next calculate the stress field incorporating the leading terms in nonlinear isotropic elastic theory. In particular, we modify the free energy such that

$$\mathcal{F}' = \mathcal{F}_{\text{linear}} + \frac{\alpha}{3}\epsilon_{ii}\epsilon_{jj}\epsilon_{kk} + \frac{\beta}{6}\epsilon_{ij}\epsilon_{jk}\epsilon_{ki} + \frac{\gamma}{4}\epsilon_{ii}\epsilon_{jk}\epsilon_{jk} \tag{5.17}$$

We insert the definition of finite strain into the new free energy to arrive at another differential equation for the displacement field as a function $u(r)$.

$$r^5(24r(\lambda + 2\mu)u''(r) + 6(\alpha + \beta + \gamma)u'(r)^5$$

$$+ 4u'(r)^3(2(8\alpha + 3(4\beta + 2\gamma + \lambda + 2\mu)) + 15r(\alpha + \beta + \gamma)u''(r))$$

$$+ 4u'(r)^2(8\alpha + 6(2\beta + \gamma + 4\lambda + 6\mu) + 9ru''(r)(2\alpha + 2\beta + 2\gamma + \lambda + 2\mu))$$

$$+ 24u'(r)(2(\lambda + 2\mu) + ru''(r)(\alpha + \beta + \gamma + 3\lambda + 6\mu))$$

$$+ 3u'(r)^4(2(6\alpha + 8\beta + 5\gamma) + 5r(\alpha + \beta + \gamma)u''(r))) + 2r^4u(r)(8(r(\alpha + 3(\beta + \lambda))u''(r)$$

$$- 3(\lambda + 2\mu)) + 3(\alpha + 3\beta)u'(r)^4 + 12(\alpha + 3\beta)u'(r)^3 + 4u'(r)^2(5\alpha + 3(6\beta + \lambda) + 3r(\alpha + 3\beta)u''(r))$$

$$+ 8u'(r)(\alpha + 3(\beta + \lambda) + 3r(\alpha + 3\beta)u''(r))) + 4r^3u(r)^2(2(r(2\alpha + 9\beta + 3\lambda)u''(r)$$

$$- 3(2\alpha + 4\beta + \gamma + 6\lambda + 6\mu)) + 3u'(r)^2(\alpha + 6\beta + r(\alpha + 3\beta)u''(r)) + 6r(\alpha + 3\beta)u'(r)u''(r))$$

$$+ 4r^2u(r)^3(2(-13\alpha - 6(5\beta + \gamma + \lambda + \mu) + r(\alpha + 6\beta)u''(r)) + (\alpha + 6\beta)u'(r)^2 - 2(\alpha + 6\beta)u'(r))$$

$$+ 2ru(r)^4(-32\alpha - 72\beta - 15\gamma + r(\alpha + 6\beta)u''(r) - 2(\alpha + 6\beta)u'(r)) - 6u(r)^5(2\alpha + 4\beta + \gamma) = 0$$

We use this nonlinear ODE to fit the pressure profile found in the experimental data using only the purely compressional third-order elastic constant $\alpha = 3.6$ Pa, leaving $\beta = \gamma = 0$. In this same fit, we set the other elastic constants $K = 0.093$ Pa and $\mu = 0.092$ Pa based on studies of hard sphere elastic constants [201]. We also set the volume change $\Delta V = -0.083$ to be the same as the experimental value. In this fit, we are able to reproduce the experimental pressure ring with one third-order elastic constant and one initial condition (u' far from the vacancy).

The value of α we find from our fit of the pressure ring is consistent with the variation of bulk modulus with packing fraction as calculated by hard sphere simulations. We can directly compare these values using the equation of state. Given the pressure of a hard sphere system as a function of packing fraction, $P(\phi)$, we can expand the elastic constants as $K(\phi_0) + K'(\phi_0)(\phi - \phi_0) + \frac{1}{2}K''(\phi_0)(\phi -$

$\phi_0)^2 + \cdots$ giving $\alpha = K'(\phi_0) = \partial_\phi(\phi\partial_\phi P)|_{\phi_0}$. Using the functional form for $P(\phi)$ we get that $\alpha(0.59) \approx 3.0$ Pa in agreement with our fit $\alpha = 3.6$ to experimental stresses using nonlinear elasticity.

Vacancy interaction   The overall sign of the interaction between vacancies must be negative as vacancies are in general attracted to areas of higher pressure just as interstitials are attracted to areas of lower relative pressure. In the case of two vacancies, the local increase in pressure around one vacancy acts as a higher pressure region for the second, causing them to mutually attract. Physically speaking, the collapse of particles towards the core of one vacancy causes particles to collide more frequently which is relieved by the negative volume change given by the second vacancy.

The above argument gives the dominant nonlinear term (the linear field of one vacancy coupling to the nonlinear pressure around another). Higher order effects can be repulsive but are smaller than this leading order term. Specifically, to calculate the attraction or repulsion of vacancies, we look at the elastic free energy, which strictly speaking is entirely entropic. In this treatment, we will be calling all entropic contributions aside from the configurational entropy the elastic energy E_{elastic}, giving us a free energy density $\mathcal{F} = E_{\text{elastic}} - TS_{\text{conf}}$. To first order, the elastic energy density is $E_{\text{elastic}} = \sigma_{ij}\epsilon_{ij}$ where σ is stress and ϵ is strain. We consider the perturbative view of the elastic free energy in the case of the interaction of two vacancies A and B, which can be expanded into three primary terms, $E_{\text{elastic}} = \sigma_{ij}^{A,L}\epsilon_{ij}^{B,L} + 2\sigma_{ij}^{A,N}\epsilon_{ij}^{B,L} + \sigma_{ij}^{A,N}\epsilon_{ij}^{B,N}$ where L indicates a linear contribution and N indications a nonlinear one. In isotropic linear elastic theory, vacancies do not interact making the first term independent of vacancy separation and leaving us with the second and third terms of the expansion. The

second term is the linear part vacancy B's quadrupole strain field sitting in the nonlinear (pressure bump) stress field of vacancy A. In this simple case, we know that the energy can be given by the vacancy quadrupole $E = \sigma_{ij}^{\text{ext}} Q_{ij}^B$, where the strain quadrupole for a vacancy is diagonal $Q_{ij}^B = \Delta V^B \delta_{ij}$ and $\sigma^{\text{ext}} = \sigma^{A,\text{N}}$ is an external stress field given by vacancy A. Therefore, the energy can be written $E = \sigma_{ii}^{\text{ext}} \Delta V^B = \sigma_{ii}^{A,\text{N}} \Delta V^B = P^{A,\text{N}} \Delta V^B$. Since $\Delta V \leq 0$ for vacancy defects, this term is negative, leading to an overall attraction, consistent with previous literature [145, 146, 57, 28]. This calculation will have higher-order corrections due to the nonlinear elastic overlap of the nonlinear pressure rings $\propto P^{A,\text{N}} P^{B,\text{N}}$ as well as nonlinear corrections to the pressure bump itself due to the presence of a second vacancy, but the qualitative behavior remains unchanged.

5.2.4 Dislocation Stress and Strain Fields

Experimental details

The dislocation is produced by templating the [100] axis on a glass coverslip at a registry 1.5% larger than the equilibrium lattice constant [218]. Particles are sedimented onto the substrate forming a single face-centered cubic crystal. As the crystal thickness reaches about 31 μm, a significant number of dislocations spontaneously nucleate and grow. We then image the three dimensional microstructure of an isolated dislocation using a confocal microscope. A schematic of these dislocations is found in the main text.

Simulated dislocation stresses

To closely simulate the particular dislocation studied in the experiment, we import the experimentally measured particle positions into the Brownian dynamics simulation. Prior to recording the stress, we relax the system to remove overlaps using a soft Hertzian potential and a large damping factor to ensure very little rearrangement. We then freeze the border particles as labeled in red in Fig. 5.14 to ensure that the topologically constrained dislocation does not migrate. After performing a time average, we find that the calculated stress field, both through virial calculation and SALSA method, are in excellent agreement with the experiment, providing a confirmation to the SALSA measurements. Importantly, since this simulation procedure only requires a single snapshot of data for the initial condition, this technique can be particularly useful in determining stresses in the experimental cases where time average is challenging. For instance, it is difficult to perform a time average in a system where the fluorescent dye photobleaches significantly or the dynamic time-scale is comparable to the time between acquisition of successive image stacks.

Continuum elastic theory

To compare against isotropic linear elasticity, we again calculate the stress field of the dislocation. The stress field of a dislocation in coordinates where z is along
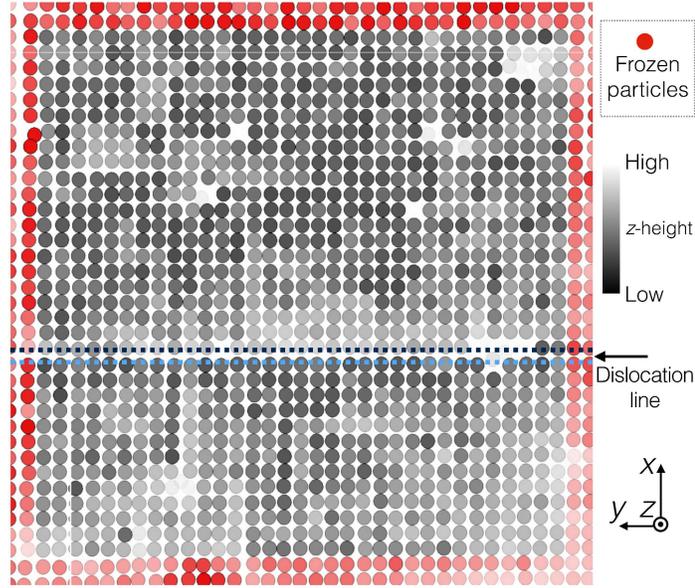
Figure 5.14: **Frozen particle border (dislocation)**. A screen-shot of our Brownian dynamics simulation of the experimental fcc crystal. In shades of gray are active hard sphere particles while in red are the frozen boundary particles. Luminance in this picture roughly indicates the position in the z direction with only a thin slice of the entire simulation shown and darker colors indicating being deeper into the page. The dislocation line is visible 1/3 from the bottom of the image running left to right as indicated by the discontinuity in particle shade between adjacent rows.

the dislocation line, is given by [124]

$$\sigma_{xx} = -y\frac{3x^2 + y^2}{(x^2 + y^2)^2}$$

$$\sigma_{yy} = y\frac{x^2 - y^2}{(x^2 + y^2)^2}$$

$$\sigma_{xy} = x\frac{x^2 - y^2}{(x^2 + y^2)^2}$$

$$\sigma_{zz} = \nu(\sigma_{xx} + \sigma_{yy})$$

Using the method known as Dislocation Extraction Algorithm (DXA), we extract the line dislocation for our partial dislocation [241]. We rotate and translate
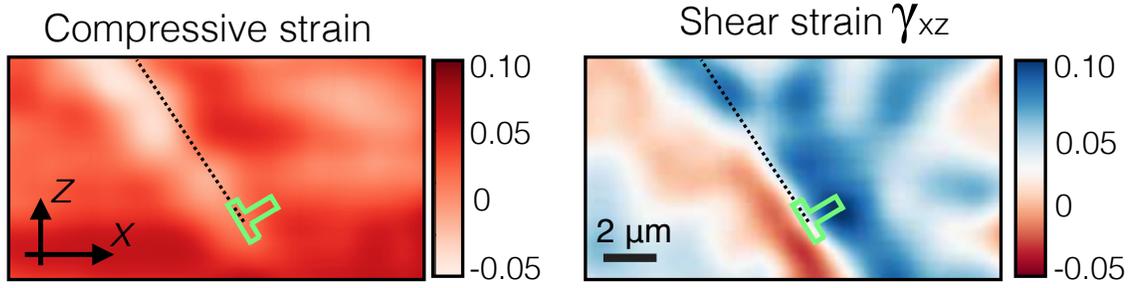
Figure 5.15: **Strain fields of a dislocation**. Experimental measurements of strain using the technique of Falk *et al.* [76] showing both compressive (left) and shear (right) strain distributions near a dislocation defect. The approximate location of the dislocation core is labeled with a (⊥). Notice that while the trend is very similar to that of the dislocation stress field (as in linear elasticity), there is a stronger divergence towards the core which is highlighted in the main text in Fig. 3.

the theoretical stress field for a single dislocation, integrating along the length of the dislocation as done in the experiment and simulation. Doing so, we find an excellent agreement with the other methods as shown in the main text Fig 3.

Strain fields

To investigate the relation between stress and strain, we determine the strain field near the dislocation. Following a previously developed algorithm [76], we measure the particle-level strain by quantifying the local affine deformation of individual particles. The compressive and shear strain (γ_{xz}) fields are shown in Fig. 5.15. We find that the strain fields show qualitatively similar features found in the stress distributions. However, as illustrated in Fig. 3(c) of the main paper, the stress-strain curve deviates from linearity near defects where strains are large.

Elastic moduli

Using the strain measurement, we analyze other experimentally accessible cubic moduli as done in the main manuscript. To that end, we compute the compressive C_{33} and shear C_{13} modulus profiles and discuss their behaviors below. We focus on the moduli associated with the strain component γ_{zz}, which shows a larger response than γ_{xx} and γ_{yy} in our experiment. For this analysis we rotate our stress and strain tensors so that the x and y axes align with the (100) and (010) axes of the cubic system. In this frame, the moduli for an cubic crystal have the following symmetries: $C_{33} = C_{22} = C_{11}$, and $C_{23} = C_{13} = C_{12}$.

To compute the compressive modulus near the dislocation core, we first determine the compressive stress σ_{zz} and strain γ_{zz}. Here, we determine the uniform background strain (due to the overall pressure arising from confinement and gravity) by matching the measured modulus to the corresponding theoretical value [80]. We then perform the same analysis used in the shear modulus calculation, and plot the compressive modulus C_{33} as a function of position $r/2a$ in Fig. 5.16(a). The region of the dislocation core is at $r/2a \approx 5.5$ (gray shade). We find that C_{33} is higher on the side with an additional half plane of particles. Similarly, we also observe a reduction in C_{33} on the other side due to the missing half plane of particles. We also find that the trend of the shear modulus C_{23} is similar to the compressive modulus C_{33} (Fig. 5.16(b)), which shows an enhancement on the left side and a reduction on the right.

The moduli we report are only calculated along a line perpendicular to the glide plane of the dislocation. The other two natural directions along the glide plane are excluded due to experimental limitations. Along the burgers vector we are limited by the noise in the stress and strain fields. Since the modulus is the
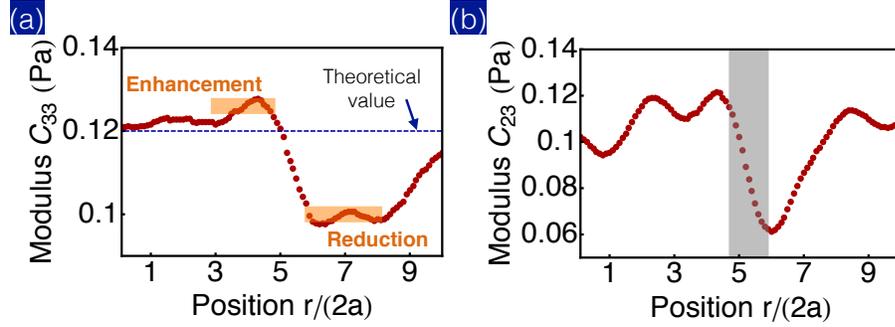
Figure 5.16: **Profile of the cubic moduli C_{33} and C_{23} for the dislocation**. (a) Compressive modulus C_{33} versus position perpendicular to the glide plane of the dislocation. (b) Shear modulus C_{23} measured along the same direction. Both moduli are calculated using the protocol described in the main manuscript.

ratio of the two, zeros remain problematic, similar to issues in deconvolution, and a new method of inference must be applied to the moduli in these regions. Along the dislocation line, we have already collapsed the data by averaging the stress and strain fields in this direction in order to reduce noise in the x-z plane. This averaging makes it infeasible to calculate the modulus variation in this direction. Due to symmetry, this direction should display a constant modulus. In the future, with a full time series of dislocation images, we can begin to look at the modulus variation near kinks and jogs along the dislocation line.

5.2.5   Polycrystal Stress Fields

Experimental details

We show a confocal image of the polycrystal we use in our experimental analysis in Fig. 5.17(a). The shown field of view is the same as the one of the stress field reported in the main manuscript. We show only a slice in the x-y plane of a complete 3D image stack. The z-interval between adjacent scan slices is 0.135 μm ~ 0.1 times the particle diameter. By matching the refractive indices of the water-glycerol mixture and silica particles, we minimize effects from the point spread function and z-axis spherical aberration, thus optimizing the image quality. We show the featured particle position with green circles in Fig. 5.17 (a). The data are visually overlaid to ensure there are no missing or repeated features.

To investigate the featuring accuracy, we plot the 1D $g(r)$ of the suspension in Fig. 5.17(b). As indicated by the sharpness of the first peak of $g(r)$, limiting our field of view to the first ten layers from the coverslip (in the optical z direction) enables us to accurately feature the particle positions. The spread of this peak arises from three contributions: featuring errors, polydispersity, and the thermal fluctuation of the equilibrium separation between particles. The location of the peak is primarily influenced by particle featuring errors and the packing fraction of the sample. While the polydispersity of this sample is $\sigma_a^2 \sim 5\%\langle a \rangle = 50$ nm, this variation in size will not shift the peak of $g(r)$ from the mean particle diameter (Fig. 5.17) unless there are spatial correlations of particle size. Therefore, at most, the averaging featuring errors should correspond to this peak shift of 50 nm.

Furthermore, we show a SEM micrograph to illustrate the roundness of the colloidal particles and the smoothness of their surfaces, see Fig. 5.17(c). As shown

186
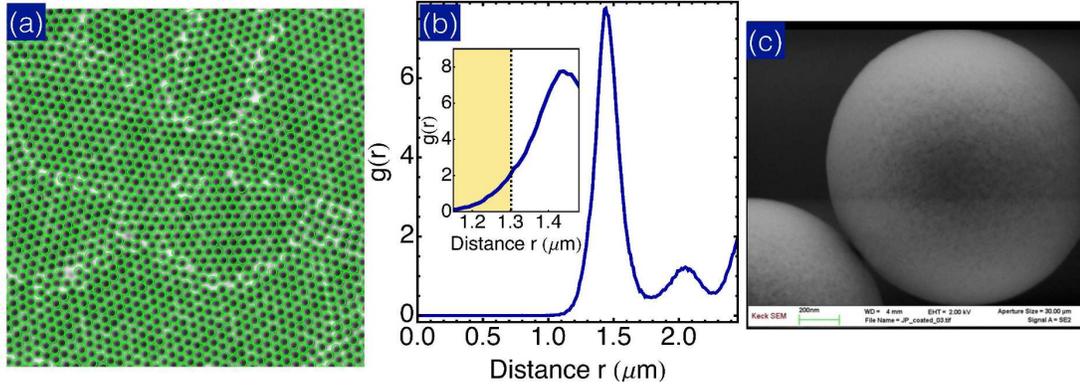
Figure 5.17: **Colloidal polycrystal sample**. (a) A representative slice of a 3D confocal image stack. Green circles illustrate the featured particle positions. This overlaid image shows that all particle positions are correctly identified without any missing particles. (b) The pair correlation function $g(r)$ calculated using featured particle positions. In the inset we show that the amount of particle overlap (yellow region) is negligible. (c) A SEM micrograph of two particles showing how the true particle size was determined in the sample.

in the SEM image, the surface roughness is less than the SEM resolution ~ 5 nm. We also use the SEM image to measure the polydispersity of the silica colloids. We find that the polydispersity of the sample is less than 3% of the particle size, consistent with the specification provided by the manufacturer.

In the stress measurement of the polycrystal, we average the calculated stress field over 50 stacks of images. The structure of the polycrystal remains unchanged within the acquisition time (~ 30 s). While we expect to observe short-time stress fluctuations arising from particle Brownian motion within the acquisition time, we do not find any significant stress fluctuations on longer time scales.

Simulated polycrystal stresses

As with the dislocation stresses, we verified that the experimental stresses are accurate using a simulation-experiment hybrid. Again we use the experimentally
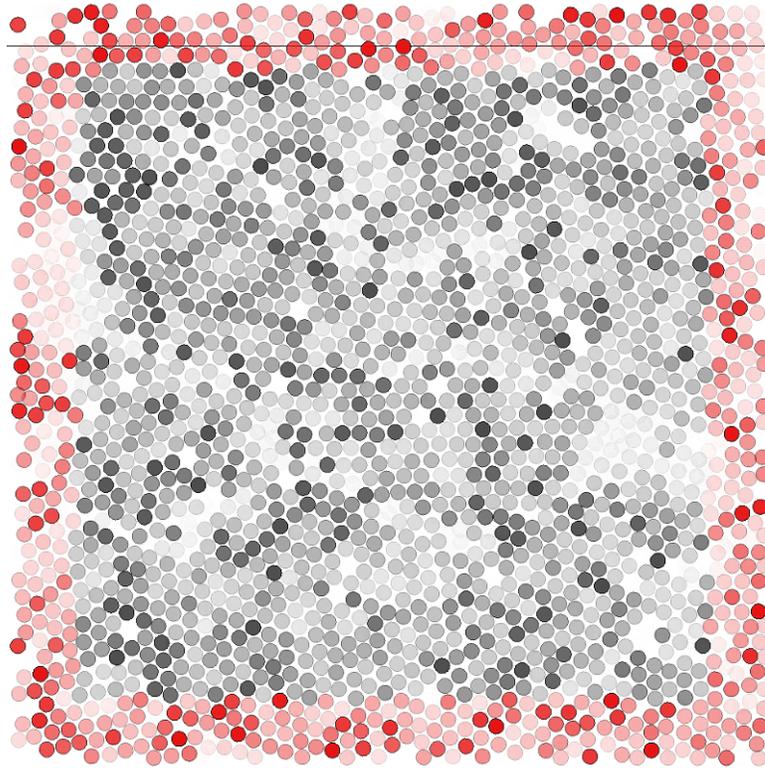
Figure 5.18: **Frozen particle border (polycrystal)**. A screen-shot of our Brownian dynamics simulation of the experimental polycrystal. In shades of gray are active hard sphere particles while in red are the frozen boundary particles. Luminance in this picture roughly indicates the position in the z direction with only a thin slice of the entire simulation shown.

featured particle positions as initial conditions for our hard sphere Brownian dynamics simulation. For each experimental snapshot, we remove overlaps and freeze boundary particles as before, then evolve the system, measuring stresses using both the true virial measurements as well as SALSA. By averaging over the various snapshots, we arrive at a stress field very similar to that found by the experimental SALSA calculation.
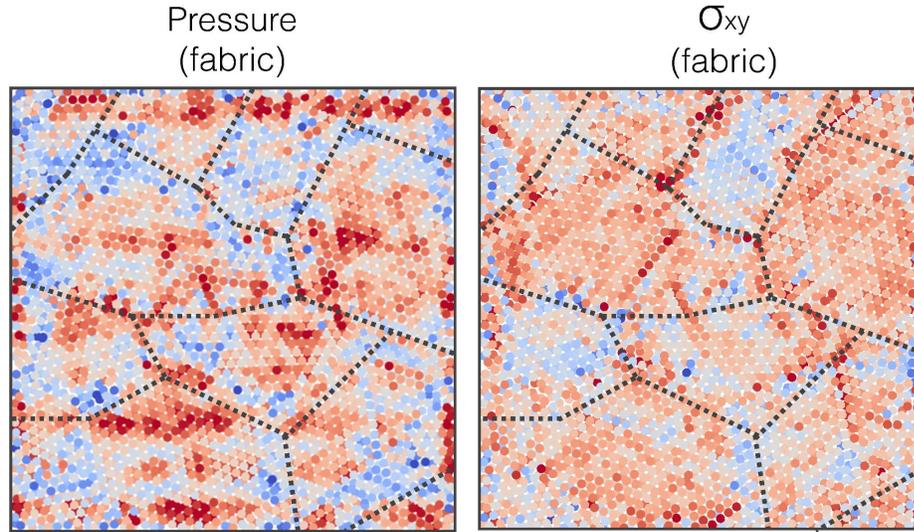
Figure 5.19: **Per-particle virials for polycrystal**. The experimental polycrystal with particles colored by the magnitude of the fabric tensor (directly proportional to particle virial). While we see large scale correlations of fabric tensor trace with grain interiors and variation of off-diagonal components between grains, the magnitude of variation is much smaller than that found in previous numerical studies of stress distributions in strained polycrystals [222]

Virial stresses

In simulation literature, the atomic-level stress has been referred to as the virial of an individual atom either normalized or un-normalized by the system volume [222, 149, 171, 267]. Here, the virial $F_i x_j$ can be considered to be a stress that does not account for the local variation of atomic (particle) free volume. To compare our experimental results with the previous simulation findings (which do not correct for local particle density fluctuations), we plot (Fig. 5.19) the un-smoothed fabric tensor (local structural anisotropy) that is calculated using SALSA. For detailed information of the fabric tensor calculation, see the section of SALSA derivation. As shown in Fig. 5.19, we find that both the pressure and shear stress fluctuations are evenly distributed throughout the polycrystal. This is in sharp

contrast to simulation results of sheared atomic polycrystals [222], which showed much stronger virial fluctuations at grain boundaries than in grain interiors.

CHAPTER 6

COLLECTIVE MOTION OF HUMANS IN MOSH AND CIRCLE PITS AT HEAVY METAL CONCERTS[1]

6.1 Heavy metal concerts

Human collective behaviors vary considerably with social context. For example, lane formation in pedestrian traffic [173], jamming during escape panic [104], and Mexican waves at sporting events [77] are emergent phenomena that have been observed in specific social settings. Here, we study large crowds ($10^2 - 10^5$ attendees) of people under the extreme conditions typically found at heavy metal concerts [260]. Often resulting in injuries [127], the collective mood is influenced by the combination of loud (130 dB [73]), fast (*blast beats* exceeding 300 beats per minute) music, synchronized with bright flashing lights, and frequent intoxication [153]. This variety and magnitude of stimuli are atypical of more moderate settings, and contribute to the collective behaviors studied here (Fig. 6.1).

Thousands of videos filmed by attendees at heavy metal concerts [6] highlight a collective phenomenon consisting of $10^1 - 10^2$ participants commonly referred to as a *mosh pit*. In traditional mosh pits, the participants (moshers) move randomly, colliding with one another in an undirected fashion (Fig. 6.2(A); see SI for video metadata). Mosh pits can form spontaneously or at the suggestion of the performing band, but in both cases, no micromanagement of individual actions is generally involved. Qualitatively, this phenomenon resembles the kinetics of gaseous particles, even though moshers are self-propelled agents that experience dissipative

---

[1]Silverberg, J. L., Bierbaum, M., Sethna, J. P., & Cohen, I. (2013). Collective motion of humans in mosh and circle pits at heavy metal concerts. Physical review letters, 110(22), 228701.

Figure 6.1: **Typical collective behavior found in a mosh pit at heavy metal concerts**. Notice that some attendees are participating (foreground), while others are not (background). Image courtesy of Ulrike Biets.

collisions and exist at a much higher density then most gaseous systems. To explore this analogy quantitatively, we watched over 10^2 videos containing footage of mosh pits on YouTube.com, obtained six that were filmed from a suitably high position to provide a clear view of the crowd, corrected for perspective distortions [204] as well as camera instability, and used PIV analysis [96] to measure the two-dimensional (2D) velocity field on an interpolated grid (Fig. 6.2(B)).

Video data of mosh pits was used to calculate the velocity-velocity correlation function c_{vv}, where we noted an absence of the spatial oscillations typically

found in liquid-like systems (Fig. 6.2(B) inset) [19]. Generally, c_{vv} was well fit by a pure exponential, and for the video used in Fig. (6.2), the decay length was 0.39 ± 0.03 m, which is approximately human shoulder width. Taken together, these findings offer strong support for the analogy between mosh pits and gases. As a further check, we examined the 2D speed distribution. Previous observations of human pedestrian traffic and escape panic led us to expect a broad distribution not well described by simple analytic expressions [110, 104]. However, the measured speed distribution in mosh pits was well fit by the 2D Maxwell-Boltzmann (MB) distribution, which is characterized by the probability distribution function $\text{PDF}(v) = (2v/T)e^{-v^2/T}$ and temperature T (Fig. 6.2(C) and inset). These observations present an interesting question: why does an inherently non-equilibrium system exhibit equilibrium characteristics?

6.2 Flocking model

Studies of collective motion in living and complex systems have found notable success within the framework of flocking simulations [61, 65, 220, 147, 105, 257, 33, 207, 256]. Thus, we use a Vicsek-like model [256] to simplify the complex behavioral dynamics of each human mosher to that of a simple soft-bodied particle we dub a Mobile Active Simulated Humanoid, or MASHer. Our model includes two species of MASHers to reflect the typical crowd at heavy metal concerts where we find both active and passive participants (Fig. 6.1, foreground and background, respectively) [268]. Active MASHers repel during collisions, exhibit self-propulsion, experience flocking interactions, and are subject to random fluctuations due to environmental
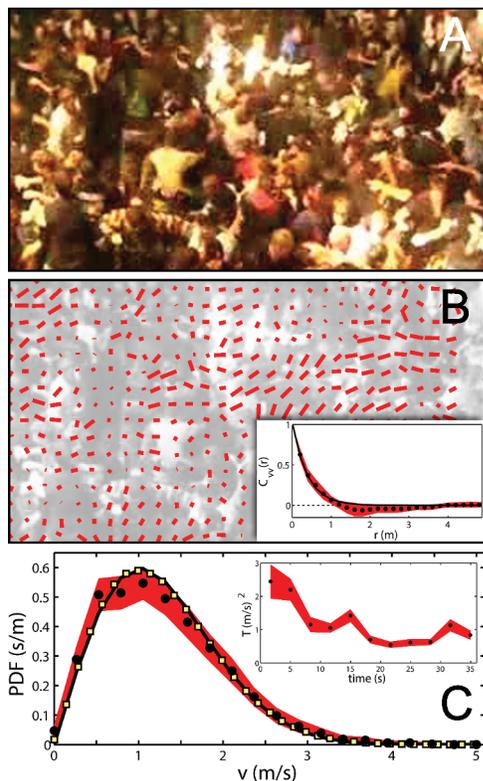
Figure 6.2: **Experimental analysis of videos of mosh pits**. (A) Single video frame illustrating a characteristic mosh pit [6]. (B) The same video image with overlaid velocity field. To facilitate comparisons with (A), this image is not corrected for perspective distortions. Inset shows the measured velocity-velocity correlation c_{vv} (solid black circles) as a function of distance r, as well as the best-fit to a pure exponential (black line, $R^2 = 0.97$). (C) The measured PDF for speed from the same video (solid black circles), the best fit to a 2D MB distribution (black line), and the speed distribution found in simulations (yellow squares). Inset shows the best-fit temperature as a function of time illustrating that an initially "hot" mosh pit "cools down". Error estimates are in red for all plots.

stimuli. These effects are modeled as forces on the ith MASHer by:

$$\mathbf{F}_i^\text{repulsion} = \begin{cases} \epsilon \left(1 - \frac{r_{ij}}{2r_0}\right)^{3/2} \hat{\mathbf{r}}_{ij}, & r_{ij} < 2r_0 \\ 0, & \text{otherwise} \end{cases} \tag{6.1}$$

$$\mathbf{F}_i^\text{propulsion} = \mu(v_0 - v_i)\hat{\mathbf{v}}_i, \tag{6.2}$$

$$\mathbf{F}_i^\text{flocking} = \alpha \sum_{j=0}^{N_i} \mathbf{v}_j \Big/ \left|\sum_{j=1}^{N_i} \mathbf{v}_j\right|, \tag{6.3}$$

$$\mathbf{F}_i^\text{noise} = \boldsymbol{\eta}_i. \tag{6.4}$$

The Hertzian repulsion force [141] has a strength ϵ, and is determined by the MASHer radius r_0, as well as the distance r_{ij} and direction $\hat{\mathbf{r}}_{ij}$ between MASHers i and j. A variant of this expression with a 5/2 power-law was tested and found to produce quantitatively similar results. The self-propulsion force has a strength μ, is aligned with the MASHer heading $\hat{\mathbf{v}}_i$, and is proportional to the difference between the current speed v_i and the preferred speed v_0. The flocking force has a strength α, and is in the direction found by vectorially averaging the headings of the N_i MASHers within a distance $r_\text{flock} = 4r_0$ of MASHer i. Consistent with previous work [65, 257, 256], this distance was fixed in our model so that $r_0 < r_\text{flock} < L$, where L is the system size. This choice minimizes the influence of finite-size effects on the flocking force [61]. Finally, the random force $\boldsymbol{\eta}_i$ is a vector whose components $\eta_{i,\lambda}$ are drawn from a Gaussian distribution with zero mean and standard deviation σ defined by the correlation function $\langle \eta_{i,\lambda}(t)\eta_{i,\kappa}(t')\rangle = 2\mu\sigma^2\delta_{\lambda\kappa}\delta(t - t')$; the noise is spatially and temporally decorrelated. Based on observational evidence, the second species in our model, passive MASHers, prefer to remain stationary and are not subject to flocking interactions or random forces. Thus, in the appropriate units, we set $v_0 = 0$, $\alpha = 0$, and $\boldsymbol{\eta}_i = \mathbf{0}$ for passive MASHers. Active MASHers have $v_0 = 1$, while α and σ were varied to explore the phase space of the model. The remaining parameters are the same for all MASHers, and were set to $\epsilon = 25$, $\mu = 0.05$, and $r_0 = 1$.

We simulated concerts with $N = 500$ MASHers at a packing fraction of $\rho = 0.94$. 30% of the population was active, while the remaining was passive. Periodic boundary conditions were used to avoid edge effects, and numerical integration of $\ddot{\mathbf{r}}_i = \mathbf{F}_i^{\text{repulsion}} + \mathbf{F}_i^{\text{propulsion}} + \mathbf{F}_i^{\text{flocking}} + \mathbf{F}_i^{\text{noise}}$ was performed with the Newton-Stomer-Verlet algorithm with cell based neighbor lists to expedite computation. Initializing the simulation with uniformly mixed populations, we found that they spontaneously phase separated with a dense region of active MASHers confined by passive MASHers. This occurs generally across parameter space, and appears to be a product of the difference in preferred speeds between populations (SI). For the parameter values studied here, this occurs in about $\sim 10^3 \times (r_0/v_0)$ time units and once formed, remains stable for greater than $10^5 \times (r_0/v_0)$ time units.

6.3   Phase diagram

We explored the model's phase space by varying α and σ for the active MASHers over the intervals $[0, 1]$ and $[0, 3]$, respectively (Fig. 6.3(A)). This led to 4.8×10^5 individual simulations sampled on 4.8×10^3 grid points. For each run, we measured the active MASHer RMS angular momentum about their center of mass $x_{cm} = (L/2\pi)\arctan(\text{Im(A)}/\text{Re(A)})$, where $L = 1.03\sqrt{\pi r_0^2 N}$ is the simulation box size, $A = \sum_{i=1}^{N_a} \exp\left(-2\pi i x_i/L\right)$, N_a is the number of active MASHers, x_i is the x position of the ith MASHer, and a similar expression holds for y_{cm}. In the low-flocking, high-noise limit, we found the angular momentum was near zero, and upon closer inspection, discovered a gas-like region (Fig. 6.3(B)) where MASHers quantitatively reproduced the statistics found in experimentally observed mosh pits (Fig. 6.2(C)).
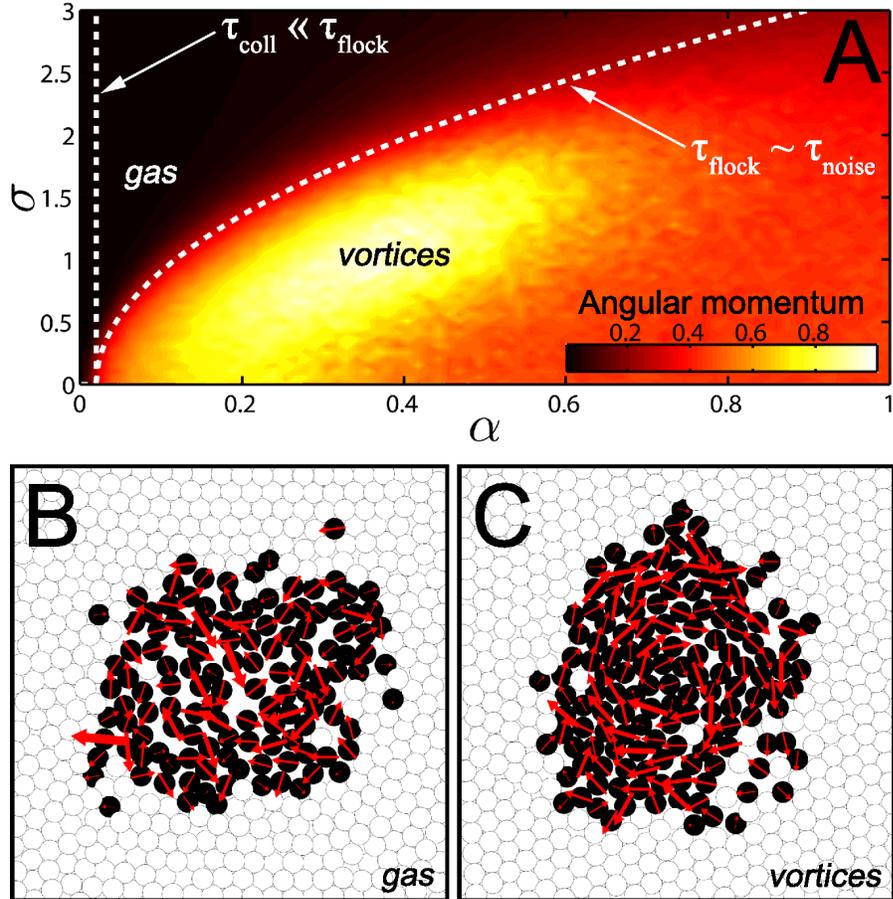
Figure 6.3: **Simulation phase diagram and example behaviors in mosh and circle pit**. (A) The RMS angular momentum of active MASHers exhibits a disordered gas-like state in the high-noise low-flocking limit. The model also predicts an ordered vortex-like state in the low-noise moderate-flocking limit. Dashed white lines correspond to the bounds of the flocking-dominated regime. (B) Active MASHers (black) are confined by passive MASHers (white), and the velocity field (red arrows) resembles that found in actual mosh pits. (C) Active MASHers spontaneously self-organize into an ordered vortex-like state. (Supplemental Movie 1 and 2.)

To interpret these results, we note that our model has three time scales: (i) the flocking time $\tau_{\text{flock}} = v_0/\alpha$, which can be found by dimensional analysis of Eq.(6.3); (ii) the noise time $\tau_{\text{noise}} = v_0^2/2\mu\sigma^2$, which can be found by calculating the amount of time required for noise to change the correlation function $\langle[v_i(\tau_{\text{noise}}) - v_i(0)]^2\rangle = 2\mu\sigma^2\tau_{\text{noise}}$ by an amount equal to the characteristic speed squared; and (iii) the collision time $\tau_{\text{coll}} = 1/(2r_0v_0\rho)$, which is the mean-free-path $(2r_0\rho)^{-1}$ divided by the speed v_0. Both noise and collisions tend to randomize motion, whereas flocking tends to homogenize motion. Thus, when $\tau_{\text{noise}} \ll \tau_{\text{flock}}$, the statistical motion of the system is dominated by random forces. The boundary given by this condition occurs when $\tau_{\text{noise}} \sim \tau_{\text{flock}}$, or rather, $\sigma \sim \sqrt{\alpha/\mu}$ (Fig. 6.3(A)). Similarly, when the $\tau_{\text{coll}} \ll \tau_{\text{flock}}$, collisions cause disordered motion. This regime is bounded by $\alpha \ll 2r_0v_0^2\rho$, which is independent of σ, and for our choice of parameters is ~ 1; empirically, we find $\alpha \sim 10^{-2}$ in agreement with this condition. This demonstrates how a non-equilibrium system of moshers can have equilibrium characteristics: random motions induced by *collisions or noise* of self-propelled agents over a sufficient time reproduce the statistics of classical gases via the Central Limit Theorem.

Conversely, when $\tau_{\text{flock}} \ll \tau_{\text{noise}}$ and τ_{coll}, the flocking term dominates active MASHer motion. With sufficiently low noise, this limit of the model predicts a highly ordered vortex-like state [26, 245] where MASHers again phase separate, but the confined active MASHers move with a large non-zero angular momentum (Fig. 6.3(C)). Remarkably, this spontaneous phase separation and vortex formation is also observed at heavy metal concerts where they are conventionally called *circle pits* (Fig. 6.4; see SI for video metadata) [6]. In simulations, we found an even distribution between clockwise (CW) and counter-clockwise (CCW) motion (when viewed from above) that switches directions at random intervals [49]. However,

observations from concerts show 5% flow CW with the remaining 95% flowing CCW ($p < 0.001$). This asymmetry is independent of regional conventions in motor vehicle traffic, as video data was collected from a variety of countries including the United State of America, the United Kingdom, and Australia. Though the origin of this effect is unknown, we speculate it may be related to the dominant handedness/footedness found in humans, as it is known to bias turning behaviors [169].

6.4 Conclusions

The collective behavior described here has not been predicted on the basis of staged experiments with humans [126, 172], making heavy metal concerts a unique model system for reliably, consistently, and ethically studying human collective motion. Currently, the most significant obstacle to further progress is the limited availability of publicly available high-quality video footage, and a general reluctance among concert organizers to allow filming at their events. Nevertheless, further studies in this unique environment may enhance our understanding of collective motion in riots, protests, and panicked crowds, as it sheds light on what collective behaviors become possible when traditional social rules are abandoned. Such studies may lead to new architectural safety design principles and crowd management strategies that limit the risk of injury at mass social gatherings [103]. For example, many heavy metal bands routinely announce during live performances "If you see someone fall down in the mosh pit, pick them back up." This simple rule is known to reduce the risk of injury by trampling, and if employed in other extreme social gatherings, would be expected to have similar social benefits. Similarly, within the MASHer model, we found that by setting the preferred speed $v_0 = 0$, all mosh and
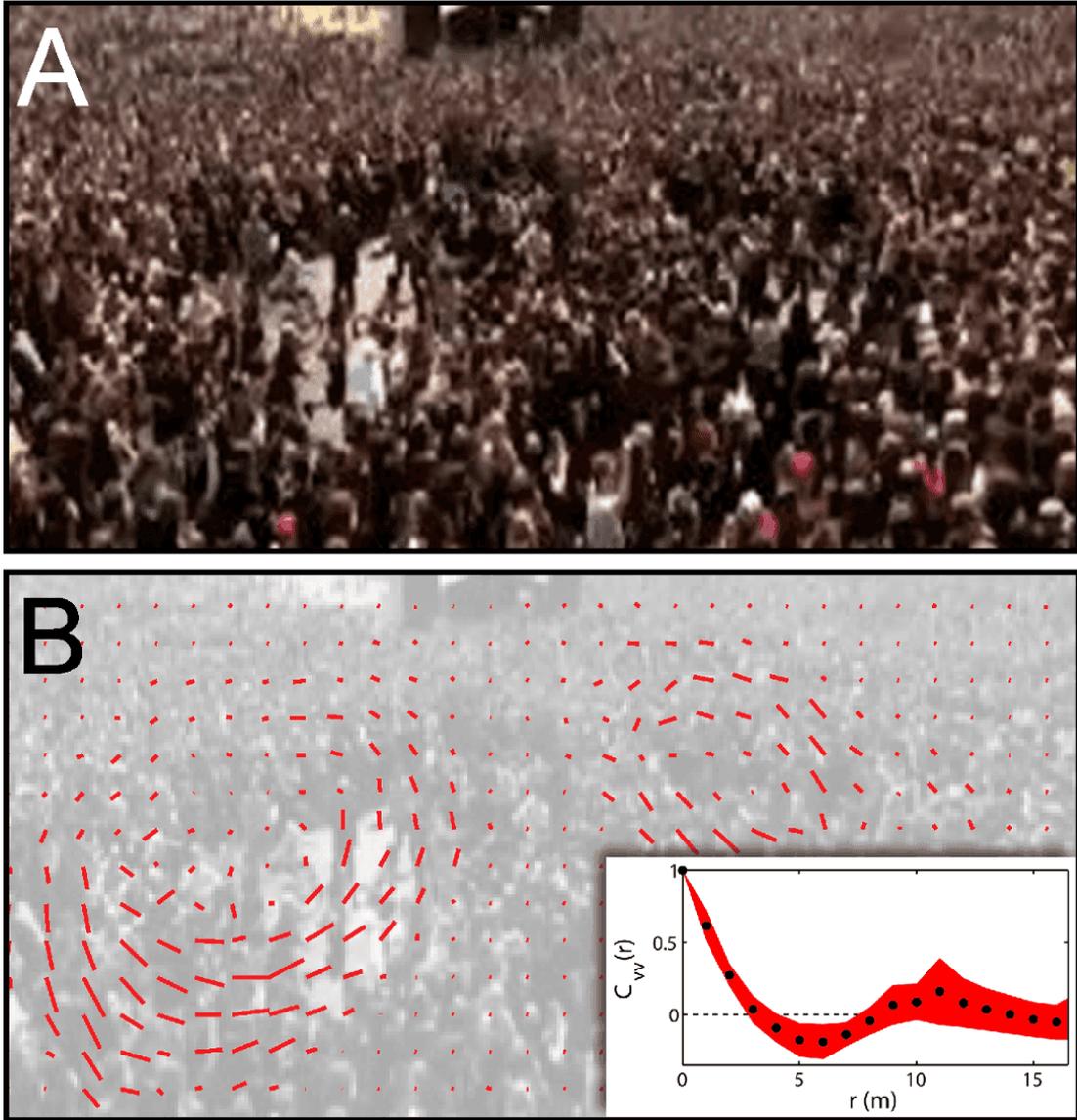
Figure 6.4: **Experimental characterization of a circle pit**. The vortex-like state predicted in simulations is also observed at heavy metal concerts, where it is called a circle pit. (A) Single video frame illustrating two side-by-side circle pits [6]. (B) The same video image with overlaid velocity field. To facilitate comparisons with (A), this image is not corrected for perspective distortions. Inset shows the measured velocity-velocity correlation c_{vv} as a function of distance r (solid black circles, error estimates in red). Note that c_{vv} is maximally negative at $r \approx 6$ m, corresponding to the approximate diameter of the left-most circle pit. Weak oscillations for $r > 6$ m are evident due to long-range correlations between the two circle pits.

circle pit behaviors ceased, suggesting an alternative approach to real-world crowd safety management.

Heavy metal concerts have the further advantage of exhibiting a rich variety of collective behaviors such as: (i) the *wall of death* (moshers split into two groups separated by an open space and, when signaled, simultaneously run at the opposing group leading to a deliberate mass collision), (ii) *pogoing* (a locally correlated but globally decorrelated collective jumping), and (iii) propagating waves in jammed attendees [234]. In addition to these broadly defined types of collective motion, there are further variations that arise when concert organizers focus on specific musical subgenres that appeal to niché audiences. For example, *hardcore pits*, *ninja pits*, and *push pits* are all variants of the traditional mosh pit with their own unique characteristics that may not, when studied in detail, be well described by Eqs.(6.1-6.4). Thus, heavy metal concerts offer many new opportunities to study the collective behaviors arising from large groups of humans in extreme social conditions.

6.5   Additional analysis

6.5.1   Video Metadata

A search on YouTube.com for the phrase "mosh pit" returns over 10^5 results. We watched over 10^2 videos that exhibited mosh pits and circle pits indicating that these collective behaviors are robust, reproducible, and largely independent of factors such as the musical subgenre, timing of performance, crowd size, arena size, suggestions from the band, time of year, and socioeconomic status of the

moshers.

For the video shown in Fig. 2 of the manuscript, the following additional data was available:

- Band: 36 Crazyfists

- Recorded on Dec. 4, 2008

- First of three bands performing

- Audience size $\sim 10^2$ attendees

- Venue: The NorVa, Norfolk, VA

- Venue capacity: 1,500

- Band encouraged moshing

For the video shown in Fig. 4 of the manuscript, the following additional data was available:

- Band: Hatebreed

- Recorded on June 11, 2006

- Second of ten bands performing

- Audience size $\sim 10^5$ attendees

- Venue: Donington Park, Leicestershire, England

- Band encouraged moshing

6.5.2 PIV Analysis

To quantify the motion of moshers in videos filmed at heavy metal concerts, we used PIV analysis to measure the 2D velocity field on an interpolated grid. In our analysis, we first correct for perspective distortions due to the camera position. This amounts to a linear transformation with two free parameters that define the skew angle and stretching factor. In turn, these parameters are set by the camera height and tilt, and can be estimated by looking at stationary objects in the field of view. The PIV algorithm then breaks up each video frame of width w and height h into a square grid spaced by an amount Δ, which is described more below. Each square tile has area Δ^2 and is centered on a position $\langle x_i, y_j \rangle$ where $i = 1 \ldots \lfloor w/\Delta \rfloor$, and $j = 1 \ldots \lfloor h/\Delta \rfloor$. To determine the velocity associated with a given tile at position $\langle x_i, y_j \rangle$ in frame f_n, the region given by $x_i - 2\Delta \ldots x_i + 2\Delta, y_j - 2\Delta \ldots y_j + 2\Delta$ in frame f_{n+1} is examined. Within this larger region, all positions of overlap are examined and a local correlation function is calculated. The region of maximum correlation is generally offset from $\langle x_i, y_j \rangle$ to $\langle x_i + \chi_i, y_j + \upsilon_j \rangle$, which in turn defines the displacement vector $\langle \chi_i, \upsilon_j \rangle$ associated with a given tile. This process is iterated for all tiles in a given frame, and all consecutive pairs of frames.

Various values for Δ were tested ranging from approximately $\ell/5$ to 2ℓ, where ℓ is the characteristic feature size in a given video, i.e., the approximate width of a human. Grid sizes between to $\ell/3$ and 1.5ℓ were found to have no substantial effect on the measured velocity field: in this range, we found consistent statistical results independent of ℓ. Outside this interval, the algorithm failed to track the motion of tiles more than 35% of the time.

The algorithm was further tested with calibration videos wherein spherical objects were translated across the field of view with small random fluctuations in

the orthogonal direction. In these tests, the motion was tracked and a series of digital "tracer particles" were placed in the measured flow field. The tracer particle positions were updated frame-by-frame according to the locally measured velocity field. The algorithm was thus validated when we found that the tracer particles followed the test objects throughout the video.

6.5.3 Phase Separation

In the main text, we note that an initially uniform mixture of active and passive MASHers spontaneously phase separate. In our observations, we see that both mosh pits and circle pits can form spontaneously, which lead us to simplify our computations by beginning with initial conditions where a circular group of active MASHers are surrounded by passive MASHers. However, we find this phase separation behavior much more general and striking. Even after removing the flocking and noise forces, self-propulsion alone is able to drive the separation of active and passive MASHers. Also, depending on our parameters, different morphologies arise with time. As shown in Supplemental Fig. 6.5, we can arrive at well-defined vortices, stringing colonies of coherent active populations, winding pathways of gas-like states, and well-defined circular gas-like states. Note that MASHer color coding is identical to that of Fig. 3 of the main text.

This patterning is reminiscent of that seen in driven granular materials. For example, clustering of grains has been reported in experiments [216] in which grains of a single type are excited by vertical vibrations. These clusters are thought to form due to the effect of high densities on dissipation of kinetic energy. We hypothesize a similar situation arises in our MASHer model due to the difference in velocities between the active and passive populations.
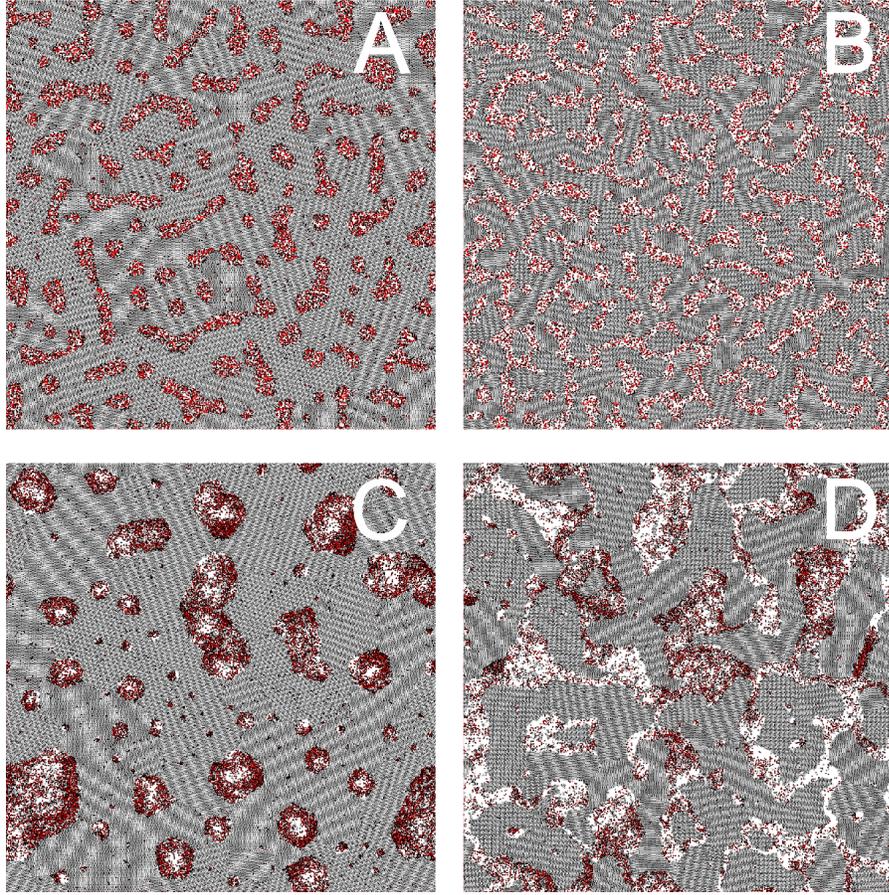
Figure 6.5: **Coarsening of mosh and circle pits in large-scale simulations**.
Simulations of 10^5 MASHers randomly assigned an active or passive
status such that 30% were active reveal phase separation behaviors that
vary with the model parameters. (A) and (B) have zero flocking and
noise ($\alpha = \sigma = 0$), while (C) and (D) have finite flocking and noise
($\alpha = 0.25, \sigma = 1$). At zero flocking, we find more homogeneously
distributed active MASHers within confined regions, whereas at finite
flocking, we find larger density fluctuations due to flocking, which acts
like a long-range attraction. (A) and (C) have a packing fraction of
$\rho = 0.83$, while (B) and (D) have a packing fraction of $\rho = 0.69$. At
higher packing fractions we see well confined groups of active MASHers,
whereas at lower packing fractions, the active MASHers form a more
network-like structure. All images represent the system after $\approx 10^4$
time steps, which is sufficient time for clusters of mosh/circle pits to
coarsen and grow.

6.5.4   Expanded MASHer Phase Diagram

The phase diagram shown in Fig. 3 of the main text illustrates a region of gas-like and vortex-like collective motions by plotting the RMS angular momentum over a range of parameter values. Here, we expand on this plot (reproduced in Fig. 6.6(A)), by showing the standard deviation of the active MASHer angular momentum (Fig. 6.6(B)). We note that the largest fluctuations arise at high values of the flocking coefficient α relative to the values associated with vortex formation. Interestingly, this region in parameter space coincides with a region where the RMS linear momentum is also large (Fig. 6.6(C)). Simulations at high flocking and low noise demonstrate the formation of active MASHer lanes that push through the passive MASHers (Supplemental Movie 3). Indeed, fluctuations in the magnitude of the linear momeuntum (Fig. 6.6(D)) increase with the linear momentum illustrating that the lane mean free path shortens at higher flocking.

In the MASHer model, we found that setting the preferred speed of the active MASHers v_0 equal to zero suppressed both mosh and circle pits. While this region of parameter space appears to reflect a calm crowd, we must be careful when extrapolating a minimal flocking model, designed to explore the physics of collective motion, to real-world circumstances. For example, the model presented here does not account for (i) individual changes in moshing preferences, (ii) that people can fall down, or (iii) groups of people can be spatially coupled by social connections. While the MASHer model appears sufficient to predict large scale collective behavior, these additional features may be important for real-world applications such as crowd safety management.
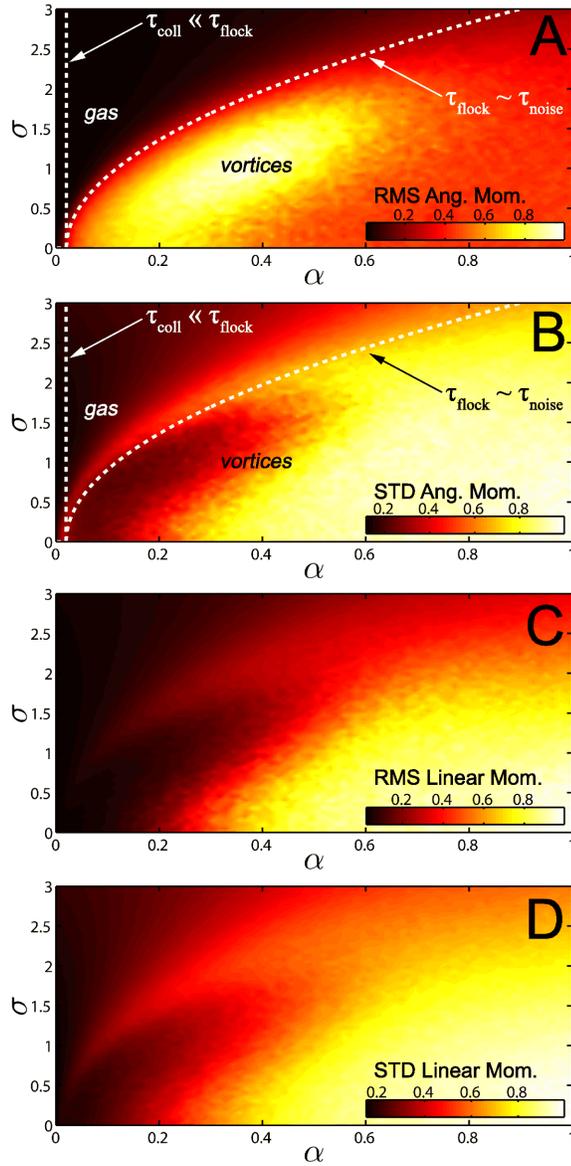
Figure 6.6: **Phase diagram of additional order parameters**. We show here
additional features of the MASHer model in the region of phase space
described in the main text. (A) Reproduction of the phase diagram
in Fig. 6.3 showing the RMS angular momentum. (B) Standard devi-
ation of the angular momentum with the identical contours from (A)
superimposed. (C) Magnitude of active MASHer linear momentum
illustrating a region of lane formation at high flocking and low noise.
(D) Standard deviation of the linear momentum magnitude illustrat-
ing that fluctuations increase with the linear momentum. All order
parameters have been normalized to their maximum value.

YOU CAN RUN, YOU CAN HIDE: THE EPIDEMIOLOGY AND STATISTICAL MECHANICS OF ZOMBIES [1]

7.1 Introduction

Zombies captivate the imagination. The idea of a deadly disease that not only kills its hosts, but turns those hosts into deadly vectors for the disease is scary enough to fuel an entire genre of horror stories and films. But at its root, zombism is just that – a (fictional) disease – and so should be amenable to the same kind of analysis and study that we use to combat more traditional diseases.

Much scholarly attention has focused on more traditional human diseases [129], but recently, academic attention has turned a bit of thought onto zombies as a unique and interesting modification of classic disease models. One of the first academic accounts of zombies was the 2009 article by Munz et al. [175], in which an early form of a compartmental model of zombism was introduced. Since then, there have been several interesting papers published including works that perform Bayesian estimations of the zombie disease parameters [263], look at how emotional factors impact the spread of zombies [186], using zombies to gain insight into models of politics [114], or into the interaction of a zombie epidemic and social dynamics [215, 170]. Additional essays can be found in two books collecting academic essays centered around zombism [44, 237]

Besides the academic papers, zombies have seen a resurgence in fiction. Of particular note are the works of Max Brooks, including a detailed *Zombie Survival*

Guide [38], as well as an oral history of the first zombie war [39] in a hypothesized post outbreak world. In both these works Brooks provides a rich source of information about zombies and their behavior. In particular, he makes the connection to disease explicit, describing zombies as the result of a hypothetical virus, *Solanum.*

Zombies form a wonderful model system to illustrate modern epidemiological tools drawn from statistical mechanics, computational chemistry, and mathematical modeling. They also form an ideal vehicle for public outreach: the Center for Disease Control uses preparation for a zombie apocalypse [188, 189] to promote emergency preparedness. In this work, we will build up to a full-scale simulation of a zombie outbreak in the continental United States, with realistic values drawn from the literature and popular culture (section 7.5, simulation accessible online [2]). Before that, we shall use statistical mechanics to scrutinize the threshold of zombie virulence that determines whether humanity survives (section 7.4). Preceding that, we shall show how methods from computational chemistry can be used to simulate every individual heroic encounter between a human and a zombie (section 7.3). But we begin by describing and analyzing a simple model of zombies (the SZR model) – the simplest and most natural generalization to the classic SIR (Susceptible-Infected-Recovered) model used to describe infectious disease spread in epidemiology.

7.2 SZR **Model**

We start with a simple model of zombies, the SZR model. There are three compartments in the model: S represents the susceptible population, the uninfected

---

[2]http://mattbierbaum.github.io/zombies-usa/

humans; Z represents the infected state, zombies; and R represents our removed state, in this case zombies that have been terminated by humans (canonically by destroying their brain so as to render them inoperable). There are two transitions possible: a human can become infected if they are bitten by a zombie, and a zombie can be destroyed by direct action by a human. There are two parameters governing these transitions: β, the bite parameter determines the rate at which a zombie will bite a human if they are in contact, and κ the kill parameter that gives the rate that a human kills the zombie. Rendered as a system of coupled differential equations, we obtain, for a particular interaction site:

$$\dot{S} = -\beta SZ \tag{7.1}$$

$$\dot{Z} = (\beta - \kappa)SZ \tag{7.2}$$

$$\dot{R} = \kappa SZ \tag{7.3}$$

Notice that these interactions are *density dependent*, in the sense that the rate at which we convert humans to zombies and kill zombies is dependent on the total count of zombies and humans in this site. This is in contrast with most models of human diseases, which frequently adopt *frequency dependent* interactions wherein S, Z, R would have been interpreted as the fraction of the population in the corresponding state.

This distinction will become stark once we consider large simulations with very inhomogeneous populations. By claiming that zombies can be modeled by a single bite parameter β that itself is a rate per person per unit time, we are claiming that a zombie in a block with 5,000 people would be one hundred times as effective at infecting new zombies as a zombie in a block with fifty people; similarly the zombie in question would be killed one hundred times faster. This would seem false for an ordinary disease like the flu, but in the case of zombies, we argue that it is

appropriate. Zombies directly seek out hosts to infect, at which point the human and zombie engage in a duel to the (un)death.

To facilitate analysis we can nondimensionalize the equations by choosing a relevant population size N, and recasting in terms of the dimensionless time parameter $\tau = t\beta N$ and dimensionless virulence $\alpha = \kappa/\beta$

$$
\begin{aligned}
\frac{dS}{d\tau} &= -\frac{SZ}{N} \\
\frac{dZ}{d\tau} &= (1-\alpha)\frac{SZ}{N} \\
\frac{dR}{d\tau} &= \alpha\frac{SZ}{N}
\end{aligned}
\tag{7.4}
$$

Unlike a traditional disease (e.g., as modeled by SIR), for the zombie model, we have a stable configuration when either the human or the zombie population is defeated ($S = 0$ or $Z = 0$). Furthermore, unlike SIR, SZR admits an analytical solution, assuming $R(0) = 0$, and with $Z_0 \equiv Z(0), S_0 \equiv S(0)$:

$$
P \equiv Z_0 + (1-\alpha)S_0
\tag{7.5}
$$

$$
\mu \equiv \frac{S_0}{Z_0}(1-\alpha) = \frac{P}{Z_0} - 1
\tag{7.6}
$$

$$
f(\tau) \equiv \frac{P\mu}{e^{\tau P/N} + \mu}
\tag{7.7}
$$

$$
Z(\tau) = P - f(\tau)
\tag{7.8}
$$

$$
S(\tau) = \frac{f(\tau)}{1-\alpha}
\tag{7.9}
$$

Given the analytical solution, it is clear to see that the sign of P governs whether there will eventually be humans or zombies in the final state. If $\alpha < 1, P > 0$, so

$$
\lim_{\tau\to\infty} f(\tau) = 0
\tag{7.10}
$$

$$
\lim_{\tau\to\infty} Z(\tau) = P = Z_0 + (1-\alpha)S_0
\tag{7.11}
$$

$$
\lim_{\tau\to\infty} S(\tau) = 0
\tag{7.12}
$$

211

and the system will always flow to a final state composed of entirely zombies and no humans, where P denotes the number of zombies that survive.

If however, $\alpha > 1$, humans are more effective at killing zombies than zombies are at biting humans. With enough zombies in the initial state, we can still convert all of the humans before they have time to kill all of the zombies.

We can recast the dynamics in terms of the variables $P \equiv Z + (1-\alpha)S$ and $\chi = S/Z$ to gain further insights. First note that:

$$\frac{dP}{d\tau} = P' = Z' + (1-\alpha)S' \tag{7.13}$$

$$= (1-\alpha)\frac{SZ}{N} - (1-\alpha)\frac{SZ}{N} = 0 \tag{7.14}$$

so P is a constant of the dynamics. As for χ:

$$\chi' = \frac{S'}{Z} - \frac{SZ'}{Z^2} \tag{7.15}$$

$$= -\frac{S}{N} - (1-\alpha)\frac{S}{N}\frac{S}{Z} \tag{7.16}$$

$$= -\frac{S}{N}\left(1 + (1-\alpha)\right)\chi \tag{7.17}$$

$$= -\frac{P}{N}\chi \tag{7.18}$$

Hence if we choose $N = |P|$, we end up with the very simple dynamics:

$$P'(\tau) = 0 \tag{7.19}$$

$$P(\tau) = P_0 = Z(\tau) + (1-\alpha)S(\tau) = Z_0 + (1-\alpha)S_0 \tag{7.20}$$

$$\chi'(\tau) = \begin{cases} -\chi & P > 0 \\ +\chi & P < 0 \end{cases} \tag{7.21}$$

$$\chi(\tau) = \frac{S(\tau)}{Z(\tau)} = \chi_0 \begin{cases} e^{-\tau} & P > 0 \\ e^{+\tau} & P < 0 \end{cases} \tag{7.22}$$

$$\chi_0 \equiv \frac{S_0}{Z_0} \tag{7.23}$$

Here we see that the dynamics is simply an exponential decay or increase in the ratio of humans to zombies $\chi = S/Z$. The final populations in either case are easy to see due to the conservation of P. If zombies win we have

$$Z_\infty = Z_0 + (1 - \alpha)S_0 \qquad (7.24)$$

And if humans win

$$S_\infty = S_0 - \frac{Z_0}{\alpha - 1} \qquad (7.25)$$

SIR model

This dynamics should be compared to the similarly nondimensionlized density-dependent SIR model:

$$\frac{dS}{d\tau} = -\frac{SI}{N} \qquad (7.26)$$

$$\frac{dI}{d\tau} = \left(\frac{S}{N} - \mu\right) I \qquad (7.27)$$

$$\frac{dR}{d\tau} = \mu I \qquad (7.28)$$

Here $\tau = t\beta N$ as above, but $\mu = \nu/(\beta N) = R_0^{-1}$, because in the SIR model our infected population recovers on its own. This is contrasted with SZR, where the process of infection and recovery have the same functional form, depending on the product SZ. This μ is the inverse of the usual R_0 parameter used to denote the infectivity of the SIR model, here used to make a closer analogy to the SZR model. It is this parameter that principally governs whether we have an outbreak or not. Unlike the α parameter for SZR which depends only on our disease constants β, κ, the relevant virulence for the density dependent SIR model (μ) has a population dependence.

Notice again that while the only stable configuration for the SIR model is when

there is no infected population ($I = 0$), the SZR model is stable when either the humans or zombies are depleted ($S = 0$ or $Z = 0$).

The SIR model does not admit a closed form analytical solution, but we can find a parametric solution by dividing the first equation by the third, revealing.

$$S(\tau) = S_0 e^{-\frac{(R(\tau) - R_0)}{\mu N}} \tag{7.29}$$

Using the observation that in the limit of infinite time, no infected population can persist, we can choose N to be the total population

$$S_0 + I_0 + R_0 = N = S_\infty + R_\infty \tag{7.30}$$

and so obtain a transcendental equation for the recovered population at long times.

$$R_\infty = N - S_0 e^{-\frac{(R_\infty - R_0)}{\mu N}} \tag{7.31}$$

Unlike the SZR model, here we see that no matter how virulent the disease is, the epidemic will be self-limiting, and there will always have some susceptibles left at the end of the outbreak. This is a sharp qualitative difference between zombies and more traditional SIR models, arising from the fact that the "recovery" of zombies is itself dependent on the presence of susceptibles.

To visually compare the difference, in Figure 7.1 we have shown deterministic trajectories for both SIR and SZR for selected parameter values.

7.3   Stochastic simulation

While most previous studies modeling zombie population dynamics have been deterministic, things get more interesting when we try to model discrete populations.
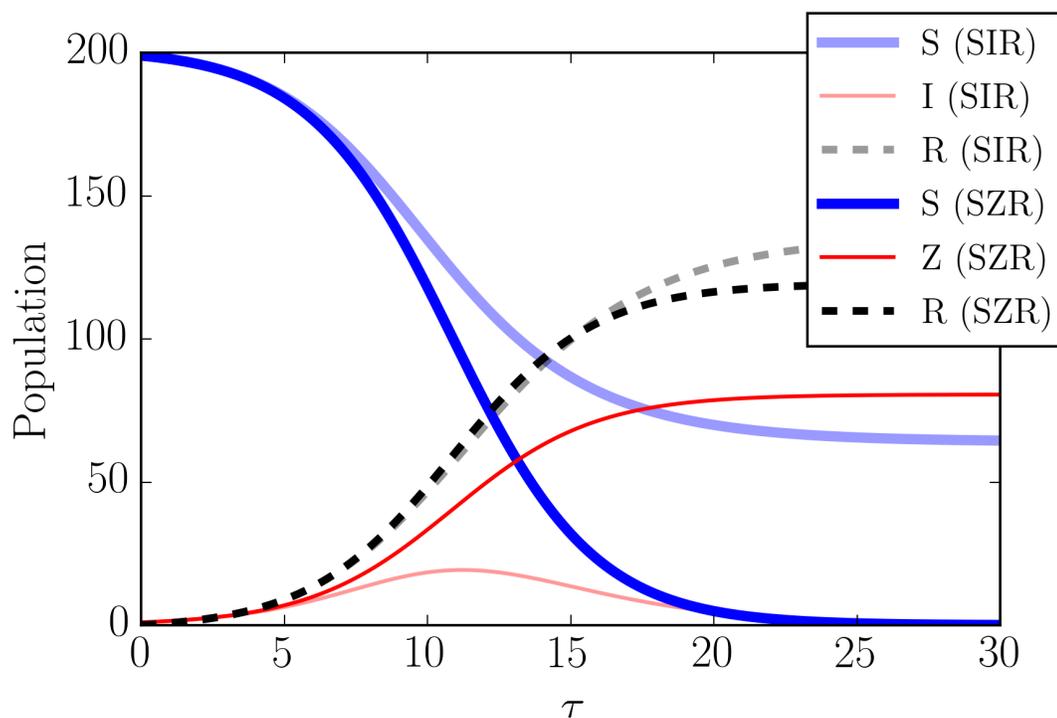
Figure 7.1: **Deterministic trajectories for the** SIR **and** SZR **models** with an initial population of 200 people, 199 uninfected and 1 infected. The (susceptible, infected, removed) population is shown in (thick blue, thin red, dashed black). The SZR results are darker lines while the SIR results are lighter lines. For both models $\tau = t\beta N$ where N was taken to be the total population. For the SZR model α was chosen to be 0.6, while for the SIR model μ was chosen to be 0.6 to show similar dynamics. Notice that in this case, in SZR the human population disappears and only zombies remain in the end, while the SIR model is self-limiting, and only a fraction of the population ever becomes infected.

By treating the number of zombies and humans as continuous variables in the last section, we are ignoring the random fluctuations that arise in small populations: even a ferociously virulent zombie infestation might fortuitously be killed early on by happy accident. Similar problems arise in chemical reactions: reactions involving two types of proteins in a cell can be described by chemical reaction kinetics evolving their concentrations (like our SZR equations 7.4), but if the number of
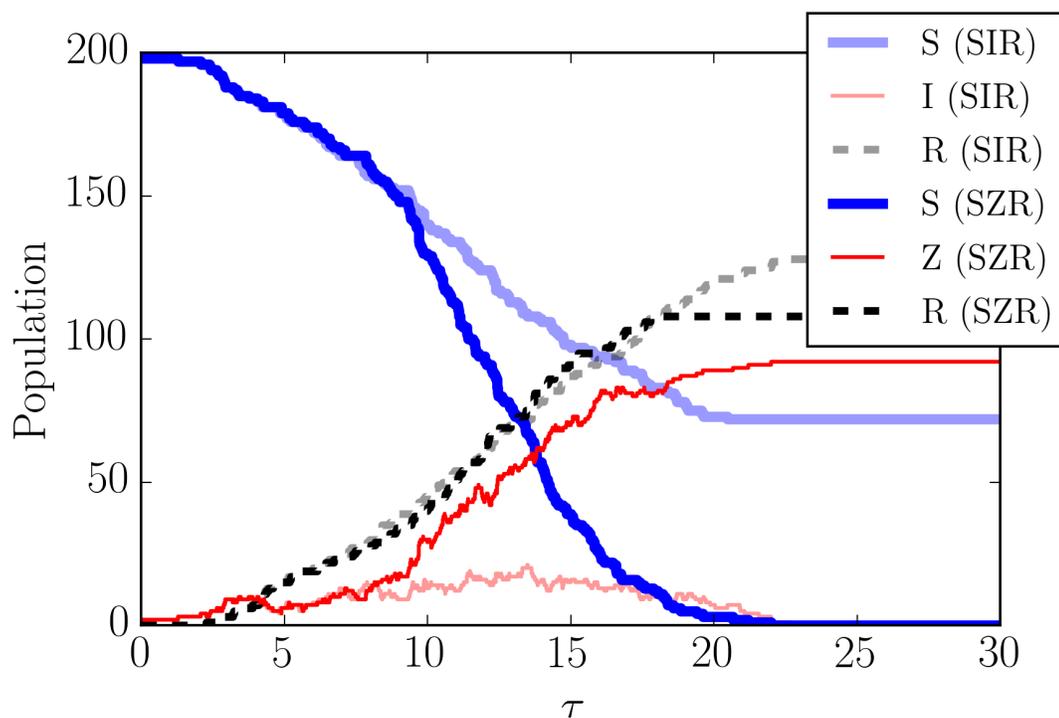
Figure 7.2: **Example Gillespie dynamics for the SIR and SZR models** with the same parameter settings as Figure 7.1. The (susceptible, infected, removed) population is shown in (thick blue, thin red, dashed black). The SZR results are dark lines while the SIR results are lighter lines. The two simulations were run with the same seed so as to match their dynamics at early times.

such proteins is small, accurate predictions must simulate the individual binary reactions (each zombie battling each human). Interpreting our SZR transitions as reaction rates, gives us a system akin to a chemical reaction with two possible transitions:

$$(S, Z) \xrightarrow{\beta SZ} (Z, Z)$$

$$(S, Z) \xrightarrow{\kappa SZ} (S, R)$$
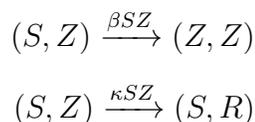
When a human and zombie are in contact, the probability of a bite in a small period of time is given by the bite rate and the size of the populations of the two

species ($\beta SZ\,dt$), and similarly for the probability of a kill. In order to efficiently simulate this dynamics, we use the Gillespie algorithm [89], which efficiently uses the computer to sequentially calculate the result of each one-on-one battle.

The stochasticity gives more character to the simulation. The fully connected continuous dynamics modeled by the differential equation is straightforward: either the humans win and kill all of the zombies, or the zombies win and bite all of the humans. While the continuous approximation may be appropriate at intermediate stages of the infection where the total population is large and there are a non-trivial number of infected individuals, we will eventually be interested in simulating an actual outbreak on an inhomogeneous population lattice, where every new site will start with a single infected individual. But even though we may be interested in modeling the outbreak case ($\alpha < 1$), we would like to allow the possibility that the humans manage to defeat the outbreak before it really takes off. The stochastic Gillespie dynamics allows for this possibility.

In Figure 7.2 we have shown an example of a single stochastic simulation using the same parameter settings as those used in Figure 7.1. The stochastic trajectory overall tracks the analytic result, but at points in the simulation there may be more or fewer zombies than anticipated if the dice fall that way.

Another implication of stochastic dynamics is that it is not always guaranteed that a supercritical ($\alpha < 1$) outbreak will take over the entire susceptible population. For the parameter settings used in Figure 7.1 and 7.2, namely $\alpha = 0.6$ with a population of 200 and one infected individual to start, the zombies win only 40% of the time. Additionally, the number of zombies we end with is not fixed, as shown in Figure 7.3.
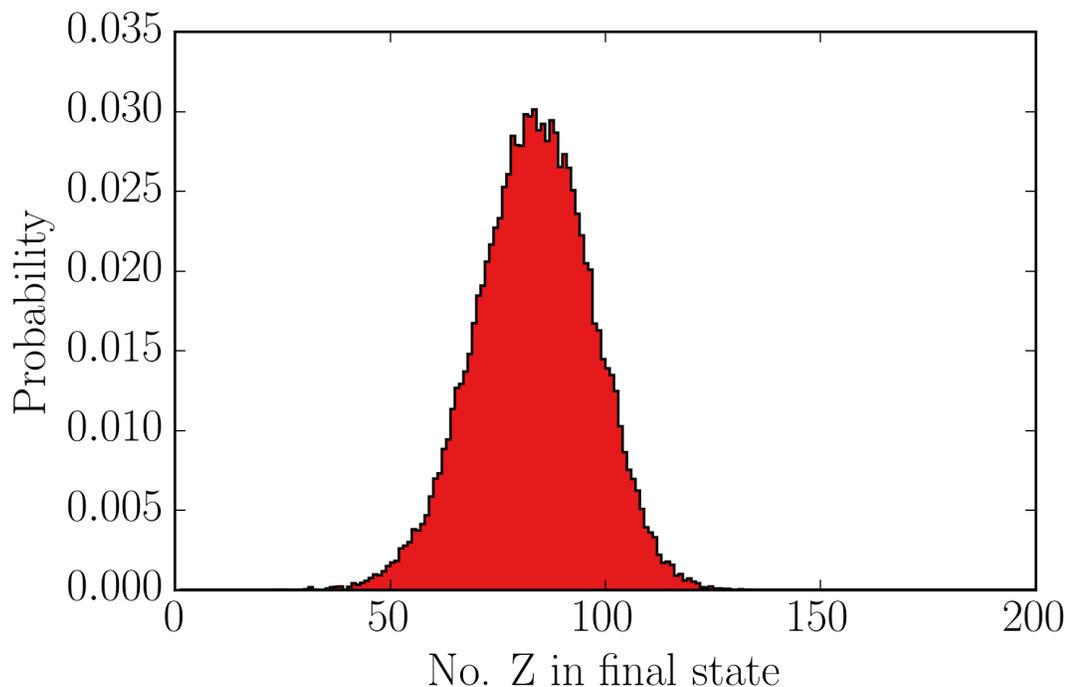
Figure 7.3: **Distribution of final states for Gillespie dynamics**. The distribution for final zombies over 100,000 stochastic trajectories with the same parameters as Figure 7.2. Not pictured are the 60% of runs that end with no zombies in the final state. Compare these to the analytical result, in which the final population of zombies would be 81 with no possibility of surviving humans.

In fact, we can solve exactly for the probability P_{ext} that an $\alpha < 1$ simulation will go extinct in the limit of large populations, using an argument drawn from the theory of *branching processes* [262]. At the very beginning of the simulation, there is only one zombie, who will be killed with probability $\kappa/(\beta + \kappa)$. If the first zombie is killed before it bites anyone, we guarantee extinction. Otherwise, the zombie will bite another human, at which point there will be two independent zombie lines that need to be extinguished, which will occur with probability P_{ext}^2.
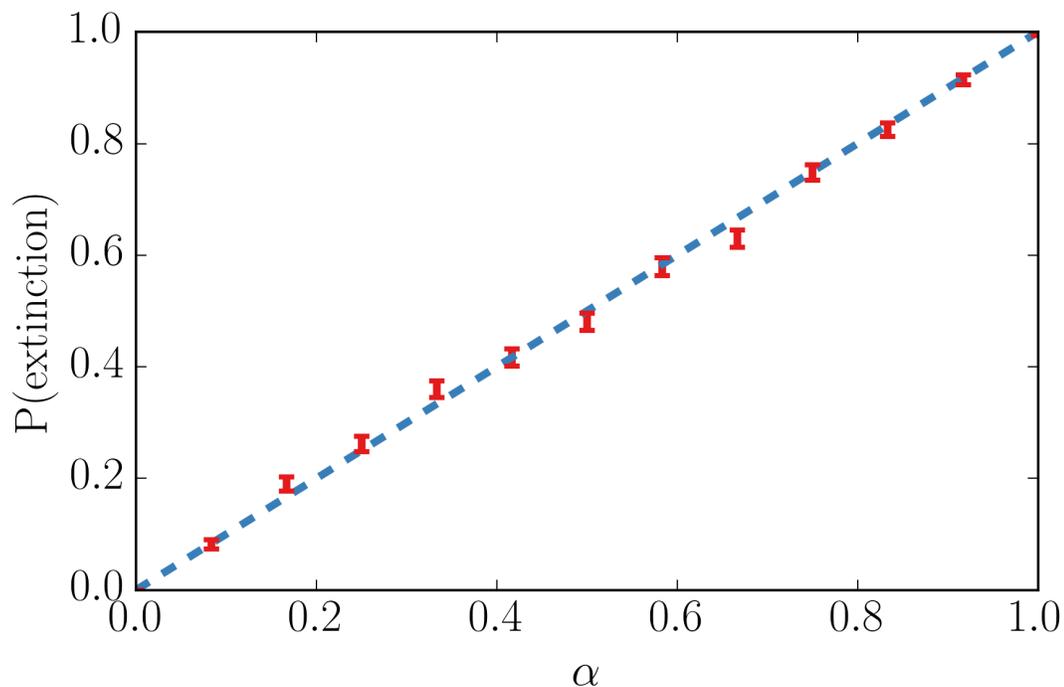
Figure 7.4: **Extinction rate of infections**. The observed fraction of simulations that end in an extinction for the zombie outbreak, for 1,000 runs of 10^4 individuals at various values of α (eqn. 7.33). The observed extinction probabilities agree with the expectation that they should go as α, here shown as the dashed line. This is the same behavior as the SIR model.

This allows us to solve:

$$P_{\text{ext}} = \frac{\kappa}{\beta + \kappa} 1 + \frac{\beta}{\beta + \kappa} P_{\text{ext}}^2 \tag{7.32}$$

$$P_{\text{ext}} = \frac{\kappa}{\beta} = \alpha . \tag{7.33}$$

The probability of extinction is just given by our dimensionless inverse virulence α. In Figure 7.4 we have shown the observed extinction probabilities for 1,000 Gillespie runs of a population of 10^4 individuals at various values of α, and overlaid our expected dependence of α.

This same extinction probability ($P_{\text{ext}} = \mu = R_0^{-1}$) is observed for the SIR model [129]. This is not a coincidence. In precisely the limit that is important for

studying the probability of an extinction event, namely at early times with very large populations, the SZR model and SIR are effectively the same, since the population of susceptibles (S) is nearly constant. Writing S as $S_0 - \delta S$, we have:

$$\frac{dZ}{d\tau} = (1 - \alpha)\frac{S_0 Z}{N} - (1 - \alpha)\frac{(\delta S)Z}{N} \tag{7.34}$$

$$\frac{dI}{d\tau} = \left(1 - \frac{\mu N}{S_0}\right)\frac{S_0 I}{N} - (\mu N + \delta S)\frac{I}{N} \ . \tag{7.35}$$

Here as $\delta S \to 0$, the two models are the same with $\alpha = \mu N/S_0$, another indication that the density dependent SIR model's virulence is dependent on population size.

To get a better sense of the effect of the stochasticity, we can look at the mean fractional population in each state for various settings of α and choices for initial population size. The results are shown in Figure 7.5.

Plotted are the fractional populations in the final state left for both the SZR model (top row) and SIR model (bottom row) for different parameter combinations of α and the initial population. In all cases, the N parameter was chosen to be 100. For each box, 1,000 independently seeded stochastic trajectories were calculated until completion. Looking at the SZR results in the top row, we can see that the dynamics is fairly independent of population size once the population size gets above around 100 individuals. The population dependence for lower population sizes is an effect of the stochasticity. We can clearly see a transition in the susceptible population near $\alpha = 1$ corresponding to where our continuous dynamics would show a sharp boundary. Here the boundary is blurred, again due to the stochasticity. The final dead zombie population R remains small for all values of α; for extremely virulent zombies $\alpha \ll 1$, very few will be killed by the humans before all of the humans are converted, while in the other extreme few zombies are created so there are few to be killed.
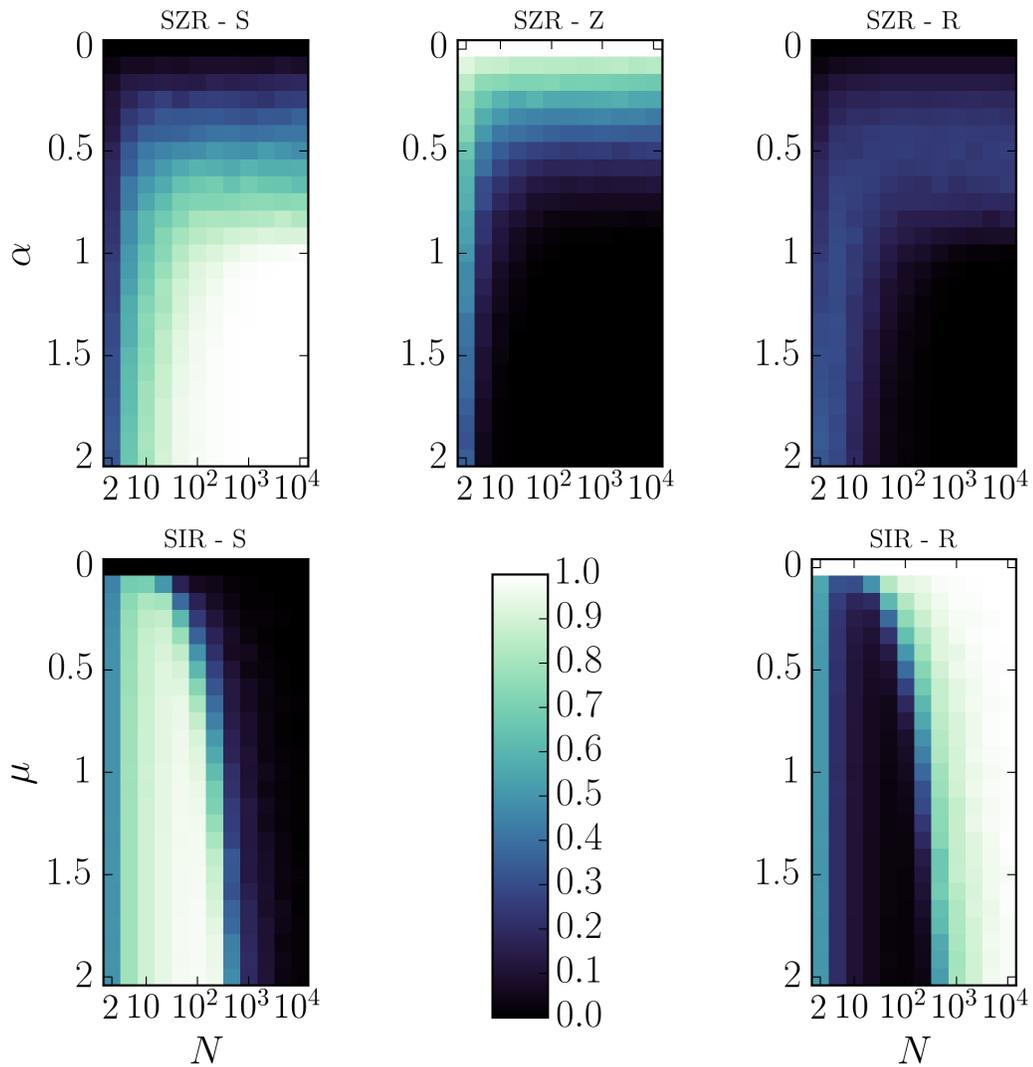
Figure 7.5: **Mean final states as a function of model parameters**. One thousand different simulations are run for each cell. Each simulation starts with a single zombie or infected individual. The runs are run until they naturally terminate, either because the susceptible population is deleted, the zombie population is gone, or there are no more infected individuals. Each cell is colored according to the mean fraction of the population occurring in each state. The top row is for SZR simulations and the bottom row is for SIR simulations. In both cases N is chosen to be 100. Here the sharp contrast between density-dependent SZR and SIR is made apparent. Notice that density-dependent SIR is very strongly population dependent.

Contrast these results with the density dependent SIR dynamics shown in the second row. There can be no infected individuals left in the end, so only the fraction of S and R in the final state are shown. The two transitions in SIR couple differently to the population of infected and susceptible. While our nondimensionalized SZR model has $Z' = (1 - \alpha)SZ/N$, our nondimensionlized SIR has $I' = (S/N - \mu)I$. This creates a very strong population dependence. The transition observed in the S population is largely independent of μ, except on the very small end. When we move to inhomogeneous population lattices this means that for the density dependent SIR model, the most important parameter governing whether a particular site has a break-out infection is the population of that site on the lattice.

7.4 Critical Behavior of Lattice Model

Until now, we have considered fully connected, well-mixed populations, where any infected individual can infect any susceptible individual with equal probability. But surely, a zombie in New York cannot bite someone in Los Angeles. Investigation of the spatial spread of infectious diseases is an important application of *network science*; social diseases spread among intimate contacts, Ebola spreads by personal contact in a network of care-givers, influenza can be spread by direct contact, through the air or by hand-to-mouth, hand-to-eye or hand-to-nose contact after exposure to a contaminated surface. For most diseases, 'long bonds' dominate the propagation to distant sites [183]; airplane flights take Ebola to new continents. Zombies do not fly airplanes, so our model is closer in spirit to the spread of certain agricultural infestations, where the disease spreads across a lattice of sites along the two-dimensional surface of the Earth (although not in those cases where

222

pathogens are transported long distances by atmospheric currents).

To begin, we will consider a two-dimensional square lattice, where each site contains a single individual. Each individual is allowed to be in one of three states: $S, Z,$ or R. The infection spreads through nearest neighbor bonds only. That is, a zombie can bite or be killed by any susceptible individuals in each of the four neighboring sites.

To make direct contact with our zombie model, the rate at which an susceptible cell is bitten is given by βZ where Z is the number of zombie neighbors (since S is one), and the rate at which a zombie site is killed is κS where S is the number of susceptible neighbors.

Because all state transitions in the SZR model depend only on Z–S contacts, for computational efficiency, we need only maintain a queue of all Z–S bonds, that is connections along which a human and zombie can interact. At each step of the simulation, one of these Z–S bonds is chosen at random, and with probability $\beta/(\beta+\kappa) = 1/(1+\alpha)$, the human is bitten, marking it as a zombie. We can then query its neighbors, and for all of them that are human, we can add a Z–S link to our queue. With probability $\kappa/(\beta+\kappa) = \alpha/(1+\alpha)$ the zombie is killed, removing any of its links to neighboring humans from the queue. This process matches the stochastic dynamics of our zombie model operating on the lattice.

Simulating zombie outbreaks on fixed lattices, there is qualitatively different behavior for small α and large α. When α is large, the zombies do not spread very far, always being defeated by their neighboring humans. When α is very small, the zombies seem to grow until they infect the entire lattice. This suggests evidence of a phase transition. Technically, the presence of a phase transition would mean that

if we could simulate our model on an infinite lattice, there should be some critical α (α_c), above which any outbreak will necessarily terminate. Below the critical value, there is the possibility (assuming the infection does not die out) of having the infection grow without bound, infecting a finite fraction of individuals in the limit that the lattice size becomes infinite. The SIR model has been demonstrated to undergo such a phase transition, and we expect the zombie model does as well.

The study of *critical phenomena* includes a series of techniques and analyses that enable us to study the properties of phase transitions even on finite lattices. A major theme of critical phase transitions is the importance of *critical points* – where a system is tuned (here by varying α) to a value separating qualitatively different behaviors (here separating low-infectivity transient zombie infestations from a potentially world-spanning epidemic). At critical points, the system can show *scale free behavior*; there is no natural length scale to the dynamics, and various physical parameters will usually be governed by *power laws* (see below).

With α chosen to be precisely at the critical value, we indeed see a giant component with fractal structure (Fig. 7.6). Note that there are holes (surviving pockets of humans) of all sizes in the figure. This reflects the proximity to the threshold: the battle between zombies and humans is so evenly matched, that one gets an *emergent scale invariance* in the survival patterns. This is in keeping with studies of the SIR model, which shows a similar critical behavior and phase transition [94].

Systems near critical points with this kind of scale invariance fall into *universality classes*. Different systems (say, a real disease outbreak and a simple computational model) can in many ways act precisely the same on large scales near their transitions (allowing us to predict behavior without knowing the details
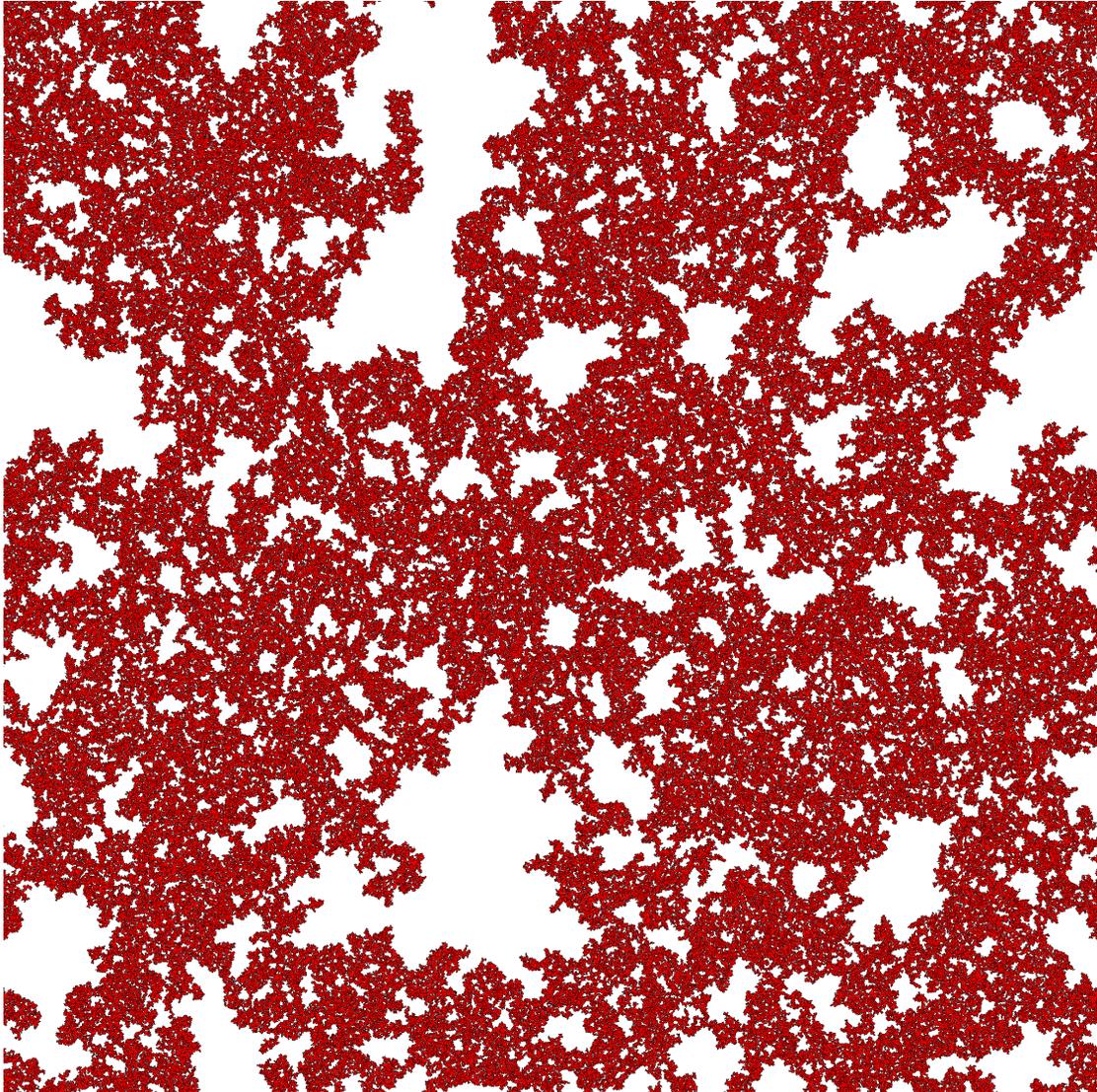
Figure 7.6: **Zombie fractal cluster at the critical point**. Example cluster resulting from the single population per site square lattice zombie model with periodic boundary conditions near the critical point $\alpha_c = 0.43734613(57)$ on a lattice of size 2048×2048. Susceptible, infected (zombie), and removed are shown in white, red, and black respectively.

of zombie-human (anti)social interactions). The SIR model on a two-dimensional lattice with a single person per site falls into the percolation universality class [43], though details of its cluster growth can differ [244]. Given that the SZR model has two second order couplings, it is of interest whether it falls into the same percolation universality class.

To extract the scaling behavior of our zombie infestation, we study the distribution $P(s, \alpha)$, the probability that a single zombie will generate an outbreak of size s at inverse virulence α. (An outbreak will be a fractal cluster in two dimensions, with ragged boundaries if it dies out before reaching the entire world.) At $\alpha = \alpha_c$ where the zombies and humans are equally matched, we have an emergent scale invariance. A large outbreak will appear to almost stop several times – it can be viewed as a sequence of medium-sized outbreaks triggering one another just before they die out. Medium-sized outbreaks are composed of small outbreaks, which are in turn composed of tiny outbreaks. At threshold, each of these scales (large, medium, small) is related to the lower scale (medium, small, tiny) in the same fashion. Let us oversimplify to say that at criticality an outbreak of size Bs is formed by what would have been B smaller outbreaks of size s which happened to trigger one another, and these in turn are formed by what would have been B outbreaks of size s/B. If the probabilities and form of this mutual triggering is the same at each scale, then it would not surprise us that many properties of the outbreaks would be the same, after rescaling the sizes by a factor of B. In particular, we expect at the critical point to find the probabilities of outbreaks of size s to be related to the probabilities at size s/B by some factor f:

$$P(s, \alpha_c) = f P(s/B, \alpha_c). \tag{7.36}$$

This formula quantifies an *emergent scale invariance* at α_c: the properties of epidemics of size s (here the probability) are rescaled versions of the properties at

a smaller scale s/B. [227] – the system is *self-similar* to itself at different scales. Eqn 7.36 is solved by $P(s, \alpha_c) \propto s^{-\tau}$, with $\tau = \log(1/f)/\log(B)$. The distribution of epidemic infection rates is a power law.

Figure 7.7 shows a thorough test of this dependence for our zombie model, following a procedure akin to that of reference [244]. We simulated a zombie outbreak on a two-dimensional lattice with periodic boundary conditions starting with a single zombie. With the outbreak sizes following a power law distribution, the probability that a site belongs to a cluster of size n_s is $P_s = sn_s$, so that at the critical point $P_s \sim s^{1-\tau}$. Integrating from s to ∞, the probability that a point belongs to a cluster of at least s in size ($P_{\geq s}$) should at the critical point itself follow a powerlaw: $P_{\geq s} \sim s^{2-\tau}$. To find our critical point α_c, we ran many simulations until our integrated cluster size distribution followed a power law, using the interpolation methods of reference [244] to get a precise estimate of the critical point.

For zombies on a two dimensional lattice, this critical point occurs at $\alpha_c = 0.43734613(57)$, the resulting integrated cluster size distribution is shown at the top of Fig. 7.7. Percolation theory predicts $\tau = 187/91$ in two dimensions, and we test that prediction in the bottom part of Fig. 7.7. Here, if we were precisely at the critical point and the SZR model is in the percolation universality class, with infinite statistics we would have asymptotically a perfectly straight line. Notice the small vertical scale: our fractional fluctuations are less than 0.1%, while our experimental results vary over several order of magnitude. The clear agreement convincingly shows that the zombie model on the two dimensional lattice is in the percolation university class.

As an additional check, we computed the fractal dimension of our clusters near
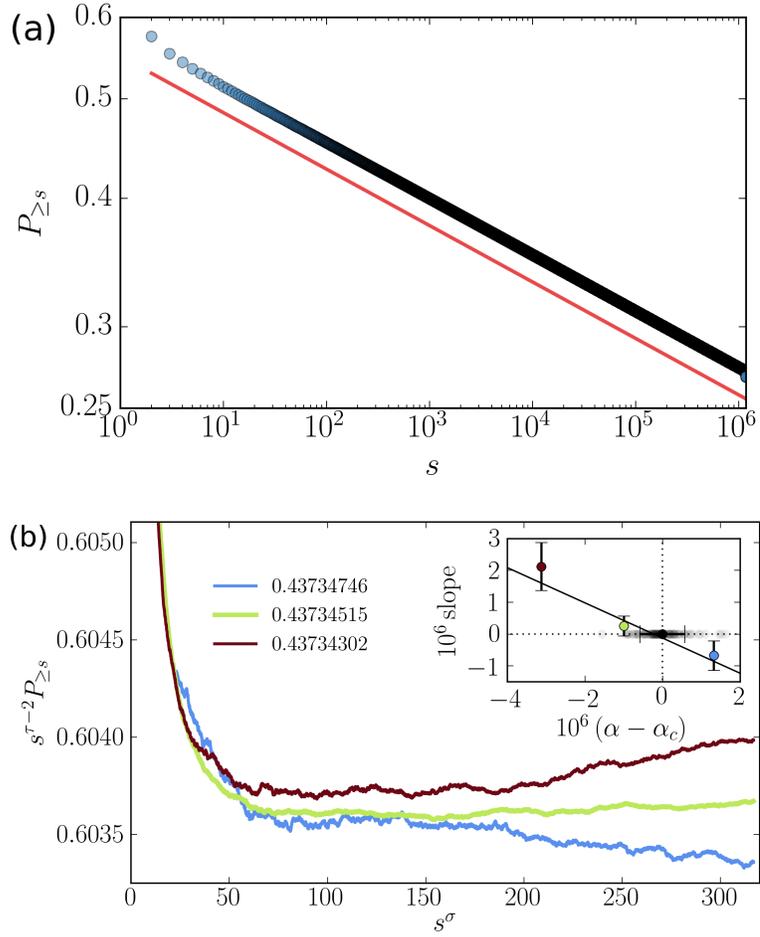
Figure 7.7: **Determination of the critical point using epidemic size distributions**. The cumulative distribution of epidemic sizes for the two dimensional zombie model near the critical virulence. The critical point found was $\alpha_c = 0.43734613(57)$. (a) The probability of a site being in a cluster of at least s in size ($P_{\geq s}$) is shown in blue circles. The fact that it forms a straight line on a log-log plot indicates that $P_{\geq s}$ is a power law, and the slope is $2 - \tau$. For comparison, the red line shows the powerlaw corresponding to the percolation critical exponent: $\tau = 187/91$. (b) Data for three different values of α near α_c, each times $s^{\tau-2}$ using the exponent from percolation theory that should make the critical point into a flat line. We plot these against s^σ, the size taken to the power law $\sigma = 36/91$, as in Ref. [244]. Notice that for $\alpha > \alpha_c$ (when the zombies lose) the large outbreaks are suppressed below the power law, and for $\alpha < \alpha_c$ (when the zombies often win) the largest outbreaks bend up. We follow Ref. [244] in estimating α_c by interpolating these slopes; the inset shows a bootstrap estimate of our error in α_c. Notice that the three curves are for very similar α – leading to excellent precision in identifying the critical point.

228

the critical point using box counting, a distribution for which is shown in Figure 7.8.

We find a fractal dimension $d_f = 1.89(4)$, compared to the exact percolation value

of $d_f = 91/48 = 1.895833$.
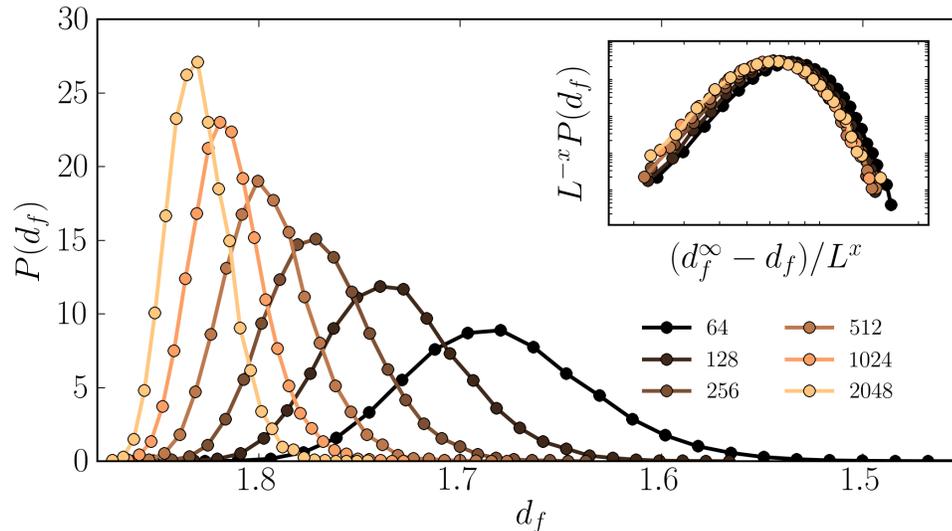
Figure 7.8: **Finite size scaling of the distribution of fractal dimensions**.
Histograms of $P(d_f|L)$, the observed fractal dimensions of the zombie
epidemic clusters as measured by box counting, measured in $L \times L$
systems sizes L varying from 64 to 2048. These extrapolate to give
a measured value consistent with the exact percolation value 91/48,
with an error of ± 0.04). The inset shows a *finite-size scaling collapse*
of these same curves. Using scale invariance arguments similar to those
in the text, one can argue that properties like our measured d_f should
take the form $P(d_f|L) \sim L^x \mathcal{P}((d_f^\infty - d_f)/L^x)$. Hence by multiplying
$P(d_f|L)$ by L^{-x} and plotting it versus $(d_f^\infty - d_f)/L^x$, the scaling theory
predicts the curves at large system sizes should all lie atop one another.
Here we estimate the critical exponent $x \sim 0.30 \pm 0.05$; we also used
the scaling collapse to extrapolate and measure d_f.

Why did we need such an exhaustive test (many decades of scaling, many digits

in our estimate of α_c)? On the one hand, a much smaller simulation could have

told us that there was emergent scale invariance and fractal behavior near the

transition; one or two decades of scaling should be convincing. However, it is often

the case that there are several universality classes with critical exponents close to

one another as in the case of percolation [18]. A small error in α_c can produce large shifts in the resulting fits for τ and d_f, demanding efficient programming and fast computers to achieve a definitive answer.

We conclude that the single person per site zombie infestation, near the critical virulence, will on long length scales develop spatial infestation patterns that are well described by two-dimensional percolation theory.

7.5 US Scale Simulation of Zombie Outbreak

Having explored the general behavior of the zombie model analytically, stochastically and on homogeneous single person lattices, we are prepared to simulate a full scale zombie outbreak.

7.5.1 Inhomogeneous Population Lattice

We will attempt to simulate a zombie outbreak occurring in the United States. This will be similar to our lattice simulation, but with an inhomogeneous population lattice. We based our lattice on code available for creating a "dot map" based on the 2010 US Census data [3]. The 2010 Census released census block level data, detailing the location and population of 11,155,486 different blocks in the United States. To cast these blocks down to a square grid, we assigned each of the 306,675,005 reported individuals a random location inside their corresponding census block, then gridded the population into a 1500×900 grid based on latitude and longitude coordinates. The resulting population lattice can be seen in

---

[3]https://github.com/meetar/dotmap

the top half of Figure 7.9. You will see the presence of many empty grids, especially throughout the western United States. This disconnects the east and west coasts in a clearly artificial pattern – our zombies in practice will gradually wander through the empty grid points. To add in lattice connectivity, we did six iterations of binary closing (an image processing technique) on the population lattice and added it to the original. The effect was to add a single person to many vacant sites, taking our total population up to 307,407,336. The resulting population map is shown in the bottom half of Figure 7.9. This grid size corresponds to roughly 3 km square boxes. The most populated grid site is downtown New York City, with 299,616 individuals. The mean population of the occupied grid sites is 420, the median population of an occupied site is 13.

7.5.2 Augmented Model

In order to more 'realistically' simulate a zombie outbreak, we made two additions to our simplified SZR model. The first was to add a latent state E (Exposed). The second was to introduce motion for the zombies. Considered as a system of differential equations, we now have:

$$\dot{S}_i = -\beta S_i Z_i \tag{7.37}$$

$$\dot{E}_i = -\nu E_i \tag{7.38}$$

$$\dot{Z}_i = \nu E_i - \kappa S_i Z_i \tag{7.39}$$

$$\dot{R}_i = \kappa S_i Z_i \tag{7.40}$$

$$\dot{Z}_i = \mu \sum_{\langle j \rangle} Z_j - \mu Z_i \tag{7.41}$$
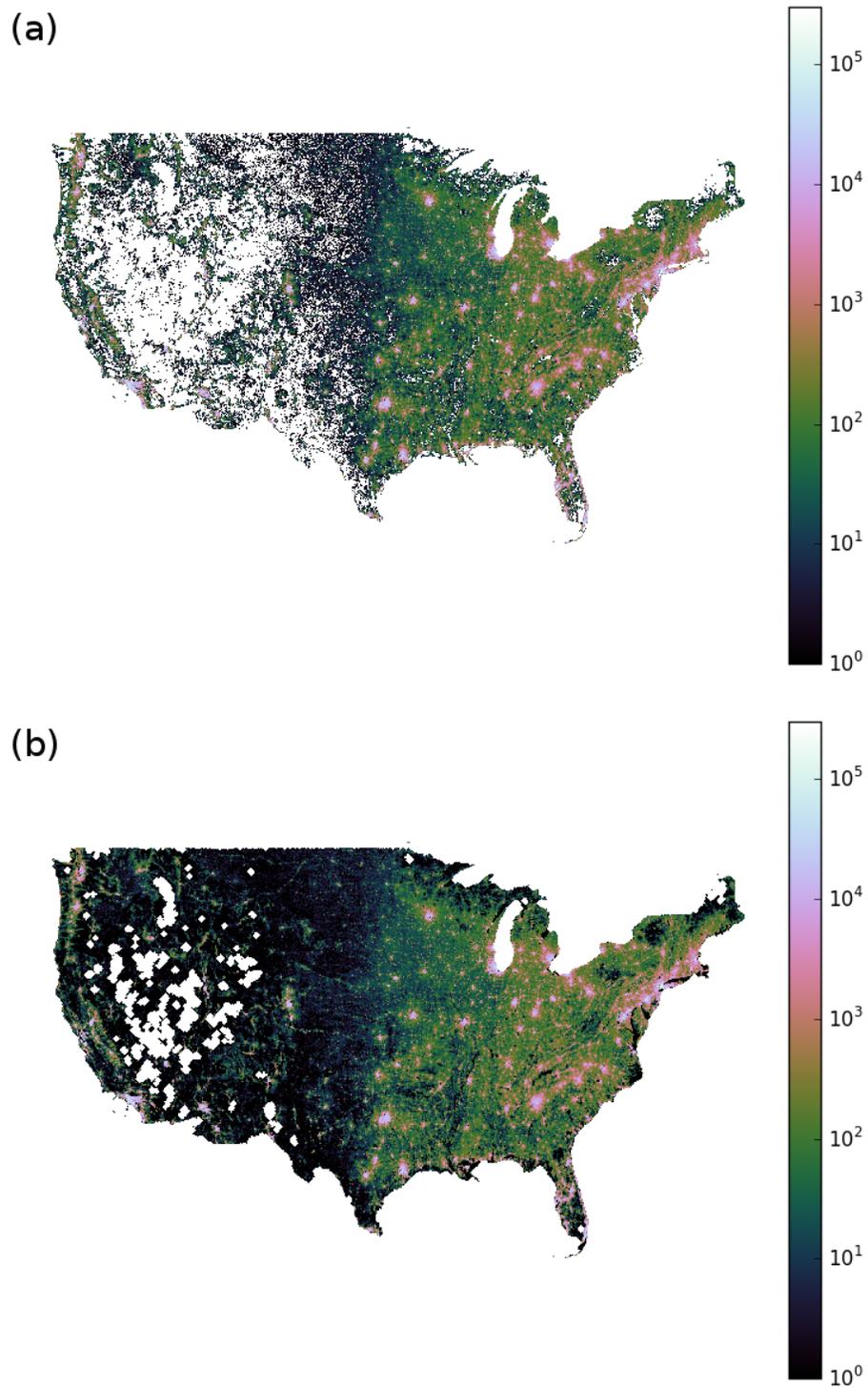
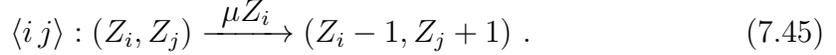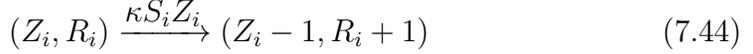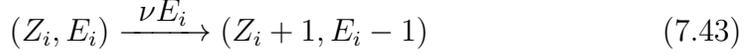Figure 7.9: **Population density of the US**. (a) A 1500×900 grid of the 2010 US Census Data showing raw data. Notice the multitude of squares with no people in them in the Western United States. (b) The resulting map after 6 steps of binary closing added to the original population.

or as a set of reactions:

$$(S_i, E_i) \xrightarrow{\beta S_i Z_i} (S_i - 1, E_i + 1) \tag{7.42}$$

$$(Z_i, E_i) \xrightarrow{\nu E_i} (Z_i + 1, E_i - 1) \tag{7.43}$$

$$(Z_i, R_i) \xrightarrow{\kappa S_i Z_i} (Z_i - 1, R_i + 1) \tag{7.44}$$

$$\langle i\,j \rangle : (Z_i, Z_j) \xrightarrow{\mu Z_i} (Z_i - 1, Z_j + 1) \ . \tag{7.45}$$
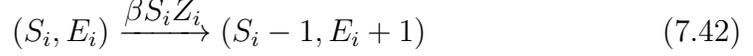
Here i denotes a particular site on our lattice. $\langle j \rangle$ denotes a sum over nearest neighbor sites, $\langle i\,j \rangle$ denotes that i and j are nearest neighbors. In this model, zombies and humans only interact if they are at the same site, but the zombies diffuse on the lattice, being allowed to move to a neighboring site with probability proportional to their population and some diffusion constant (μ). We assume that the humans do not move, not only for computational efficiency, but because, as we will see, the zombie outbreaks tend to happen rather quickly, and we expect large transportation networks to shut down in the first days, pinning most people to their homes. The addition of a latent state coincides with the common depiction that once a human has been bitten, it typically takes some amount of time before they die and reanimate as a zombie. If a human is bitten, they transition to the E state, where at some constant rate (ν) they convert into the zombie state.

To choose our parameters we tried to reflect common depictions of zombies in movies. The work of Witkowski and Blais [263] performed a Bayesian fit of a very similar SZR model to two films, *Night of the Living Dead*, and *Shawn of the Dead*. In both cases, the observed α was very close to 0.8. This means that the zombies in the films are 1.25 times more effective at biting humans than the humans are at killing the zombies. We will adopt this value for our simulation. For our latent state, we adopt a value close to that reported for *Shawn of the Dead*, namely a half-life of 30 minutes. To set our movement parameter, we estimate that zombies

move at around 1 ft/sec. (Note that metric units are uniformly used in science. We use the parochial US units of feet in homage to the popular culture from which we draw our data.) To estimate the rate at which the zombies will transition from one cell to the next, we assume that the zombies behave like a random gas inside the cell, so that the probability that a zombie will cross a cell boundary is roughly $\frac{1}{4}\frac{Z}{L^2}Lv\Delta t$, that is, one-fourth of the zombies within $v\Delta t$ of the edge will move across that edge in a small amount of time. This suggests a value of μ of 0.0914 /hr. This corresponds to an average time between transitions of around 11 hours, which for a zombie stumbling around a 3 km block agrees with our intuitions. Finally, to set a rate for our bite parameter, we similarly assume that the zombies are undergoing random motion inside the cell at 1 ft/sec, and they interact with a human anytime they come within 100 feet. We can then estimate the rate at which humans and zombies will interact as $SZ\frac{Rv\Delta t}{L^2}$, which corresponds to a choice of β of around 3.6×10^{-3} /hr. Another way to make sense of these parameter choices is to ask how many susceptible individuals must be in a cell before a single zombie has a higher rate for biting a human than transitioning to a neighboring cell. For our choice of parameters, this gives

$$N\beta = 4\mu \implies N \sim 102 . \tag{7.46}$$

This corresponds to a low population density of ~ 11 people/km^2, again agreeing with our intuition. All of our parameter choices are summarized in Table 7.1.

7.5.3 Simulation Details

To effectively simulate an outbreak at this scale, we employed the Next Reaction Method of [86]. We maintained a priority queue of all possible reactions, assigning

β	3.6×10^{-3} /hr/person
α	0.8
κ	$\alpha\beta$
η	2 /hr
μ	0.0914 /hr

Table 7.1: **The parameters chosen for our US-scale simulations of a zombie outbreak**. These parameters were chosen to correspond with standard depictions of zombies and simple physical estimations explained in the main text.

each the time at which the reaction would take place, an exponentially distributed random number with scale set by the rate for the reaction. At each time step of the simulation, we popped the next reaction off of the queue, and updated the state of the relevant squares on our grid. Whenever population counts changed, we of course needed to update the times for the reactions that depend on those population counts. This method remained efficient for simulating the entire US. However, at late times a large amount of simulation time was spent simulating the diffusion of the zombies back and forth between highly populated states. We could have achieved additional computational efficiency by adopting the time dependent propensity function approach of Fu et al. [83].

7.5.4 Results

With the simulation in place, we are now in a position to simulate a full scale zombie outbreak. We first consider an outbreak that began with one in every million individuals starting in the Exposed (E) state in the United States. For a single instance the overall populations are shown in Figure 7.10. This looks similar to the analytical outbreaks we saw in Figure 7.1, but with a steeper rate of initial infection and some slight perturbations to the curves. The total population curves

however hide most of the interesting features. In Figure 7.11 we attempt to give a sense of how this outbreak evolves, showing the state of the United States at various times after the outbreak begins.
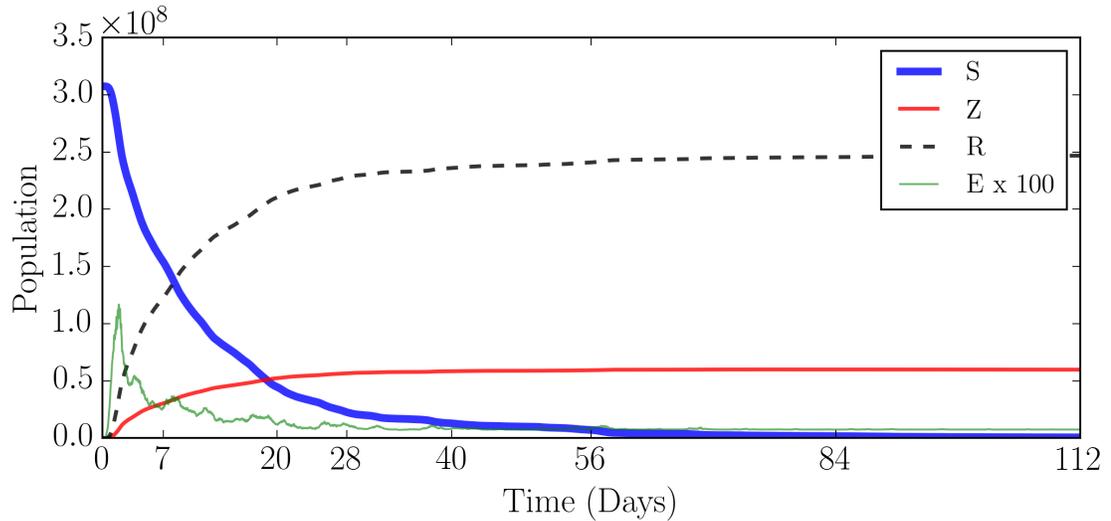
Figure 7.10: **Sample trace of an epidemic across the U.S.** The S (thick solid blue), Z (solid red), R (dashed black), and E (light thin green) populations as a function of time for a full scale zombie outbreak in the continental United States starting with one in every million people infected. The exposed population (E) has been magnified by a factor of 100.

As you can see, for the parameters we chose, most of the United States population has been turned into zombies by the first week, while the geographic map does not necessarily seem all that compelling. In the early stages of the outbreak, while the population is roughly homogeneous, the zombie plague spreads out in roughly uniform circles, where the speed of the infection is tied to the local population density. Infestations on the coasts, with their higher population density, have spread farther than those near the center of the country. After several weeks, the map exhibits stronger anisotropy, as we spread over larger geographical areas and the zombie front is influenced by large inhomogeneities in population density.
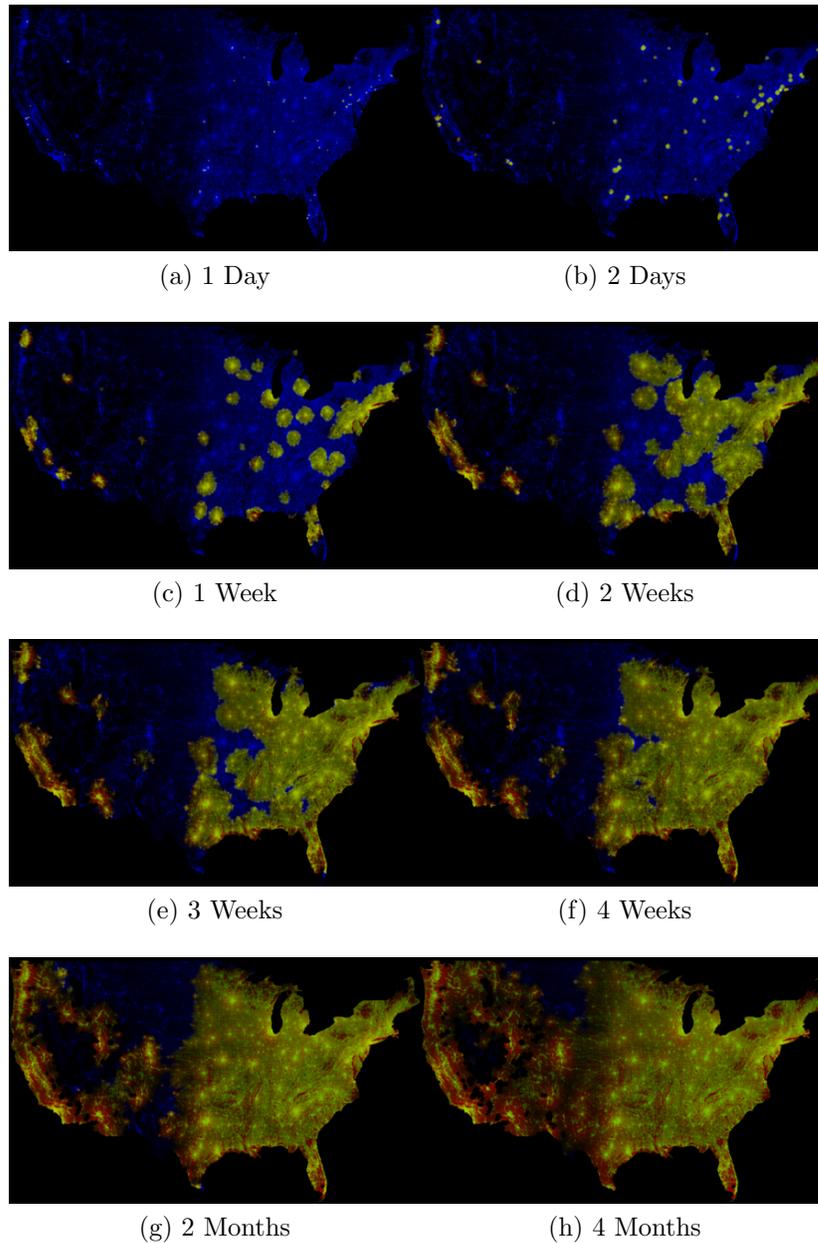
Figure 7.11: **Simulation of a zombie outbreak in the continental United States**. Initially one in every million individuals was infected at random. Results are shown above at (a) one day, (b) two days, (c) one week, (d) two weeks, (e) three weeks, (f) four weeks, and (g) two months after the outbreak begins. Shown here are the population of susceptible individuals (S) in blue, scaled logarithmically, zombies in red and removed in green. All three channels are superimposed. A movie version of this outbreak is available in the supplemental materials online [7].

After four weeks, much of the United States has fallen, but it takes a very long time for the zombies to diffuse and capture the remaining portions of the United States. Even four months in, remote areas of Montana and Nevada remain zombie free.

To investigate the geographical characteristics of the outbreak, we must move beyond a single instance of an outbreak and study how different regions are affected in an ensemble of outbreaks. If it takes a month to develop and distribute an effective vaccine (or an effective strategy for zombie decapitation), what regions should one locate the zombie-fighting headquarters? We ran 7,000 different 28-day zombie outbreaks in the continental United States starting with a single individual. A single instance of one of these outbreaks originating in New York City is shown in Figure 7.12.
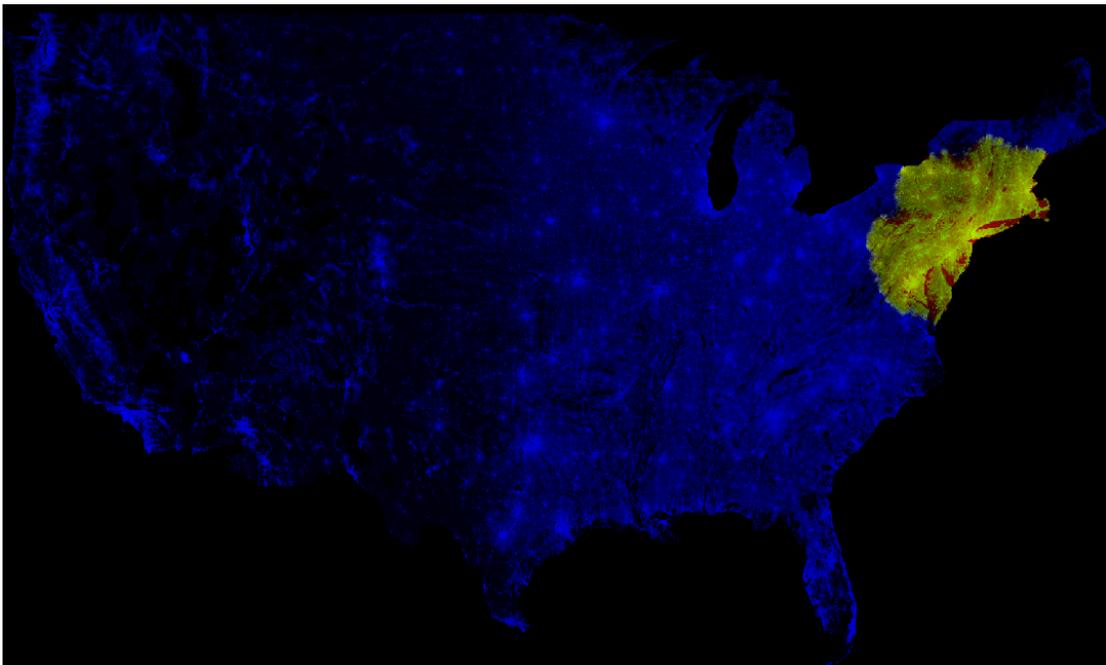
Figure 7.12: **Status of the United States 28 days after an outbreak** that started in New York City. Here blue represents humans, red represents zombies and green represents dead zombies. The three color channels have been laid on top of one another.
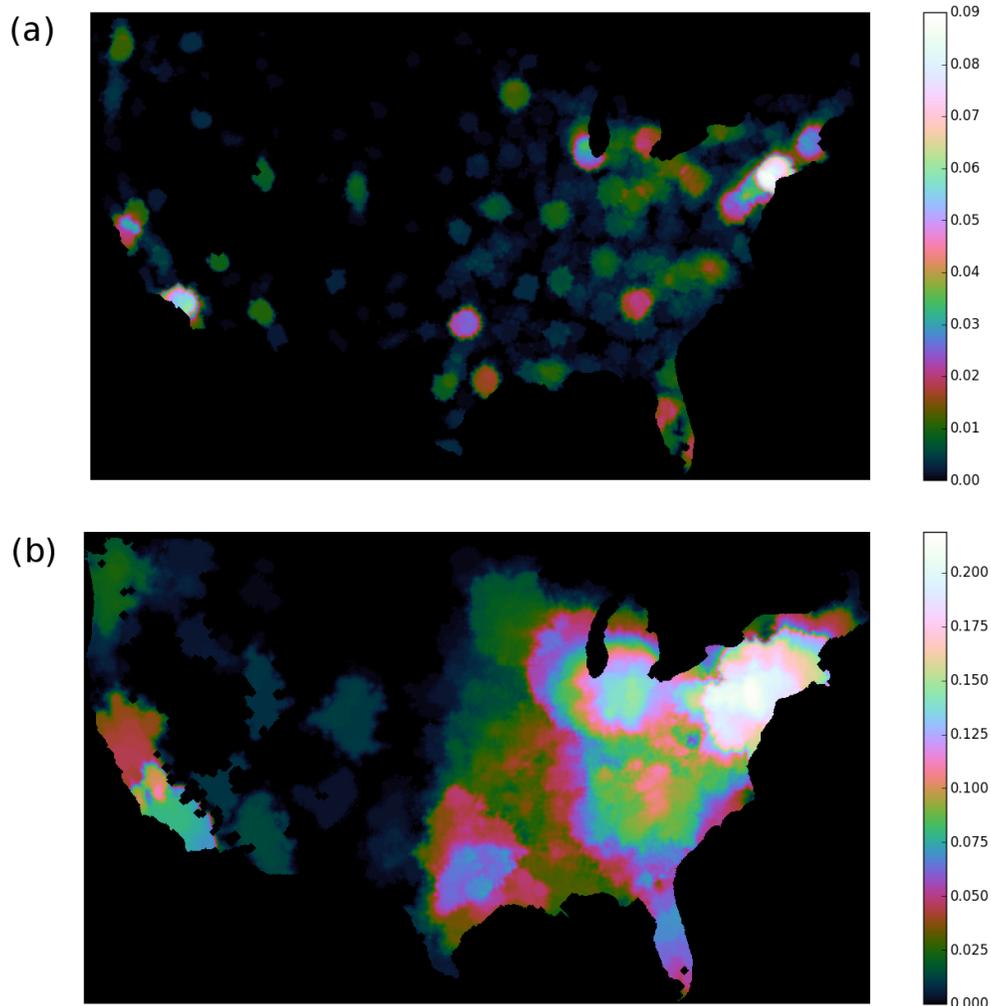
Figure 7.13: **Average infection rate from US scale runs**. In both cases, the plot shows the probability of being infected in that square after an epidemic that originates from a single infected individual chosen at random from the total population. The top figure (a) is the probability of being infected after 7 days, while the bottom plot (b) is after 28 days. In total, this represents 7,000 simulated runs starting from a single individual. The top plot represents the 1,467 outbreaks that lasted at least 7 days, the bottom plot represents 1,458 outbreaks that lasted at least 28 days.

By averaging over all of these runs, we can start to build a zombie danger map, as shown in Figure 7.13. In the top plot, we show the probability that the given cell is overrun by zombies after seven days. Here you can clearly see that there are certain regions – those surrounding populous metropolitan areas – that are at a greater risk. This is partly because those regions have lots of individuals who could potential serve as patient zero, and partly due to the rapid spread of zombies in those areas. In the bottom plot, we plot the probability that the cell is overrun, but at the 28 day mark.

After 28 days, it is not the largest metropolitan areas that suffer the greatest risk, but the regions located between large metropolitan areas. For instance, in California it is the region near Bakersfield in the San Joaquin Valley that is at the greatest risk as this area will be overrun by zombies whether they originate in the San Francisco area or the Los Angeles / San Diego area. The area with the greatest one month zombie risk is north eastern Pennsylvania, itself being susceptible to outbreaks originating in any of the large metropolitan areas on the east coast.

7.6   Conclusion

Zombies offer a fun framework for introducing many modern concepts from epidemiology and critical phenomena. We have described and analyzed various zombie models, from one describing deterministic dynamics in a well-mixed system to a full scale US epidemic. We have given a closed form analytical solution to the well-mixed dynamic differential equation model. We compared the stochastic dynamics to a comparable density-dependent SIR model. We investigated the critical behavior of the single person per site two-dimensional square lattice zombie model

and demonstrated it is in the percolation universality class. We ran full scale simulations of a zombie epidemic, incorporating each human in the continental United States, and discussed the geographical implications for survival.

While this work is predicated on a fictional infestation, one might ask whether there are any phenomena in the real world that behave in a manner similar to our modeled zombie outbreaks. As noted, the SZR model requires that susceptible hosts directly participate in the removal of zombie hosts from the infectious population, leading to runaway outbreaks as susceptible hosts are depleted. One might imagine a similar phenomenon for infectious diseases that require medical intervention to be suppressed; as medical personnel themselves become infected (as has sadly happened to a considerable degree during the recent Ebola outbreak in West Africa), they become less able to stem the spread of infection. (Medical personnel, however, represent only a small fraction of all susceptible hosts, so a refinement to an SZR-type model would be required to account for this.) One might also imagine SZR-like dynamics in the spread of ideas and opinions: a person spreading a controversial opinion in a population, for example, might be able to sway some converts, but is also likely to meet resistance and counter-arguments, which act to reduce infectivity and perhaps ultimately stop the spread.

We hope our systematic treatment of an imaginary disease will provide a useful and inspiring teaser for the exciting fields of statistical mechanics, network science, and epidemiology.

7.7 Acknowledgments

CHAPTER 8

OPENKIM PROCESSING PIPELINE: A VIRTUAL MACHINE CLOUD-BASED AUTOMATIC MATERIALS PROPERTY COMPUTATION ENGINE[1]

8.1 Introduction

We introduce the OpenKIM Pipeline as a computational infrastructure for studying material properties as predicted by interatomic potentials. In this project we develop a set of tools built on top of Open Source Software packages that run simulation codes against interatomic potentials. The results of these couplings ultimately allow researchers to decide which material properties a given model is able to predict accurately (to transfer to new problems) and why.

As computational resources become more powerful, cheaper, and more prevalent, the use of numerical simulations is becoming increasingly popular and important in the understanding and prediction of material properties. Often, in order to reduce computational complexity, interatomic potentials are developed and used as a coarse-grained description of how atomic species interact. However, the predictive power of these simulations hinges delicately on a number of factors: the form of the model, specific model parameters, physical properties under scrutiny, and the simulation method to name a few. For example, when calculating the mechanical properties of crystals, a researcher would want to find a model which is able to accurately reproduce experimental values for the lattice constant, elastic constants, and defect properties. However, when performing simulations of liquids,

---

[1]Matthew Bierbaum, Alexander A. Alemi, James P. Sethna, Ryan S. Elliott, Ellad Tadmor, Trevor Wennblom, & Daniel Karls

the viscosity, surface tension, and pair correlation are more relevant and must be matched to reference data. A researcher must then determine these properties for many models (from literature or calculations) and decide which model best for their application. Additionally, the development of an interatomic model or even a parametrization for a single material is a daunting task, requiring months of work to create and verify, and warranting at least one publication afterward. Many times, the code used to create the results presented in the publication is not archived along with the paper making it difficult for others to acquire, having to correspond directly with the authors. Aside from communication, the code may not even be archived and set aside by the author at all so that the results cannot ever be reproduced. This leaves other researchers to independently implement and test interatomic potentials based on the description found in research papers, adding greatly to the barrier to adoption.

OpenKIM aims to solve these scientific and practical issues of material simulations that use interatomic models by means of an all-encompassing framework [242]. KIM is a secure, stable repository to store simulation codes (Tests) and interatomic potentials (Models), a programming API to mediate their interaction, and a set of resources to study the results of their coupling, including a user interface to download all information associated with the project. Together, these services provide a method to reliably store, share, and study interatomic models. Therefore, along with the actual Tests and Models, a major responsibility of the project is to create the Test Results which come from running each Test with each Model, as applicable. Making these results available for systematic study is central to the goals of the OpenKIM project. The calculation of these Test Results is the responsibility of the OpenKIM processing Pipeline.

8.2 The Pipeline

The processing Pipeline is a set of distributed computers that constitute the computational resources of the OpenKIM project. The main responsibility of the Pipeline is to calculate the results of Test-Model couplings and return the results to the central repository for archival storage. The form of its implementation is guided by three main design goals which are based on the goals of the project and the resources available to complete it:

- **Provenance** - ability to track the origin of and perfectly recreate every Test Result. This includes not only the source for the Test and Model, but the entire software suite including all shared libraries, compiler version, and input data.

- **Flexibility** - run on most hardware in many different physical locations with varying network constraints.

- **Ease of development** - utilize standard software packages and protocols to minimize the development and maintenance time.

To this end, the Pipeline is a set of computers separated into specific roles that run a simple suite of custom software on a standardized set of Virtual Machines (VM). To maintain separation of responsibilities between the Pipeline and the rest of the project, there is a Gateway machine which interacts with the outside world as well as the main KIM resources. The Gateway maintains a job queue, a database of results and KIM objects, stores a Pipeline-local copy of the repository, and runs a website that allows both users and the Web Application to interact with the Pipeline. Additionally, the Gateway hosts a MongoDB [54] database of

Test Results and Reference Data that is made available publicly to query through a query API hosted on the OpenKIM resources [2]. Overall, the Gateway receives job submission requests and handles the results generated from those requests.

Connected to the Gateway is a set of Directors and Workers, the two machines interior to the Pipeline. The Director connects to the Gateway's queue and repository to decide which Test-Model pairs are jobs that need to be completed. The Director's main task is to fill the work queue with Test-Model couplings that need to be run. The Workers consume these jobs from the Gateway's queue and send back the results to the Gateway. The Gateway, as intermediary, sends these results back to the main KIM repository.

A graphical web user interface is hosted on the Gateway machine, allowing external monitoring of the queue, as well as submission and deletion of jobs as shown in Fig. 8.2.

8.3   Software and Virtual machine

The Director and Worker machines are based on a set of virtual machine (VM) images maintained as a part of the KIM project. The VM is built on top of an installation of modern Linux configured with standard scientific computing packages. On top of this base VM, a custom package manager installs other software packages including the OpenKIM API [242], LAMMPS [197], ASE [20], BLAS [144], and more, whose snapshots are stored by the OpenKIM project. Each software package stores a set of dependencies that it requires to work and is kept under version control so that older versions may be used at any time to reproduce old

---

[2]`query.openkim.org`
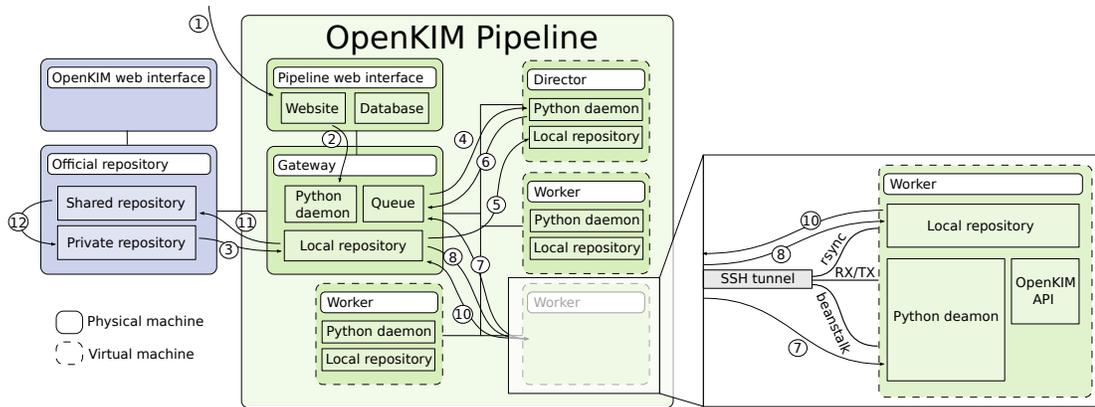
Figure 8.1: **The architecture of the OpenKIM Pipeline**. The Pipeline consists of a central Gateway machine which connects to a set of Directors and Workers, as well as a web interface. When a new KIM object is uploaded to the OpenKIM portal website, the Pipeline is notified in order to update any Test Results that might have changed. (1) First, an API request is sent to the web interface for the Pipeline which is directly (2) passed on to the Gateway machine. The Gateway (3) syncs its local repository with the official repository and (4) passes on the update to a Director through the queue. The Director (5) syncs its local repository as well and finds all relevant jobs that need to be sent to the queue, including those dependencies which have become stale. (6) The Director places all of these jobs onto the queue after which (7) Workers grab individual jobs and start their execution. (8) A Worker grabs only the files needed to complete a job from the Gateway, compiles, templates, and runs the job. (10) Results or errors are sent back to the Gateway. (11) The Gateway passes along these results to the shared section of the official repository. Once they are accepted as official Test Results, (12) they are moved into the official repository.

results. When a Worker receives a request to compute a Test-Model coupling, it first downloads and activates all dependencies associated with the Test and Model, as listed in the configuration files of those KIM items. These configuration files list both static and dynamic dependencies. Static dependencies include other KIM items (drivers, tests, etc) or software packages and dynamic dependencies are limited to the results of other Test-Model couplings and Reference data. In this way, a Test Result can be completely specified by the Test, Model, VM, software packages (with versions), and Test Result and Reference Data ids.

The use of a VM addresses all three of the goals of the KIM project. Since the software on the VM is held fixed, the runtime environment for every Test Result is uniform. When updates to the base machine must be made, the VM is versioned and the old one archived, meaning that old Test Results can always be reproduced. For flexibility, these standard machines make it easy to run on most hardware, given that virtualization technologies are supported. Finally, these same VMs can be used as development machines when creating tests and models, providing a lower startup cost when trying to join the project.

The final layer of software is a set of custom Python scripts which allow these various machines to coordinate with one another to launch jobs and record the output of the Test-Model couplings over a network architecture.

8.4 Network architecture

The virtual machines that comprise Workers and Directors talk to each other as mediated by the Gateway through several different protocols. However, every communication channel is ultimately tunneled through `SSH` [266] via port forwarding. While this decreases the speed of communication due to encryption, we gain several advantages: (1) security through standard open source software `SSH` (2) connections are always possible since the `SSH` connection is outgoing to the Gateway machine (this is not an issue for firewalls, etc) (3) if new ports are introduced, networks and firewalls do not need to be updated.

As of version 1.0 of the Pipeline, three protocols are used to communicate between every VM in the project. Files are synced between local repositories and the Gateway using `rsync` [250], the remote sync utility. The list of jobs, results, and er-

rors are kept and communicated through `Beanstalk` [206], a lightweight networked priority queue. A third transmission / reception line is opened through `ZeroMQ` [12], allowing direct polling requests in the publisher / subscriber paradigm. Each connection is made to the localhost and forwarded to the Gateway box through `SSH`. Therefore, each VM does not directly communicate with each other, rather only through indirect communication with the Gateway. Security issues can be addressed simply through breaking one connection rather than all-to-all.

8.5 Job workflow

Here we track a job from initial submission to results appearing on the frontend website. A shorter version of this account can be tracked in Fig 8.1 by following the descriptions and flow arrows.

A job submission starts at the Pipeline website. An authenticated API is exposed which allows for job submission through a RESTful [273] URL that is formatted as

https://pipeline.openkim.org/jobs/update/KIM_CODE

For this workflow, let's assume that a PUT request is sent indicating that there was an update in the official repository for the KIM object Example__TE_000000000000_000. Immediately, the website sends this update to the Gateway, indicating that an external update was initiated. The Gateway syncs its local repository using the official repository and forwards the update notification to a Director machine via the `Beanstalk` queue.

A Director (randomly picked due to queue assignment) picks up the notification
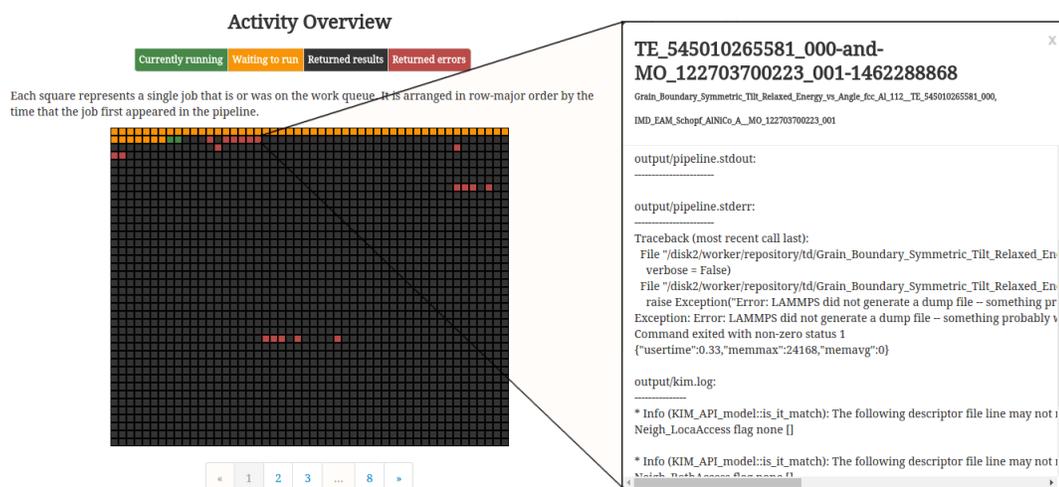
Figure 8.2: **Screen capture of the Pipeline's web interface**. In the left panel we see green, yellow, black, and red squares corresponding to different jobs in the queue which are running, waiting to run, completed, and completed with errors, respectively. In the right panel we show sample output from a Test Model coupling that has produced an error. The `stdout`, `stderr`, `stdin`, and `kim.log` files are provided through this interface to investigate runtime problems with Tests that have been submitted.

and syncs its local repository with the one that the Gateway recently updated. The Director then proceeds to find all relevant Test Results that should be brought up to date due to this change by resolving dependencies. To do so, it first finds all relevant couplings by utilizing the matching logic in the OpenKIM API, asking which of the Test-Model pairs can run together (see further discussion in Sec. 8.6. The Director inserts each of these jobs onto the `Beanstalk` queue run by the Gateway and waits for more updates.

For each job on the queue, a Worker selects the job and begins the execution process. First, it syncs both the Test and Model as well as their dependencies including Test Driver and Model Driver from the Gateway's repository. It then executes the Test, keeping track of errors as returned by exit codes. If there were no errors, the result is transfered to the Gateway's repository via `rsync` and a

250

message placed on the `Beanstalk` queue that a Test Result is waiting. In the case of an error, the `stdout` and `stderr` files are synced to the Gateway and an error message is placed in the queue.

The Gateway then receives the message that a new result or error is present and syncs the result to the shared region of the OpenKIM repository. The OpenKIM editor then decides whether the result is accepted, and if so, moves it into the official repository. A message of an accepted result is placed on the Gateway's queue. Upon reception of the new result message, the result is placed into the database and made available to the public.

There are several subtleties in the story that is presented here. For example, the Gateway receives many types of messages at once. How is the order of execution determined and are all jobs successfully included in the queue? For each message, only the relevant files are synced from the official repository. In this way, the order does not matter, though they will be addressed in the order that they are received. What if a Model is updated while a Test that it couples to is in the middle of being updated? An update to a Test in fact gets versioned at the official repository and the version number is incremented creating a new KIM ID. Therefore, the older version of the Test will be coupled along with the new one to the updated Model. In this way, these types of race conditions should not occur.

We show an instantaneous view of the Pipeline queue as displayed by the Gateway's website [3] in Fig. 8.2. In the left panel of the figure we see green, yellow, black, and red squares each denoting a job that is currently running, waiting to run, completed, and completed with errors. In the right panel we show the error logs associated with a particular Test Model coupling highlighted by the inset. We

---

[3] `http://pipeline.openkim.org/activity` accessed on May 4, 2016

can see by contents of `output/pipeline.stderr` that the LAMMPS simulator did not produce a valid dump file. Along with this graphical view, the entire job history is stored in the MongoDB database for provenance.

8.6   Dependency resolution

While the previous section outlines how a single job is resolved in the Pipeline, we must also consider how jobs are related to one another and how dependencies are automatically resolved. In particular, in this section we describe how the Directors, in concert, traverse the (perhaps circular) dependency tree of jobs.

There are two stages of dependency resolution depending on the type of update that the Director receives. If the update received is a Test, Model, Test Driver, or Model Driver then the Director first determines all possible matches as given by the OpenKIM API. For each of these pairs, the Director verifies that all dynamic dependencies are resolved by checking that all necessary Test Results are present in the database. If these dependencies are not finished then the current pair is postponed and the dependency pair analyzed by the same method. This process proceeds recursively until the current Test-Model pair's dependencies are resolved and the pair is submitted as a job. The jobs that were passed over in this process are handled in the next stage of the dependency resolution. The other stage of dependency resolution is when the Director receives a notification that a Test Result has been completed. In this stage, only one layer of the job network is analyzed as the Director finds Test-Model pairs that require this Test Result as a dynamic dependency and submit them if they have not yet been completed. This step resolves the dependencies that were skipped during the first stage. In this

way, we are able to resolve dependencies using queries local in the dependency network and iteratively calculate all Test Results without supervision.

This dependency search introduces a large computational overhead to job submission. Therefore, as new KIM objects are submitted to the Pipeline, the Directors create a cache of links in the job dependency network which is stored in a local database. Using this cache, we are able to speed up dependency resolution by a factor of 1000.

8.7 Status and conclusions

As of the writing of this manuscript, the Pipeline contains 6287 Tests, 35 Test Drivers, 355 Models, and 37 Model Drivers which are been used by the Pipeline to calculate 47784 Test Results. As a highly parallelized infrastructure we are able to calculate these results in only a few hours of wall time. To demonstrate the amount of information in the Pipeline's database and the ease of access, we show three different Python scripts which directly download and plot various views of the Test Results calculated by the Pipeline. In Fig. 8.3 we plot 3620 lattice constants as measured by energy minimization for all species and models present in OpenKIM. When plotted by atomic number the trends of the periodic table can clearly be seen. In Fig. 8.4 we show the density of states of an fcc Al crystal as measured for 26 different Models. Finally, in Fig. 8.5 we show the cohesive energy versus lattice constant curve for these same Models. The code used to create these figures can be found in the listings following the figures.

Overall, we have developed a distributed infrastructure to support the calculation of Test Results for the OpenKIM project. To do so, we have created a
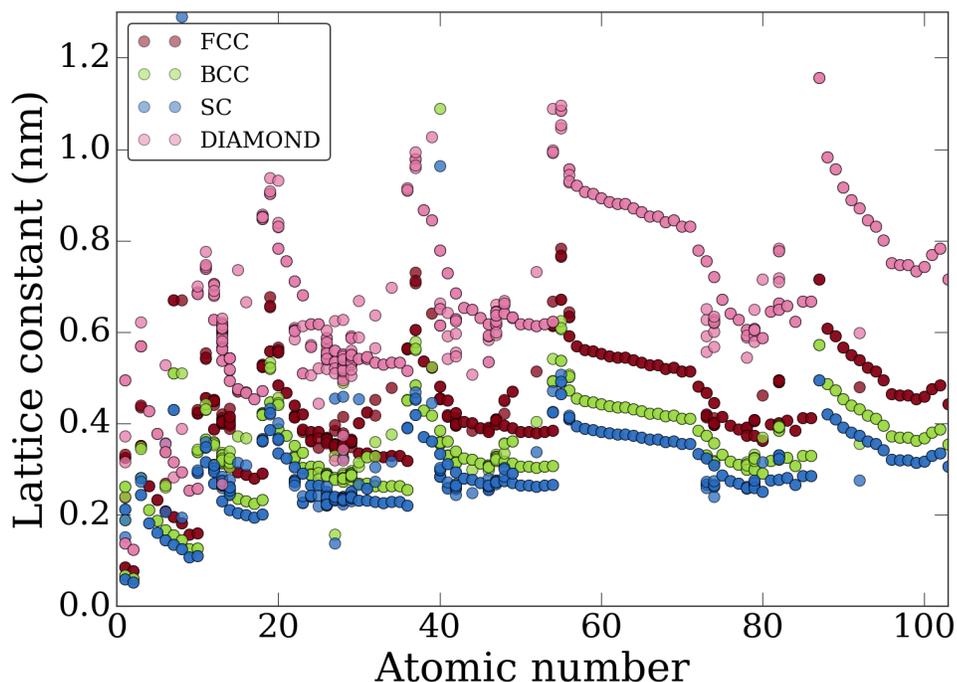
Figure 8.3: **Example OpenKIM figure – lattice constants**. Here we query the OpenKIM public MongoDB database to retrieve all lattice constant results associated with the Test Driver prefixed by LatticeConstantCubicEnergy. We plot the lattice constants of both the stable as well as the unstable cubic structures for all potentials available in the repository. Trends in size with atomic number are clearly seen in all lattice types. The code that produced this graphic is found in Listing 8.1.

networked queueing and communication system built on standard software packages and network protocols on top of a modern Linux Virtual Machine. This network consists of a interface Gateway as well as a set of Directors and Workers, which calculate jobs to complete and as well as the result of these jobs, respectively. The design choices we have made support the project wide goals of provenance, flexibility, and ease of development.

```python
import pylab
import requests
from json import dumps

from ase.data import atomic_numbers, chemical_symbols

results = requests.post(
    "https://query.openkim.org/api",
    data={
        "project": dumps([
            "a.si-value",
            "species.source-value",
            "short-name.source-value"
        ]),
        "query": dumps({
            "meta.runner.kimcode": {
                "$regex":"^LatticeConstantCubicEnergy"
            }
        }),
        "database": "data"
    }
).json()

lattices = ["fcc", "bcc", "sc", "diamond"]
colors = ["#850113", "#9BDA42", "#2A6DC2", "#E57BAA"]

for color, lattice in zip(colors, lattices):
    for i, symbol in enumerate(chemical_symbols):
        lbl = lattice.upper() if i == 0 else None
        x = atomic_numbers[symbol]
        y = []
        for r in results:
            if r[1][0] == symbol and r[2][0] == lattice:
                y.append(r[0]/1e-9)
        pylab.plot([x]*len(y), y, "o", c=color, alpha=0.5, label=lbl)

pylab.xlim(0, 103)
pylab.ylim(0, 1.3)
pylab.xlabel("Atomic number")
pylab.ylabel("Lattice constant (nm)")
pylab.legend(loc="best", prop={"size": 16})
```
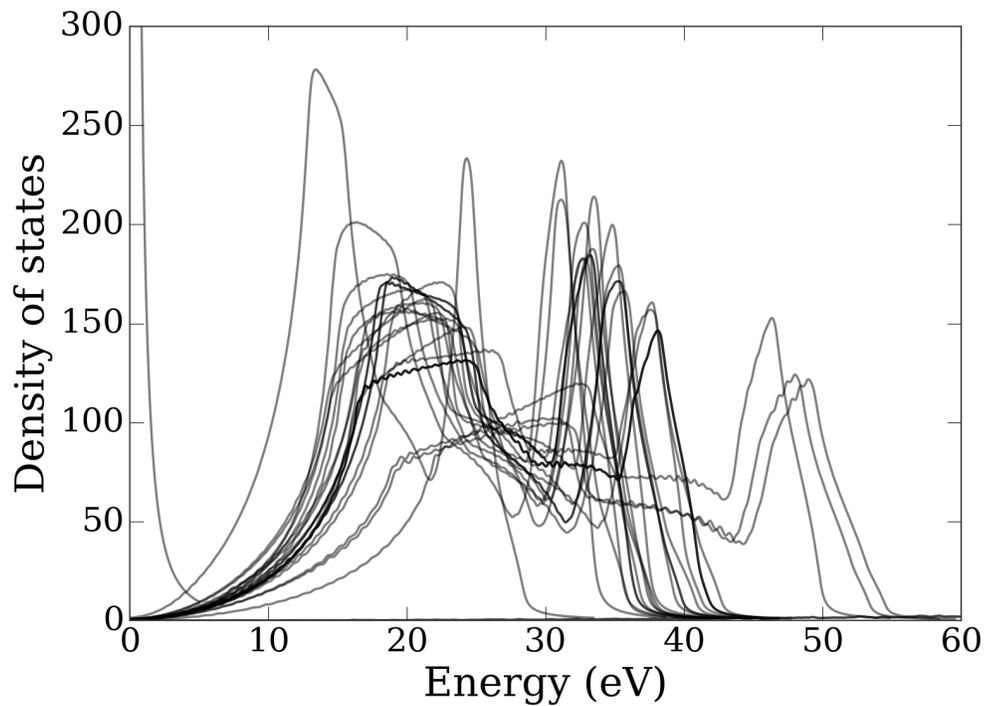
Figure 8.4: **Example OpenKIM figure – density of states**. Here we plot the density of states versus energy as measured by the PhononDispersionCurve Test for an fcc lattice of Aluminum for each Model in the OpenKIM repository. This information was retreived from the Pipeline Mongo database through the query API available at `query.openkim.org` as shown in Listing 8.2.

Listing 8.2: Code listing that produces Fig. 8.4

```python
import pylab
import requests
from json import dumps

results = requests.post(
    "https://query.openkim.org/api",
    data={
        "project": dumps([
            "meta.subject.kimcode",
            "energy.source-value",
            "density-of-states.source-value"
        ]),
        "query": dumps({
            "property-id": {
                "$regex": "/phonon-dispersion-dos-cubic-crystal-npt"
            },
            "meta.runner.kimcode": {
                "$regex": "^PhononDispersionCurve"
            },
            "short-name.source-value": "fcc",
            "species.source-value": "Al"
        }),
        "database": "data"
    }
).json()

for result in results:
    name, energy, dos = result
    pylab.plot(energy, dos, "k-", lw=1.7, label=name, alpha=0.5)

pylab.xlabel("Energy (eV)")
pylab.ylabel("Density of states")
pylab.xlim(0, 60)
pylab.ylim(0, 300)
```
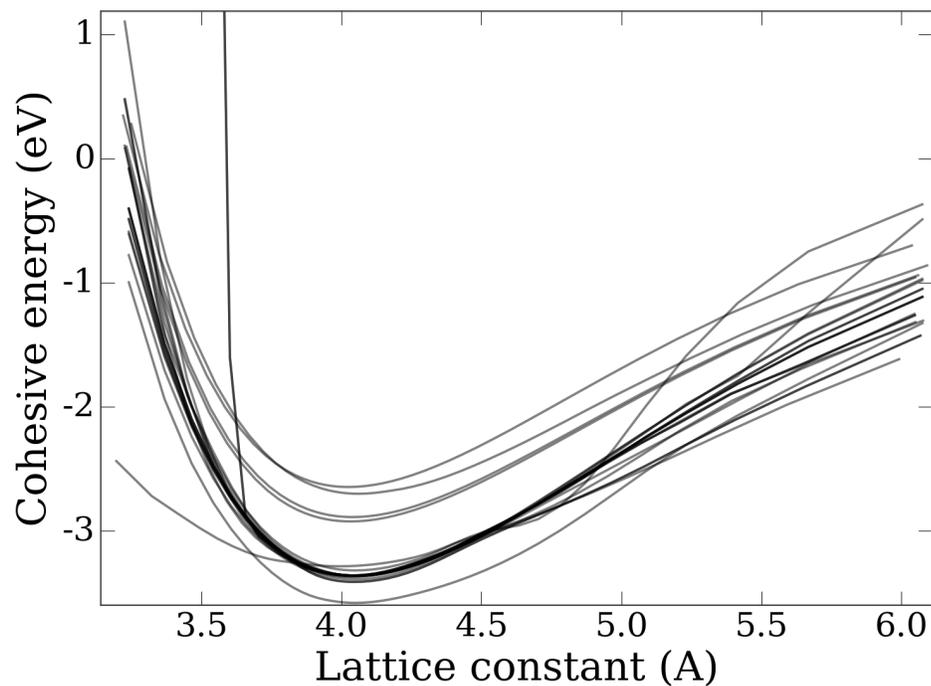
Figure 8.5: **Example OpenKIM figure – cohesive energy**. Here we plot the cohesive energy versus lattice constant for fcc Al as measured for by every Model in the repository in a method similar to that in Fig. 8.4. The code to produce this figure can be found in Listing 8.3.

Listing 8.3: Code listing that produces Fig. 8.5

```python
import pylab
import requests
import numpy as np
from json import dumps

results = requests.post(
    "https://query.openkim.org/api",
    data={
        "project": dumps([
            "meta.subject.kimcode",
            "a.source-value",
            "cohesive-potential-energy.source-value"
        ]),
        "query": dumps({
            "property-id": {
                "$regex": "cohesive-energy-relation-cubic-crystal"
            },
            "meta.runner.kimcode": {
                "$regex": "^CohesiveEnergyVsLatticeConstant"
            },
            "short-name.source-value": "fcc",
            "species.source-value": "Al"
        }),
        "database": "data"
    }
).json()

for result in results:
    name, a, energy = result
    pylab.plot(
        a, -np.array(energy), "k-", lw=1.7,
        label=name, alpha=0.5
    )

pylab.xlabel("Lattice constant (A)")
pylab.ylabel("Cohesive energy (eV)")
pylab.xlim(3.14, 6.11)
pylab.ylim(-3.6, 1.19)
```

BIBLIOGRAPHY

[1] Y.S. Chen, W. Choi, S. Papanikolaou, M. Bierbaum, and J.P. Sethna, *Plasticity Tools*, http://www.lassp.cornell.edu/sethna/Plasticity/Tools/.

[2] Dislocatons which cancel at the macroscale may be geometrically necessary at the mesoscale. The distinction between GND and SSD (statistically stored dislocations, also called geometrically unnecessary dislocations) depends on the coarse-graining length scale. At the mesoscopic scale of dislocation pattern formation, GNDs dominate the dynamics.

[3] In Fourier space, the zero mode of β^p can not be obtained from the GND density field. In the context of the correlation function of β^p, the zero mode is eliminated because we substract fields at different sites before correlating.

[4] Stefano Zapperi and Michael Zaiser (private communication).

[5] In these analyses of TEM micrographs, the authors must use an artificial cut-off to facilitate the analysis. This arbitary scale obscures the scale-free nature behind the emergent dislocation patterns.

[6] 10^5 videos are available on YouTube.com illustrating mosh pits and circle pits. Some notable examples include these.

[7] See Supplemental Material at URL for a movie rendition of the simulated zombie outbreak in Figure 7.11.

[8] A. Acharya. A model of crystal plasticity based on the theory of continuously distributed dislocations. *J. Mech. Phys. Solids*, 49(4):761–784, 2001.

[9] A. Acharya. Driving forces and boundary conditions in continuum dislocation mechanics. *Proc. R. Soc. A*, 459(2034):1343, 2003.

[10] A. Acharya. Constitutive analysis of finite deformation field dislocation mechanics. *J. Mech. Phys. Solids*, 52(2):301–316, 2004.

[11] A. Acharya and A. Roy. Size effects and idealized dislocation microstructure at small scales: predictions of a phenomenological model of mesoscopic field dislocation mechanics: Part i. *J. Mech. Phys. Solids*, 54(8):1687–1710, 2006.

[12] Faruk Akgul. *ZeroMQ*. Packt Publishing, 2013.

[13] BJ Alder, WG Hoover, and DA Young. Studies in molecular dynamics. v. high-density equation of state and entropy for hard disks and spheres. *The Journal of Chemical Physics*, 49(8):3688–3696, 1968.

[14] G. P. Alexander, B. G. Chen, E. A. Matsumoto, and R. D. Kamien. *Phys. Rev. Lett.*, 104:257802, 2010.

[15] M. P. Allen, M. A. Warren, and M. R. Wilson. *Phys. Rev. E*, 57:5585, 1998.

[16] Ahmed M Alsayed, Mohammad F Islam, J Zhang, Peter J Collings, and Arjun G Yodh. Premelting at defects within bulk colloidal crystals. *Science*, 309(5738):1207–1210, 2005.

[17] S. M. Anthony and S. Granick. Image analysis with rapid and accurate two-dimensional gaussian fitting. *Langmuir*, 25:8152–8160, 2009.

[18] N Araújo, P Grassberger, B Kahng, KJ Schrenk, and RM Ziff. Recent advances and open challenges in percolation. *The European Physical Journal Special Topics*, 223(11):2307–2321, 2014.

[19] NW Ashcroft. Liquid metals. *Scientific American*, 221:72–82, 1969.

[20] Sune R Bahn and Karsten W Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Computing in Science & Engineering*, 4(3):56–66, 2002.

[21] B Bakó and I Groma. Computer simulations of a three-dimensional ising ferromagnet with quenched disorder. *Phys. Rev. B*, 60(13):9228–9231, Oct 1999.

[22] B. Bakó, I. Groma, G. Gyorgyi, and G. T. Zimanyi. Dislocation glasses: Aging during relaxation and coarsening. *Phys. Rev. Lett.*, 98(7):075701, 2007.

[23] Botond Bakó and Wolfgang Hoffelner. Cellular dislocation patterning during plastic deformation. *Phys. Rev. B*, 76(21):214108, Dec 2007.

[24] Stanley L Barnes. A technique for maximizing details in numerical weather map analysis. *Journal of Applied Meteorology*, 3(4):396–409, 1964.

[25] AR Bausch, MJ Bowick, A Cacciuto, AD Dinsmore, MF Hsu, DR Nelson, MG Nikolaides, A Travesset, and DA Weitz. Grain boundary scars and spherical crystallography. *Science*, 299(5613):1716–1718, 2003.

[26] Sepideh Bazazi, Karin S Pfennig, Nils Olav Handegard, and Iain D Couzin. Vortex formation and foraging in polyphenic spadefoot toad tadpoles. *Behavioral ecology and sociobiology*, 66(6):879–889, 2012.

[27] Ted Belytschko, Yury Krongauz, Daniel Organ, Mark Fleming, and Petr Krysl. Meshless methods: an overview and recent developments. *Computer methods in applied mechanics and engineering*, 139(1):3–47, 1996.

[28] CH Bennett and BJ Alder. Studies in molecular dynamics. ix. vacancies in

hard sphere crystals. *The Journal of Chemical Physics*, 54(11):4796–4808, 1971.

[29] T. H. Besseling, M. Hermes, A. Kuijk, B. de Nijs, T-S. Deng, M. Dijkstra, A. Imhof, and A. van Blaaderen. Determination of the positions and orientations of concentrated rod-like colloids from 3d microscopy data. *J. Phys. Condens. Matter*, 27:194109, 2015.

[30] Thijs Herman Besseling, Jissy Jose, and A BLAADEREN. Methods to calibrate and scale axial distances in confocal microscopy as a function of refractive index. *Journal of microscopy*, 257(2):142–150, 2015.

[31] E. Betzig et al. Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, 313:1642–1645, 2006.

[32] Dapeng Bi, Jie Zhang, Bulbul Chakraborty, and RP Behringer. Jamming by shear. *Nature*, 480(7377):355–358, 2011.

[33] William Bialek, Andrea Cavagna, Irene Giardina, Thierry Mora, Edmondo Silvestri, Massimiliano Viale, and Aleksandra M Walczak. Statistical mechanics for natural flocks of birds. *Proceedings of the National Academy of Sciences*, 109(13):4786–4791, 2012.

[34] P. Boltenhagen, O. Lavrentovich, and M. Kleman. *J. Phys. II France*, 1:1233, 1991.

[35] P. Boltenhagen, O. Lavrentovich, and M. Kleman. *Phys. Rev. A*, 46:R1743, 1992.

[36] L. Bourdon, M. Kleman, and J. Sommeria. *J. de Phys.*, 43:77, 1982.

[37] J. F. Brady. The rheological behavior of concentrated colloidal dispersions. *J. Chem. Phys.*, 99:567, 1993.

[38] Max Brooks. *The Zombie Survival Guide: Complete protection from the living dead.* Broadway books, 2003.

[39] Max Brooks. *World War Z.* Querido's Uitgeverij BV, Em., 2013.

[40] A. Buchel and J. P. Sethna. *Phys. Rev. Lett.*, 77:1520, 1996.

[41] A. Buchel and J. P. Sethna. *Phys. Rev. E*, 55:7669, 1997.

[42] Wei Cai, Athanasios Arsenlis, Christopher R Weinberger, and Vasily V Bulatov. A non-singular continuum theory of dislocations. *Journal of the Mechanics and Physics of Solids*, 54(3):561–587, 2006.

[43] John L Cardy and Peter Grassberger. Epidemic models and percolation. *Journal of Physics A: Mathematical and General*, 18(6):L267, 1985.

[44] Andrew Cartmel et al. *Mathematical Modelling of Zombies.* University of Ottawa Press, 2014.

[45] P. M. Chaikin and T. C. Lubensky. *Principles of condensed matter physics.* Cambridge University Press, Cambridge, 1995.

[46] P.M. Chaikin and T.C. Lubensky. *Principles of Condensed Matter Physics.* Cambridge University Press, Cambridge, England, 1995.

[47] S. Chatterjee and S. L. Anna. *Soft Matter*, 8:6698, 2012.

[48] S. Chatterjee and S. L. Anna. *Phys. Rev. E*, 85:011701, 2012.

[49] Hanshuang Chen and Zhonghuai Hou. Noise-induced vortex reversal of self-propelled particles. *Physical Review E*, 86(4):041122, 2012.

[50] Ke Chen, Wouter G Ellenbroek, Zexin Zhang, Daniel TN Chen, Peter J Yunker, Silke Henkes, Carolina Brito, Olivier Dauchot, Wim Van Saarloos, Andrea J Liu, et al. Low-frequency vibrations of soft colloidal glasses. *Physical review letters*, 105(2):025501, 2010.

[51] Y S Chen, W Choi, S Papanikolaou, and J P Sethna. (manuscript in preparation).

[52] Yong S. Chen, Woosong Choi, Stefanos Papanikolaou, and James P. Sethna. Bending crystals: Emergence of fractal dislocation structures. *Phys. Rev. Lett.*, 105(10):105501, Sep 2010.

[53] Xiang Cheng, Jonathan H McCoy, Jacob N Israelachvili, and Itai Cohen. Imaging the microscopic structure of shear thinning and thickening colloidal suspensions. *Science*, 333(6047):1276–9, September 2011.

[54] Kristina Chodorow. *MongoDB: the definitive guide.* " O'Reilly Media, Inc.", 2013.

[55] W Choi, Y S Chen, S Papanikolaou, and J P Sethna. to be published in Comput. Sci. Eng. e-print arXiv:1105.5351.

[56] W Choi, Y S Chen, S Papanikolaou, and J P Sethna. (manuscript in preparation).

[57] Fabio Cinti, Tommaso Macrì, Wolfgang Lechner, Guido Pupillo, and Thomas Pohl. Defect-induced supersolidity with soft-core bosons. *Nature communications*, 5, 2014.

[58] N. A. Clark and A. J. Hurd. *J. Physique*, 43:1159, 1982.

[59] N. A. Clark and R. B. Meyer. *Appl. Phys. Lett.*, 22:493, 1973.

[60] P Coles and F Lucchin. *Cosmology: the Origin and Evolution of Cosmic Structures.* Wiley, Chichester, 1995.

[61] Iain D Couzin, Jens Krause, Richard James, Graeme D Ruxton, and Nigel R Franks. Collective memory and spatial sorting in animal groups. *Journal of theoretical biology*, 218(1):1–11, 2002.

[62] J. C. Crocker and D. G. Grier. Methods of digital video microscopy for colloidal studies. *J. Colloid and Interface Science*, 179:298, 1995.

[63] John C Crocker and David G Grier. Methods of digital video microscopy for colloidal studies. *Journal of colloid and interface science*, 179(1):298–310, 1996.

[64] P. A. Cruz, M. F. Tomé, I. W. Stewart, and S. McKee. *J. Non-Newtonian Fluid Mech.*, 165:143, 2010.

[65] András Czirók, Eshel Ben-Jacob, Inon Cohen, and Tamás Vicsek. Formation of complex bacterial colonies via self-generated vortices. *Physical Review E*, 54(2):1791, 1996.

[66] LC DaSilva, L Cândido, L da F Costa, and Osvaldo N Oliveira Jr. Formation energy and interaction of point defects in two-dimensional colloidal crystals. *Physical Review B*, 76(3):035441, 2007.

[67] P. G. de Gennes. *Solid State Commun.*, 10:753, 1972.

[68] P. G. de Gennes and J. Prost. *The Physics of Liquid Crystals.* Claredon Press, Oxford, 1993.

[69] M. Delaye, R. Ribotta, and G. Durand. *Phys. Lett.*, 44A:139, 1973.

[70] TJ Delph. Local stresses and elastic constants at the atomic scale. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 461(2058):1869–1888, 2005.

[71] D M Dimiduk, C Woodward, R LeSar, and M D Uchic. Scale-free intermittent flow in crystal plasticity. *Science*, 312(5777):1188, 2006.

[72] Anthony D Dinsmore, Eric R Weeks, Vikram Prasad, Andrew C Levitt, and David A Weitz. Three-dimensional confocal microscopy of colloids. *Applied Optics*, 40(24):4152–4159, 2001.

[73] AB Drake-Lee. Beyond music: auditory temporary threshold shift in rock musicians after a heavy metal concert. *Journal of the royal society of medicine*, 85(10):617–619, 1992.

[74] G. Durand. *CR Acad. Sci.*, 275B:629, 1972.

[75] Mark S Elliot and Wilson CK Poon. Conventional optical microscopy of colloidal suspensions. *Advances in Colloid and Interface Science*, 92(1):133–194, 2001.

[76] ML Falk and JS Langer. Dynamics of viscoplastic deformation in amorphous solids. *Physical Review E*, 57(6):7192, 1998.

[77] Illés Farkas, Dirk Helbing, and Tamás Vicsek. Social behaviour: Mexican waves in an excitable medium. *Nature*, 419(6903):131–132, 2002.

[78] David R Foss and John F Brady. Brownian dynamics simulation of hard-sphere colloidal dispersions. *Journal of Rheology (1978-present)*, 44(3):629–651, 2000.

[79] F. C. Frank. *Discuss. Faraday Soc.*, 25:19, 1958.

[80] Daan Frenkel and Anthony JC Ladd. Elastic constants of hard-sphere crystals. *Physical review letters*, 59(10):1169, 1987.

[81] G. Friedel and F. Grandjean. *Bull. Soc. Franc. Minér.*, 32:192, 409, 1910.

[82] C Friesen and CV Thompson. Reversible stress relaxation during precoalescence interruptions of volmer-weber thin film growth. *Physical review letters*, 89(12):126103, 2002.

[83] Jin Fu, Sheng Wu, Hong Li, and Linda R Petzold. The time dependent propensity function for acceleration of spatial stochastic simulation of reaction–diffusion systems. *Journal of Computational Physics*, 274:524–549, 2014.

[84] S. Fujii, Y. Ishii, S. Komura, and C. Y. D. Lu. *Europhys. Lett.*, 90:64001, 2010.

[85] S. Fujii, S. Komura, Y. Ishii, and C. Y. D. Lu. *J. Phys. Condens. Matter*, 23:235105, 2011.

[86] Michael A Gibson and Jehoshua Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *The journal of physical chemistry A*, 104(9):1876–1889, 2000.

[87] J Gil Sevillano. Patterns in heavily deformed metals. *Phys. Scr.*, T49B:405, 1993.

[88] J Gil Sevillano, E Bouchaud, and L P Kubin. The fractal nature of gliding dislocation lines. *Scr. Metall. Mater.*, 25:355–360, 1991.

[89] Daniel T Gillespie, Andreas Hellander, and Linda R Petzold. Perspective:

Stochastic algorithms for chemical kinetics. *The Journal of chemical physics*, 138(17):170901, 2013.

[90] M. A. Glaser, R. Malzbender, N. A. Clark, and D. M. Walba. *J. Phys.: Condens. Matter*, 6:A261, 1994.

[91] Shreyas Gokhale, K Hima Nagamanasa, V Santhosh, AK Sood, and Rajesh Ganapathy. Directional grain growth from anisotropic kinetic roughening of grain boundaries in sheared colloidal crystals. *Proceedings of the National Academy of Sciences*, 109(50):20314–20319, 2012.

[92] D Gomez-Garcia, B Devincre, and L P Kubin. Dislocation patterns and the similitude principle: 2.5 D mesoscale simulations. *Phys. Rev. Lett.*, 96(12):125503, 2006.

[93] Robert Gracie, Jay Oswald, and Ted Belytschko. On a new extended finite element method for dislocations: core enrichment and nonlinear formulation. *Journal of the Mechanics and Physics of Solids*, 56(1):200–214, 2008.

[94] Peter Grassberger. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 63(2):157–172, 1983.

[95] I Groma and B Bakó. Dislocation patterning: From micro- to mesoscale description. *Phys. Rev. Lett.*, 84(7):1487, 2000.

[96] John Grue, Philip LF Liu, and Geir K Pedersen. *PIV and water waves*, volume 9. World Scientific, 2004.

[97] T. Gruhn and S. Hess. *Z. Naturforsch.*, 51a:1, 1996.

[98] P. Hähner. On the foundations of stochastic dislocation dynamics. *Appl. Phys. A*, 62(5):473–481, 1996.

[99] Peter Hähner, Karlheinz Bay, and Michael Zaiser. Fractal dislocation patterning during plastic deformation. *Phys. Rev. Lett.*, 81(12):2470, 1998.

[100] Jean-Pierre Hansen and Ian R McDonald. *Theory of simple liquids*. Elsevier, 1990.

[101] N. Hansen, X. Huang, and G. Winther. Effect of grain boundaries and grain orientation on structure and properties. *Metall. Mater. Trans. A*, 42:613–625, 2011.

[102] Robert J Hardy. Formulas for determining local properties in molecular-dynamics simulations: Shock waves. *The Journal of Chemical Physics*, 76(1):622–628, 1982.

[103] Dirk Helbing, Illés Farkas, and Tamas Vicsek. Simulating dynamical features of escape panic. *Nature*, 407(6803):487–490, 2000.

[104] Dirk Helbing, Anders Johansson, and Habib Zein Al-Abideen. Dynamics of crowd disasters: An empirical study. *Physical review E*, 75(4):046109, 2007.

[105] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical review E*, 51(5):4282, 1995.

[106] S. Hell, G. Reiner, C. Cremer, and E. H. K. Stelzer. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *J. Microsc.*, 169:391, 1993.

[107] S Hell, G Reiner, C Cremer, and Ernst HK Stelzer. Aberrations in confocal fluorescence microscopy induced by mismatches in refractive index. *Journal of microscopy*, 169(3):391–405, 1993.

[108] S. W. Hell. Far-field optical nanoscopy. *Science*, 316:1153–1158, 2007.

[109] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Opt. Lett.*, 19:780, 1994.

[110] LF Henderson. The statistics of crowd fluids. *Nature*, 229:381–383, 1971.

[111] S. T. Hess, T. P. K. Girirajan, and M. D. Mason. Ultra-high resolution imaging by fluorescence photoactivation localization microscopy. *Biophys. J.*, 91:4258–4272, 2006.

[112] D. Hilbert and S. Cohn-Vossen. *Geometry and imagination.* AMS Chelsea Publishing, Providence, Rhode Island, 1999.

[113] J P Hirth and J Lothe. *Theory of Dislocatons.* Wiley, New York, 1982.

[114] Ronald Hochreiter and Christoph Waldhauser. Zombie politics: Evolutionary algorithms to counteract the spread of negative opinions. *arXiv preprint arXiv:1401.6420*, 2014.

[115] A. Honglawan, D. A. Beller, M. Cavallaro, R. D. Kamien, K. J. Stebe, and S. Yang. *Adv. Mater.*, 23:5519, 2011.

[116] A. Honglawan, D. A. Beller, M. Cavallaro, R. D. Kamien, K. J. Stebe, and S. Yang. *Proc. Natl. Acad. Sci. USA*, 110:34, 2013.

[117] M. Huang. *Meissner effects, vortex core states, and the vortex glass phase transition.* Ph.D. thesis, Cornell University, 1991. http://www.lassp.cornell.edu/sethna/pubPDF/HuangPhD.pdf.

[118] Pinshane Y Huang, Simon Kurasch, Jonathan S Alden, Ashivni Shekhawat, Alexander A Alemi, Paul L McEuen, James P Sethna, Ute Kaiser, and

David A Muller. Imaging atomic rearrangements in two-dimensional silica glass: Watching silica's dance. *science*, 342(6155):224–227, 2013.

[119] D A Hughes, D C Chrzan, Q Liu, and N Hansen. Scaling of misorientation angle distributions. *Phys. Rev. Lett.*, 81(21):4664, 1998.

[120] D. A. Hughes and N. Hansen. Microstructural evolution in nickel during rolling from intermediate to large strains. *Metall. Trans. A*, 24(9):2022–2037, 1993.

[121] D A Hughes and N Hansen. Graded nanostructures produced by sliding and exhibiting universal behavior. *Phys. Rev. Lett.*, 87(13):135503, 2001.

[122] D A Hughes, Q Liu, D C Chrzan, and N Hansen. Scaling of microstructural parameters: misorientations of deformation induced boundaries. *Acta Mater.*, 45(1):105–112, 1997.

[123] Derek Hull and David J Bacon. *Introduction to dislocations*, volume 257. Pergamon Press Oxford, 1984.

[124] Derek Hull and David J Bacon. *Introduction to dislocations*, volume 37. Elsevier, 2011.

[125] William TM Irvine, Vincenzo Vitelli, and Paul M Chaikin. Pleats in crystals on curved surfaces. *Nature*, 468(7326):947–951, 2010.

[126] Motonari Isobe, Dirk Helbing, and Takashi Nagatani. Experiment, theory, and simulation of the evacuation of a room without visibility. *Physical Review E*, 69(6):066132, 2004.

[127] Timothy Janchar, Chris Samaddar, and David Milzman. The mosh pit expe-

rience: Emergency medical care for concert injuries. *The American journal of emergency medicine*, 18(1):62–63, 2000.

[128] Y Kawasaki and T Takeuchi. Cell structures in copper single-crystals deformed in the [001] and [111] axes. *Scr. Metall.*, 14(2):183–188, 1980.

[129] Matt J Keeling and Pejman Rohani. *Modeling infectious diseases in humans and animals.* Princeton University Press, 2008.

[130] K. J. Kidney, G. McKay, and I. W. Stewart. *Mol. Cryst. Liq. Cryst.*, 438:263, 2005.

[131] Andrew King, G Johnson, D Engelberg, W Ludwig, and J Marrow. Observations of intergranular stress corrosion cracking in a grain-mapped polycrystal. *Science*, 321(5887):382–385, 2008.

[132] M. Kleman and O. D. Lavrentovich. *Soft matter physics: An introduction.* Springer-Verlag, New York, 2003.

[133] M. Kleman, C. Meyer, and Yu. A. Nastishin. *Philos. Mag.*, 86:4439, 2006.

[134] Kröner. Continuum Theory of Defects. In R Balian, M Kleman, and J P Poirier, editors, *Physics of Defects, Les Houches, Session XXXV, 1980*, pages 215–315. North-Holland: Amsterdam, 1981.

[135] D. Kuhlmann-Wilsdorf. Theory of workhardening 1934-1984. *Metall. Mater. Trans. A*, 16:2091–2108, 1985.

[136] D. Kuhlmann-Wilsdorf. Leds: Properties and effects of low energy dislocation structures. *Mater. Sci. Eng.*, 86:53–66, 1987.

[137] D. Kuhlmann-Wilsdorf and N. Hansen. Geometrically necessary, incidental and subgrain boundaries. *Scr. Metall. Mater.*, 25(7):1557–1562, 1991.

[138] A Kurganov, S Noelle, and G Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *SIAM J. Sci. Comput.*, 23(3):707–740, 2001.

[139] Rei Kurita, David B Ruffner, and Eric R Weeks. Measuring the size of individual particles from three-dimensional imaging experiments. *Nature communications*, 3:1127, 2012.

[140] L D Landau and E M Lifshitz. *Theory of Elasticity*. Pergamon Press, 2nd ed. edition, 1970.

[141] LD Landau and EM Lifshitz. Theory of elasticity, 3rd, 1986.

[142] Erik Lange, Jose B Caballero, Antonio M Puertas, and Matthias Fuchs. Comparison of structure and transport properties of concentrated hard and soft sphere fluids. *The Journal of chemical physics*, 130(17):174903, 2009.

[143] R. G. Larson. *The Structure and Rheology of Complex Fluids*. Oxford University Press, 1999.

[144] Chuck L Lawson, Richard J. Hanson, David R Kincaid, and Fred T. Krogh. Basic linear algebra subprograms for fortran usage. *ACM Transactions on Mathematical Software (TOMS)*, 5(3):308–323, 1979.

[145] Wolfgang Lechner and Christoph Dellago. Defect interactions in two-dimensional colloidal crystals: vacancy and interstitial strings. *Soft Matter*, 5(14):2752–2758, 2009.

[146] Wolfgang Lechner and Christoph Dellago. Point defects in two-dimensional colloidal crystals: simulation vs. elasticity theory. *Soft Matter*, 5(3):646–659, 2009.

[147] Naomi E Leonard, Tian Shen, Benjamin Nabet, Luca Scardovi, Iain D Couzin, and Simon A Levin. Decision versus compromise for animal groups in motion. *Proceedings of the National Academy of Sciences*, 109(1):227–232, 2012.

[148] V. a. Levashov, J. R. Morris, and T. Egami. Viscosity, Shear Waves, and Atomic-Level Stress-Stress Correlations. *Phys. Rev. Lett.*, 106(11):115703, March 2011.

[149] Valentin A Levashov, James R Morris, and Takeshi Egami. Viscosity, shear waves, and atomic-level stress-stress correlations. *Physical review letters*, 106(11):115703, 2011.

[150] Lyle E Levine, Bennett C Larson, Wenge Yang, Michael E Kassner, Jonathan Z Tischler, Michael A Delos-Reyes, Richard J Fields, and Wenjun Liu. X-ray microbeam measurements of individual dislocation cell elastic strains in deformed single-crystal copper. *Nature materials*, 5(8):619–622, 2006.

[151] Danilo B. Liarte, Matthew Bierbaum, Muxin Zhang, Brian D. Leahy, Itai Cohen, and James P. Sethna. Flow dynamics of focal conic domains in smectic-A liquid crystals. http://www.lassp.cornell.edu/sethna/LiquidCrystals/Smectics/, 2014.

[152] Danilo B. Liarte, Matthew Bierbaum, Muxin Zhang, Brian D. Leahy, Itai Cohen, and James P. Sethna. 2015.

[153] Megan SC Lim, Megan SC Lim, Margaret E Hellard, Megan SC Lim, Margaret E Hellard, Jane S Hocking, Megan SC Lim, Margaret E Hellard, Jane S Hocking, Campbell K Aitken, et al. A cross-sectional survey of young peo-

ple attending a music festival: associations between drug use and musical preference. *Drug and alcohol review*, 27(4):439–441, 2008.

[154] S Limkumnerd and J P Sethna. Shocks and slip systems: Predictions from a mesoscale theory of continuum dislocation dynamics. *J. Mech. Phys. Solids*, 56(4):1450–1459, 2008.

[155] Surachate Limkumnerd and James P. Sethna. Mesoscale theory of grains and cells: Crystal plasticity and coarsening. *Phys. Rev. Lett.*, 96(9):095503, 2006.

[156] Surachate Limkumnerd and James P. Sethna. Stress-free states of continuum dislocation fields: Rotations, grain boundaries, and the nye dislocation density tensor. *Phys. Rev. B*, 75(22):224121, 2007.

[157] N. Y . C. Lin, J. H. McCoy, X. Cheng, B. Leahy, J. N. Israelachvili, and I. Cohen. A multi-axis confocal rheoscope for studying shear flow of structured fluids. *Rev. Sci. Instrum.*, 85:033905, 2014.

[158] Neil YC Lin, Sushmit Goyal, Xiang Cheng, Roseanna N Zia, Fernando A Escobedo, and Itai Cohen. Far-from-equilibrium sheared colloidal liquids: Disentangling relaxation, advection, and shear-induced diffusion. *Physical Review E*, 88(6):062309, 2013.

[159] Neil YC Lin, Jonathan H McCoy, Xiang Cheng, Brian Leahy, Jacob N Israelachvili, and Itai Cohen. A multi-axis confocal rheoscope for studying shear flow of structured fluids. *Review of Scientific Instruments*, 85(3):033905, 2014.

[160] A B Lopes, F Barlat, J J Gracio, J F Ferreira Duarte, and E F Rauch. Effect of texture and microstructure on strain hardening anisotropy for aluminum

deformed in uniaxial tension and simple shear. *Int. J. Plasticity*, 19:1–22, 2003.

[161] V.S. L'vov. Scale invariant theory of fully developed hydrodynamic turbulence-hamiltonian approach. *Phys. Rep.*, 207(1):1–47, 1991.

[162] David JC MacKay. *Information theory, inference and learning algorithms.* Cambridge university press, 2003.

[163] R Madec, B Devincre, and L P Kubin. Simulation of dislocation patterns in multislip. *Scr. Mater.*, 47(10):689–695, 2002.

[164] R. J. Mandle, E. J. Davis, C. T. Archbold, S. J. Cowling, and J. W. Goodby. *J. Mater. Chem. C*, 2:556, 2014.

[165] J. D. Martin, C. L. Keary, T. A. Thornton, M. P. Novotnak, J. W. Knutson, and J. C. W. Folmer. *Nature Mater.*, 5:271, 2006.

[166] C. Meyer and M. Kleman. *Mol. Cryst. Liq. Cryst.*, 437:111, 2005.

[167] M C Miguel, A Vespignani, S Zapperi, J Weiss, and J R Grasso. Intermittent dislocation flow in viscoplastic deformation. *Nature*, 410(667):71, 2001.

[168] D P Mika and P R Dawson. Polycrystal plasticity modeling of intracrystalline boundary textures. *Acta Mater.*, 47(4):1355–1369, 1999.

[169] Christine Mohr, T Landis, HS Bracha, and P Brugger. Opposite turning behavior in right-handers and non-right-handers suggests a link between handedness and cerebral dopamine asymmetries. *Behavioral neuroscience*, 117(6):1448, 2003.

[170] Bruno Mota. Optimum survival strategies against zombie infestations-a population dynamics approach. *Bulletin of the American Physical Society*, 59, 2014.

[171] C Mottet, G Rossi, F Baletto, and R Ferrando. Single impurity effect on the melting of nanoclusters. *Physical review letters*, 95(3):035501, 2005.

[172] Mehdi Moussaïd, Dirk Helbing, Simon Garnier, Anders Johansson, Maud Combe, and Guy Theraulaz. Experimental study of the behavioural mechanisms underlying self-organization in human crowds. *Proceedings of the Royal Society of London B: Biological Sciences*, pages rspb–2009, 2009.

[173] Mehdi Moussaïd, Dirk Helbing, and Guy Theraulaz. How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888, 2011.

[174] H Mughrabi, T Ungar, W Kienle, and M Wilkens. Long-range internal-stresses and asymmetric x-ray line-broadening in tensile-deformed [001]-orientated copper single-crystals. *Philos. Mag. A*, 53(6):793–813, 1986.

[175] Philip Munz, Ioan Hudea, Joe Imad, and Robert J Smith? When zombies attack!: mathematical modelling of an outbreak of zombie infection. *Infectious Disease Modelling Research Progress*, 4:133–150, 2009.

[176] I. Mura. On dynamic problems of continuous distribution of dislocations. *Int. J. Eng. Sci.*, 1(3):371–381, 1963.

[177] T Mura. *Micromechanics of Defects in Solids*. Kluwer Academic Publishers, Dordrecht, 2nd ed. edition, 1991.

[178] A Ian Murdoch. A critique of atomistic definitions of the stress tensor. *Journal of Elasticity*, 88(2):113–140, 2007.

[179] AI Murdoch and D Bedeaux. Continuum equations of balance via weighted averages of microscopic quantities. *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 445(1923):157–179, 1994.

[180] K. Nakamura and T. Akahane. *Jpn. J. Appl. Phys.*, 29:L1157, 1990.

[181] M. J. Nasse and J. C. Woehl. Realistic modeling of the illumination point spread function in confocal scanning optical microscopy. *J. Opt. Soc. Am. A*, 27:295–302, 2010.

[182] R. M. Neal. Slice sampling. *Ann. Stat.*, 31:705–767, 2003.

[183] Mark EJ Newman, I Jensen, and RM Ziff. Percolation and epidemics in a two-dimensional small world. *Physical Review E*, 65(2):021904, 2002.

[184] Kerstin N Nordstrom, E Verneuil, PE Arratia, Anindita Basu, Zheng Zhang, Arjun G Yodh, Jerry P Gollub, and Douglas J Durian. Microfluidic rheology of soft colloids above and below jamming. *Physical review letters*, 105(17):175701, 2010.

[185] Katja C Nowack, Eric M Spanton, Matthias Baenninger, Markus König, John R Kirtley, Beena Kalisky, Christopher Ames, Philipp Leubner, Christoph Brüne, Hartmut Buhmann, et al. Imaging currents in hgte quantum wells in the quantum spin hall regime. *Nature materials*, 12(9):787–791, 2013.

[186] Felipe Nunez, Cesar Ravello, Hector Urbina, and Tomas Perez-Acle. A rule-based model of a hypothetical zombie outbreak: Insights on the role of emotional factors during behavioral adaptation of an artificial population. *arXiv preprint arXiv:1210.4469*, 2012.

[187] J F Nye. Some geometrical relations in dislocated crystals. *Act. Metall.*, 1(2):153–162, 1953.

[188] U.S. Department of Health, Human Services Centers for Disease Control, and Prevention. Preparedness 101: Zombie pandemic. `http://www.cdc.gov/phpr/zombies/`.

[189] U.S. Department of Health, Human Services Centers for Disease Control, and Prevention. Preparedness 101: Zombie apocalypse. `http://blogs.cdc.gov/publichealthmatters/2011/05/preparedness-101-zombie-apocalypse/`, 2011.

[190] C. W. Oseen. *Trans. Faraday Soc.*, 29:883, 1933.

[191] W Pantleon. On the distribution function of disorientations in dislocation cell structures. *Scr. Metall.*, 35:511–516, 1996.

[192] W Pantleon. On the statistical origin of disorientations in dislocation structures. *Acta Mater.*, 46:451–456, 1998.

[193] R. Pathasarathy. Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods*, 9:724–726, 2012.

[194] P J E Peebles. *Principles of Physical Cosmology.* Princeton University Press, Princeton, 1993.

[195] Yi Peng, Ziren Wang, Ahmed M Alsayed, Arjun G Yodh, and Yilong Han. Melting of colloidal crystal films. *Physical review letters*, 104(20):205703, 2010.

[196] See-Eng Phan, William B Russel, Zhengdong Cheng, Jixiang Zhu, Paul M Chaikin, John H Dunsmuir, and Ronald H Ottewill. Phase transition, equa-

tion of state, and limiting shear viscosities of hard sphere dispersions. *Physical Review E*, 54(6):6633, 1996.

[197] Steve Plimpton. Fast parallel algorithms for short-range molecular dynamics. *Journal of computational physics*, 117(1):1–19, 1995.

[198] WCK Poon. The physics of a model colloid–polymer mixture. *Journal of Physics: Condensed Matter*, 14(33):R859, 2002.

[199] Wilson CK Poon, Eric R Weeks, and C Patrick Royall. On measuring colloidal volume fractions. *Soft Matter*, 8(1):21–30, 2012.

[200] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, Cambridge, third edition, 2007.

[201] Sander Pronk and Daan Frenkel. Large difference in the elastic properties of fcc and hcp hard-sphere crystals. *Physical review letters*, 90(25):255501, 2003.

[202] A Pumir and E D Siggia. Finite-time singularities in the axisymmetrical 3-dimension Euler equations. *Phys. Rev. Lett.*, 68(10):1511–1514, 1992.

[203] Alain Pumir and Eric D. Siggia. Development of singular solutions to the axisymmetric Euler equations. *Phys. Fluids A*, 4(7):1472–1491, 1992.

[204] Charlotte Py, Emmanuel De Langre, Bruno Moulia, and Pascal Hémon. Measurement of wind-induced motion of crop canopies from digital video images. *Agricultural and forest meteorology*, 130(3):223–236, 2005.

[205] S. Ram, E. S. Ward, and R. J. Ober. Beyond rayleigh's criterion: A resolution

measure with application to single-molecule microscopy. *Proc. Natl. Acad. Sci. U.S.A.*, 103:4457–4462, 2006.

[206] Keith Rarick. Beanstalk. <http://kr.github.io/beanstalkd/>, 2007-2016.

[207] Craig W Reynolds. Flocks, herds and schools: A distributed behavioral model. In *ACM SIGGRAPH computer graphics*, volume 21, pages 25–34. ACM, 1987.

[208] J M Rickman and Jorge Viñals. Modelling of dislocation structures in materials. *Philos. Mag. A*, 75:1251–1262, 1997.

[209] Ian Robinson and Ross Harder. Coherent x-ray diffraction imaging of strain at the nanoscale. *Nature materials*, 8(4):291–298, 2009.

[210] D Rosenbaum, PC Zamora, and CF Zukoski. Phase behavior of small attractive colloidal particles. *Physical review letters*, 76(1):150, 1996.

[211] C. S. Rosenblatt, R. Pindak, N. A. Clark, and R. B. Meyer. *J. Physique*, 38:1105, 1977.

[212] A Roy and A Acharya. Finite element approximation of field dislocation mechanics. *J. Mech. Phys. Solids*, 53(1):143–170, 2005.

[213] A. Roy and A. Acharya. Size effects and idealized dislocation microstructure at small scales: predictions of a phenomenological model of mesoscopic field dislocation mechanics: Part ii. *J. Mech. Phys. Solids*, 54(8):1711–1743, 2006.

[214] M. J. Rust, M. Bates, and X. Zhuang. Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nat. Methods*, 3:793, 2006.

[215] Evelyn Sander and Chad M Topaz. The zombie swarm: Epidemics in the presence of social attraction and repulsion. *Mathematical Modelling of Zombies*, page 265, 2014.

[216] MV Sapozhnikov, IS Aranson, and JS Olafsen. Coarsening of granular clusters: Two types of scaling behaviors. *Physical Review E*, 67(1):010302, 2003.

[217] M Saxlová, J. Kratochvil, and J. Zatloukal. The model of formation and disintegration of vein dislocation structure. *Mater. Sci. Eng. A*, 234:205–208, 1997.

[218] Peter Schall, Itai Cohen, David A Weitz, and Frans Spaepen. Visualization of dislocation dynamics in colloidal crystals. *Science*, 305(5692):1944–1948, 2004.

[219] Peter Schall, Itai Cohen, David A Weitz, and Frans Spaepen. Visualizing dislocation nucleation by indenting colloidal crystals. *Nature*, 440(7082):319–323, 2006.

[220] Volker Schaller, Christoph Weber, Christine Semmrich, Erwin Frey, and Andreas R Bausch. Polar patterns of driven filaments. *Nature*, 467(7311):73–77, 2010.

[221] L. Schermelleh, R. Heintzmann, and H. Leonhardt. A guide to super-resolution fluorescence microscopy. *J. Cell Biol.*, 190:165–175, 2010.

[222] Jakob Schiøtz, Francesco D Di Tolla, and Karsten W Jacobsen. Softening of nanocrystalline metals at very small grain sizes. *Nature*, 391(6667):561–563, 1998.

[223] J H Schmitt, J V Fernandes, J J Gracio, and M F Vieira. Plastic behaviour of

copper sheets during sequential tension tests. *Mater. Sci. Eng. A*, 147:143–154, 1991.

[224] C Schwink. Flow-stress dependence on cell geometry in single-crystals. *Scr. Metall. Mater.*, 27(8):963–968, 1992.

[225] J. P. Sethna. *Statistical Mechanics: Entropy, Order Parameters, and Complexity.* Oxford University Press, Oxford, 2006.

[226] J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, 410:242, 2001.

[227] J. P. Sethna, K. A. Dahmen, and C. R. Myers. Crackling noise. *Nature*, 410:242–250, 2001.

[228] J. P. Sethna and M. Kleman. *Phys. Rev. A*, 26:3037, 1982.

[229] James P. Sethna, Valerie R. Coffman, and Eugene Demler. Scaling in plasticity-induced cell-boundary microstructure: Fragmentation and rotational diffusion. *Phys. Rev. B*, 67:184107, 2003.

[230] Peter J Shaw and David J Rawlins. The point-spread function of a confocal microscope: its measurement and use in deconvolution of 3-d data. *Journal of Microscopy*, 163(2):151–165, 1991.

[231] Colin JR Sheppard and Min Gu. Aberration compensation in confocal microscopy. *Applied optics*, 30(25):3563–3568, 1991.

[232] Colin JR Sheppard, Min Gu, Keith Brain, and Hao Zhou. Influence of spherical aberration on axial imaging of confocal reflection microscopy. *Applied optics*, 33(4):616–624, 1994.

[233] S. Shojaei-Zadeh and S. L. Anna. *Langmuir*, 22:9986, 2006.

[234] Leonardo E Silbert, Andrea J Liu, and Sidney R Nagel. Normal modes in model jammed systems in three dimensions. *Physical Review E*, 79(2):021308, 2009.

[235] I. I. Smalyukh, S. V. Shiyanovskii, and O. D. Lavrentovich. Three-dimensional imaging of orientational order by fluorescence confocal polarizing microscopy. *Chem. Phys. Lett.*, 336:88–96, 2001.

[236] C. S. Smith, N. Joseph, B. Rieger, and K. A. Lidke. Fast, single-molecule localization that achieves theoretically minimum uncertainty. *Nat. Methods*, 7:373–375, 2010.

[237] Robert Smith? *Braaaiiinnnsss!: From Academics to Zombies.* University of Ottawa Press, 2011.

[238] Th. Soddemann, G. K. Auernhammer, H. Guo, B. Dünweg, and K. Kremer. *Eur. Phys. J. E*, 13:141, 2004.

[239] Chaoming Song, Shlomo Havlin, and Hemán A Makse. Self-similarity of complex networks. *Nature*, 433:392–395, 2005.

[240] I. W. Stewart, F. M. Leslie, and M. Nakagawa. *Q. Jl. Mech. Appl. Math*, 47:511, 1994.

[241] Alexander Stukowski and Karsten Albe. Extracting dislocations and non-dislocation crystal defects from atomistic simulation data. *Modelling and Simulation in Materials Science and Engineering*, 18(8):085001, 2010.

[242] Ellad B Tadmor, RS Elliott, JP Sethna, RE Miller, and CA Becker. Knowledgebase of interatomic models (kim), 2011.

[243] Ellad B Tadmor and Ronald E Miller. *Modeling materials: continuum, atomistic and multiscale techniques*. Cambridge University Press, 2011.

[244] Tânia Tomé and Robert M Ziff. Critical behavior of the susceptible-infected-recovered model on a square lattice. *Physical Review E*, 82(5):051921, 2010.

[245] John Toner, Yuhai Tu, and Sriram Ramaswamy. Hydrodynamics and phases of flocks. *Annals of Physics*, 318(1):170–244, 2005.

[246] P Török, SJ Hewlett, and P Varga. The role of specimen-induced spherical aberration in confocal microscopy. *Journal of microscopy*, 188(2):158–172, 1997.

[247] Alejandro Torres-Sánchez, Juan M Vanegas, and Marino Arroyo. Examining the mechanical equilibrium of microscopic stresses in molecular simulations. *Physical review letters*, 114(25):258102, 2015.

[248] Mark K Transtrum and James P Sethna. Improvements to the levenberg-marquardt algorithm for nonlinear least-squares minimization. *arXiv preprint arXiv:1201.5885*, 2012.

[249] VERONIQUE Trappe, V Prasad, Luca Cipelletti, PN Segre, and DAVID A Weitz. Jamming phase diagram for attractive particles. *Nature*, 411(6839):772–775, 2001.

[250] Andrew Tridgell and Paul Mackerras. rsync. https://rsync.samba.org/, 1996-2016.

[251] M D Uchic, D M Dimiduk, J N Florando, and W D Nix. Sample dimensions influence strength and crystal plasticity. *Science*, 305(5686):986, 2004.

[252] T Ungár, L S Tóth, J Illy, and I Kovács. Dislocation structure and work hardening in polycrystalline ofhc copper rods deformed by torsion and tension. *Acta Metall.*, 34:1257–1267, 1986.

[253] Alfons Van Blaaderen, Rene Ruel, and Pierre Wiltzius. Template-directed colloidal crystallization. *Nature*, 385(6614):321–324, 1997.

[254] S. Varadhan, A.J. Beaudoin, and C. Fressengeas. Coupling the dynamics of statistically distributed and excess dislocations. In *Proceedings of Science, SMPRI2005*, page 004, 2006.

[255] M Vergassola, B Dubrulle, U Frisch, and A Noullez. Burgers' equation, devil's staircases and the mass distribution for large-scale structures. *Astron. Astrophys.*, 289:325–356, 1994.

[256] Tamás Vicsek, András Czirók, Eshel Ben-Jacob, Inon Cohen, and Ofer Shochet. Novel type of phase transition in a system of self-driven particles. *Physical review letters*, 75(6):1226, 1995.

[257] Tamás Vicsek and Anna Zafeiris. Collective motion. *Physics Reports*, 517(3):71–140, 2012.

[258] T. D. Visser and S. H. Wiersma. Electromagnetic description of image formation in confocal fluorescence microscopy. *J. Opt. Soc. Am. A*, 11:599–607, 1994.

[259] D. Walgraef and E.C. Aifantis. On the formation and stability of dislocation patterns–I: One-dimensional considerations. *Int. J. Eng. Sci.*, 23(12):1351–1358, 1985.

[260] Robert Walser. *Running with the devil: Power, gender, and madness in heavy metal music.* Wesleyan University Press, 1993.

[261] J A Wert, X Huang, G Winther, W Pantleon, and H F Poulsen. Revealing deformation microstructures. *Materials Today*, 10(9):24–32, 2007.

[262] Wikipedia. Branching process — wikipedia, the free encyclopedia. https://en.wikipedia.org/wiki/Branching_process, 2015. [Online; accessed 29-July-2015].

[263] Caitlyn Witkowski and Brian Blais. Bayesian analysis of epidemics-zombies, influenza, and other diseases. *arXiv preprint arXiv:1311.6376*, 2013.

[264] David C. Wright and N. David Mermin. Crystalline liquids: the blue phases. *Rev. Mod. Phys.*, 61:385–432, Apr 1989.

[265] Anand Yethiraj and Alfons van Blaaderen. A colloidal model system with an interaction tunable from hard sphere to soft and dipolar. *Nature*, 421(6922):513–517, 2003.

[266] Tatu Ylonen and OpenBSD team. Openssh. http://www.openssh.com/, 1999-2016.

[267] Kenji Yoshimoto, Tushar S Jain, Kevin Van Workum, Paul F Nealey, and Juan J de Pablo. Mechanical heterogeneities in model polymer glasses at small length scales. *Physical review letters*, 93(17):175501, 2004.

[268] Sang Koo You, Dae Hyuk Kwon, Yong-ik Park, Sun Myong Kim, Myung-Hoon Chung, and Chul Koo Kim. Collective behaviors of two-component swarms. *Journal of theoretical biology*, 261(3):494–500, 2009.

[269] Klaus Zahn, Axel Wille, Georg Maret, Surajit Sengupta, and Peter Nielaba. Elastic properties of 2d colloidal crystals from video microscopy. *Physical review letters*, 90(15):155506, 2003.

[270] Michael Zaiser, Karlheinz Bay, and Peter Hähner. Fractal analysis of deformation-induced dislocation patterns. *Acta Mater.*, 47:2463–2476, 1999.

[271] B. Zhang, J. Zerubia, and J-C. Olivo-Marin. Gaussian approximations of fluorescence microscope point-spread functions. *Appl. Optics*, 46:1819–1829, 2007.

[272] H. Zocher. *Trans. Faraday Soc.*, 29:945, 1933.

[273] Michael Zur Muehlen, Jeffrey V Nickerson, and Keith D Swenson. Developing web services choreography standardsthe case of rest vs. soap. *Decision Support Systems*, 40(1):9–29, 2005.

†Credits on next page. . .

Credits

Director :	JAMES P. SETHNA
Producers :	MARGARET BIERBAUM
	WAYNE BIERBAUM
	JAMES P. SETHNA
Executive Producer :	JAMES P. SETHNA

Matt Bierbaum :	AS HIMSELF
Barry the Loner :	ALEX ALEMI
Barry the Whale :	ALEX ALEMI
Barry the Talking Computer :	ALEX ALEMI
Jeff Goldblum :	CORKY WHARTON
Chim Richalds :	JARED MAXSON
Not a Scone :	ANDY BOHN
Tommy Pickles :	JOHN STOUT
Dolly the Llama :	KAYLA CROSBIE
Chris Earls :	BRIAN LEAHY
Angela Lansbury :	SAM WHITEHEAD
Barry Gibb :	BEN SAVITZSKY
Reevbot :	YARIV YANAY
Pam Smosen :	SAM POSEN
Labradoodle, Butternut :	COLIN CLEMENT
Labradoodle, Butternut 2 :	JEFF GOLDBLUM (CORKY WHARTON)
Stuntman for Mr. Wharton :	KATE BIERBAUM
Stuntman for Mr. Bierbaum :	CRAIG DAUGHERTY

Production assistant :	VERONICA PILLAR	Boy :	ANDY BOHN
Key Grip :	BRIAN KOOPMAN	Best Boy :	ANDY BOHN
Gaffer :	NIC EGGERT	Best Best Boy :	ANDY BOHN
Catering :	GREG STIEHL	Onsite entertainment :	DANIEL CITRON
Catering :	JESSE HOPKINS	Stunt Coordinator :	ALEX MELLNIK