Order and Jamming on Curved Surfaces

A dissertation submitted by

Christopher J. Burke

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Physics

TUFTS UNIVERSITY

August 2016

Adviser: Timothy Atherton

Abstract

Geometric frustration occurs when a physical system's preferred ordering (e.g. spherical particles packing in a hexagonal lattice) is incompatible with the system's geometry. An example of this occurs in arrested relaxation in Pickering emulsions. Pickering emulsions are emulsions (e.g. mixtures of oil and water) with colloidal particles mixed in. The particles tend to lie at an oil-water interface, and can coat the surface of droplets within the emulsion (e.g. an oil droplet surrounded by water.) If a droplet is deformed from its spherical ground state, more particles adsorb at the surface, and the droplet is allowed to relax, then the particles on the surface can become close packed and prevent further relaxation, arresting the droplet in a non-spherical shape. The resulting structures tend to be relatively well ordered with regions of highly hexagonal packings; however, the curvature of the surface prevents perfect ordering and defects in the packing are required. These defects may influence the stability of these structures, making it important to understand how to predict and control them for applications in the food, cosmetic, oil, and medical industries.

In this work, we use simulations to study the ordering and stability of sphere packings on arrested emulsions droplets. We first isolate the role of surface geometry by creating packings on a static ellipsoidal surface. Next we perform simulations which include dynamic effects that are present in the experimental Pickering emulsion system. Packings are created by evolving an ellipsoidal surface towards a spherical shape at fixed volume; the effects of relaxation rate, interparticle attraction, and gravity are determined. Finally, we study jamming on curved surfaces. Packings of hard particles are used to study marginally stable packings and the role curvature plays in constraining them. We also study packings of soft particles, compressed beyond marginal

stability, and find that geometric frustration plays an important role in determining their mechanical properties.

Contents

List of Figures

Acknowledgements

I want to thank, first and foremost, my advisor Tim Atherton for being a tremendously supportive mentor who has invested himself in all aspects of my development as a scientist. I also want to thank Patrick Spicer at the University of New South Wales, who provided the inspiration for this research and with whom we have enjoyed a fruitful collaboration.

I want to thank the Tufts University Department of Physics and Astronomy for the opportunity to pursue my Ph.D. and for providing a wonderful community in which to do this.

I want to thank my parents Patrick and Patricia who have supported me through every decision I've made, as well the rest of my loving family.

Finally, I want to thank all of my friends at Tufts and elsewhere who have helped me get through grad school, especially Jeremy, Eric, Erin, Josh, Andrew, Dave, Ozlem, Marc, Anna, Kacie, Billy, Kyle, Peter, Joe, Dan, Mike, Andrew, and Brian.

Order and Jamming on Curved Surfaces

Chapter 1

Introduction

THE goal of this thesis is to explore the role that geometric frustration plays in packings of spheres on ellipsoidal surfaces, which are used to model arrested Pickering emulsion droplets. We want to understand the role geometry plays in disrupting order and creating defects in these packings, how the order is affected by the dynamics of droplet formation, and how the geometry affects jamming, i.e. the mechanical rigidity of these packings.

In emulsions, which are mixtures of two immiscible fluids such as water and oil, it is energetically favorable to minimize the surface area at the interface between the two fluids. This promotes coalescence of droplets of the same fluid, leading to the fluids separating. It also creates a tendency for a droplet of one of the fluids surrounded by the other to take on a spherical shape, as this is the shape that minimizes surface area for a given volume. This becomes more complicated in a *Pickering emulsion*, which is an emulsion containing colloidal particles. In a Pickering emulsion, it is energetically favorable for particles to adsorb at fluid-fluid interfaces in order to reduce interfacial tension, thereby helping to stabilize the emulsion.

One specific phenomenon which can occur in a Pickering emulsion is *arrested relaxation*. It is possible to deform a droplet from its spherical ground state, increasing the surface area and allowing more particles to adsorb at its surface. When the droplet is allowed to relax back towards a spherical shape, the particles become more densely packed. If the initial surface coverage was high enough the particles form a close packed layer, arresting the surface evolution and resulting in a non-spherical droplet coated by a shell of particles.

This phenomenon is relevant in crude oil extraction, as well as for applications in medicine, cosmetics, and the food industry. It leads to a system with a rich microstructure that lies at the intersection of several fields of study within physics. If the particles are all spheres of the same size, the resulting structures consist of a relatively ordered packing on a curved surface; however, perfect ordering is incompatible with curvature due to the phenomenon of *geometric frustration*. Dense packings of particles may not be perfectly rigid; some applied load or thermal motion may lead to displacements which allow particles more room to move. If a packing is stable against these types of motions, it is *jammed*. *Packing problem*s have been studied in a wide variety of systems, from the infinite packing limit down to highly confined packings. Typically the optimal arrangement of a collection of particles depends strongly on its constraints.

In this thesis we explore how these factors influence the structure of packings of particles on ellipsoidal surfaces to better understand the structures produced through arrested relaxation. We will begin by giving further background on the physics mentioned above in chapter 2. In chapter 3 we will detail the numerical methods used to perform the studies presented in this thesis.

Chapter 4 will present results on packings of particles created on static

ellipsoidal surfaces, in order to understand the influence of surface geometry on particle ordering. We investigate how anisotropic surface curvature influences the placement of topological defects, how it influences the formation of excess defects, whether it causes directional ordering of chains of defects, and finally we carry out a search for highly symmetric packings which are found among smaller packings. Chapter 5 will then look at packings formed on surfaces which evolve at a constant volume in order to understand how the dynamics of the packing formation affect the resulting packing, with a focus on the effects of relaxation rate, interparticle attraction, and gravity.

In chapter 6 we study the rigidity of the arrested structures to determine whether they are jammed. We find that while the initially arrested structures are not jammed, they can be further evolved to reach a jammed state, and surface curvature plays and important role on these jammed packings. In chapter 7 we begin looking at bidisperse packings of particles on spheres (i.e. packings with particles of two different sizes) in order to understand how packing structure and particle ordering are affected by curvature in bidisperse systems.

Chapter 8 briefly describes an experimental realization of arrested Pickering emulsions, which our collaborator Patrick Spicer's group is able to create and study. The thesis concludes in chapter 9 with a summary of results, a list of remaining questions, and prospects for future work.

An appendix which briefly explores the idea of particles constrained to curved surfaces by a soft potential is included. Additionally, papers which I have published or submitted for review, but which are not directly related to the rest of this thesis, are included as appendices for the sake of completeness.

Chapter 2

Background

2.1 Order and curvature

PACKING particles on a curved surface is a problem that leads to geometric frustration: packing monodisperse (equal-size) spheres on a flat surface favors an orderly hexagonal lattice, but the curvature of the surface frustrates this ordering, introducing defects. A classic example of this is the Thompson problem[2], which asks: what is the arrangement of like-signed electrical charges constrained to a sphere which minimizes their energy? Solutions to this problem tend to have large regions of hexagonal lattice, but require defects. Defects are topological in nature and can be understood in terms of deformations in the underlying hexagonal lattice. There are two types of defects possible in a hexagonally ordered system: disclinations and dislocations. Disclinations correspond to an angular slice missing from, or an extra slice added to, a hexagonal lattice: upon traversing a closed loop around a disclination, one will have passed through less than or greater than six angular portions of the lattice (fig. 2.1a and b). This deficit or excess can be quantified as a topological charge, $q = (2\pi - \theta)/(\pi/3)$, where θ is the angular

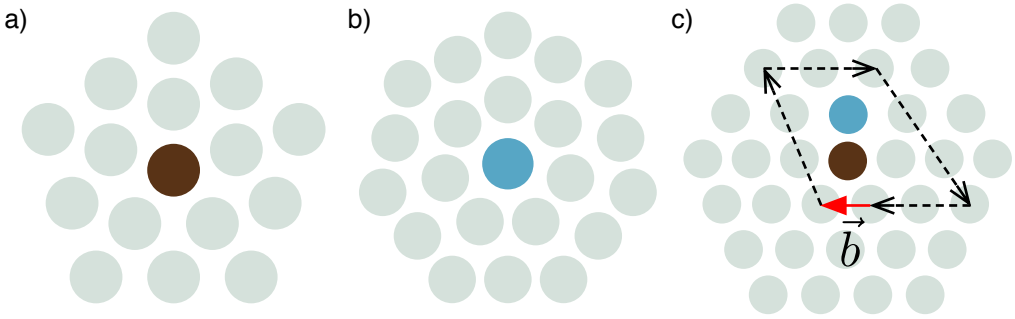


Figure 2.1: a) A positive disclination, with 5-fold coordination. b) A negative disclination, with 7-fold coordination. c) A dislocation, consisting of adjacent positive and negative disclinations. By traversing a path consisting of equal and opposite vectors along each lattice direction, one does not follow a closed loop. This mismatch is given by the Burger's vector, $\vec{b}$.

portion of the lattice traversed.

Dislocations correspond to a translational mismatch after traversing a closed loop. After circling a dislocation by moving in directions parallel to the lattice vectors, as in fig. 2.1c, one will find that the number of lattice sites traversed along one of the directions will not reduce to zero. The lattice vector corresponding to this mismatch is known as the Burger's vector. Dislocations can be understood as pairs of positive and negative disclinations, and have zero net topological charge.

Another way to characterize topological defects is in terms of particle coordination numbers. The coordination number refers to the number of neighbors a particle has. Given a suitable definition for neighbors, the topological charge is given by $q = 6 - c$ where c is the coordination number. In Euclidean space one can utilize the Voronoi diagram of a collection of particles. A Voronoi diagram is a collection of polygons, one for each particle, such that any given point in space lies within to the polygon of the particle to which that point is closest (in terms of Euclidean distance to the center of the particle.)[3] For each particle, the coordination number is given by the number of faces its

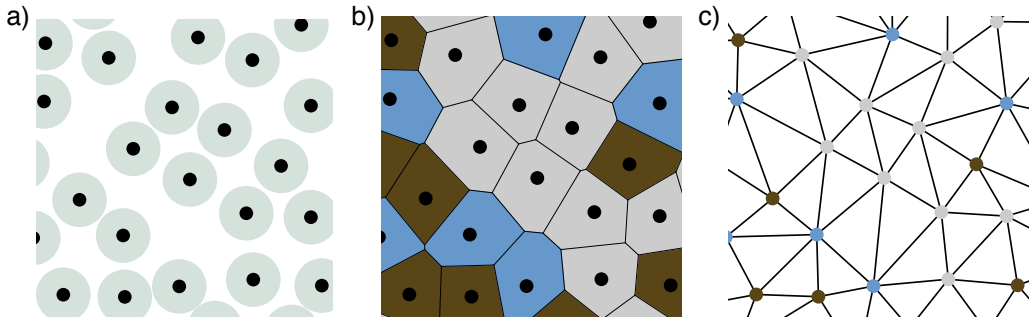


Figure 2.2: a) A set of randomly placed particles with centroids highlighted. b) The Voronoi diagram of the same set of points. The Voronoi cells are colored by coordination number: brown is 5, grey is 6, blue is 7. b) The Delaunay triangulation of the same set of points. Vertices are again colored by coordination number.

Voronoi cell shares with neighboring cells (i.e. polygons which share a face). Thus, for example, a positive topological charge will have five neighbors, and a negative charge will have seven neighbors. Another useful concept is the Delaunay triangulation. The Delaunay triangulation is the dual graph to the Voronoi diagram. To construct it from the Voronoi diagram, the center of each cell becomes a vertex, and an edge connects each pair of vertices which correspond to neighboring cells. Thus, a particle's topological charge is determined by the number of edges connected to its corresponding vertex. Fig. 2.2 illustrates this duality.

Mechanically speaking, defects are associated with strain in a crystalline structure. In flat space, a crystal's ground state contains no defects. Defects can be induced by thermal fluctuations, frozen in kinetically during the crystal's formation, induced by mechanical strain, or created spontaneously in non-equilibrium internally driven systems (i.e. active matter)[4, 5, 6, 7]. In a curved space, however, there must be strain in the lattice to accommodate curvature. More specifically, the Gaussian curvature of a surface is associated with stretching deformations and couples to the density of defects[8]. For a 2D

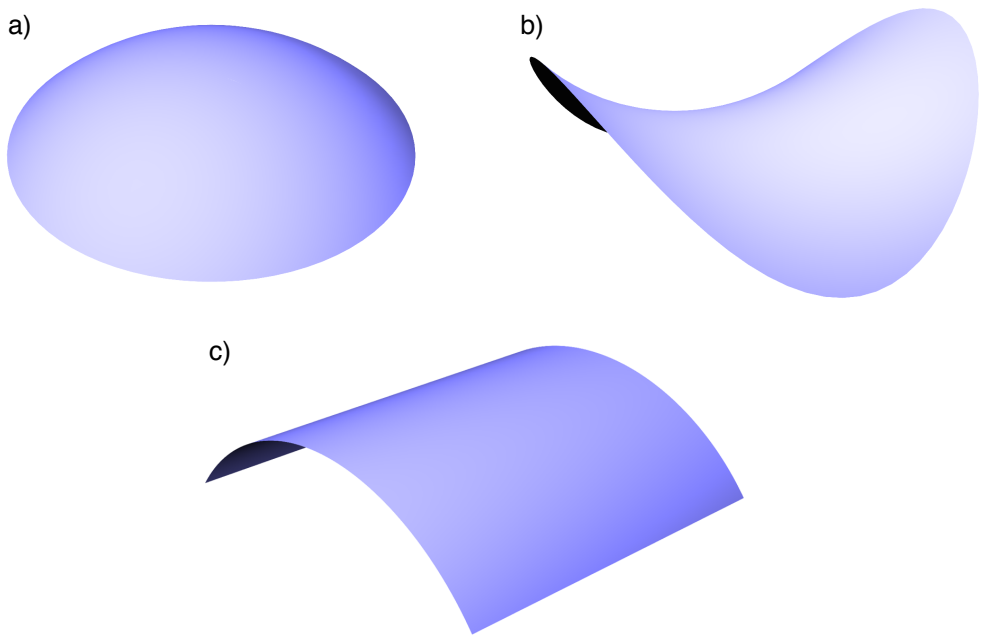


Figure 2.3: Surfaces with different types of curvature: a) positive Gaussian curvature; b) negative Gaussian curvature; c) zero Gaussian curvature but non-zero mean curvature.

surface, a curvature tensor can be defined at each point on the surface and its eigenvalues, κ_1 and κ_2, are the local principal curvatures, i.e. the maximum and minimum curvatures along orthogonal directions. These curvatures are the inverse of the local radii of curvature. From the principal curvatures, two averages can be constructed. The arithmetic mean gives the mean curvature, $H = \frac{\kappa_1+\kappa_2}{2}$ and the geometric mean gives the Gaussian curvature, $K = \kappa_1\kappa_2$. Note that these are local variables, defined for each point on the surface. Fig. 2.3 illustrates surfaces with different types of curvature. Fig. 2.3a has positive Gaussian curvature and fig. 2.3b has negative Gaussian curvature. Fig. 2.3c has zero Gaussian curvature but non-zero mean curvature. Gaussian curvature is associated with in-plane stretching deformations while mean curvature is associated with out-of-plane bending deformations. An intuitive example of this is seen in carbon fullerenes. A flat sheet of graphene will be made of hexagonal interstices without any defects. Carbon nanotubes are essentially rolled

up sheets of graphene, and as such no defects need to be introduced to accommodate their curvature, which is purely mean curvature and not Gaussian. Buckyballs, on the other hand, are spherical and possess Gaussian curvature, so they must have twelve pentagonal faces, like a soccer ball.

One way to understand the coupling of defects to curvature is in terms of the elastic theory of two-dimensional thin membranes. The goal of such a theory is to relate the strain induced by a defect to the stresses and strains in the surrounding membrane by minimizing the elastic energy. A membrane is treated as a continuum rather than as a collection of discrete particles, while the defects themselves are point-like sources of strain. For a thin membrane, the stress tensor σ can be related to a scalar potential known as the Airy stress function, χ, which is defined such that, in Cartesian coordinates,

$$\sigma_{xx} = \frac{\partial^2 \chi}{\partial y^2},\ \sigma_{yy} = \frac{\partial^2 \chi}{\partial x^2},\ \sigma_{xy} = \sigma_{yx} = -\frac{\partial^2 \chi}{\partial x \partial y}.$$

It can then be show that[8][9]

$$\frac{1}{K_0}\nabla^4 \chi(\vec{x}) = s(\vec{x}) - K(\vec{x}) \tag{2.1}$$

where K_0 is the 2D Young's modulus and s is the density of dislocations,

$$s(\vec{x}) = \frac{\pi}{3g(\vec{x})} \sum_{i=1}^{N} q_i \delta(\vec{x} - \vec{x}_i),$$

where the q_i are the charges of each disclination, the $\vec{x}_i$ are their positions, and $g(\vec{x})$ is the conformal factor (the square root of the determinant of the metric). Eq. 2.1 is an inhomogeneous biharmonic equation where the the Gaussian curvature and the defect density act as source terms. In this way, the Gaussian curvature can be thought of as a geometric charge. Disclinations tend to be

attracted to regions of like-signed Gaussian curvature in order to minimize the source term. This makes intuitive sense, as positive Gaussian curvature results in an angular deficit when traversing a closed loop (i.e. the the circumference of a circular path of radius r will be less than $2\pi r$), while positive disclinations correspond to an angular deficit in the lattice, and oppositely for negative curvature and defects. However, the specific placements of defects will be highly nontrivial, as not only are biharmonic equations difficult to solve, but eq. 2.1 determines only the stress function for a given defect configuration, and an energy minimization must be performed to find the defect configuration. The system energy is given by

$$E = \int \mathrm{d}A\chi(\vec{x}) \left(s(\vec{x}) - K(\vec{x})\right),$$

which must be minimized with respect to both the number of defects and the position of those defects.

The connection between angular deficit and topological defects has deeper implications for hexagonal order on curved surfaces. For a surface of a given topology, the net defect charge is determined by the topology of the surface. Specifically, $\sum q_i = 6\chi_e$ where χ_e is the Euler characteristic of the surface. The Euler characteristic is determined by the number of handles in the surface, e.g. a sphere has $\chi_e = 2$, a torus has $\chi_e = 0$, a double torus has $\chi_e = -2$, etc. Thus, an ellipsoid or any other surface which is topologically equivalent to a sphere will have a net defect charge of 12.

This topological constraint relates to the Gauss-Bonnet theorem which states that the integrated Gaussian curvature is a topological constant of a surface:

$$\int_M K(\vec{x})dA + \int_{\partial M} k_g(\vec{x})ds = 2\pi\chi_e$$

where the first and second integrals are over the surface manifold and its boundary, respectively, and k_g is the geodesic curvature of the boundary. A discrete analogue applies the the Delaunay triangulation of a packing on a surface (or rather a Delaunay-like triangulation, as the Delaunay triangulation does not generalize trivially to curved surfaces; this will be addressed in the *Methods* chapter.) A triangulation of a set of points is a polyhedron, and the Gaussian curvature can be thought of as discretely localized to the vertices. The Gaussian curvature is then given by the angular deficit at each vertex, i.e. 2π minus the sum of the angles of the corners surrounding that vertex. The discrete version of the Gauss-Bonnet theorem for a closed surface is then

$$\sum \alpha_i = 2\pi\chi_e$$

where α_i is the angular deficit at each vertex.

This also relates to Euler's theorem, which relates the number of vertices V, edges E, and faces F of a polyhedron to its topology:

$$V - E + F = \chi_e.$$

For a triangulation, each face has 3 edges, and each edge is shared between two faces, giving us $2E = 3F$. For a hexagonally ordered triangulation, most vertices will be connected to 6 edges, with deviations at the defects, so a given vertex will have $e_i = 6 - q_i$ edges. Each edge is also shared between two vertices, so $\sum_{i=1}^{V} 6 - q_i = 2E$. Substituting these relations between V, E, and F back into Euler's theorem we recover the desired result

$$\sum q_i = 6\chi_e.$$

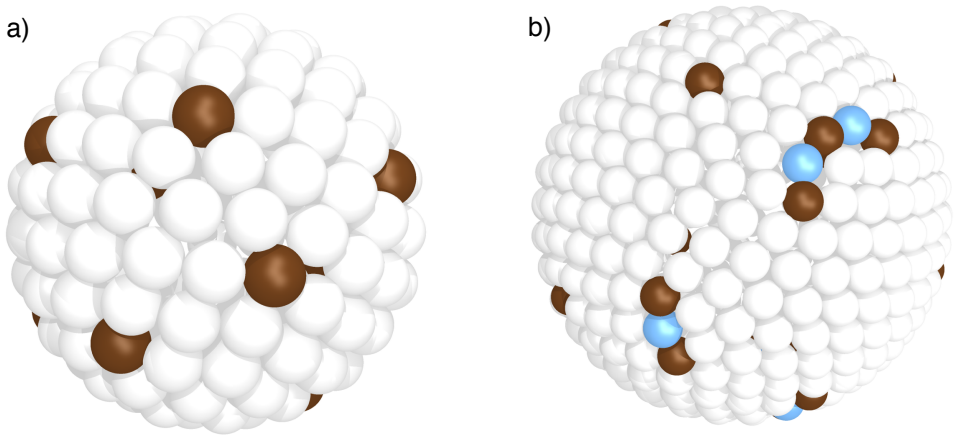


Figure 2.4: a) A packing on a sphere with the minimal number of defects required by topology. b) A packing on a sphere exhibiting scars.

While the surface topology constrains the net charge of a packing, it does not determine the total number of defects. Excess dislocations, neutral in charge, can appear which act to screen out high strains in the lattice. For example it has been predicted[9] and experimentally observed[10] in packings on spheres that, as the system size increases (i.e. the surface radius becomes much larger than the particle radius), excess defects begin appearing in the form of "scars", chains of dislocations extending out from topologically required disclinations. For small packings, only the 12 required disclinations appear, as in fig. 2.4a. Above a critical system size, however, scars begin to form, and the length of the scars increase linearly with system size. A packing with scars is shown in fig. 2.4b. A number of studies have also been done which examine topological defects on a number of other surface geometries, include tori, Gaussian bumps, catenoids, and paraboloids[11, 12, 13, 14]. In this thesis, we will be extending this understanding of the effect of surface geometry on topological defects by looking at packings on ellipsoids, which have variable Gaussian curvature. While defects on spheres are well studied, it has not previously been studied how the anisotropy in curvature on an ellipsoid will

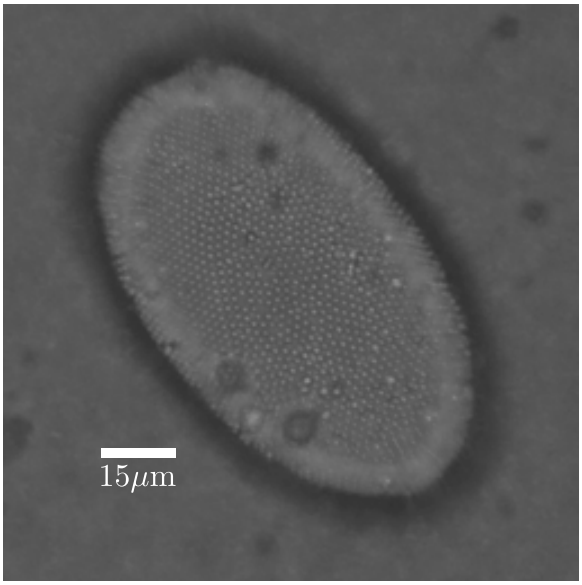


Figure 2.5: An arrested Pickering emulsion droplet, imaged using optical microscopy. Image courtesy Patrick Spicer, University of New South Wales.

effect the distribution of defects, which will be one of the goals of this thesis.

2.2 Arrested relaxation

Emulsions are mixtures of two immiscible fluids. A high surface tension between the two fluids prevents them from mixing by penalizing interfacial area between them. For example, a single droplet of one fluid surrounded by the other will tend to take on a spherical shape, as a sphere is the shape which minimizes the surface area of a given volume. A droplet deformed from a spherical shape will relax back towards the spherical ground state. Another manifestation of this is in droplet coalescence. If two droplets come into contact, they will begin to coalesce into a single droplet. Initially, this droplet is highly deformed, being made up of two spheres connected by a capillary bridge, and will relax to a single larger sphere. Through this mechanism, the fluids of an emulsion can separate out into two bulk phases.

One way to prevent separation in emulsions is by including colloidal par-

ticles, i.e. solid particles on the length scale of one micron. An emulsion with colloids mixed in is known as a Pickering emulsion[15]. In a Pickering emulsion, particles tend to adsorb at the fluid-fluid interface. This happens because it reduces interfacial energy: the particle-fluid surface tensions are low compared to the fluid-fluid surface tension, and a particle at the interface reduces the fluid-fluid interfacial area. These particles are strongly bound to the interface: the binding energy can be as high as $10^7 k_B T$[16]. These particles stabilize emulsions by preventing droplet coalescence, as they make it difficult for droplets to come into contact and initiate coalescence.

One interesting phenomenon that occurs in Pickering emulsions, which is the motivation for this thesis, is arrested relaxation. If a droplet is deformed from its spherical ground state, its surface area increases. This allows more particles to become adsorbed at the fluid-fluid interface. When the droplet is then allowed to relax the particles will become more densely packed as the area reduces, and if the initial density of surface particles is high enough, eventually the particles will become packed together and their elastic repulsion prevents further relaxation of the droplet. This results in a particle packing covering a droplet arrested in a non-spherical shape.

This process has been realized in several experimental systems. A spinning drop tensiometer can be used, in which the droplet is spun and centrifugal forces cause the droplet to elongate along the axis of rotation, followed by particle adsorption and droplet relaxation[17]. The deformation can also be achieved by an applied electric field[18]. A third mechanism is through coalescence: if two droplets with a large enough initial surface coverage coalesce, the initial coalescence contact acts as the deformation process and the following relaxation drives arrest[19].

This thesis will explore the microstructure of arrested Pickering emulsion

droplets, and particularly how the curvature of the droplet surface affects the particle packing by exploring the ideas outlined in the rest of this chapter. We investigate the coupling of topological defects to surface curvature, the role of the dynamics of droplet relaxation on the arrested packing, and the rigidity of the final packing, i.e. whether or not it is jammed.

2.3 Jamming

The concept of jamming has been used in condensed matter physics to understand systems including granular materials, foams, and glasses. Generally, jamming deals with packings of particles that are mechanically stable and have solid properties in the bulk, but are not necessarily ordered. Within the literature, there are multiple approaches employed to study jamming. One approach considers packings of soft particles and their properties near the threshold of mechanical stability[20]. The other considers hard particles and focuses on whether they are mechanically stable[21]. Each approach has its own advantages and can offer insight into the structure and stability of particle packings. In this work we employ both approaches.

Much work on jamming in soft particles has focused on the concept of the jamming point, which is a point in the parameter space of density, temperature, and applied stress at which jamming is well defined. At zero temperature and zero applied stress, there is a critical density ϕ_c above which packings are stable and below which they are not (fig. 2.6). This is the jamming point. At this point, packings are marginally stable to an infinitesimal applied stress.

Packings at the jamming point have a number of generic features. They are typically disordered. This can be measured by an order parameter: in 3D,

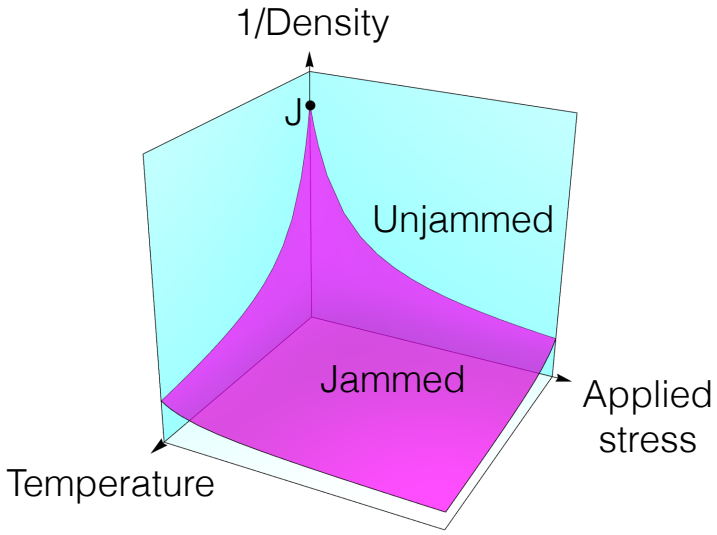


Figure 2.6: A schematic of the jamming phase diagram. For a packing of particles, in the parameter space of temperature, applied stress, and inverse density, the jamming point J occurs at zero temperature and applied stress, at a critical packing density.

a common way to quantify ordering is by using the spherical harmonics Y_{lm}:

$$Q_6 = \left(\frac{4\pi}{13} \sum_{m=-6}^{6} \left| \frac{1}{N_b} \sum_{i=1}^{N_b} Y_{6m}(\theta_i, \phi_i) \right|^2 \right)^{1/2}$$

where N_b is the number of neighbors of a given particle and the second sum is over these neighbors. θ_i and ϕ_i are the angular positions of the neighbors with respect to the given particle. This parameter looks at a local measure of the angular distribution of a particles neighbors; a global average can then be taken. Often, this quantity is normalized by the value of an ideal packing like a face-centered cubic (FCC) crystal: $Q = Q_6/Q_6^{\mathrm{FCC}}$. Thus, one way to understand a disordered packing is as highly non-crystalline. An important order parameter for 2D systems is the hexatic order parameter:

$$\psi_6 = \left| \frac{1}{N_b} \sum_{i=1}^{N_b} \exp(i6\theta_i) \right|$$

which measures the degree to which, on average, the neighbors of a particle are evenly spaced at $\pi/3$ angles around it as in a hexagonal lattice.

Jammed packings, in the disordered, soft particle sense, are also *isostatic*. That is, they have the minimum number of interparticle contacts required for stability (hence, marginal stability). In order for a packing to be stable, the number of constraints (contacts) must equal the degrees of freedom. For a packing of N spheres in n dimensions, there are Nn degrees of freedom. If each particle has, on average, Z contacts with neighboring particles, there are $ZN/2$ constraints (because each contact is shared between two particles). Setting the number of degrees of freedom equal to the number of contacts, we find $Z = 2n$. Thus, a jammed packing of spheres will have $Z = 4$ in 2D and $Z = 6$ in 3D. This is in contrast with crystalline packings which are hyperstatic (they have more contacts than required for mechanical stability.) For example a hexagonal packing in 2D has $Z = 6$ and an FCC packing in 3D has $Z = 12$. In studying soft-particle jamming, particles are typically modeled as having a compact interaction potential. That is, particles exert a repulsive force on each other when they are in contact, and otherwise exert no force on each other.

Note that contacts are not necessarily the only type of constraints on a packing. For example, in packings of ellipsoidal particles, the relatively flat sides of ellipsoids can prevent rotations, implicitly imposing additional constraints on the packing in addition to the contacts which prevent translation[22]. These rotational constraints are higher order non-linear constraints, as they do not prevent infinitesimal rotations but do prevent finite rotations.

Another area of interest in soft particle jamming is how the properties of a packing scale with the density around the jamming point[23]. The average contact number per particle, at the jamming point, has an isostatic value

Z_c. Above the jamming point, the increase in the contact number scales as $Z - Z_c \propto (\phi - \phi_c)^{0.5}$. Below the jamming point, packings are not rigid, and this is evident in the fact that they have zero elastic moduli and they have an abundance of "floppy modes", i.e. collective particle motions which do not increase the energy of the packing. Even at and above the jamming point, jammed solids have a high density of low frequency dynamical eigenmodes, deviating from the behavior of crystalline solids.

Above the jamming point, packings have elastic properties which behave much differently than those of crystalline solids. In calculating a bulk or shear modulus for these systems, one can calculate an instantaneous modulus or an infinite-time modulus: the instantaneous modulus is calculated immediately after applying a perturbation, but the infinite-time modulus is calculated after applying a perturbation and then minimizing the systems energy. For an affine system like a crystal, this will not make a difference. For packings near the jamming point it is observed that both the instantaneous bulk modulus B_0 and the infinite time modulus B_∞ scale like$\sim (\phi - \phi_c)^{\alpha-2}$ (where α is a parameter determining the particle interactions), as would also be expected for a crystal. However, $B_\infty < B_0$, i.e. the scaling laws have a different coefficient, indicating that jammed packings are not affine. The shear modulus behaves even more non-affinely. The instantaneous shear modulus scales as $G_0 \sim (\phi - \phi_c)^{\alpha-2}$ but the infinite-time shear modulus scales as $G_\infty \sim (\phi - \phi_c)^{\alpha-1.5}$. Note that the ratio $G_\infty/B_\infty = (\phi - \phi_c)^{0.5}$ does not depend on the particle interaction.

The hard particle jamming literature defines jamming in terms of what types of particle motions are possible in a packing. Packings can be classified into a hierarchy of jamming categories which have varying degrees of stringency[24]. The least stringent jamming category is local jamming: a packing is locally jammed if, for any given particle, that particle is unable to

move if all other particles are held fixed. The next, more stringent category is collective jamming: a packing is collectively jammed if there are no possible collective particle motions while the boundary is held fixed. Finally, strict jamming occurs when a packing is stable to both collective motions of particles and volume-non-increasing deformations of the packing boundary. Once a packing has been classified according to jamming category, quantities of interest including order parameters and average contact number can be calculated.

The concept of a jamming polytope is useful for analyzing a packing and its possible motions. For a packing of N spheres, the packing is specified by the collection of the positions of the sphere centers, $\mathbf{R} = \{r_1, \ldots, r_N\}$. For a given configuration $\mathbf{R}$, there is a neighborhood of points in configuration space which are accessible from the current configuration. For packings at low density, the configuration can access the majority of configuration space as the particles can move freely and slip past one another. As the density increases, the packing will become confined to a smaller region of configuration space. For an ideal jammed packing, the available configuration space will consist only of a single point, the jammed configuration $\mathbf{R_J}$. In practice, numerical methods cannot produce ideal packings, as there will always be some spacing between particles due to limits on numerical precision. Thus, a jammed packing will have a small neighborhood of configurations available to it, $\mathcal{J}_{\Delta\mathbf{R}}$, made up of the jammed configuration plus possible displacements, $\mathbf{R_J} + \{\Delta\mathbf{R}\}$. This neighborhood can be linearized and approximated as a high dimensional polyhedron, known as the jamming polytope $\mathcal{P}_{\Delta\mathbf{J}}$[25]. The jamming polytope concept is also valid when a packing is not actually jammed, and in this case it can be used as a numerically convenient way to explore possible unjamming motions, by searching for large or unbounded displacements within the jamming polytope[26]. Details on this procedure will be given in the *Methods*

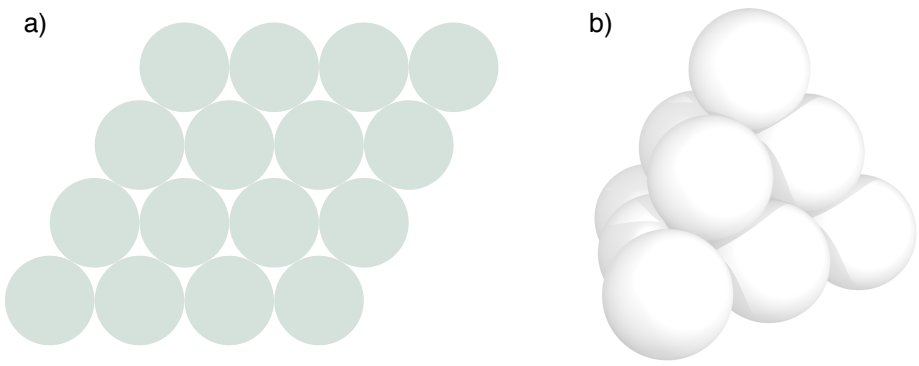


Figure 2.7: a) A hexagonal packing in 2D. b) An FCC packing in 3D.

chapter. It is important to note that

One of the goals of this thesis will be to determine whether arrested packings are jammed, and the role that surface curvature plays in jammed packings. We explore the new concept of metric jamming, appropriate for packings on deformable curved surfaces, and we show that while arrested packings are not metric jammed, they can be unjammed and further surface evolution can push them towards a metric jammed state. Additionally, we show that the curvature of the surface plays an important role in constraining the metric jammed packings. We also show that monodisperse packings of soft particles on a curved surface, which are relatively well ordered, share many properties with disordered jammed systems.

2.4 Packing

The densest packing of spheres in flat 2D space is a hexagonal lattice (fig. 2.7a). In 3D it is a face-centered-cubic (FCC) lattice (fig. 2.7b) and its stacking variants (produced by translating layers within an FCC lattice). Kepler first conjectured that the FCC lattice is the densest 3D sphere packing, but despite this result seeming rather intuitive, it was not proven until surprisingly recently

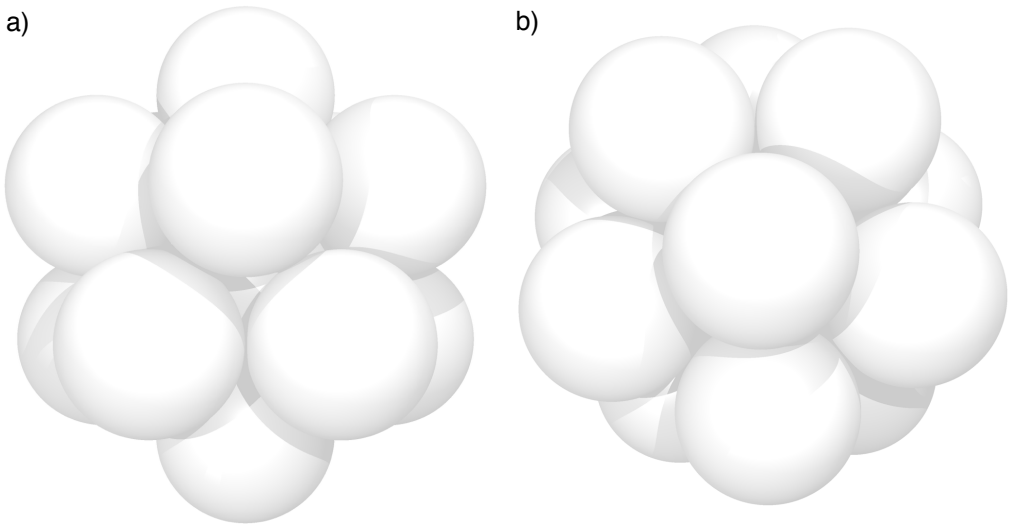


Figure 2.8: Two views of a locally-dense icosahedral packing arrangement, which is incompatible with long-range order.

in 2005[27]. Generically, questions about optimal packings are very difficult to answer rigorously. In this work, we do not investigate optimal packings (i.e. maximally dense packings), but rather we are interested in the ensemble of packings produced through a given packing protocol, and how changes in the protocol correlate to the properties of the packings produced. In this section we briefly review some results from the packing literature.

In packings of monodisperse spheres in 3D, it is not uncommon to produce a highly noncrystalline packing. In 2D however, packings tend to be very crystalline. The reason for this is that the locally densest packing (i.e., how can one pack particles densely around a single particle) in 2D is a hexagonal arrangement, which is the same as the globally densest packing arrangement. In 3D however, the locally densest packing is icosahedral (fig. 2.8). This is incompatible with long-range FCC ordering. This incompatibility makes it more difficult kinetically for a packing to reach an FCC state.

The above is true for monodisperse spheres. However, introducing polydispersity, i.e. including spheres of different sizes, further promotes disor-

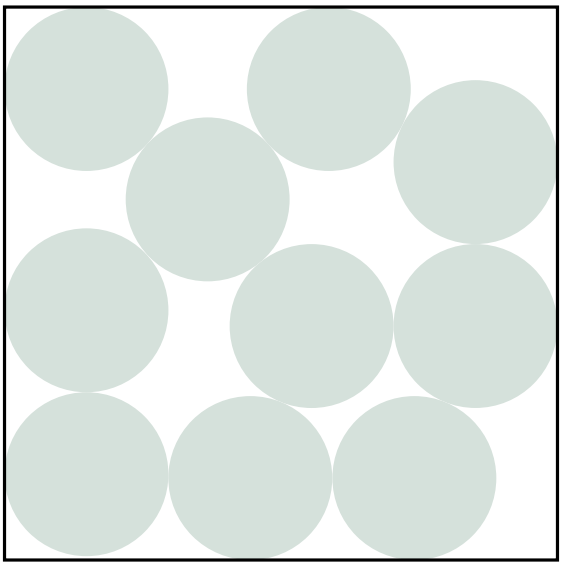


Figure 2.9: The optimal configuration for 10 particles packed into a square[1].

der. While in 3D it is still possible to achieve crystallization of monodisperse spheres, polydispersity makes this much more difficult. In 2D as well, where crystallization is the norm for monodisperse packings, polydispersity makes it much harder to achieve crystallization[28].

In 3D, where packings of monodisperse spheres typically do not crystallize, the concept of "random close packing" (RCP) has been used to describe this phenomenon. It has been observed that, using a remarkably diverse variety of packing protocols, random packings of hard spheres typically have a density of $\phi = 0.64$[29]. It has been argued, however, that the RCP concept is not well defined. It has been suggested that it should be replaced with the concept of the Maximally Random Jammed state (MRJ)[30]. Given a packing protocol and choosing an appropriate order parameter, an ensemble of packings can be produced within a continuum of packing fraction and order parameter. The MRJ is defined as the jammed packing with the lowest achievable order parameter.

In some systems, typically where the packing size is not very large com-

pared to the particle diameter, it is seen that compatibility between the packing and its boundary conditions become important. This effect is known as commensurability. For instance, the optimal packing for ten particles packed into a square has a very specific and irregular configuration, as shown in fig. 2.9[1]. Another example occurs in packings of spheres on a cylinder[31], where it is possible to get an ideal hexagonal packing if the ratio of the particle diameter and the cylinder diameter are commensurate; otherwise, an oblique lattice (a stretched hexagonal lattice) can occur, or a line slip defect can occur (a hexagonal lattice, with a mismatch between two rows which wraps around the cylinder.) Whether an oblique lattice or a line slip occurs depends on the particle interaction potential: softer particles tend to form an oblique lattice, and harder particles tend to form a line slip defect.

In this thesis we draw connections between these previously observed packing phenomena and packings on curved surfaces. Given that an ellipsoid is a compact surface, the packings must obey periodic boundary conditions and it is expected that commensurability is important, especially at low particle numbers. One may also wonder whether these constraints, combined with the strain induced by the curvature, influences the formation of line slip or oblique packing structures. Another question which is addressed is the influence of curvature in bidisperse packings. Both curvature and bidispersity tend to frustrate the ordering of sphere packings and combining these features has interesting effects.

Chapter 3

Methods

Here we describe in detail the computational methods employed in this work. These consist of algorithms for generating particle packings as well as analysis techniques for studying these packings.

The first algorithm is an inflation packing algorithm, which is used to create packings of particles on static surfaces. The second is a dynamical packing algorithm, which simulates the packing of particles on an evolving surface, incorporating additional physical effects including interparticle attraction and gravity. The third is a soft-particle Monte Carlo algorithm which uses the Metropolis algorithm to generate low energy particle packings. The final algorithm is another soft-particle energy minimization, which uses gradient descent to find stable particle configurations for particles with an arbitrary interparticle potential. This is used to generate packings of soft particles, as well as to study the stability of hard particle packings by attaching a soft effective potential.

The analysis techniques employed are used to study the qualities of the packings produced and to quantify their ordering. We use a Delaunay triangulation to identify topological defects, employing the software Meshlab in

combination with Morpho. A hexatic order parameter is calculated through standard techniques. We also identify "rattlers" in the packings, i.e. particles which are not locally jammed and thus have no bearing on the stability of a packing. Finally, we employ a linear program to find motions which unjam a packing, and we combine this with the soft particle gradient descent scheme in order to quickly find unjamming motions.

3.1 Simulations

3.1.1 Inflation packing algorithm

We employ a stochastic inflation packing algorithm inspired by the Lubachevsky-Stillinger (LS) algorithm, which is known to yield packings of high coverage fraction[32]. The central concept is that particles are constrained to a fixed surface, and their radii are increased, thus increasing the packing fraction, until the particles become arrested, as illustrated in fig. [need to add figure showing an initial and final configuration from inflation algorithm]. The main difference between the LS algorithm and this inflation algorithm is that the LS algorithm is event-driven: between collisions, particles travel at a constant velocity, and based on this the next collision is predicted and particle positions and velocities are updated at each collision. Our inflation algorithm employs Brownian dynamics. This is motivated by the physical system we are modeling, in which particles diffuse at a fluid-fluid interface.

In a given simulation, a fixed ellipsoidal surface, either prolate or oblate, is chosen with aspect ratio a and the length of the semi-minor axis is fixed to be unity in dimensionless units. Particles are modeled as monodisperse hard spheres of radius r which is slowly increased during the simulation. The number of particles N is specified and particles are deposited at the start of

the simulation by random sequential adsorption (i.e. each particle is placed randomly, one at a time, with a uniform probability distribution across the surface, and any placement that results in overlap between particles is rejected and placement is attempted at another random position) such that the center of each particle is constrained to lie on the surface of the ellipsoid. Initially, r is such that the packing fraction is $\phi = 0.05$.

The algorithm proceeds by two kinds of moves: i) Monte Carlo *diffusion steps* where particles are moved randomly along the surface and ii) *inflation steps* where the radius of all particles is increased by δr. In each diffusion step, N individual Monte Carlo moves of randomly chosen particles are attempted. The step size is chosen randomly using a Gaussian distribution, as described below. Only moves that do not result in overlap are accepted, with overlaps checked for in the 3D configuration frame.

The moves are performed in the 2D space of conformal surface parameters (u, v). Conformal coordinates ensure a radially symmetric probability distribution for moving a certain arc length s in any tangential direction from the current location. This is because the local tangent vectors taken with respect to either conformal coordinate have the same magnitude. This magnitude can vary across the surface, and this is taken into account by scaling the tangent vectors by the conformal factor (defined below) when taking steps during the simulation. The surface is parametrized as,

$$x(\theta, \phi) = (x_0 \sin\theta \cos\phi, x_0 \sin\theta \sin\phi, z_0 \cos\theta), \tag{3.1}$$

where $x_0 = 1$, $z_0 = a$ for prolate surfaces and $x_0 = a$, $z_0 = 1$ for oblate surfaces. The determinant of this metric is,

$$g(\theta) = \frac{1}{2} x_0 \sin(\theta)^2 \left(z_0^2 + x_0^2 + \left(z_0^2 - x_0^2\right) \cos(2\theta)\right), \tag{3.2}$$

the conformal factor is given by $\sqrt{g(\theta)}$ and the conformal parameter u is given by the integral of the conformal factor such that $u = 0$ at the equator,

$$u(\theta) = \int_{\pi/2}^{\theta} \sqrt{g(\theta')} d\theta', \tag{3.3}$$

which can be inverted to find $\theta(u)$. We do an approximate inversion by calculating $u(\theta)$ for values of θ from 0 to π in increments of $\pi/100$ and using a 19th order polynomial least squares fit on these points, enforcing equality between the fit and exact values at the endpoints $\theta = 0$ and $\theta = \pi$. The conformal coordinate v is simply $v(\phi) = \phi$.

Given the definitions above, diffusion steps are taken as follows. An unscaled step size is chosen for each direction, Δu_o and Δv_0, from a normal distribution with variance 1. These are scaled by the simulation step size σ and by the inverse of the conformal factor to give step sizes in the (u, v) conformal space:

$$\Delta u = \frac{\sigma \Delta u_0}{\sqrt{g(\theta(u))}} \tag{3.4}$$

$$\Delta v = \frac{\sigma \Delta v_0}{\sqrt{g(\theta(u))}}. \tag{3.5}$$

These steps are used to update the previous u and v coordinates of the particle, which are then transformed to the θ and ϕ coordinates as explained above. Finally, the surface parametrization eq. 3.1 is used to give the particle coordinates in the 3D configuration space.

Because θ must have a value between 0 and π, we take the following step if it falls outside this range at any point. If u is greater than $u(0)$ (less than $u(\pi)$), we set $u = 2u(0) - u$ ($u = 2u(\pi) - u$) and $v =$ mod $(v + \pi, 2\pi)$, i.e. we allow the particle to pass over the coordinate singularity at the poles of the

surface.

As the particles diffuse, σ is varied in order to more efficiently explore relevant areas of configuration space (leading to large steps when the configuration is loosely packed and smaller, more relevant steps when tightly packed.) The initial value of σ scales with the square root of the ellipsoid surface area A, $\sigma_{init} = 1 \times 10^{-4}\sqrt{\frac{A}{4\pi}}$. After each time step, the fraction of attempted moves that were accepted is calculated. The length scale σ is then decreased by 1% if the acceptance fraction is < 0.5 and increased by 1% otherwise; σ is reset after each inflation (described below) to its initial value. Bounds are imposed such that $1 \times 10^{-6} < \sigma < 1$. Adjusting σ leads to improved performance of the algorithm as the particles can diffuse more when they are less densely packed and take smaller steps (which are more likely to be accepted) when they are more densely packed. We do this as it is known that adaptive algorithms lead to packings of higher density[33]. We emphasize that in this work the Monte Carlo approach is used as an optimization strategy; it is not intended to, and indeed cannot, replicate the physical process by which the structures form since the σ updates are non-Markovian.

After $M = 100$ diffusion steps, an inflation step is performed where the particle radius is increased slightly ("inflated") either by a specified fixed amount $\Delta r = 1 \times 10^{-5}\sqrt{\frac{A}{4\pi}}$ or by the half of the largest amount allowed that would not result in the overlap of any pair of particles, whichever is smaller.

The halting criteria for these simulations is as follows: every $L = 100$ inflation steps, the relative change in coverage fraction $\Delta\phi$ is calculated. If this is less than a specified value $\Delta\phi_{tol} = 10^{-4}$ then the simulation halts.

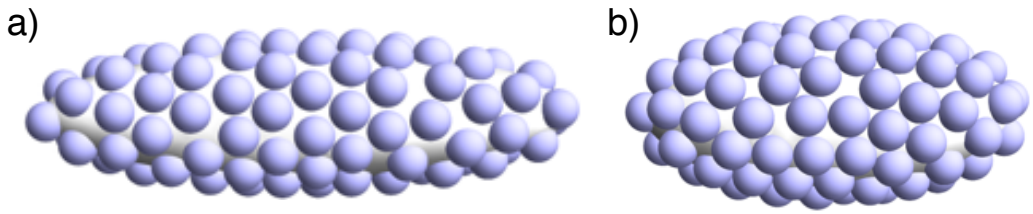


Figure 3.1: An example of a configuration produced using the dynamic packing algorithm. a) The initial configuration, with the surface in an elongated state. b) An arrested packing, produced by relaxing the ellipsoidal surface.

3.1.2 Dynamic Packing Algorithm

In order to understand dynamical effects on particle ordering, we perform simulations which model the physics of the droplet shape evolution. In these simulations, the particles have a fixed radius, and the surface on which they are adsorbed evolves while the particles are constrained to move with it, as illustrated in fig. 3.1. Thus it is the decrease in area of the surface that drives packing, as in the corresponding physical system.

Throughout a simulation, the surface relaxes from an initial aspect ratio to a sphere at constant volume, with the full time the unimpeded relaxation would take setting a relaxation timescale τ_r. As this occurs, particles experience three forces (in addition to the surface constraint): Brownian diffusion, interparticle attraction, and gravity, and they evolve under Stokes dynamics, i.e. non-inertial dynamics where $\frac{d\vec{x}}{dt} = \mu\vec{F}$ where μ is a the particle mobility. Each of these forces has an associated timescale, and the strength of each force is specified by setting the ratio of its timescale to the relaxation timescale. The diffusion timescale τ_d is the time it would take, on average, for a single particle, on a flat surface unimpeded by other particles, to diffuse a distance equal to its diameter. Due to the linear dependence of the mean squared displacement on time for diffusive processes, we have $2r = \sqrt{2D\tau_d}$ (where D is the diffusion constant) so

$$\tau_d = \frac{2r^2}{D}.$$

The attraction timescale τ_a is the time it would take a particle to move one particle diameter, under the attractive force felt from a particle with which it was in contact, if that force was held fixed. Thus, under Stokes dynamics and based on the force given by eq. 3.7 below, $2r = \frac{\alpha}{2r}\tau_a$ (where α is a constant determining the attraction strength) so

$$\tau_a = \frac{4r^2}{\alpha}.$$

The gravity timescale t_g is the time it would take a particle to move one particle diameter under the influence of gravity. Again from Stokes dynamics, $2r = g\tau_g$ (where g is a measure of the effect of gravity) so

$$\tau_g = \frac{2r}{g}.$$

Note that α and g are measures of the relative effect of a force on a particle. They depend, in principle, on the particle mass and radius.

For a given simulation, τ_d/τ_r, τ_a/τ_r, and τ_g/τ_r are specified (or attraction and/or gravity may be ignored, in which case $t_a = \infty$ or $t_g = \infty$.) The diffusion constant is held the same across simulations, and thus τ_r, α, and g are determined.

N particles of fixed radius r are initially placed randomly on an ellipsoid of aspect ratio a, using a random sequential adsorption scheme, with their centroids evenly distributed over the ellipsoid's area. If there is any overlap, the overlap is removed by a gradient descent relaxation, where particles interact through the following pair potential:

$$V_{\text{overlap}} = \begin{cases} r^2 - rx & x < r \\ 0 & x \geq r \end{cases}$$

where x is the center-to-center distance between particles. This gives a constant force proportional to the particle size if particles are overlapping, and no force otherwise. The gradient descent uses a timestep of $\Delta t_{\text{overlap}} = 0.001$, i.e. two completely overlapping particles, in the absence of any other overlaps, would take 1000 steps to remove overlap. If it takes more than 1×10^6 steps to remove all overlap between particles, the initial state is considered to be overcrowded and the simulation is halted. After overlap is removed the particles are then allowed to diffuse for $t = 4\tau_d$, using timesteps such that the average step distance will be 1/100 of the particle diameter.

Throughout the simulation particles evolve according to a Langevin equation, neglecting inertial effects, with the surface constraint enforced using Lagrange multipliers. The position of a particle i is updated to its unconstrained position $\vec{x}'$ by,

$$\vec{x}_i'(t + \Delta t_p) = \vec{x}_i(t) + \Delta t_p \left(\vec{F}_g + \sum_{i \neq j} \vec{F}_{ij} \right) + \vec{\eta}_i \sqrt{2D\Delta t_p} \tag{3.6}$$

where $\vec{F}_g = -g\hat{y}$ is the downward force on a particle due to gravity (the ellipsoid symmetry axis is oriented horizontally, perpendicular to gravity), $\vec{F}_{ij}$ is the attractive force of particle j acting on particle i, $\vec{\eta}_i$ is a noise term applied to particle i sampled from a Gaussian distribution with unit variance along the tangent plane of the surface at the location of particle i, D is the diffusion constant, and Δt_p is the particle movement timestep.

Particles centroids are constrained to the ellipsoidal surface, and this is enforced using Lagrange multipliers[**?**]. We define a constraint function $\sigma(\vec{x})$ such that the constraint is satisfied when $\sigma(\vec{x}) = 0$. For particles on an ellipsoid,

$$\sigma(\vec{x}) = \frac{x^2}{a_1} + \frac{y^2}{a_1} + \frac{z^2}{a_2} - 1 = 0.$$

where a_1 and a_2 are the semi-major and minor axis lengths of the ellipsoid. By requiring this constraint to be satisfied to first order, we can update the position to more closely follow the constraint,

$$\vec{x}''(t + \Delta t_p) = \vec{x}'(t + \Delta t_p) - \lambda \nabla \sigma(\vec{x}(t))$$

where λ is a Lagrange multiplier given by,

$$\lambda = \frac{\sigma(\vec{x}'(t + \Delta t))}{\nabla \sigma(\vec{x}'(t + \Delta t)) \cdot \nabla \sigma(\vec{x}(t))}.$$

Note that this amounts to applying a constraint force along the surface normal. By repeatedly using the original position $\vec{x}(t)$ and the most recent improved position $\vec{x}''(t + \Delta t_p)$ in place of $\vec{x}'(t + \Delta t_p)$, the constrained position can be calculated iteratively until $\sigma = 0$ up to a specified tolerance, for which we use machine precision in this work.

Attractive particles interact through a screened attractive log potential, such that the force between particles is given by

$$\vec{F}_{ij} = \begin{cases} -\alpha \frac{\vec{r}_{ij}}{|\vec{r}_{ij}|^2} \left(1 - \frac{s}{2r}\right) & 0 < s < 2r \\ 0 & s \geq 2r \end{cases} \tag{3.7}$$

where $\vec{r}_{ij} = \vec{x}_i - \vec{x}_j$ is the particle center-to-center separation, $s = |\vec{r}_{ij}| - 2r$ is the particle surface-to-surface separation, and α is a factor determining the strength of the attraction, which is set as described above. The force is screened so that particles in contact feel the full unscreened force, and the force decays linearly as the particle separation increases, falling off with a decay length equal to twice the particle diameter.

The initial particle movement timestep is set to $\Delta t_p = 1$, and the diffusion

constant is chosen so that a purely stochastic step would, on average, have a step size $\Delta x = \frac{2r}{1000}$ in a given direction, giving $D = (\Delta x)^2$. This determines the total simulation time, $\tau_r = \frac{2r^2}{D(\tau_d/\tau_r)}$. In order to enforce hard particle constraints and mimic particle reflections upon collision, individual particle steps which result in overlap between particles are rejected. To this end, all forces are calculated at the beginning of a given particle timestep, and particle steps are attempted in a random order.

Throughout the simulation, the surface is relaxed such that it maintains an ellipsoidal shape, the volume is conserved, and the surface area follows an exponential decay. The decay has the form $A = c + d\exp(-\lambda t)$ where c and d are chosen such that when the surface reaches a spherical shape at $t = \tau_r$, the area has decreased by 90% of the total amount the decay would have led to in the limit $t \to \infty$ were it possible to continue an exponential decay (i.e. at $A = c + d/10$); particles are then reprojected to the nearest point on the surface. Where reprojection causes overlap of particles, the particle positions are collectively evolved along the surface using the gradient descent method that is also employed during the initial particle displacement, described above.

An adaptive timestep scheme is used to evolve the system. Surface evolution steps are made less often than particle movement steps. Initially, the particle movement timestep is set at $\Delta t_p = 1$, and the surface evolution timestep is set such that it would take $N_r = 10^5$ surface relaxation steps to relax to a spherical shape, so $\Delta t_r = \tau_r/N_r$. We also require that there is at least one diffusion step per relaxation step: if the number of diffusion steps per relaxation step $\frac{\tau_r}{N_r \Delta t_p}$ is less than 1, then we set $\Delta t_p = \Delta t_r$. Upon a relaxation step, if the gradient descent scheme takes more than 10^4 steps to undo any overlap, then the simulation is reset to its state after the previous surface relaxation step, and the timesteps Δt_r and $\Delta t_{\text{overlap}}$ are all decreased by half and the

simulation continues.

As the packing becomes closely packed, Δt_p is adjusted based on the average interparticle spacing $\langle s \rangle$. To determine the average spacing, the separation between the three nearest neighbors of each particle are identified(as a particle needs at least three contacts to be constrained), and the average of these distances is taken. We then set the particle timestep so that the average particle displacement due to diffusion only will be 10 times this spacing, $\Delta t_p = 100\langle s \rangle^2/(2D)$. We choose this relatively large timestep because lower timesteps cause the simulations to run for a prohibitively long time. Having investigated the acceptance ratio for particle moves over the course of the simulation, we find that it levels out at a value of approximately 0.1 when the typical interparticle spacing $\langle s \rangle \approx 10^{-4}$. This is acceptable, as the packing is at a high density at this point and its evolution will be driven mainly by surface evolution and the associated particle overlap gradient descent steps, though some movement due to diffusion and other forces is still allowed. We also set a minimum timestep, corresponding to a particle spacing of $\langle s \rangle = 10^{-7}$.

The system is considered arrested and the simulation halts when the relaxation timestep are reduced to $1/2^{14}$ of their original value, i.e. after 14 timestep reductions. An outline of the algorithm is given in algorithm 3.1.

3.1.3 Soft Particle Monte Carlo Algorithm

In order to investigate packings of soft particles, we employ two algorithms, the first of which is a Monte Carlo Metropolis algorithm based on the inflation packing algorithm. We use a simulated annealing method to find near-optimal energy minima, although it should be noted that the energy as a function of the particle coordinates is highly complicated with many local minima and global energy optimization is not guaranteed.

Algorithm 3.1 Dynamic packing algorithm

```
if continuing from a previous simulation then
  read data
else
  place particles randomly
  equilibrate particles
end if
while surface has not finished relaxing and packing is not jammed do
  if time t > next relaxation time then
    update surface shape
    project particles onto new surface
    attempt to use gradient descent to remove particle overlaps
    if overlaps are succesfully undone then
      save configuration and set next relaxation time
    else
      reset configuration to state after previous relaxation
      reduce relaxation timestep
      if the relaxation timestep has been reduced 14 times then
        the packing is jammed; halt the simulation
      end if
    end if
  end if
  calculate terministic and stochastic forces on each particle
  in a random order, integrate the forces on each particle, rejecting any
  move that results in overlap.
  calcualte typical interparticle spacing and adjust the simulation timestep
  increment time t
end while
```

In order to test potentials of different softness, the interparticle potentials are set as either $U_{int} = d^{-1}$ or $U_{int} = d^{-6}$ (where d is the center-to-center distance between particles). The particles diffuse similarly to the inflation algorithm with two differences: the average step size σ is constant for all moves, and moves are accepted or rejected using a Metropolis scheme[34], with acceptance probability

$$P = \begin{cases} 1 & \Delta U < 0 \\ \exp(-\Delta U/k_B T) & \Delta U > 0 \end{cases} \tag{3.8}$$

where ΔU is the change in the system energy after a single particle move, k_b is the Boltzmann constant, and T is the temperature. The initial temperature is set by using a rough estimate of what the energy of a single particle in the final configuration will be assuming six-fold ordering and that nearest neighbor interactions dominate: $T_0 = 6U_{int}(2r_{est})/k_B$, where $r_{est} = \sqrt{A/N}$ is an estimate of the average particle separation. The system is annealed by multiplying the temperature by 0.99 after every 100 sets of diffusion moves until $\exp(-\Delta U/k_B T) \to 0$ within machine precision. After every 100 sets of diffusion moves, the change in energy is recorded and the simulation halts once this change in energy is less than 1×10^{-16}.

3.1.4 Soft Particle Gradient Descent Algorithm

We use a second, deterministic soft-particle packing algorithm to study the stability of hard particle packings, as well as to generate and minimize the energy of soft particle packings.

The algorithm takes as input an inter-particle energy function, the associated force function, and the ellipsoid axis lengths. It can also take a pre-

existing particle configuration, or can generate a random one using random sequential adsorption by using a radius such that a specified coverage fraction ϕ is achieved ($\phi = 0.35$ in our simulations).

The algorithm then proceeds using either a gradient descent method or a conjugate gradient method, as chosen for the specific application. Given a particle configuration, the force vector on each particle is calculated, which taken together as a $3n$-dimensional vector correspond to the negative gradient of the configuration energy, $\vec{F}$. For gradient descent, steps are taken along the direction of the gradient. For conjugate gradient, steps are taken along a direction $\vec{S}$ chosen as follows. If it is the first step, set $\vec{S}_1 = \vec{F}$. Otherwise, the new step direction is $\vec{S}_i = \vec{F}_i + \beta_i \vec{S}_{i-1}$, where $\vec{F}_i$ is the current gradient, and β_i is given by

$$\beta_i = \frac{\vec{F}_i \cdot \vec{F}_i}{\vec{F}_{i-1} \cdot \vec{F}_{i-1}}.$$

The intuition behind this approach is that information from previous steps is used to build up information about the curvature of the objective function. This way, for instance, if the minimum is in a long narrow valley, the minimization will not waste time hopping back and forth along the valley, but will take steps along the length of the valley. For an N-dimensional objective function, new curvature information is no longer useful after N steps, so the process must be reset periodically to use the pure gradient. We do this every $N/3 = n$ steps.

A move is taken in the direction of the force, choosing a timestep Δt using a line search, i.e. a given a fixed direction, a minimum (or approximate minimum) is found along that direction. For each move, an unconstrained step $\Delta\vec{x} = \vec{F}\Delta t$ is performed which is then projected onto the surface using Lagrange multipliers, as described in 3.1.2 above. The line search is first performed using the golden bracket method to bound the timestep around an

energy minimum, and then a parabolic interpolation is used to approximate the minimum within that bracket[34].

The golden bracket method works as follows: the current energy E_0 is calculated. Given an initial guess for the timestep Δt_{a0}, the configuration energy E_a is found at that timestep (where the energy is calculated after both taking a gradient step and enforcing the surface constraint). A larger timestep $\Delta t_{b0} = \Phi \Delta t_{a0}$ (where $\Phi = \frac{1+\sqrt{5}}{2}$ is the golden ratio) is then checked, with the resulting energy E_b. If $E_a > E_0$, then the timesteps are too large and the timesteps $\Delta t_{b0} = \Delta t_{a0}$ and $\Delta t_{a0} = \Delta t_{a0}/\Phi$ are checked. If $E_b < E_a$, then the timesteps are too small and the timesteps $\Delta t_{a0} = \Delta t_{b0}$ and $\Delta t_{b0} = \Delta t_{b0}\Phi$ are checked. This adjustment of the timestep is repeated until $E_a < E_0$ and $E_a < E_b$, i.e. a timestep Δt_{b0} is found such that there must be a local energy minimum at some time $0 < \Delta t < \Delta t_{b0}$. The golden ratio is used to adjust the timesteps as it can be shown that this is the most efficient adjustment factor[34].

Once a minimum has been bracketed, a quadratic interpolation is used on the bracketing points to approximate the minimum within the bracket. We do this for only one iteration, as we find that multiple iterations does not provide a significant increase in performance.

Once the timestep has been found, the configuration is evolved based on the magnitude and direction of the gradient, as well as the timestep: $\Delta\vec{x} = \vec{F}\Delta t$. After each step, the energy of the configuration is recorded. If the percent change in energy is below a specified tolerance $\epsilon_{\Delta E}$ generally taken to be machine precision, $(E^n - E^{n-1})/E^{n-1} < \epsilon_{\Delta E}$, then the algorithm is determined to have found a local minimum of the energy landscape and is halted.

3.2 Packing analysis

3.2.1 Particle Triangulation

Topological defects in a system with six-fold ordering are identified based on a triangulation of the particle positions. In such a triangulation, bonds between neighboring particles are identified, and these bonds become the edges in a graph in which triplets of neighboring particles are connected by edges which form triangles. These triangles must tile the space in which the packing is imbedded. For a given particle, the number of neighbors to which it is connected is counted giving the coordination number c, and the topological charge quantifies the deviation from six-fold ordering, $q = 6 - c$.

For a perfect 2D hexagonal lattice, this triangulation trivially consists of equilateral triangles. For more general packings in a flat 2D space, a Delaunay triangulation can be used as a unique way to identify the topology of a packing. In a Delaunay triangulation, a circle which circumscribes a given triangle will not contain any vertices other than the three which belong to that triangle, as illustrated in fig. 3.2a. Delaunay triangulations also have the property that they maximize the minimum angle of all triangles. This property can be used to build a Delaunay triangulation from a generic triangulation of a set of points by the process of edge-flipping. Given two triangles which share an edge (and together make up a quadrilateral) the shared edge can be removed and an edge joining the other two vertices of the quadrilateral can be inserted. If this edge flip increases the minimum angle within the two triangles it is accepted (as in fig. 3.2b), otherwise it is rejected. This edge-flipping process can be iterated over pairs of adjacent triangles until no more improvement is possible, resulting in a Delaunay triangulation.

Delaunay triangulations, however, do not generalize straightforwardly to

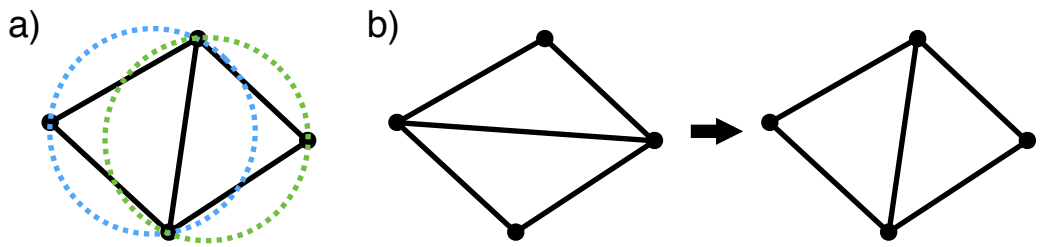


Figure 3.2: A small Delaunay triangulation of four points. a) The two triangles making up the triangulation can each be circumscribed by circles which contain only the three vertices of their respective triangles and no other vertices. b) Given an initial triangulation which is not a Delaunay triangulation, an edge-flip can be performed to recover the Delaunay triangulation, which consists of triangles which clearly have larger minimum angles.

sets of points on curved surfaces. In this work, we construct well-defined triangulations through a combination of a ball-pivoting algorithm (available in the software *Meshlab*) and a modified edge-flipping algorithm.

The ball-pivoting algorithm[35] is based on the intuition provided by the first property of Delaunay triangulations explained above, i.e. that each triangle can be circumscribed by a circle which contains no other points. The algorithm uses a sphere of radius r and begins by placing the sphere so that it is in contact with three points and contains no other points. It proceeds by rotating the sphere about one edge of the contained triangle until the surface of the sphere hits another point. The algorithm continues in this manner, rotating the sphere about the edges that make up the boundary of the triangulation until the triangulation is complete. Occasionally the algorithm will not find a closed triangulation. We check this by making sure all points are included in the triangulation, and that the defect charge indicates that the mesh has the correct topology, i.e. $c = 12$. If not, the algorithm is run again with a random ball radius until a complete triangulation with the correct topology is found.

Once the ball-pivoting algorithm has been used to build an initial triangulation, the software *Morpho* is used to perform an equiangulation through

edge-flipping. This edge-flipping process is similar to the one described above for Delaunay triangulations, in that it proceeds by iterating over each edge in the triangulation and attempting to flip it until no more improvement can occur. Here, however, the criterion for accepting or rejecting a flip is different. An equiangulation norm which penalizes triangles which deviate from an equilateral triangle is defined as

$$L = \sum_{i=1}^{6} (\frac{\pi}{3} - \theta_i)^2$$

where the sum is taken over the six angles of the two triangles adjacent to a given edge. Any edge-flip which causes this norm to increase, or which causes the edge length to increase is rejected. In this way, both edge flipping methods penalize long, narrow triangles with sharp angles.

3.2.2 Hexatic Order Parameter

Hexatic order is characterized by short range six-fold rotational symmetry[4]. Hexatic order is often defined across an entire system; however we are often interested in how ordering varies spatially and so we calculate a local hexatic order parameter. The local hexatic order is defined for a single particle as

$$\Psi_6 = \sum_{n=1}^{N} \exp 6i\theta_n$$

where the sum is taken over the neighbors of a particle, and θ_n is the angle between some reference direction (here taken to be along the local tangent vector in the polar direction) and the bond connecting the central particle and its neighbor. This situation for one particle and its neighbors is illustrated in fig. 3.3. There is some ambiguity in defining the neighbors of a particle. Depending on the context, we use one of two definitions: we either take the

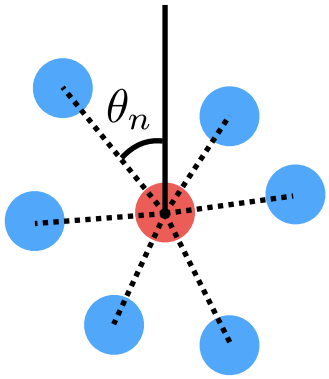


Figure 3.3: A schematic illustrating the calculation of a local hexatic order parameter. The separation vectors from a given particle (red) to its neighbors (blue) are found, and the angles from some reference direction are found and used to calculate the order parameter. The hexatic order parameter is equivalent to the amplitude of the sixth Fourier mode of the angular positions of a particle's neighbors.

six nearest particles to be the neighbors, or we use a particle separation cutoff, so that all particles with a distance $x < 2r + \delta$ are considered neighbors. Once the local hexatic order has been calculated for each particle in a packing, the spatial average over some regions can be calculated.

3.2.3 Rattler Removal

Generically, one finds that an arrested packing contains some number of particles which are not held in place by their neighbors, which are known as rattlers[26]. In the language of the jamming literature, these packings are not locally jammed, though they may contain a locally jammed sub-packing which excludes the rattlers. In assessing the stability of a packing, rattlers do not contribute mechanical contacts and thus are irrelevant. For this reason, it is desirable to remove rattlers for certain packing analyses.

In order for a particle to be locally jammed in n dimensions, it must have $n + 1$ contacts which are not in the same n-dimensional hemisphere. This is illustrated in $2D$ in fig. 3.4. We use this criteria to check whether individual

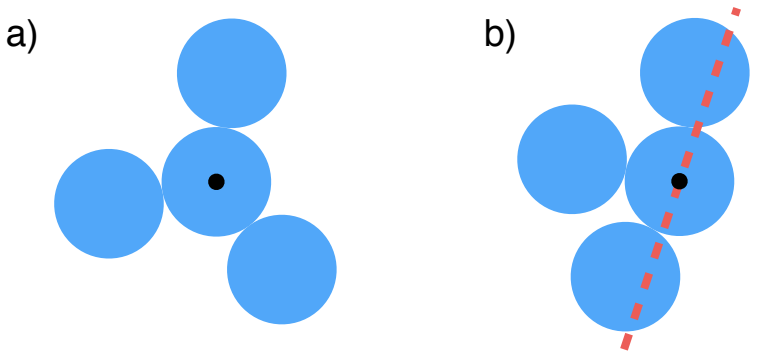


Figure 3.4: a) A particle which is locally jammed, as it is trapped between three neighbors. b) A particle which is in contact with three neighbors, but which is not locally jammed because the contacts are all within the same hemisphere.

particles are rattlers. In a given packing, each particle is checked one-by-one. For a given particle, all neighbors within some center-to-center distance $(1+\delta)2r$ are identified, where $\delta = 10^{-6}$ is used, as it is slightly higher than the typical inter-particle spacing. If there are less than three neighbors, the particle is immediately identified as a rattler. Otherwise, for each neighbor, the separation vector from the central particle to that neighbor is found, and projected onto the local tangent plane of the central particle. The angles between these separation vectors are then found, and if any angle is greater than π radians, the particle is recognized as a rattler.

Because removing a rattler which is in contact with another particle may then cause the other particle to become a rattler, the rattler removal program repeatedly sweeps over all particles until it completes a sweep without removing any rattlers. Once this occurs, the rattler removal program halts and a rattler free, locally jammed packing has been produced. It is also possible that the rattler removal program will remove all particles in the case that the packing did not contain a locally jammed sub-packing.

3.2.4 Linear Program for Unjamming

As discussed in the background chapter, hard particle packings can belong do different jamming categories. Locally jammed packings are stable against movements of single particles. A packing with all rattlers removed will be locally jammed. Packings can also be collectively jammed, in which case there are no collective motions of particles which can unjam the packings, and they can be strictly jammed, in which case collective motions combined with boundary deformations cannot unjam the packing. As will be demonstrated in chapter [], packings can also be metric-jammed, in which case collective particle motions plus a deformation of the underlying metric cannot unjam the packing.

In order to uncover collective unjamming motions, we employ a linear programming approach developed by Donev[26], which we have extended to account for curved surface constraints. Essentially, Donev's algorithm finds allowed particle motions under the application of a random load to the particle configuration being tested. This is done while enforcing the appropriate constraints (i.e. particles cannot overlap) to linear order, and also ensuring the displacement found is bounded. We extend this algorithm by enforcing that particle motions are perpendicular to the surface normal.

More precisely, for a particle configuration $\mathbf{R}$ the algorithm aims to find an allowed unjamming displacement $\Delta\mathbf{R}$. In order to find allowed motions of the particles, a random load $\mathbf{F}$ is applied to the packing. This load consists of a random force applied to each particle, whose cartesian components are sampled from a Gaussian distribution of unit variance. The goal of the algorithm is to then maximize the virtual work done by the load:

$$\max_{\Delta\mathbf{R}} \mathbf{F}^{\mathrm{T}}\Delta\mathbf{R}. \tag{3.9}$$

This optimization is of course subject to the constraints that particles cannot overlap, and they must move along the surface. These constraints can be linearized, giving a linear programming problem which can be solved to find allowed unjamming displacements.

In order to prevent overlap, neighboring particles must obey the equation

$$(\Delta\mathbf{r}_i - \Delta\mathbf{r}_j)^{\mathrm{T}}\mathbf{u}_{i,j} \leq \Delta l_{i,j} \tag{3.10}$$

where $\Delta\mathbf{r}_i$ is the displacement of particle i, $\mathbf{u}_{i,j}$ is the unit vector separating particles i and j, and $\Delta l_{i,j}$ is the interparticle separation between particles i and j. This equation is only applied to particle pairs which are close to one another and may come into contact ($\Delta l_{i,j}/2r < \delta$ where δ is typically chosen to be around 0.01). This equation requires that two particles cannot move towards each other by an amount greater than their initial separation. This equation can be rewritten by defining a rigidity matrix $\mathbf{A}$: each column of $\mathbf{A}$ corresponds to a neighboring particle pair $\{i, j\}$ and each row corresponds to a particle coordinate. The only nonzero elements in a given column $\{i, j\}$ are in the rows for particles i and j: the rows for particle i will contain $\mathbf{u}_{i,j}$ and the rows for j will contain $-\mathbf{u}_{i,j}$. Then, eq. 3.10 can be written

$$\mathbf{A}^{\mathrm{T}}\Delta\mathbf{R} \leq \Delta\mathbf{l} \tag{3.11}$$

where $\Delta\mathbf{R}$ is a vector containing the particle displacements $\Delta\mathbf{r}_i$ and $\Delta\mathbf{l}$ is a vector containing the $\Delta l_{i,j}$ for each particle pair.

To satisfy the surface constraint to first order, the particle displacements must be along the local tangent plane of the surface. This condition is satisfied when

$$\mathbf{N}^{\mathrm{T}}\Delta\mathbf{R} = 0 \tag{3.12}$$

where $\mathbf{N}$ is the collection of surface normal vectors at each of the particle positions.

Finally, because spheres and ellipsoids with axial symmetry have trivial rotational symmetries, we constrain the first moment of the motion $\Delta\mathbf{R}$ about the symmetry axes to suppress trivial motions. For an ellipsoid symmetric about the z-axis, the sum of the cross products of the displacements Δr_i with their radial components of the displacement vector from the z-axis $r_i^{\perp z}$ must be zero. Or,

$$(\mathbf{R}^{\perp z})^{\mathrm{T}} \times \Delta\mathbf{R} = 0 \tag{3.13}$$

where $\mathbf{R}^{\perp z}$ is the vector composed of the radial components of the particle positions.For a packing on a sphere, the corresponding constraints are added for rotations about the x- and y-axes.

One final constraint is imposed. Because the surface constraint is imposed only to first order, large displacements will move particles off of the surface. As such, we require that each component of $\Delta\mathbf{R}$ has a magnitude less than or equal to 0.01 (keeping in mind that that semi-major and minor axis lengths of the surface are of order 1):

$$|\Delta\mathbf{R}_i| \leq 0.01. \tag{3.14}$$

Eq. 3.9 combined with the constraints of eqs. 3.11-3.14 together make up a linear program which can be solved through standard software, e.g. *Mathematica* to find motions which may unjam an arrested particle packing.

One finds, however, that because the linearization gives a result which only approximately satisfies the constraints, care is needed in order to apply the unjamming motion and find configurations which are clearly unjammed while still obeying the constraints. In order to do this, a combination of the linear

program and the soft-particle gradient descent method described in section 3.1.4 is used.

First, a motion $\Delta\mathbf{R}$ found by the linear program is applied to move the particles along the surface. A move $\mathbf{R}' = \mathbf{R} + \alpha\Delta\mathbf{R}$, where initially $\alpha = 1$, is attempted, with the surface constraint enforced using Lagrange multipliers as in subsection 3.1.2. If this new configuration has an overlapping particles, the move is undone and attempted again with α reduced by half. This is repeated until an updated configuration with no overlap is found.

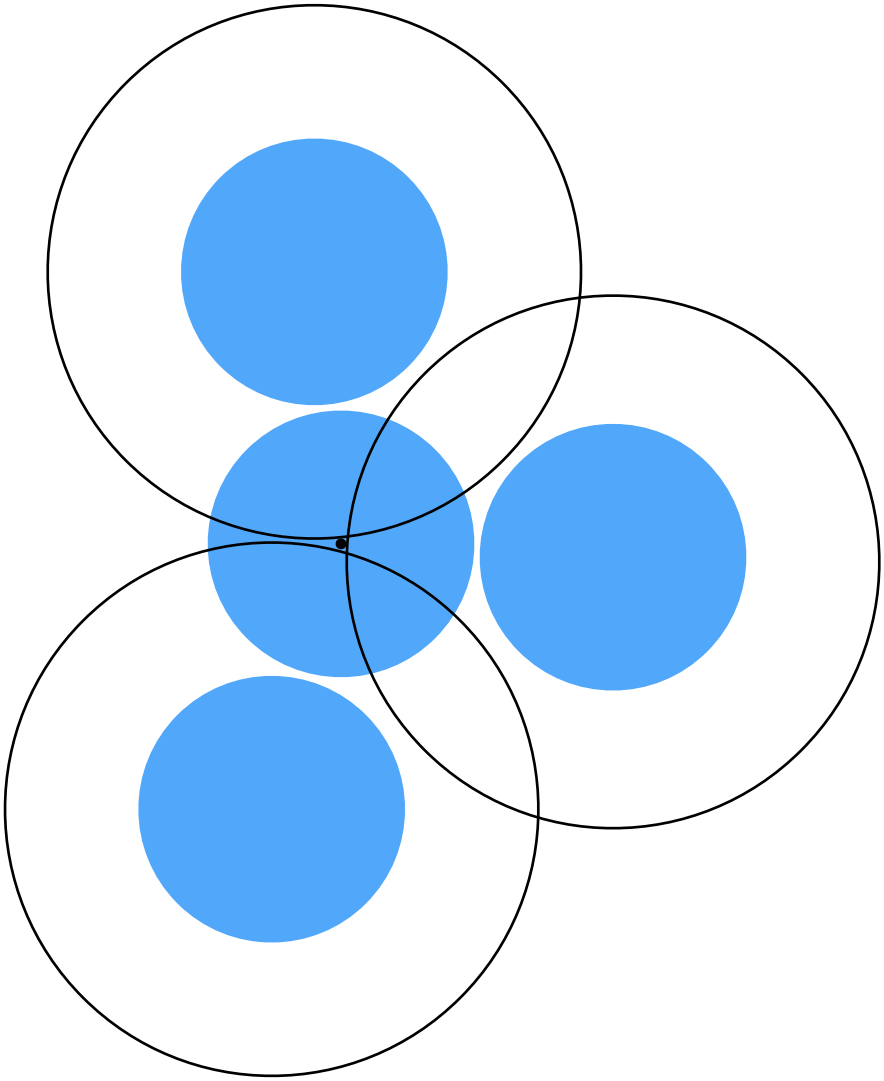
After applying the unjamming motion, the soft-particle gradient descent algorithm is applied to the new configuration for 20 gradient descent steps. For this gradient descent method, a pairwise potential of the form

$$V = \begin{cases} \infty & x \leq 2r \\ (x - x_c)^2 \left(\log(x - 2r) - \log(x_c)\right) & 2r < x < x_c \\ 0 & x_c \leq x \end{cases}$$

is used where x is the center-to-center particle distance and $x_c = 2.1r$ is a distance cutoff beyond which particles do not interact. Note that the logarithmic form of the potential diverges as particles come in contact with one another, preserving the hard-particle constraint.

The linear program and gradient descent methods can be applied back-to-back repeatedly to find significant unjamming motions. The combination of both algorithms consistently finds unjamming motions much more quickly than repeatedly applying linear program motions alone, as illustrated in fig. 3.6. The reason for this can be understood in terms of jamming polytopes. A jamming polytope is the region in configuration space available to a given packing from its current state. Fig. 3.5 gives an intuitive picture of what is

occurring in terms of the jamming polytope of a single particle surrounded by three other fixed particles. In the linear program, solutions are constrained to a linearized approximation of the jamming polytope. If a packing is near the border of its polytope, the linearized polytope may be quite restricted. However, after the gradient descent algorithm is applied, a packing will be moved closer to the center of its jamming polytope. From here, the linearized polytope will contain portions of configuration space which are closer to unjamming.



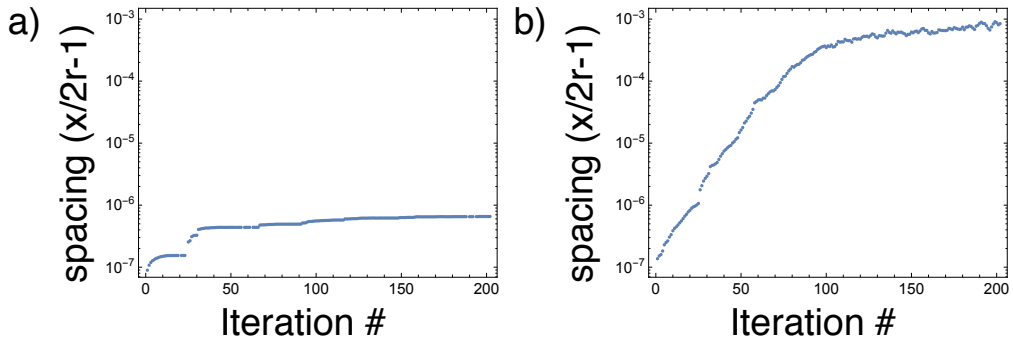


Figure 3.6: Typical particle spacing (in units of the particle diameter) versus unjamming iteration for a) the linear program only and b) the linear program combined with gradient descent steps. After the same number of iterations, while the linear program alone has led to some improvement, the combined linear program and gradient descent method has increased the interparticle spacing by several orders of magnitude. Here, the typical particle spacing is calculated by finding the particle surface-to-surface distance from each particle to its three nearest neighbors, and then taking the geometric mean of all of these distances.

Chapter 4

The role of curvature in the sphere packings on a static geometry

SPHERE packings on the surface of an arrested emulsion droplet are created through a complex dynamic process, and a variety of factors will influence their formation. These include the droplet shape evolution and forces such as interparticle attraction and gravity acting on particles. These effects will drive the evolution of the system and thus also affect the final particle ordering. They can, however, can be controlled to some extent and their influence can be reduced. A more fundamental influence which cannot be reduced is the droplet geometry. The final packing exists on a curved 2D manifold, and the curvature and topology of this manifold are incompatible with perfect ordering.

In this chapter we investigate the influence of geometry by creating packings on static surfaces. We do this by inflating particles on a fixed surface until they can inflate no more. As inflation proceeds, particles are diffusing as

well. The diffusion occurs on a shorter timescale than inflation — this allows particles to more effectively explore their configuration space, and leads to higher quality packings on average (as measured by quantities such as packing fraction and defect count. This idea will be explored in more depth in chapter 5.)

The surfaces studied are all ellipsoids. This geometry is chosen for several reasons. First, ellipsoidal droplets are straightforward to produce experimentally, for example by simply shearing a spherical droplet. From a theoretical standpoint, ellipsoids have interesting properties while still being relatively easy to describe mathematically. An ellipsoid is simply a sphere scaled along one axis. This results in varying curvature along the surface, as well as curvature anisotropy (i.e. different curvatures along different directions.)

The introduction of these interesting curvature features raises a number of questions about the packings that will be produced. How will defects be distributed across the surface, both in terms of defect charge and total defect count? How will defects in excess of the topologically required ones be introduced? What effect will curvature variation and anisotropy have on scars? Will it influence their length or orientation? How will the symmetry of the surface influence the symmetry of the packing? Another important feature of these packings is their hard-sphere nature — will hard sphere packings differ from soft-sphere packings?

We study the structure of the defects in the resulting packings, as identified by a triangulation of the particle coordinates. The spatial distribution of defects across the surface is measured and correlated with the Gaussian curvature of the surface. Excess defects in the form of scars are investigated as a function of particle number and aspect ratio. The orientation of scars along the surface is measured in order to identify couplings of scars to the

curvature anisotropy. We also measure the length of scars, finding that the scar transition seen on spherical surfaces is softened by surface anisotropy. We introduce a heuristic model to explain this. Finally, from a fine grained search of the particle number and surface aspect ratio parameter space, we identify a wide variety of symmetric particle packings.

To perform these analyses, we created two datasets consisting of packings of varying particle number and aspect ratio. One data set was used for studying the curvature-defect coupling and scar length, which consisted of packings with aspect ratio a varying from 1.2 to 4.0 in increments of 0.2 (for both the prolate and oblate cases: we consider the aspect ratio to be the ratio of the semi-major to semi-minor axis.) The particle number n was varied from 10 to 800 in increments of 10. Additional prolate packings were generated to study scar orientation, from aspect ratio 4.2 to 8.0 in increments of 0.2, from particle number 710 to 800 in increments of 10. 50 configurations were generated for each pair of parameters. The second data set was used for studying symmetry, where we are interested in lower particle numbers and a more fine-grained search of the parameter space. This data set consisted of packings with aspect ratio varying from 1.1 to 4.0 in increments of 0.1, and particle number varying from 3 to 200 in increments of unity. 80 configurations were generated for each pair of parameters.

4.1 Defect Distribution

We first examined the distribution of the defects as a function of the aspect ratio. Defect locations were determined by assigning a defect charge $q = 6 - c$ to each particle, where c is the coordination number determined from the Delaunay triangulation of the particle positions (see *Methods* chapter). The surface

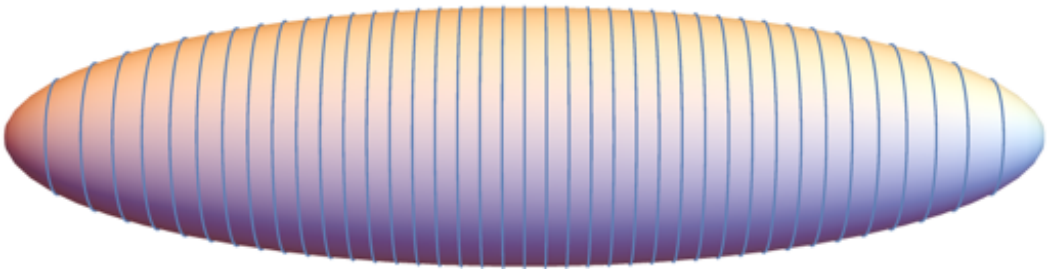


Figure 4.1: Illustration of the partitioning of the surface into equal area, axisymmetric regions.

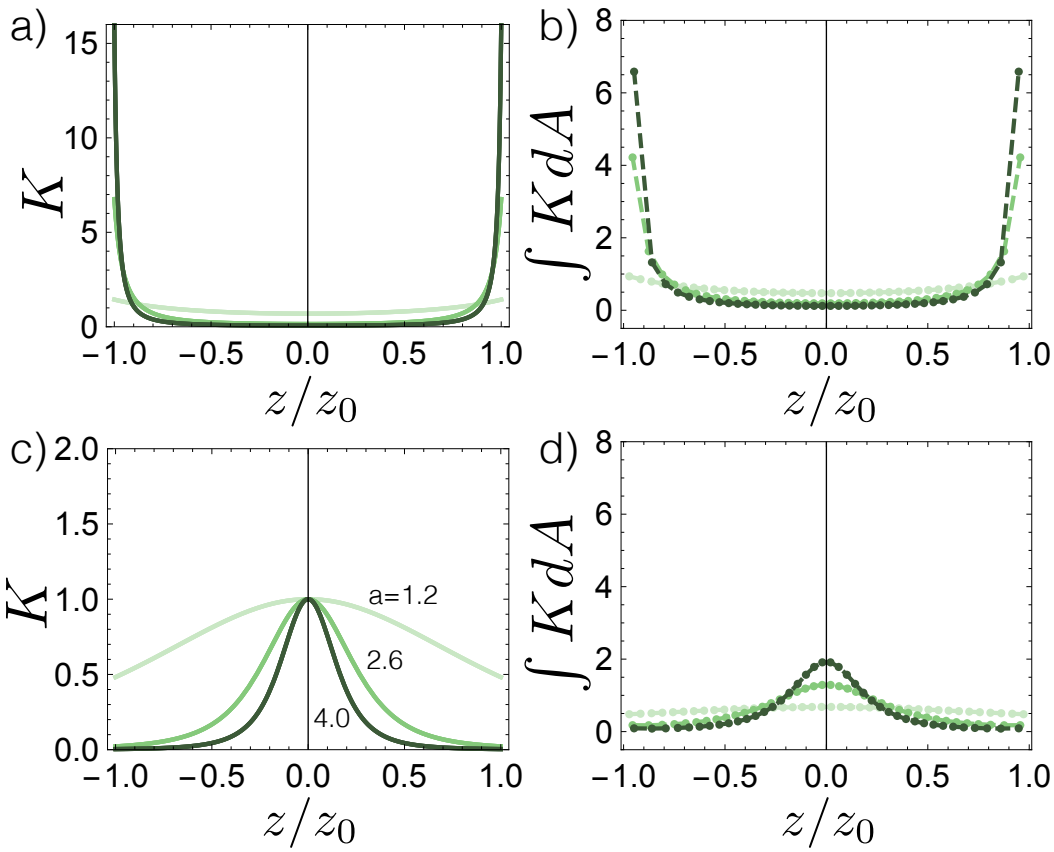


Figure 4.2: (a) Gaussian curvature as a function of position along the symmetry axis of a prolate ellipsoid for varying aspect ratios. (b) The integrated Gaussian curvature in a series of equal area bins (as illustrated in fig. 4.1) for a prolate ellipsoid. (c) and (d) show the same quantities for oblate ellipsoids. In each plot, the light green curve corresponds to an aspect ratio of 1.2, medium green corresponds to 2.6, and dark green corresponds to 4.0.

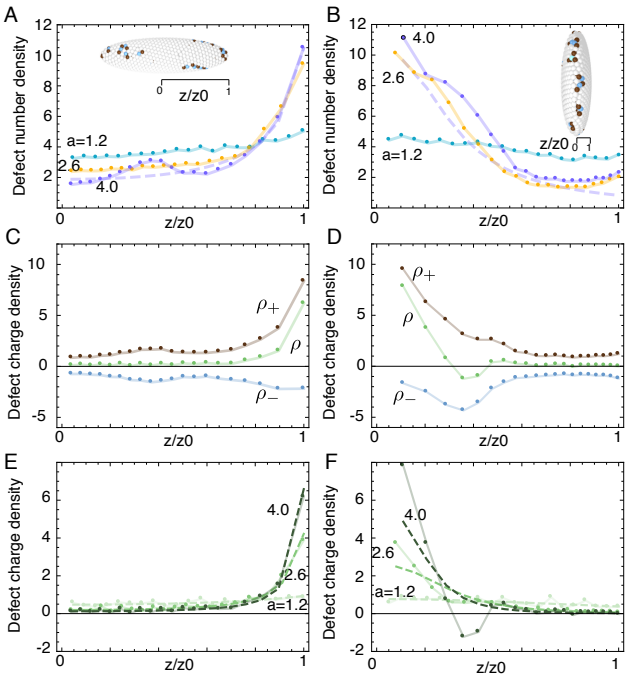


Figure 4.3: Defect number density for (A) prolate and (B) oblate ellipsoids with $n = 800$ varying aspect ratio: blue is 1.2; yellow 2.6; purple 4.0. Points with solid lines represent simulation data. Dashed lines represent the prediction of the model in eq. 4.2 for surfaces of aspect ratio 4.0. Note the small secondary peak near $z/z_0 = 0.4$ at $a = 4$ in the prolate case. Example configurations of $a = 4$ are shown as insets. Defect charge density is shown for (C) prolate and (D) oblate ellipsoids of $a = 4$. The green points represent the net charge density, and the brown and blue points represent the density of positive and negative defects, respectively. The secondary peak in (A) is also visible in the positive and negative charge densities in (C). In (D), there is a net negative defect charge density near $z/z_0 = 0.4$, despite the Gaussian curvature being positive. Net defect charge densities for different aspect ratios are compared to the integrated Gaussian curvature (dashed lines) for (E) prolate and (F) oblate ellipsoids. In all plots, densities are given in units of defect number or defect charge per equal-area segment, averaged over the ensemble of simulation results, with symmetric segments on opposite halves of a surface being combined. Lines are guides to the eye.

was partitioned into equal-area axisymmetric regions as illustrated in fig. 4.1 and the number of defects in each region was counted. Each segment has a different average Gaussian curvature, varying monotonically from the equator to the poles, as illustrated in fig. 4.2. Regions near the poles having larger curvature for prolate and the reverse for oblate ellipsoids. In fig. 4.3A for prolates and fig. 4.3B for oblates, the defect number density is shown as a function of the axial position z/z_0 averaged over the ensemble of simulations at fixed aspect ratio and particle numbers ranging from $710 < n < 800$. Generically, it is apparent that defect number density increases with the Gaussian curvature, as expected. For prolate ellipsoids at low aspect ratio, the defect number density increases monotonically with respect to the average K in each region. At higher aspect ratios, there is a small secondary peak in segments with low Gaussian curvature. We verified this occurs for other ranges of particle numbers $n > 210$.

In order to understand this, we plot separate defect charge densities for positive and negative defects in fig. 4.3C, as well as the net defect charge density. The anomalous peak is apparent in both the separate positive and negative defect charge densities, but not in the net defect charge density, indicating that the excess defects are taking the form of neutral dislocations or scars.

In fig. 4.3B, we see that for oblate ellipsoids, the defect density again increases near the more highly curved regions. Fig. 4.3D reveals, however, that the coupling between defect charge and curvature is again complicated: while there is a peak in positive defects at the highly positively curved edge of the surface, there is a high density of negative defects surrounding this, and the net defect charge density is actually negative near $z/z_0 = 0.4$.

Further investigation reveals that this is a finite size effect. For $n = 800$ and

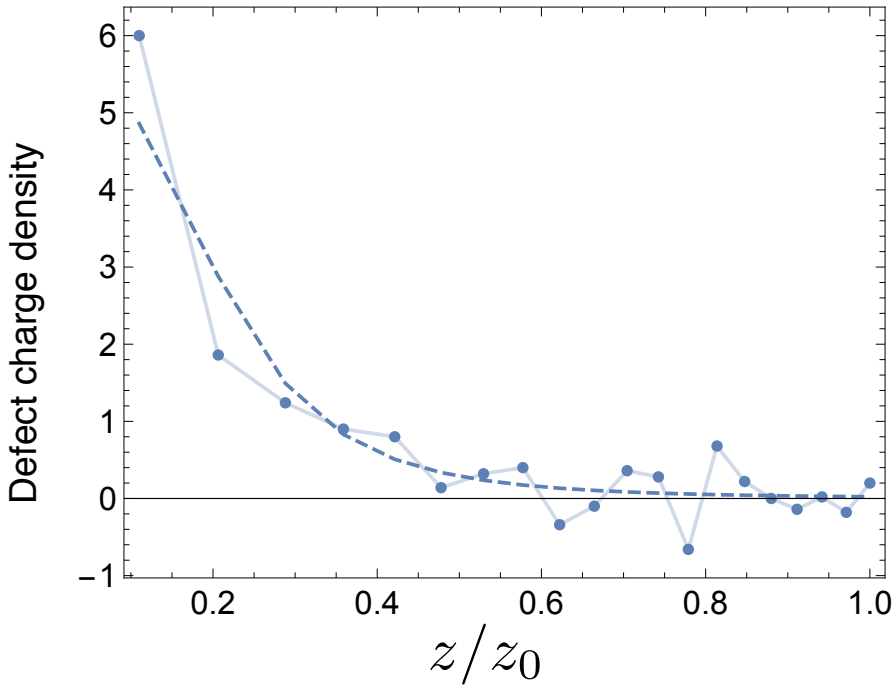


Figure 4.4: For packings of $n = 4000$ on oblate ellipsoids of $a = 4.0$, observed net charge density (points) and predicted net charge density (dashed line). The observed defect density much more closely matches the model than does that for $n = 800$.

$a = 4$ with 20 surface segments, the lateral size of the segments is smaller than the particle radius (near the equator, an arc length of 0.1136 compared to $r = 0.1976$). Upon visual inspection, we see that on highly oblate ellipsoids, scars tend to form along the highly curved equator (this is confirmed quantitatively in section 4.2). These scars tend to zigzag such that positive defects occur slightly closer to the equator.

To remove the finite size effect, 50 simulations were run with $n = 4000$ and $a = 4.0$ (oblate). The results are shown in fig. 4.4. With the particle radius now smaller than the size of the surface segments ($r = 0.08866$), the defect charge density now follows the expected trend.

Previous studies[14] have shown that the defect charge present on a curved

surface is determined by the integrated Gaussian curvature of that surface, such that

$$\frac{1}{\pi/3}\int KdA = \sum_{i=1}^{n} q_i. \tag{4.1}$$

To test whether this model is consistent with the results seen here, we plot the integrated curvature, normalized as in eq. 4.1, in each of the equal-area segments described above, and compare this with the net defect charge in those sections. The results for prolate ellipsoids of different aspect ratios are shown in fig. 4.3*E* as dashed lines, and there is good agreement between the model and the simulation data. The result for oblate ellipsoids with $n = 800$, shown in fig. 4.3*F*, do not show agreement. This is unsurprising, given the negative net defect charge present in the simulation results. The results for $n = 4000$, with the finite size effect removed, do show agreement.

This model can be extended to attempt to account for excess dislocations which occur in addition to the topologically required core disclinations. As discussed in the background chapter, excess dislocations appear in the form of scars in packings of particles on spherical surfaces when R/r, the ratio of surface radius to particle radius, is above a critical value, and above this value the scar length grows linearly with R/r[9, 10]. On a non-spherical surface such as an ellipsoid, there is not a single surface radius R, but one can estimate an effective local surface radius based on the Gaussian curvature as $1/\sqrt{K}$ (because $K = \frac{1}{r_1 r_2}$ where r_1 and r_2 are the local principal radii of curvature). One can then assume that for each core disclination (whose surface density is predicted by K), there is a scar made up of some number of dislocations n_s given by $1/(r\sqrt{K})$. We fit the scar length data for hard particles packed on spheres (discussed below in section 4.3, illustrated in fig. 4.8) to obtain a

function $n_s(R/r)$ and use this to model the defect number density ρ_n as,

$$\rho_n = \frac{K}{\pi/3}\left(1 + 2n_s\left(\frac{1}{r\sqrt{K}}\right)\right). \tag{4.2}$$

This model at first glance seems like a promising candidate to explain the non-monotonic nature of the observed defect distributions on prolate ellipsoids, as there is a competition between high disclination density and low scar length in regions of high K, and low disclination density and high scar length in regions of low K. Upon calculating the defect density, shown in fig. 4.3A, we see that, while this model accurately predicts the magnitude of the defect number density across much of the surface for prolate ellipsoids, it fails to capture the anomalous bump. The model does not accurately predict the defect number density for oblate ellipsoids; it under-predicts it across the entire surface (fig. 4.3B).

One possible explanation for the secondary peak in the defect distribution on highly prolate ellipsoids is that it is due to the hard-particle nature of the packings. To test this, we generate packings of soft particles and investigate the defect distribution. We employ the gradient-descent energy minimization program described in the methods chapter to generate packings of soft particles, with interaction energy of the form $V(r) = \frac{1}{r^6}$. 500 simulations were run, and the 50 results with the lowest energy were analyzed. The resulting defect number density is shown in fig. 4.5. There is no secondary bump in the soft particle defect distribution, but we also see other differences between the soft and hard particle packings. The soft particle defect distribution has a much sharper drop-off and then begins to level off as one moves away from the ends of the ellipsoid. The model in eq. 4.2 can be used to predict the defect number density of soft particles, by using the data for scar length on spherical

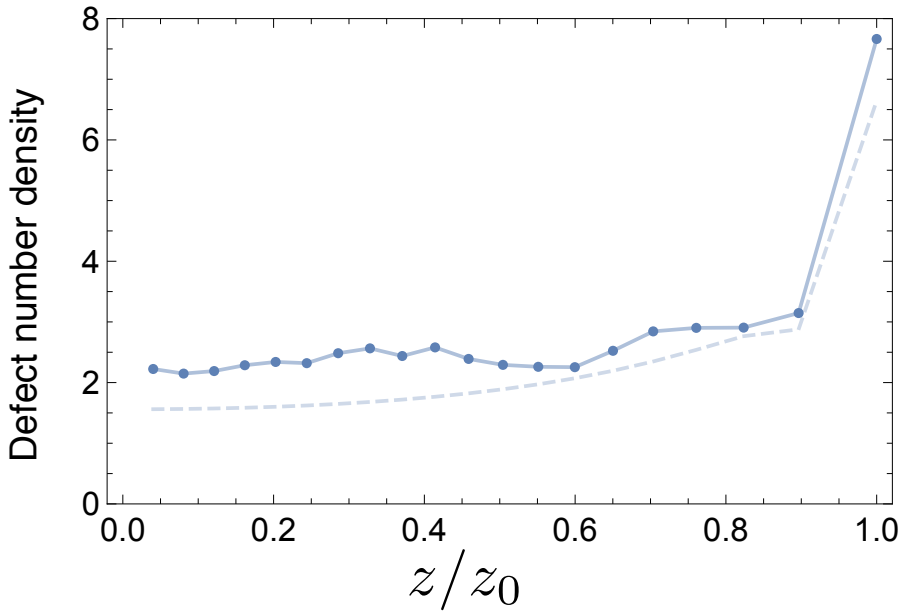


Figure 4.5: Observed defect number distribution (points) and model (dashed line). The model predicts the sharp change in the distribution near $z/z_0 = 0.9$. However, it under-predicts the number of defects closer to $z/z_0 = 0$. No secondary peak is observed.

packings of soft particles, shown in fig. 4.8 in section 4.3. The prediction is plotted alongside the observed distribution in fig. 4.5. It does predict the lack of a secondary peak and the sharper drop off. However, it under-predicts the defect density over much of the flatter regions of the surface, making it difficult to compare to the hard particle case and determine whether the secondary peak is due to the hard particle nature of the packings.

4.2 Scar Orientation

We next determine whether the scars are oriented by the curvature anisotropy of the surface. To do so, we consider a local scar orientational distribution function (ODF) $f(\alpha)$ where the angle α is measured locally in the tangent plane of the surface, relative to the polar direction (i.e. the direction pointing

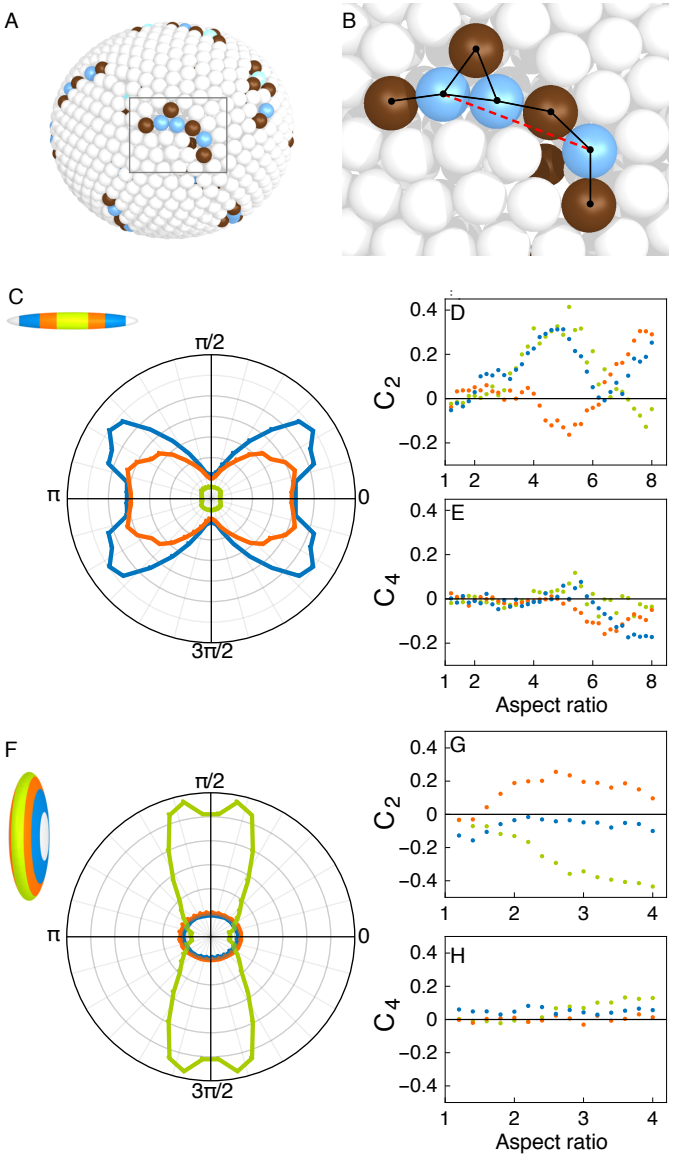


Figure 4.6: Orientation of the scars relative to the curvature anisotropy. (A) A configuration with a typical scar. (B) Close-up of the scar. Black lines show edges in a graph comprising the scar. The red dashed line shows a chain of length 3. Results are shown for (C-E) prolate and (F-H) oblate ellipsoids. The C_2 (D, G) and C_4 (E, H) order parameters for prolate and oblate ellipsoids, respectively, are plotted as a function of aspect ratio for different regions along the symmetry axis of the ellipsoid: green corresponds to the center, orange to the mid-region, and blue to the ends. (C) and (F) show the ODF of chains in the center, mid-regions, and ends of the ellipsoid, respectively, for prolate ellipsoids of aspect ratio 8 in (C) and oblate ellipsoids of aspect ratio 4 in (F). Insets of (C) and (F) illustrate the regions used for spatial binning.

along the uniaxial axis of the ellipsoid). The ODF may be expanded as a Fourier series,

$$f(\alpha) = \sum_n C_n \cos(n\alpha). \tag{4.3}$$

The average value of the first two non-zero coefficients, $C_2 = \langle cos(2\alpha) \rangle$ and $C_4 = \langle cos(4\alpha) \rangle$, were calculated for our ensemble of packings (odd coefficients are zero due to the symmetry of scars under rotation by π). These quantities are order parameters for orientational order as they vanish if the scars align isotropically with the curvature. C_2 quantifies nematic order, i.e. uniaxial orientational order and C_4 quantifies quadrupolar order.

To determine the scar orientation, we studied contiguous chains of defects as shown in fig. 4.6*A* and *B*. Given a packing and its Delaunay triangulation, the neighboring defects around each defect are identified. These adjacent pairs become the edges of graphs of contiguous defects. Two defects are identified as the ends of a *chain* of length l if they are within a connected graph of defects and the shortest path between them contains l edges. Note that a single defect graph can contain multiple chains. For example, the graph in fig. 4.6 contains 4 chains of $l = 3$. Once a chain of length l is identified, its orientation relative to the local principal directions—i.e. the polar and azimuthal tangent vectors $\vec{t}_\theta$ and $\vec{t}_\phi$, respectively (see eq. 3.1 in the Methods chapter for the parametrization of the surface)—is calculated thus: given a pair of chain endpoints, their separation vector is projected onto the local tangent plane at each endpoint, giving components along $\vec{t}_\theta$ and $\vec{t}_\phi$. These components are then averaged between the endpoints, and the angle α that the resulting vector makes with $\vec{t}_\theta$ is recorded as the orientation of the chain. The z-component of the midpoint of each chain (the average of it endpoints taken in 3D space) is recorded as the chain's position and this is used to examine how the coupling varies across the surface.

The analysis was applied to an ensemble of simulation results as follows: For a given aspect ratio, the orientations of all chains of length l are collected across simulations with $N \in [710, 800]$ in increments of ΔN=10 (with 50 results at each N resulting in 500 simulations per aspect ratio). Order parameters C_2 and C_4 are then calculated from this ensemble. Because the curvature anisotropy varies with the z-coordinate along the surface, results can be divided according to their position in order to measure effects due to the variation in curvature across the surface. In our analysis, we exclude scars in the regions near the poles which make up 10% of the surface area as here the curvature tensor is degenerate and the alignment is undefined. The remainder of each half of the surface is broken into three equal-area, azimuthally symmetric regions, as illustrated in the insets of fig. 4.6 *C* and *F*, and data from symmetric regions on opposite halves of the ellipsoid are combined. A chain length of l=3 was used as this is long enough to capture scar behavior while having enough chains for statistical purposes. Shorter chain lengths show a weaker tendency to orient.

The behavior exhibited by prolate ellipsoids is rather complicated, as seen in the plots of order parameter versus aspect ratio in fig. 4.6 *D* and *E*. In the center region near the equator, scars are nematic along the $\vec{t}_\theta$ direction between aspect ratio 3.6 and 6. At higher aspect ratio this center region is very flat, leading to fewer scars, and so any orientational order is insignificant. In the mid-regions between the equator and poles, scars become nematic along the $\vec{t}_\phi$ direction at aspect ratio 4.4, and then transition to nematic along the $\vec{t}_\theta$ direction at aspect ratio 6.4. Scars near the poles show nematic order along $\vec{t}_\theta$ above aspect ratio 2, although this order peaks near aspect ratio 5, then drops to C_2=0 at aspect ratio 6.4 before increasing again. Interestingly, scars on highly prolate ellipsoids can also show C_4 order. This appears in the mid

regions above aspect ratio 5.2, and in the end regions above aspect ratio 6.

The chain ODFs for prolate ellipsoids of aspect ratio $a = 8$ in fig. 4.6C illustrate the trends that appear at high aspect ratio. It is apparent from the green curve that that there are few chains in the relatively flat center of the ellipsoid. The orange curve shows a high degree of nematic order directed along the polar direction in the mid-region, and the blue curve for the ends shows nematic order along the polar direction, as well as a peak between the directions of principal curvature, which is indicative of negative C_4 order.

The case of scar orientation on oblate ellipsoids is more straightforward. The order parameters are plotted as a function of aspect ratio for different azimuthally symmetric regions across the surface, in fig. 4.6 G and H. Scars at the equator exhibit a high degree of nematic order in the $\vec{t}_\phi$ direction, which increases linearly with aspect ratio up to $a = 4$. This is unsurprising, because the curvature on highly oblate ellipsoids is localized to a nearly one-dimensional region around the equator of the ellipsoid, so one expects the scars to form there, aligned along the equator. There is also a small degree of C_4 ordering. In the regions midway between the equator and poles, there is a weak coupling of scars along the $\vec{t}_\theta$ direction. These trends are illustrated for $a = 4$ in fig. 4.6F: the green curve for the edges displays a peak near the azimuthal direction, whereas the orange and blue curves show that there are fewer chains without much order in the flatter regions.

While the scar orientation results for the oblate case are easily understood, the ordering of the scar orientation on prolate ellipsoids is far more complicated. The orientation varies greatly depending on chain position and ellipsoid aspect ratio. Especially surprising is the emergence of C_4 ordering, which corresponds to a tendency for chains to align in a direction intermediate to the directions of principal curvatures.

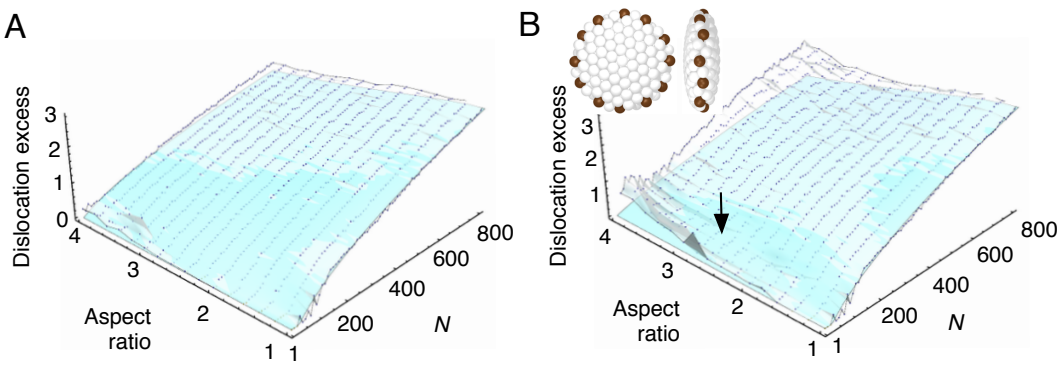


Figure 4.7: The number of excess dislocation defects per scar on (A) prolate ellipsoids and (B) oblate ellipsoids. Points and white surface represent the simulation data. The blue surface represents the prediction of the model in eq. 4.6. For low aspect ratio near 1, there is a clear scar transition, which is not present at aspect ratios far from 1. The inset in (B) shows a highly commensurate oblate packing with $n = 140$ and $a = 2.6$. Note that data for oblate ellipsoids with $n = 10$, $a \geq 2.0$ and $n = 20$, $a \geq 3.0$ has been excluded.

4.3 Scar Transition

As discussed in the background chapter, packings of spheres on spherical surfaces exhibit a transition: For low particle numbers, only the twelve defects required by topology are present; above a critical particle number n_c, it is favorable for larger defect structures to occur, typically chains of scars extending from a core disclination. Increasing n above n_c leads to a monotonic increase in average scar length[9, 10].

From our simulation results of packings with $10 \leq n \leq 800$, we calculated the average number of excess dislocations per topologically required disclination for each (a, n). Defects were weighted in the analysis by the absolute value of their charge. Given that there are two disclinations per dislocation, and 12 core disclinations, the number of excess dislocations per scar is calculated thus,

$$n_d = \frac{1}{2}\left(\frac{\sum_i |q_i|}{12} - 1\right), \tag{4.4}$$

where the sum is taken over all defects. This quantity captures the same information as the scar length but is easier to calculate, as individual scars are often not well defined.

Results of the analysis are displayed in fig. 4.7. Prolate ellipsoids [fig. 4.7*A*] show the experimentally observed behavior for low aspect ratio: for $n < 100$ particles there are few excess defects, but at higher particle numbers there is a roughly linear increase in the number of excess defects. As aspect ratio increases, however, the transition is softened such that there is a smooth increase in excess defects with n. This is reminiscent of how applied fields soften phase transitions[**?**]; here the anisotropy of the curvature seems to play a similar role.

The oblate packings show the same trends [fig. 4.7*B*]. There is, however, an additional feature that stands out. At $n = 140$, $a > 2$, there is a set of nearly scar-free configurations. This is due to commensurability, as the particle number and surface geometry for these cases are compatible with a highly symmetric packing with only the minimally required defects, as seen in the inset of fig. 4.7*B*. Similar commensurability issues occur in other systems, e.g. sphere packings on cylinders[31].

The model for defect number density developed above in the *Defect Distribution* subsection can be applied to predict the average scar length for a given aspect ratio and particle number. For each point in our parameter space, we integrate eq. 4.2, estimating r by assuming a packing fraction of $\phi = 0.86$, where the packing fraction is given by

$$\phi = \frac{n\pi r^2}{A}, \tag{4.5}$$

where A is the area of the underlying surface. This formula assumes that the

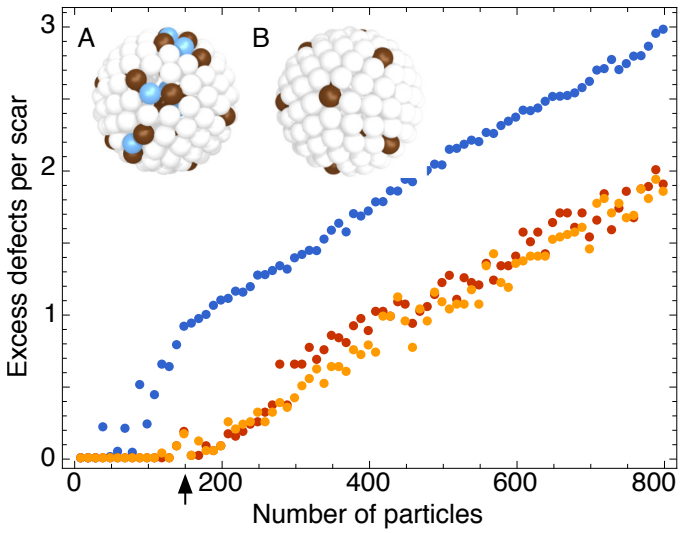


Figure 4.8: Excess dislocations per scar as a function of particle number for hard (blue) and soft $V = 1/d$ (orange) and $V = 1/d^6$ (red) interactions. Inset (A) is a hard particle packing and inset (B) is a soft particle packing. The arrow indicates the particle number of the inset packings.

area on the surface taken up by a particle is equal to its projection onto a flat surface. More comments on this approximation can be found in the next section. Our result for scar length is given by,

$$n_d = \frac{1}{2}\left(\frac{\int \rho_n dA}{12} - 1\right). \tag{4.6}$$

The results are plotted in fig. 4.7A and B, alongside the simulation data. For prolate ellipsoids, we see excellent agreement. Perhaps most importantly, the scar transition is softened at high aspect ratios, as in the simulation results. For oblate ellipsoids, the model does not fit the simulation results quite as well; it tends to under-predict the scar length, especially at higher aspect ratio. It does, however, exhibit softening of the scar transition.

A striking difference between these results and those from a previous study is that here, for hard particles, the transition occurs at a lower particle number; in ref. [10] it was seen at $n_c \approx 400$ using colloidal particles with a soft repulsive

interaction. We therefore performed simulations (see the *Methods* chapter) using two different particle pair potentials, $V = d^{-1}$ and $V = d^{-6}$ (where d is the interparticle separation), the results of which are shown in fig. 4.8. For soft particle packings, we take the average scar length of the five lowest energy configurations obtained out of an ensemble of 50. For the hard spheres, $n_c \approx 80$, while for the two soft potentials the transition occurs around $n_c \approx 200$ (which appears to be within the uncertainty of the result presented in ref. [10]). The defect number increases at the same rate with respect to particle number for both soft potentials. This supports the conclusion in ref. [10] that, for soft particles, the scar transition does not depend on the specific form of the particle potential. For hard particles we have quantitatively different behavior. Visual inspection of hard and soft sphere configurations reveals that hard sphere configurations possess gaps (fig. 4.8*A*). It is rare to find a lone disclination; it is much more common to find a disclination attached to one dislocation (i.e. a small 5-7-5 scar) adjacent to a gap in the packing. This is not seen in soft particle configurations (fig. 4.8*B*), as the energy penalty is too high, rather a particle can be squeezed to fill in the gaps. The fact that hard particle packings tend to have gaps makes them especially suitable for chemical functionalization as described in ref. [36].

4.4 Packing Fraction and Symmetry

We now turn to how the packing fraction varies with respect to particle number and ellipsoid aspect ratio. To simplify the calculation we make the approximation, valid for large N, that the area covered by a particle is its projection onto a flat 2D surface,

$$\phi = \frac{n\pi r^2}{A}, \tag{4.7}$$

where A is the area of the underlying surface. We checked the validity of this estimate by numerically integrating the area of intersection between the surface and the spheres on oblate surfaces of aspect ratio 4.0, and found that the difference between our estimate and the true value is very small: using the projected area underestimates the packing fraction by approximately 1% for packings with $n = 100$ and 0.1% for packings with $n = 800$.

For large n, the packing fraction increases slightly with aspect ratio. This is because for large a the curvature—and hence the defects—are mainly localized to the poles on prolate surfaces or the equator on oblate surfaces and so more of the surface can be covered by the planar hexagonal packing, consistent with the results of the above sections on the *Defect Distribution* and *Scar Transition*. For low n, the opposite tends to be true; the packing fraction decreases with aspect ratio. However, the trend is more complex and the packing fraction is sensitive to both n and a at low n. Visual inspection of these configurations reveals that for specific combinations of n and a, the packings have a high degree of symmetry, suggesting a commensurability effect, such as that seen in the *Scar Transition* subsection above.

To identify these commensurate combinations, we conducted a more thorough search for symmetric packings using the second data set. An arbitrary packing must break the ellipsoidal symmetry group of the surface and hence must belong to some finite subgroup of $D_{\infty h}$; most packings at high particle number do so trivially, retaining only the identity element. Defining a suitable inner product (A, B) that measures the distance between two packings, a packing possesses a symmetry $\mathcal{C}$ if $(A, \mathcal{C}A) = 0$ where $\mathcal{C}$ is a group element of $D_{\infty h}$. The elements $\mathcal{C}$ can be constructed from the group generators: i) an infinitesimal rotation about the ellipsoid symmetry axis; ii) spatial inversion, and iii) a rotation by π about an axis perpendicular to the symmetry axis.

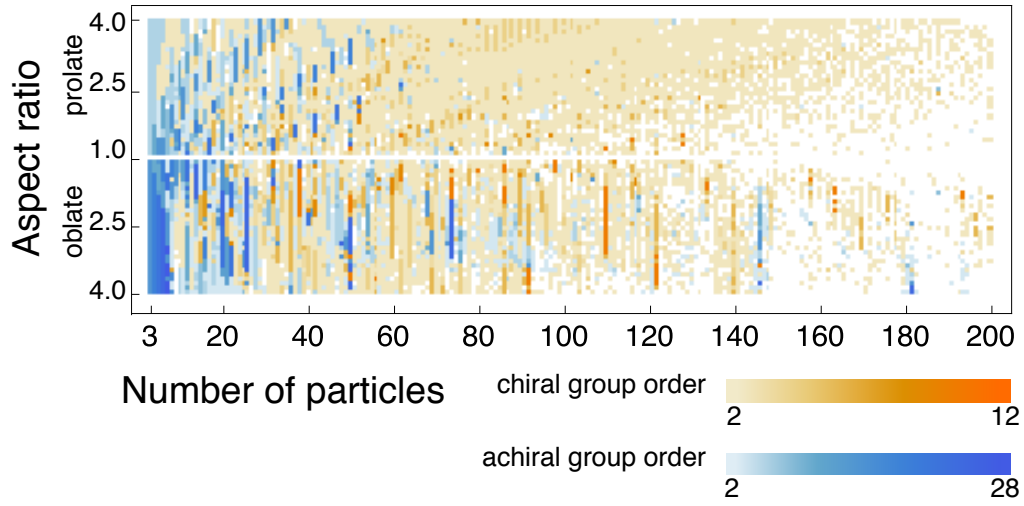


Figure 4.9: The symmetry landscape for packings with varying particle number and aspect ratio, using a symmetry norm cutoff of 0.1. Coloring indicates the chirality and the order of the largest symmetry group found. Orange represents chiral packings and blue represents achiral packings. The boldness of the color corresponds to the order of the packing's symmetry group as shown in the key. Note that packings whose only symmetry is the identity are colored white to distinguish them as being trivially symmetric. Note that no data is shown for $a = 1$ (spheres) as the spherical symmetry group is not a subgroup of $D_{\infty h}$.

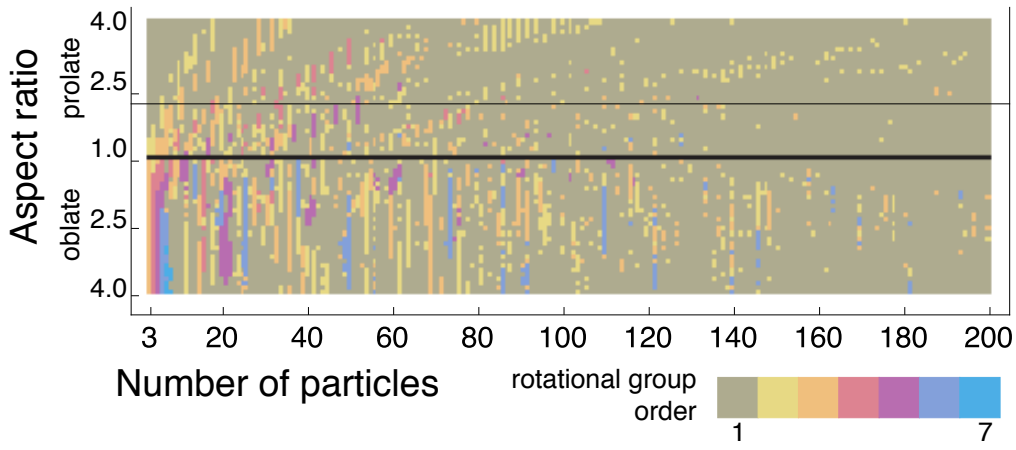


Figure 4.10: The same parameter space as in 4.9, colored by the degree of rotational symmetry of each configuration about its ellipsoidal symmetry axis.

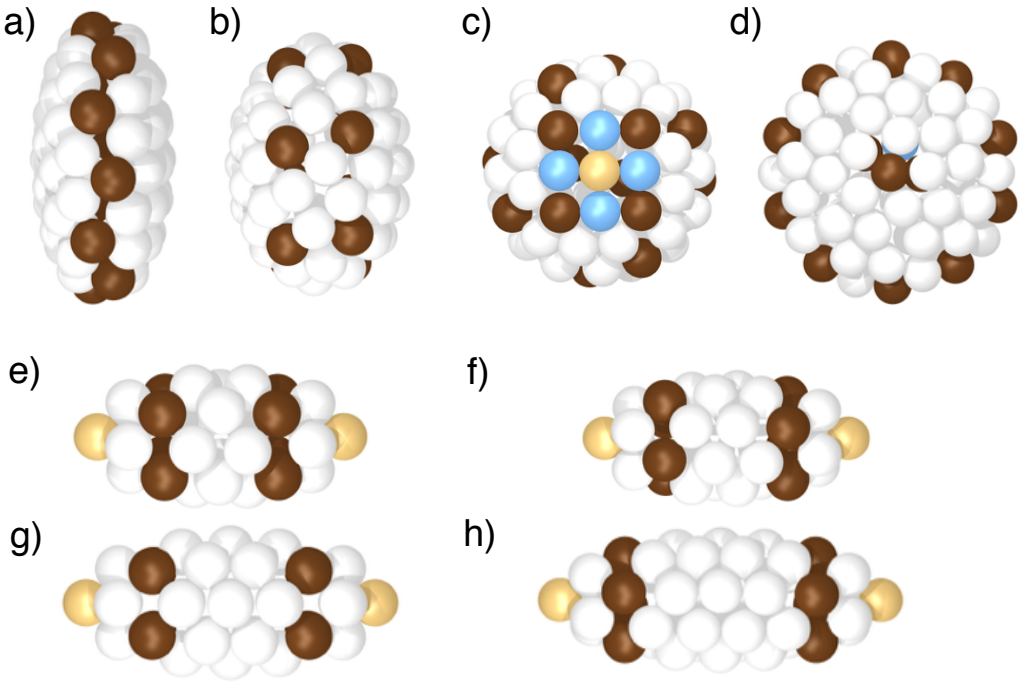


Figure 4.11: A selection of symmetric packings: (a) an achiral packing with $n = 74$, $a = 2.5$; (b) a chiral packing with $n = 74$, $a = 1.5$ — note that (a) and (b) have the same particle number, but show different chirality for different aspect ratio; (c) a packing with fourfold rotational symmetry with $n = 69$, $a = 1.4$; (d) a packing with fivefold symmetry $n = 76$, $a = 2.4$. (e-h) A series of packings which occur in the diagonal band of fourfold rotational symmetry in the top left of fig. 4.9 and 4.10 e) $n = 30$, $a = 2.4$; f) $n = 34$, $a = 2.5$; g) $n = 38$, $a = 2.7$; h) $n = 46$. Note that Light brown particles have coordination $c = 4$.

We used a norm (A, B) defined such that,

$$(A, B) = \sqrt{\frac{1}{n}\sum_{i}^{n}\left(\frac{\min_j \left|\vec{a}_i - \vec{b}_j\right|}{r}\right)^2}, \tag{4.8}$$

where the $\vec{a}_i$ and $\vec{b}_j$ are the positions of particles in packings A and B, respectively: for each particle in A, the closest particle in B is found and the separations between these pairs are divided by the particle radius. The root mean square of these normalized separations is then taken as the inner product. From this, together with the group generators, all symmetries such that $(A, \mathcal{C}A) \leq \epsilon$,a threshold separation were found. From this catalog of symmetries, for a particular configuration the appropriate group was determined. From a collection of configurations with a given (n, a), the most symmetric configuration was chosen by the following procedure. First, the configurations with the largest symmetry group were identified. Then, for each of these configurations, the symmetry group element with the highest symmetry norm was identified and the configuration with the minimum highest symmetry norm was chosen as the most symmetric.

The results of this analysis are displayed in fig. *4.9* showing the order and chirality of the symmetry group of the best packing for each combination of particle number and aspect ratio. The degree of rotational symmetry for each packing is shown in fig. 4.10. One striking feature is that, for certain particle numbers, long vertical stripes appear in the plots representing commensurate aspect ratios for that particle number. Furthermore, low n favors achiral packing while chiral packings occur more often for higher particle number. We do, however, see packings that are either chiral or achiral at the same particle number, as shown in fig. 4.11 a and b. For prolates the stripes occupy a narrow range of aspect ratio and occur in band-like sequences described by a

straight line $a = mn$ with some slope m. Each of these sequences corresponds to a different degree of rotational symmetry n_r, and the particle numbers in the sequence are separated by n_r. Inspecting the configurations in a single sequence, the difference between a configuration with n particles and the next with $n+n_r$ particles is that an additional row of n_r particles has been inserted in the space created by the longer aspect ratio. This is illustrated by a sequence of configurations with fourfold rotational symmetry in fig. 4.11(e-h).

For oblate ellipsoids, the symmetric configurations for n particles occur at a much broader range of aspect ratios and symmetric configurations are observed at much higher n and tend to have six-fold rotational symmetry. The reason for this is that the high curvature at the end of the prolate ellipsoids accommodates n_r-fold defects at the poles, and these appear to determine the rotational symmetry for the entire configuration; for oblates, the poles have low curvature and promote hexagonal packing, hence causing six-fold rotational symmetry to be more common. Interestingly, other degrees are present including $n_r = 4$ and $n_r = 5$ and these configurations contain regions of highly oblique packings (fig. 4.11 c and d).

In general, these symmetric packings are notable because they contain a high degree of hexagonal ordering over much of their surface, with evenly spaced defects throughout. This high degree of regularity should provide stability to the packed structure, and reduce the likelihood of failure from irregularly spaced defects.

4.5 Discussion

In this chapter we explored the structure of hard spheres packed onto static ellipsoidal surfaces. It was observed that the defect charge density is given

by the Gaussian curvature of the surface, as expected. The defect number density is roughly predicted by a heuristic model that combines the influence of the Gaussian curvature on core disclination density with the influence of Gaussian curvature on scar length. Unexplained by this model, however, is the appearance of a secondary peak in the defect number distribution.

The coupling of scar orientation and curvature was investigated and found to be highly nontrivial for prolate ellipsoids; it is much simpler on oblate ellipsoids. For prolate ellipsoids, we see varying degrees of nematic and quadrupolar orientational order both for varying surface aspect ratio and in different regions across the surface. On oblate ellipsoids, the majority of scars occur at the highly curved rim of the surface and lie along this rim.

The above heuristic model is also used to predict the average scar length across an ellipsoidal surface. The model accurately predicts the softening of the scar transition. However, it begins to fail for packings on highly oblate ellipsoids. We also see that hard particle and soft particle packings differ in their scar transition: hard particles lead to a scar transition at lower particle number due to their tendency not to accommodate lone 5-fold disclinations.

Finally we identified symmetric packings that occur across a range of particle numbers and aspect ratios. This search revealed a variety of packings in which the particle number is commensurate with the surface geometry, leading to highly symmetric packings. Several trends were identified, including several families of similar prolate packings, and the fact that symmetric packings can occur at higher particle number for oblate packings.

Chapter 5

The influence of dynamics on arrested emulsion droplets

5.1 Introduction

Having investigated sphere packings on static ellipsoids to understand the influence of geometry in, we now turn our attention to the dynamical effects involved in the creation of an arrested droplet in a Pickering emulsion. A number of factors may influence the formation of these structures, moving the problem beyond the scope of a standard sphere packing problem. The most obvious difference is the fact that the surface of an emulsion droplet is changing shape at a constant volume as it evolves. Another way to express this is that the metric of the space in which the packing lies is changing. Thus, a different packing protocol must be used to generate these packings and it is well known that different packing protocols can result in packings with different qualities[**?**].

In addition to the surface evolution, the physics of particle motion may also play an important role in determining the structure of the packing. Colloids on

the order of a micron will undergo Brownian motion, so the rate at which they diffuse compared to the timescale at which the surface relaxes could affect the quality of the packing. In the static simulations, particles were highly diffusive and were able to rearrange as the packing density increased. It may be the case the the surface relaxation happens much faster than particle diffusion, preventing particles from rearranging and causing a less ordered packing.

Other forces may act on particles as well. In the experimental system, it is observed that particles are attracted to one another. This causes "rafts" to form which are then pushed together. It is also observed that gravity tends to pull particles downward, creating an uneven initial coverage. These forces which alter the ordering and distribution of particles before arrest may have an influence on the evolution of the packing and the final arrested state.

In this chapter, we will investigate how these dynamical effects and their relatives strengths affect the structure of the arrested packing. We do this using packings generated by the dynamic packing algorithm outlined in section 3.1.2. To briefly review, these simulations take as input an initial surface aspect ratio a, the number of particles n, and the particle radius r. To specify the system dynamics, a timescale associated with each dynamical effect is also given: diffusion timescale τ_d, attraction timescale τ_a, and gravity timescale τ_g. These timescales determine the typical time it would take for a particle to move a distance of one particle diameter under the influence of the associated force, given in units of the relaxation timescale (i.e. the time it takes for the surface to evolve from its initial shape to a sphere). A shorter timescale corresponds to a stronger force. The surface evolves such that its area follows an exponential decrease. More details can be found in the *Methods* chapter.

5.2 Relaxation rate

To investigate the effect of relaxation rate on packings with purely diffusive particles (i.e. no attraction or gravity) a dataset was generated consisting of 2500 packings with $n = 800$, $a = 8$, and $r = 0.148316$. This radius was chosen so that arrest would occur at an aspect ratio of roughly 4. The diffusion timescale ranged from $\tau_d = 1$ down to $\tau_d = 4^{-4}$, at five values of τ_d each differing by a factor of four, with 500 configurations at each value.

We first examine the distribution of packing fractions (as defined by 4.7) at which arrest occurs for each τ_d. As the final aspect ratio is directly related to the density of the arrested packing, the aspect ratio is one indicator of the quality of the packing. Because of this, one might expect that for higher τ_d, when the particles have less time to rearrange by diffusion, lower quality packings will form resulting in a lower packing fraction at arrest, invoking an analogy to quenching a system by rapid cooling leading to the formation of a glassy state[37]. This is in fact what we find. Histograms of the final packing fraction are shown in fig. 5.1a for each τ_d. We see that the mean packing fraction shifts to a higher value as τ_d decreases, i.e. a slower relaxation rate results in arrest at a higher packing fraction. Note, however, that this effect is subtle: the size of the shift of the mean is roughly as large as the width of the distribution at fast relaxation rates. We see also that as relaxation rate decreases, the width of the distribution of aspect ratios gets narrower.

We can also consider the quality of the packing in terms of the number of defects. Fig. 5.1b shows a scatter plot of the packing fraction and total defect number at arrest for packings produced through fast and slow relaxation. We see the same trend for defect number as we do for packing fraction: fast relaxation results in more defects, and a wider distribution of defect numbers. Because of the wide range in the quality of packings produced by fast relax-

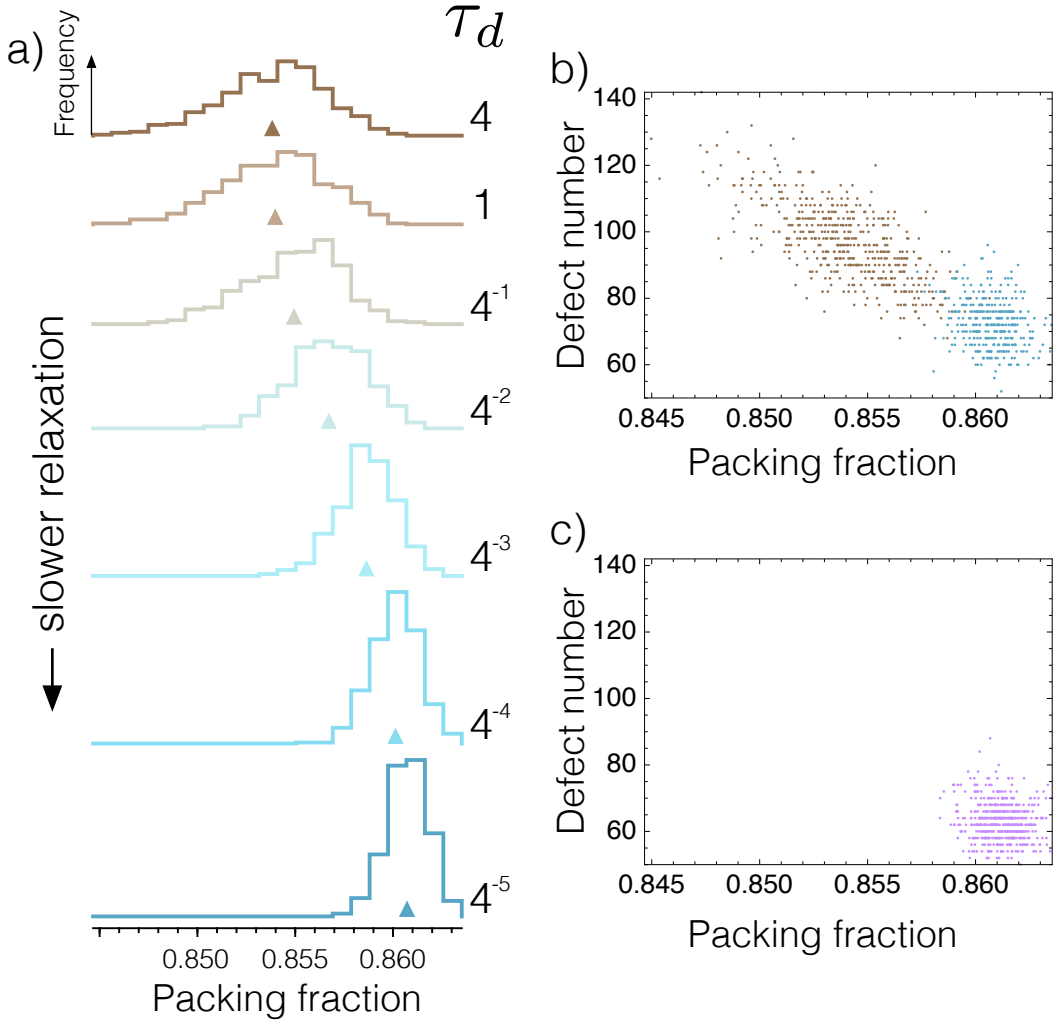


Figure 5.1: a) For varying relaxation rates (as indicated by τ_d), histograms of the packing fraction at arrest. We see that as the relaxation rate slows, higher packing fractions are achieved and the distribution of aspect ratios becomes narrower. b) A scatter plot of packing fraction versus total defect number for fast relaxation ($\tau_d = 4$, brown) and slow relaxation ($\tau_d = 4^{-5}$, blue). The two quantities are correlated, though the variation is low enough for the slow relaxation case the the correlation is not apparent within the resulting configuration. c) Packing fraction versus defect number for packings on a static geometry produced by the inflation algorithm, from the dataset studied in chapter 4

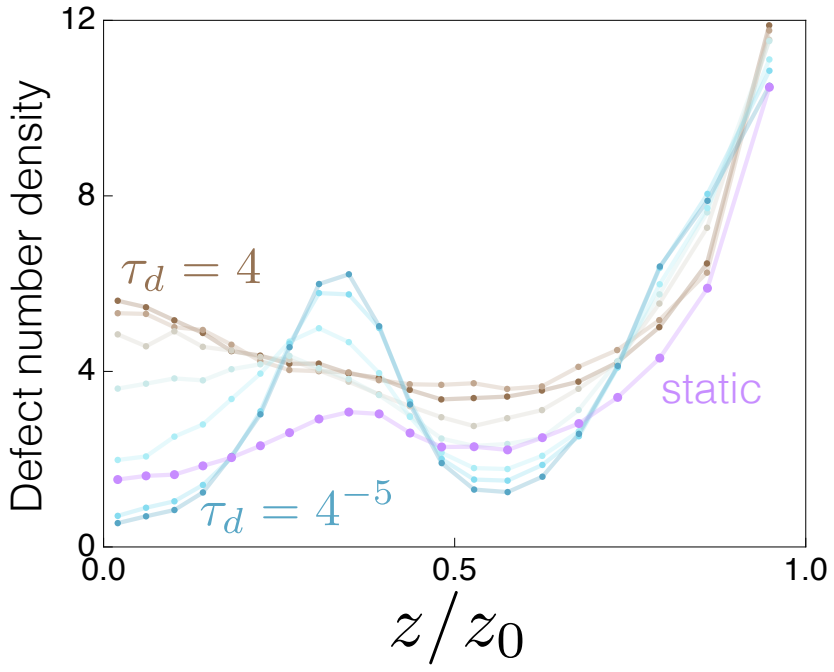


Figure 5.2: For varying relaxation rates (as indicated by τ_d) we plot number density of topological defects along the surface (from the center to the ends) averaged over 500 configurations. We also include the defect density produced on a static geometry by the inflation algorithm, from chapter 4 for comparison. We see that for slow relaxation rates (e.g. $\tau_d = 4^{-5}$) the defect distribution shows qualitative similarities to that produced by the inflation algorithm, with secondary peaks between the center and the ends, though it differs in magnitude. For fast relaxation (e.g. $\tau_d = 1$) there is a single secondary peak in the defect distribution near the center of the droplet.

ation, the correlation between packing fraction and defect number is apparent. Within the set of packings produced through slow relaxation, this correlation is not apparent, though taken together with the fast relaxation results, the correlation becomes even more clear. We also show the

Next we investigate the distribution of topological defects as a function of relaxation rate, using the same type of analysis as in chapter 4. The average defect number density, as a function of axial position z/z_0, is plotted in fig. 5.2 for each τ_d, as well as as the results for the inflation algorithm presented in

chapter 4. A clear qualitative difference is present in the defect distributions for different τ_d. For low τ_d, where the relaxation proceeds slowly compared to diffusion, we see qualitatively the same features in the defect distribution that we see for packings on a static surface produced by the inflation algorithm. However, the distribution is not exactly the same: in packings produced by the relaxation algorithm in the slow relaxation limit the secondary peak in the defect density is much more exaggerated. The overall defect density is also lower around the secondary peak, compared to the inflation algorithm. This suggests that, even in the slow relaxation limit which one might expect to approach the static limit, the dynamics of the surface evolution has an influence on the spatial distribution of ordering.

For higher τ_d i.e. faster relaxation rates, the secondary peaks in the defect distribution shift towards the center of the droplet and become one peak at the center of the droplet. In order to understand this shift, we track the evolution of the hexatic order as surface relaxation occurs. To do this, we calculate a local measure of the hexatic order for each particle as defined by the expression

$$\psi_6^i = \frac{1}{N_b^i} \sum_{j=1}^{N_b^i} \exp(i6\theta_{ij})$$

where i is a particle label, N_b^i is the number of neighbors of particle i (determined by a center-to-center cutoff distance, here 1.3 times the particle diameter), and θ_{ij} is the angle of the line connecting particle i to particle j with respect to the local surface coordinate frame. We split the surface into 20 segments with boarders of equal spacing along the ellipsoid symmetry axis, and take the average of the local hexatic order of the particles within each bin. We then average these averages over a set of 100 simulations, and plot the order across the surface at equally spaced timesteps (note that each simu-

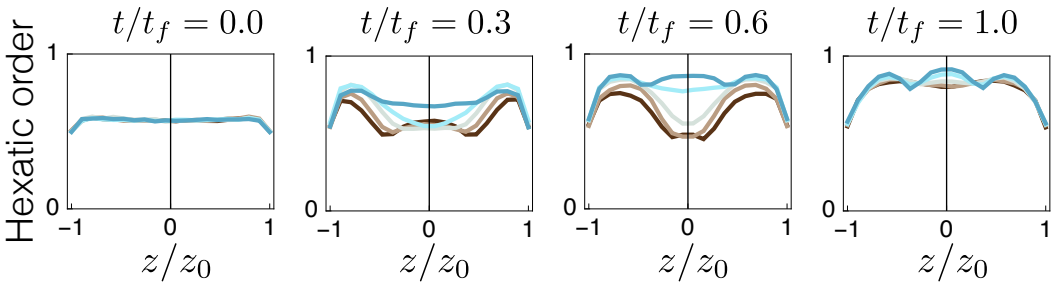


Figure 5.3: The local average hexatic order across the surface for different τ_d throughout surface evolution. t_f is the time at which arrest occurs. τ_d ranges from 4^{-0} (fast relaxation, brown) to 4^{-4} (slow relaxation, blue). For fast relaxation, order propagates in from the ends, while for slow relaxation order is allowed to develop more uniformly.

lation arrests at a slightly different time, but there is a small enough variation in this time that taking equally spaced timesteps for each simulation and then average the results at each step across simulations is sufficient for investigating the average behavior.)

These results are shown in fig. 5.3 for four timesteps, for $\tau_d = 1$ to $\tau_d = 4^{-4}$. We see that generically, the order first increases most rapidly towards the ends of the droplets (though not at the ends, where geometric frustration dominates.) This is because as the surface begins to relax, the ends of the droplet are pulled inward rapidly. The surface area near the ends shrinks as the ends are pulled inward while the surface area grows towards the center of the droplet as the center expands. Order propagates inward as a crystalline growth front expands from the droplet ends. For slower relaxation, we don't see a clear growth front. Instead, ordering develops more uniformly across the surface as it evolves. Essentially, the higher relative rate of diffusion allows particles to rearrange themselves more efficiently as the density increases, promoting order more evenly across the surface. The slower rate of relaxation allows diffusion to dominate and more closely mimics the static surface limit, i.e. the inflation algorithm. For fast relaxation, this rearrangement does not have time to occur

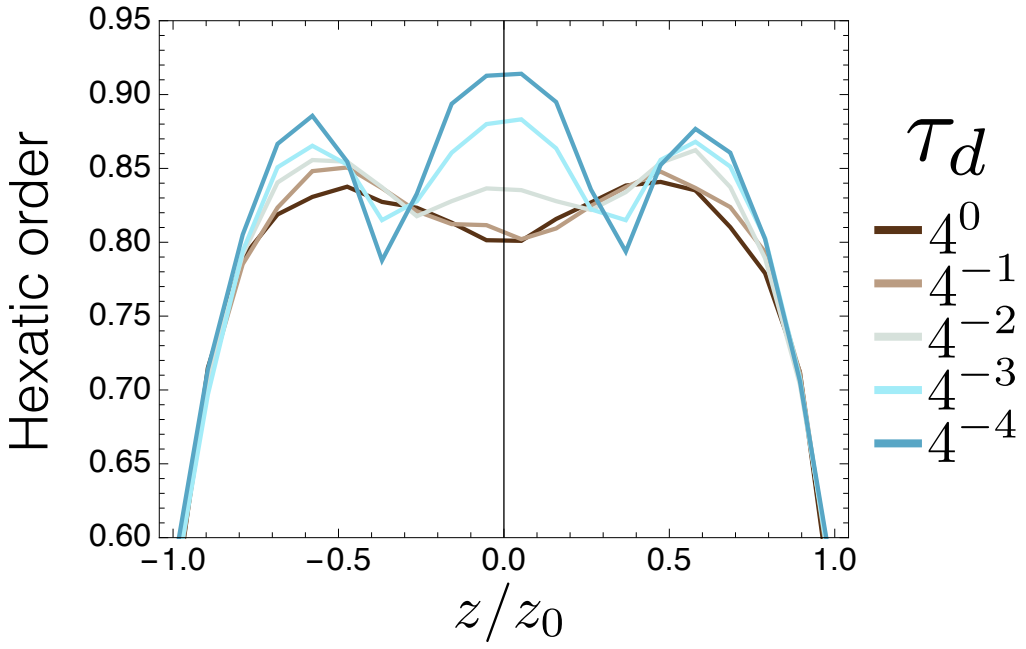


Figure 5.4: The local average hexatic order across the surface for different τ_d at arrest, scaled to emphasize variation. The hexatic order shows the same patterns as the defect number density. For fast relaxation, the order is typically lower across most of the surface, and single secondary minimum in order is seen at the center of the droplet. For slow relaxation, there are two secondary minima in the order between the center and the ends of the droplet.

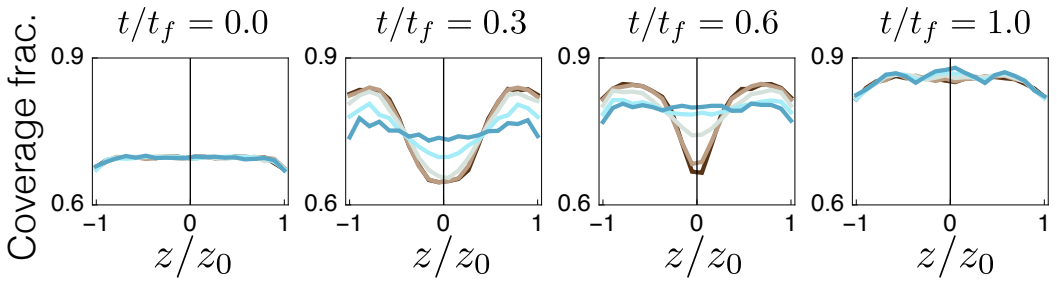


Figure 5.5: The coverage fraction across the surface for varying τ_d throughout surface evolution. t_f is the time at which arrest occurs. τ_d ranges from 4^{-0} (fast relaxation, brown) to 4^{-4} (slow relaxation, blue). For fast relaxation, particles are densified first towards the ends, while for slow relaxation particles are allowed to diffuse towards the center as relaxation occurs.

and the center is still disordered as the crystalline growth front moves in. In the end, this center becomes more ordered to a significant degree, but comparing the final distribution of hexatic order, we see that for the fast relaxation case the relative disorder has been quenched in to some extent. The distribution of hexatic order at arrest is shown in fig. 5.4, scaled to emphasize the differences.

It is instructive to see how the local particle density evolves during relaxation as well. In fig. 5.5 we show, for four timesteps, the packing fraction in bins across the surface (averaged using the same method as was used for hexatic order.) We see that, similar to the hexatic order, the density increases the most rapidly at the ends for fast relaxation. A high density front then moves in from the edges towards the center until the density is nearly uniform across the surface. This effect is much more subtle for slow relaxation. Note that in the final arrested state, the density distribution across the surface shows the same qualitative features as the hexatic order distribution for each relaxation rate, verifying the correlation between them.

5.3 Attractive particles

In experiments, as discussed in chapter 8, it is observed that particles are attracted to one another, often forming well-ordered aggregates which are pushed together as relaxation proceeds. To investigate the effect attraction has, we include attractive interactions between particles as described in the chapter 3. To reiterate, the attractive force is given by an inverse-distance potential with a linear screening factor. It takes the form

$$\vec{F}_{ij} = \begin{cases} -\alpha \frac{\vec{r}_{ij}}{|\vec{r}_{ij}|^2}\left(1 - \frac{s}{2r}\right) & 0 < s < 2r \\ 0 & s \geq 2r \end{cases}$$

where $\vec{r}_{ij} = \vec{x}_i - \vec{x}_j$ is the particle center-to-center separation, $s = |\vec{r}_{ij}| - 2r$ is the particle surface-to-surface separation, and α is a factor determining the strength of the attraction, which is set by fixing an attractive timescale τ_a related to the the time it would take a particle to move a distance of one diameter under the influence of attraction from another particle (using the force felt between two particles in contact) with $\tau_r = 1$:

$$\tau_a = \frac{4r^2}{\alpha}.$$

We run simulations with both weak and strong diffusion, at $\tau_d = 4^{-1}$ and $\tau_d = 4^{-3}$ respectively. At each τ_d, we run simulations at five values of τ_a: for weak diffusion,τ_a varies from 4^{-1} to 4^{-5}, and for strong diffusion τ_a varies from 4^{-3} to 4^{-7}, so that the ratio of attraction to diffusion strengths is the same for both strong and weak diffusion. Thus it is again the relaxation rate that is different for the different τ_d. The remaining parameters are the same as in the diffusive case, $n = 800$, $a = 8$, and $r = 0.148316$. We run 200 simulations

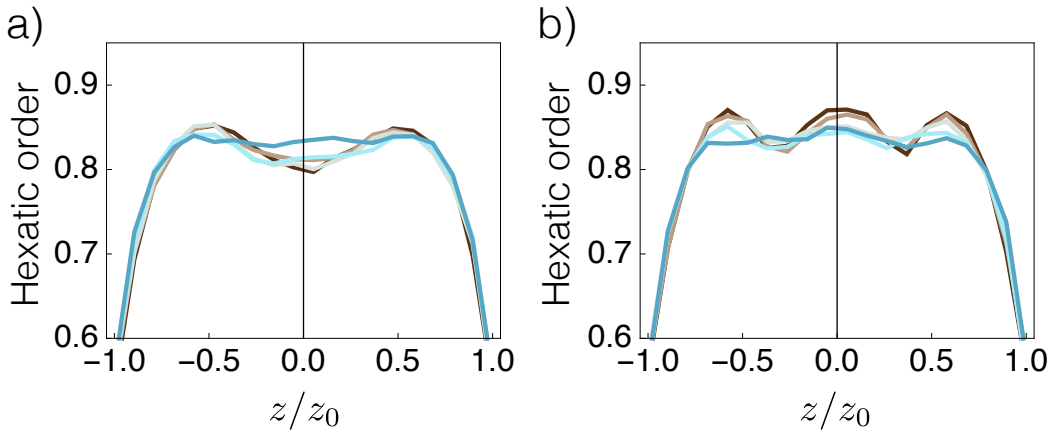


Figure 5.6: The local average hexatic order along the surface at arrest for different attraction strengths for (a) fast relaxation ($\tau_d = 4^{-1}$) and (b) slow relaxation ($\tau_d = 4^{-3}$). In both cases, we see that as the influence of attraction in increased relative to diffusion, the ordering becomes more uniform across the surface, and we also see ordering of the same magnitude regardless of relaxation rate.

for each combination of diffusion and attraction timescale parameters.

Fig. 5.6 shows the distribution of hexatic order across the surface for the strongest and weakest attraction strengths, for weak diffusion (a) and strong diffusion (b). The weak attraction limit recovers the purely diffusive limit. The strong attraction case, for both weak and strong diffusion, shows much more uniform hexatic order.

To investigate this, we again look at the evolution of the hexatic order as surface relaxation proceeds, as shown in fig. 5.7. We see that initially (after a period of equilibration but before surface relaxation), the hexatic order is very high across the entire surface. This is consistent with previous studies of crystal growth on curved surfaces[38], where attractive particles crystallizing on a curved surface form defect-free structures with a high perimeter-to-area ratio. We display an example of an equilibrated, pre-relaxation surface in fig. 5.8. Because the initial surface coverage is high, our structure is not totally defect free, but the presence of gaps allows for fewer defects. As relaxation proceeds,

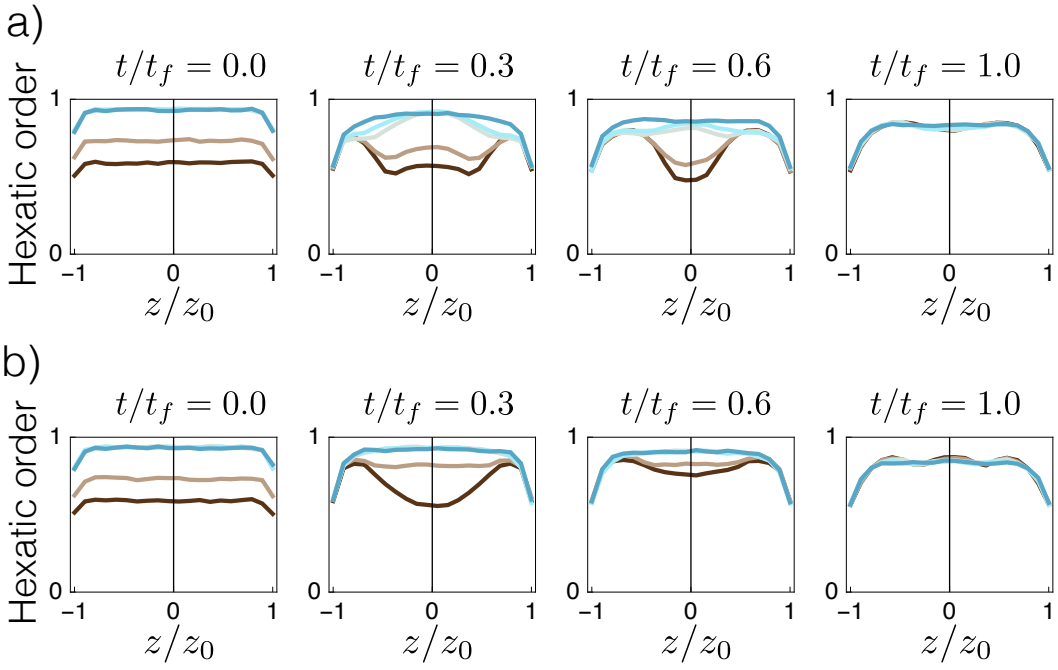


Figure 5.7: The local average hexatic order across the surface for varying (a) fast relaxation ($\tau_d = 4^{-1}$) and (b) slow relaxation ($\tau_d = 4^{-3}$) throughout surface evolution. t_f is the time at which arrest occurs. τ_a ranges from 4^{-1} (brown) to 4^{-5} (blue) for fast relaxation and 4^{-3} (brown) to 4^{-7} (blue) for slow relaxation. For both relaxation rates, the strong attraction cases are very ordered across the surface initially and this uniform order is largely preserved (slightly less so towards the ends for the fast relaxation case.)

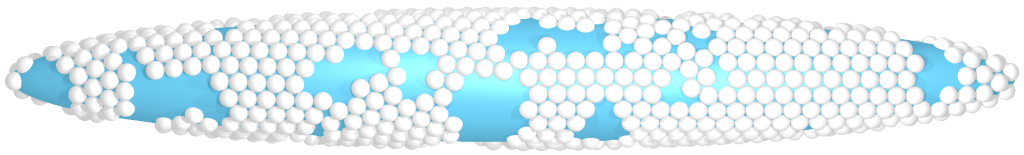


Figure 5.8: An image of an equilibrated configuration with $\tau_d = 4^{-1}$ and $\tau_a = 4^{-5}$. Large gaps are present within a highly ordered packing, removing the need for defects within the packing.

the order begins lowering near the ends of the droplet, where area is being depleted and the crystal structure is being pushed and deformed by droplet evolution. The variation in hexatic order across the packing is, however, much smaller than in the purely diffusive case: the strong attraction makes it difficult for particles to rearrange and the initial uniform hexatic order is only slightly disrupted. In the final arrested state, the hexatic order has been reduced across the entire surface, as the increased packing fraction no longer allows gaps in the packing which reduce the need for defects, introduces geometric frustration of the packing. On average, the order is lower in the strongly attractive case as compared to the diffusive case, again because attraction prevents particles from rearranging and arresting in a higher-density, higher-order state, analogous to the purely diffusion, fast relaxation case.

5.4 Gravity

In experiments, it is also observed that gravity plays an important role: before relaxation, particles tend to accumulate at the bottom of the droplet. As this initial crowding will affect how the particle configuration evolves and eventually arrests, we perform relaxation simulations with gravity acting on the particles.

Gravity acts on each particle with a force acting downward (which we define to be in a direction perpendicular to the ellipsoid symmetry axis) with a magnitude g which is determined by setting a gravity timescale,

$$\tau_g = \frac{2r}{g}.$$

This timescale determines how long it would take a particle to fall a distance of $2r$ under the influence of gravity only (if it were undeterred by other particles

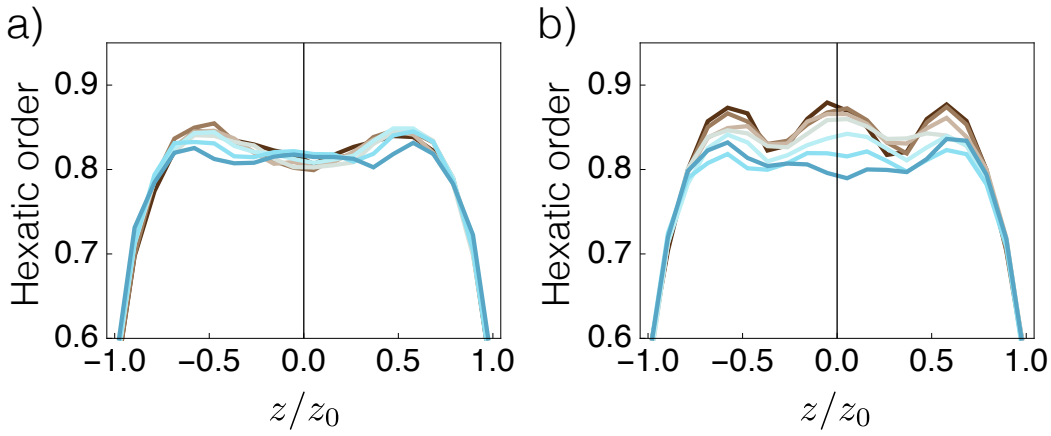


Figure 5.9: The local average hexatic order across the surface for varying (a) fast relaxation ($\tau_d = 4^{-1}$) and (b) slow relaxation ($\tau_d = 4^{-3}$) throughout surface evolution. t_f is the time at which arrest occurs. τ_a ranges from 1 (brown) to 4^{-6} (blue) for fast relaxation and 4^{-2} (brown) to 4^{-8} (blue) for slow relaxation.

or the droplet surface), again with $\tau_r = 1$.

As before for attraction, we run simulations with both weak and strong diffusion, at $\tau_d = 4^{-1}$ and $\tau_d = 4^{-3}$ respectively. At each τ_d, we run simulations at seven values of τ_g: for weak diffusion,τ_g varies from 1 to 4^{-6}, and for strong diffusion τ_a varies from 4^{-2} to 4^{-8}. The remaining parameters are the same as in the diffusive case, $n = 800$, $a = 8$, and $r = 0.148316$. We run 200 simulations for each combination of diffusion and gravity timescale parameters.

In fig. 5.9 we show the distribution of hexatic order along the surface for arrested droplets under the influence of gravity for both strong diffusion (a) and weak diffusion (b). We see in both cases that strong gravity has the effect of lowering the order slightly and leading to more uniform ordering. This is because, similar to the attractive particle case, gravity has the effect of promoting early ordering by packing particles together towards the bottom of the surface. This is seen in fig. 5.10 where we show a packing which has been equilibrated before relaxation, leaving a large gap at the top of the packing.

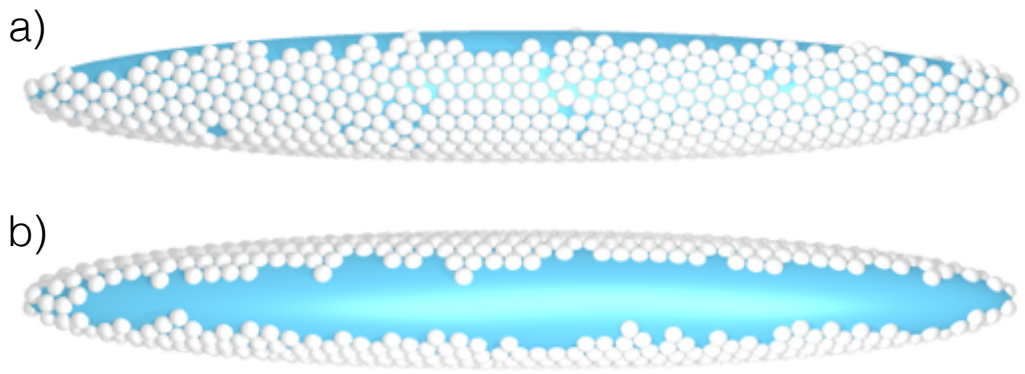


Figure 5.10: a) Side view and b) top view of an equilibrated configuration with $\tau_d = 4^{-1}$ and $\tau_g = 4^{-6}$.

As the surface evolves, this even ordering is preserved and "locked in" to some extent at arrest.

In fig. 5.11 we show the final distribution of hexatic ordering, this time averaged in bins in azimuthal segments around the ellipsoid symmetry axis in order to investigate how the broken azimuthal symmetry introduced by gravity affects the azimuthal ordering. We see that for strong gravity the order is increased at the bottom of the droplet ($\phi = \pm\pi$) and decreased at the the top of the droplet ($\phi = 0$). This is consistent with the fact that gravity promotes ordering early towards the bottom of the surface (fig. 5.10). A relatively well ordered packing covers some fraction of the surface, with an empty gap at the top of the packing. As the surface approaches arrest, the boundary of the packing closes in on itself and because gravity suppresses the effect of diffusion, the particles at the boundary are unable to rearrange themselves and are left in a relatively disordered state. We see, comparing the weak and strong diffusion cases (which should be interpreted as fast and slow relaxation, respectively, as the relative strengths of gravity and diffusion are kept the same between the two) that slower relaxation promotes a larger variation in order. This is because for slower relaxation, the dominant effect

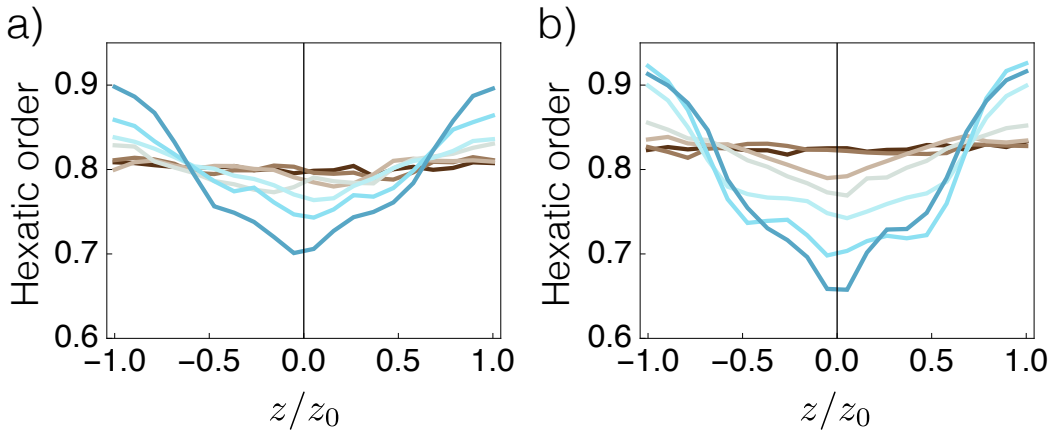


Figure 5.11: The local average hexatic order around the surface for varying (a) fast relaxation ($\tau_d = 4^{-1}$) and (b) slow relaxation ($\tau_d = 4^{-3}$) throughout surface evolution. t_f is the time at which arrest occurs. τ_a ranges from 1 (brown) to 4^{-6} (blue) for fast relaxation and 4^{-2} (brown) to 4^{-8} (blue) for slow relaxation.

(gravity in this case) has a larger influence.

5.5 Discussion

By including and varying the relative strengths of different dynamic effects, we see a variation in the ordering of the resulting arrested packings.

In the limit of slow relaxation, in which diffusion is the primary dynamic effect driving the packing evolution, the distribution of hexatic order is determined primarily by the surface geometry (although, interestingly, we see that surface evolution does have an effect even for very slow relaxation.) As the influence of relaxation rate, attraction, or gravity is increased, however, these effects dominate the particle ordering. In each case, the dynamic effect promotes earlier ordering over all or some of the surface: a fast relaxation rate promotes order near the ends, gravity promotes order near the bottom, and attraction promotes uniform ordering. As the surface evolves and eventually arrests, a signature of this earlier ordering remains: lower order persists in the

areas in which order was not promoted. In all non-diffusive cases, the overall average hexatic order is lower than for the diffusive case. This is because diffusion allows the packing to better explore its configuration space, and on average arresting in a more ordered state.

Chapter 6

Metric jamming on arrested emulsion droplets

JAMMING refers to the concept of rigidity in particle packings. Packings at low densities are not rigid; they are not stable under applied stress and they can flow. As the density is increased and particles are forced into contact, the packing becomes constrained and eventually rigid as the packing jams. In the context of hard particle packings, a jammed packing is truly rigid - no further compression is possible. Thus, for hard particle packings, the relevant questions are whether a packing truly jammed (and to what extent - multiple categories of jamming are possible based on the types of particle motions which are or are not allowed), and if it is jammed, what is the structure of the packing and its contact network? Packings of soft particles can, of course, be compressed beyond the jamming point. In this case, it is interesting to ask how the packings behave as they are compressed. In particular, jamming often deals with packings of disordered particles, so the contrast between disordered jammed packings and crystalline materials has been a central focus.

In this chapter the rigidity of arrested packings on relaxing ellipsoids is addressed. The surface relaxation packing protocol produces structures which are densely packed and appear stable over short timescales, but due to the non-equilibrium nature of these systems, one may wonder whether any further particle motions and surface relaxation are possible, i.e., whether they age. This question can be approached in terms of the concept of jamming. There are well established protocols for testing for jamming in packings in Euclidean space, which we have adapted for packings on curved surfaces. By applying these protocols to find unjamming motions, we can determine whether our packings are jammed, and if not, we can unjam them and continue to evolve the surface until true jamming has been reached.

It is important to be precise about what we mean by "true jamming" in this system, as compared to what it means for systems that have been studied in the past[21]. Jamming has been studied in packings in Euclidean space, where the most extreme category of jamming is strict jamming, in which a packing is stable to both collective motions of particles and deformations of the packing boundary. Here we are studying packings in a compact non-Euclidean space and as such there is no boundary to deform. However, the underlying metric of the space changes as surface relaxation occurs. Thus, we propose a new jamming classification called "metric jamming" in which a packing is stable to both collective particle motions as well as changes in the underlying metric.

In jammed sphere packings in N dimensional Euclidean space, it is well understood that in order for a packing to be stable, the average contact number Z must be at least $2N$, i.e. each particle must have on average $2N$ contacts with its neighbors[39]. In the case of marginally stable jammed packings, it is found that the contact number is exactly $2N$, and hence the packings are isostatic. For monodisperse sphere packings on a 2D ellipsoidal surface,

which tend to be largely crystalline, one then might expect a contact number of $Z \geq 4$. Interestingly, we do not see this. The metric jammed packings produced here appear to be hypostatic, i.e. $Z < 4$. In effect, the curvature of the surface is imposing extra constraints on the packing.

A note on terminology: we are careful to distinguish between "arrested" and "jammed" packings. Arrested packings are stable to surface evolution. The surface evolution imposes a specific stress field on the packing. The output of the surface relaxation packing simulations are arrested packings. Jammed packings are characterized according to the jamming categories outlined in the *Background* chapter and are stable to more general stresses.

In addition to studying packings of hard particles which approach jamming from an under-constrained state, we also study packings of soft particle and their properties as the density is increased above the jamming point. Packings of monodispersed particles in flat 2D space tend to be highly crystalline, unless extreme measures are taken to avoid crystallization[40]. However, due to the geometric frustration imposed by the curved surface, some degree of disorder is induced, as well as strain in the crystalline regions of the packing, and it is not clear what effect this will have on the mechanical properties of soft-particle packings above the jamming point.

In this chapter, we first deal with the details of identifying and treating "rattlers" during the analysis of packings. We then show that a repeated unjamming and surface relaxation scheme produces jammed packings, and that these packings are indeed metric jammed. Ee show that these metric jammed packings have a deficit in the number of interparticle contacts which grows linearly with the number of particles, causing the packings to appear hypostatic when they are actually stable. The contact deficit is due to constraints imposed by the curvature of the surface. Finally, soft particle packings at and

above the jamming point are studied. We focus on their elastic properties above the jamming point, and compare monodisperse and bidisperse packings to understand the extent to which frustrated monodisperse packings behave like disordered packings.

6.1 Identifying rattlers

As discussed in the *Background* chapter, jammed packings of hard particles typically contain rattlers, which are individual particles that are not locally constrained by their neighbors, while the rest of the packing excluding these rattlers is jammed. Because these particles are under-constrained, they are able to move and the contacts they make with other particles do not contribute to the overall stability of the packing.

For a numerically produced packing, there will necessarily be finite gaps between particles which are effectively in contact, due to limitations on the numerical precision of the packing algorithm. Ideally there would be a clearly defined interparticle distance below which particles are in contact, and above which particles are not in contact. After determining this contact length scale, rattlers can be identified based on whether they have sufficient contacts, or based on the distance available for them to move. Using the latter method, one must choose a cutoff for the size of the displacement a particle can make before it is considered at rattler. One expects this rattler-distance to coincide roughly with the contact length scale, and indeed we see that this distance is of similar magnitude but somewhat higher. We find, however, that the contact and rattler length scales themselves are somewhat ambiguous.

A metric used to determine the contact length scale the can be found in ref [41]. A plot of average contact number Z versus contact cutoff δ (where δ

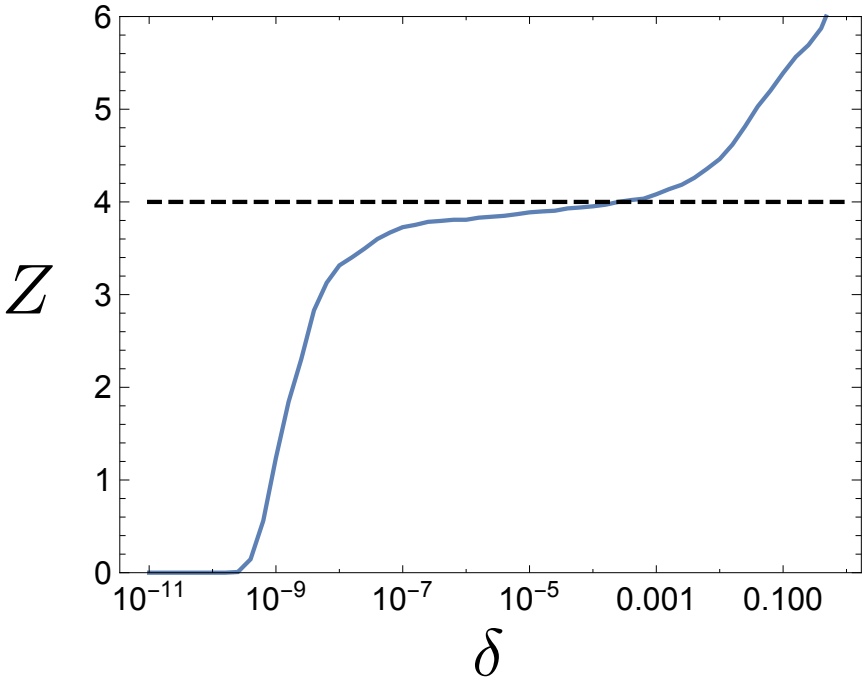


Figure 6.1: The average contact number per particle versus contact cutoff distance for an initially arrested packing. The curve rises quickly from $\delta = 10^{-9}$ to 10^{-8} and then slowly increases without a clear plateau.

is the interparticle separation divided by the particle diameter), for a jammed packing with a well defined contact length scale, will show a very distinct plateau: $Z = 0$ for low δ, and Z quickly rises to the isostatic value at the contact length scale and levels off. The sizes of the contacts are roughly log-normally distributed around this value.

For packings on ellipsoids we find that, while there is a clear peak in the distribution of contact distances, Z slowly and steadily increases as δ increases above the initial large jump, as seen in fig. 6.1 for a packing of $n = 800$. Thus, the packing is somewhat sensitive to the chosen contact cutoff. This will affect which particles are identified as rattlers and which are part of the jammed packing.

To identify rattlers, one must choose a cutoff δ_r for the distance a particle

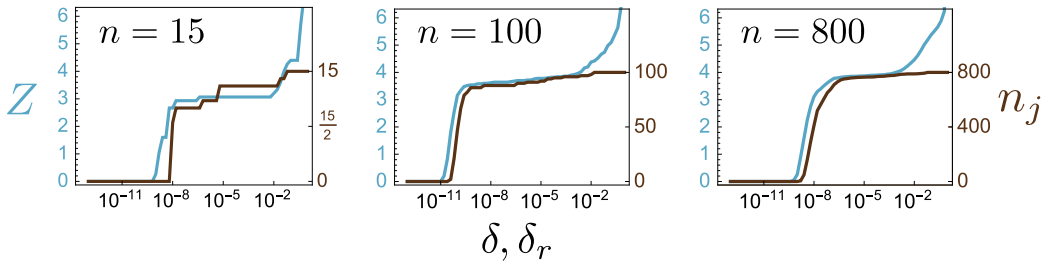


Figure 6.2: For particle numbers n=10, 100, and 800, a comparison of the average contact number as a function of the contact cutoff and the number of locally jammed particles as a function of the rattler cutoff. We see that generally, the ratter length scale is slightly higher than the contact length scale across all particle numbers.

is able to move (normalized by the particle diameter) before it is considered a rattler. In order to determine what this cutoff should be, we use the linear programming rattler protocol described in the *Methods* chapter to find the largest distance each particle can move, with all other particles held fixed. For a given δ_r we count the number n_j of particles which are locally jammed; i.e. cannot move more than this cutoff distance. One might expect that this distance should be the same as the contact cutoff distance: the distance available for a jammed particle to move should be on the same order as the spacing between particles.

We plot n_j as a function of δ_r, and compare this to Z as a function of the contact cutoff δ. Results are shown in fig. 6.2 for configurations with varying n, with vertical axes scaled so that $n_j = n$ coincides with $Z = 4$. We find that the rattler length scale, i.e. the length scale where the the majority of the particles in the packing becomes locally jammed, is slightly higher than the contact length scale, typically by a factor of 2.5. This is reasonable based on the fact that a particle must be held in place by at least 3 contacts, and the allowed displacement will be at least as large as the largest gap among these contacts. Interestingly, the rattler and contact length scales are the smallest

for packings with particle numbers near $n = 100$, with typical values near $\delta = 10^{-10}$. At higher and lower particle numbers, the rattler and contact length scales are typically near $\delta = 10^{-8}$.

While the complication of the tail in the distribution of contact lengths persists, we see that the majority of contacts occur near some length scale in each packing. Although this length scale varies from packing to packing, the largest is below $\delta = 10^{-7}$. As such, we set a cutoff distance for both contacts and rattlers of $\delta = \delta_r = 10^{-6}$ which is safely above the range of distances in which we see the contact and rattler length scales, but not so high that we include "contacts" with spuriously large interparticle distances.

Once rattlers are identified, they can be excluded from the contact counting analysis. This must be done with care, however. We find that while rattlers themselves have freedom to move a significant distance and are thus unstable, they can still contribute to the stability of the rest of the jammed packing. This is demonstrated by applying a standard method of rattler removal which is considered suitable for ideal packings[26]. In this method, the rattler identification scheme is applied and the rattlers identified based on a set distance cutoff are removed, and then this is repeated (as some of the initially found rattlers may have been preventing other particles from moving) until no more rattlers are found. For our packings we find that generically, this iterative scheme will find more and more rattlers each iteration as particles are removed and more space is opened up, until the entire packing is dismantled.

However, this is not because the packing is not actually jammed (the linear program unjamming technique confirms this as discussed below), but because of the finite nature of the interparticle gaps. It is possible, for instance, that a rattler may be able to move some significant distance in one direction, while still remaining within the contact cutoff distance with particles which are po-

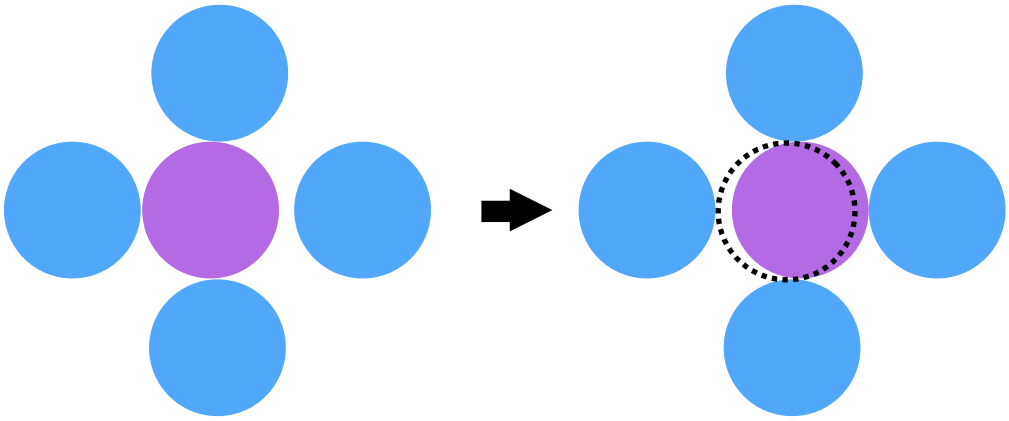


Figure 6.3: A schematic of a rattler (purple) which is under-constrained but still provides constraints on nearby particles. While the purple particle can move by some distance left and right, it is still always within some much smaller distance of the particles above and below it. Thus there exists some contact cutoff for which this particle is considered a rattler but still provides constraints on the rest of the packing.

sitioned in the perpendicular direction. This is illustrated schematically in fig. 6.3.

Because of these, when counting contacts, we do not remove rattlers from the packing. Rather, we count the contacts other particles make with rattlers, but we do not include rattlers and their contact numbers when calculating the average contact number. Thus the constraints imposed by rattlers are accounted for, but because rattlers are unstable we do not consider their degrees of freedom.

6.2 Metric jamming

To test for jamming, we apply the unjamming protocol described in the *Methods* chapter, which consists of a series of particle moves along directions found by the unjamming linear program as well as conditioning steps using the soft-particle energy minimization. Generically, we find that the initial arrested

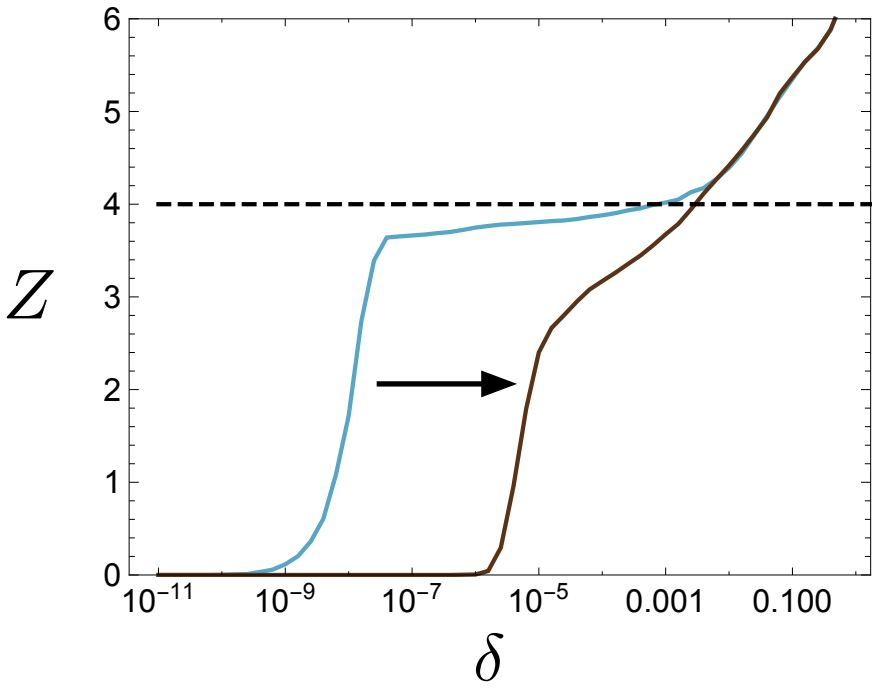


Figure 6.4: For a packing of $n = 800$, Z versus δ curve for the initially arrested packing (blue) and after an unjamming step (brown). There is a shift of 4 orders of magnitude in the spacing before and after unjamming, indicating that the initially arrested packing was not jammed.

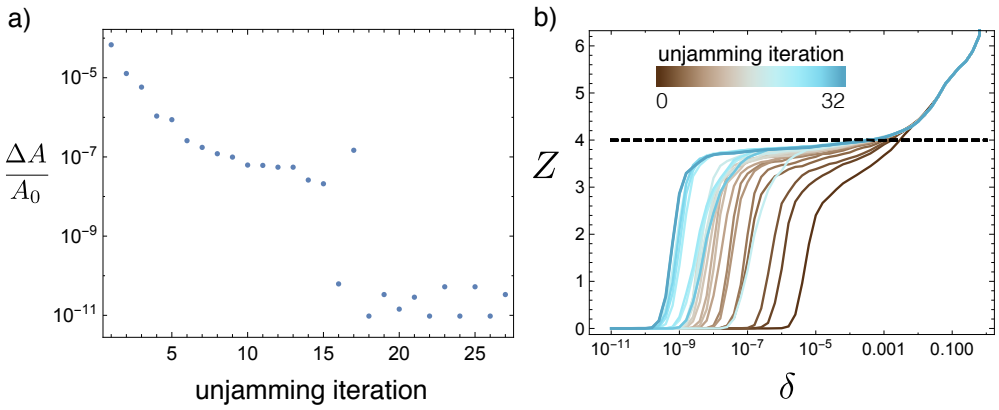


Figure 6.5: a) The decrease in surface area, as a fraction of the area at initial arrest, for a series of unjamming and relaxation steps. Most of the decrease occurs in the first few steps. b) Z vs δ curves after a series of unjamming steps. The earlier unjamming steps uncover significant unjamming motions

packings generated by the surface relaxation simulations are not collectively jammed. Significant unjamming displacements can be found resulting in an increase in the average spacing between particles as demonstrated by the Z versus δ curves in fig. 6.4.

After unjamming a packing, surface relaxation can be performed until the packing is arrested again. By repeatedly applying unjamming and surface relaxation steps, we can move the packing closer to a jammed state. To measure the progress made by this process, we calculate the decrease in ellipsoid surface area after each relaxation step, as a fraction of the initially arrested surface area (an example for a packing of $n = 800$ is given fig. 6.5a). For packings of $n > 100$, it is not uncommon to find unjamming motions which allow for a 1% decrease in surface area. For lower n the unjamming motions are typically much smaller, resulting in fractional decreases in surface area on the order of 10^{-8}. For $n = 800$, we see that the first iteration of unjamming and relaxation results in a decrease in ellipsoid surface area of between 0.05% and 1%. Successive iterations produce further decrease in the surface area, i.e. produce

denser packings, though the size of the successive relaxation steps tend to be much smaller. The decrease in relaxation resulting from the first iteration accounts for, on average, 72% of the total decrease in area. We consider packings to be jammed once no further significant surface relaxation can occur. We find that the decrease in surface area between steps tends to approach a value below 10^{-10}, representing convergence up to numerical limitations. To determine when our configurations are jammed, we run the iterative unjamming and relaxation process until 10 iterations in a row result in a fractional area decrease of less than 10^{-8}. In addition to the linear programming unjamming scheme, a full minimization of the soft external potential can be performed, which also finds significant unjamming motions but is typically very slow. A full performed in place of the linear program after iterations which result in an area decrease of less than 10^{-8}, but are never performed twice in a row.

To investigate the structure of the packing after each unjamming step, we plot Z versus δ for each iteration, as in fig. 6.5b. We see that after the initial unjamming step, the typical interparticle contact spacing is $\delta = 10^{-5}$. Successive iterations push the spacing to lower values of δ. Note that the curve does not shift monotonically towards lower δ. This is due to the iterations which employ the full energy minimization. The earlier full-minimization steps find relatively large unjamming motions which lead to a larger increase in spacing. Closer to jamming, the full-minimization steps result in more even spacing between particles, leaving the curves for these steps slightly to the right of the final curves for linear programming steps. Regardless, the Z versus δ curves converge to a static shape by the time the packing is jammed as determined by the criteria explained above, further confirming that these packings are indeed jammed.

Given the convergence of the surface evolution and the Z vs δ curves, we

can be confident that this method produces jammed packings, and furthermore these packings are metric jammed. In ideal packings, the unjamming linear program can find only unjamming motions for packings which are not locally or collectively jammed. To unjam a packing which is collectively jammed but not metric jammed would require motions of the particles along the direction of the evolving surface. However, due to the numerical nature of our packings, any packing is only jammed up to some numerical tolerance. Given this, a packing which is collectively jammed but not metrically jammed will allow motions with sizes on the order of this tolerance both in the unjamming and in the relaxation phase. By repeatedly applying these phases as we do, any metric unjamming motions should be uncovered after several iterations. Because we consider our packings jammed when repeated unjamming and relaxation no longer accomplishes anything, we can be confident that our packings are metrically jammed

6.3 Apparent hypostaticity: constraints induced by surface curvature

After confirming that a metric jammed state has been found, we identify rattlers and exclude them from the average contact number calculations (while still counting the contacts that other particles make with them.) Plotting Z versus δ without rattlers, in fig. 6.6, we see that at the value of $\delta = 10^{-6}$ chosen as the cutoff for rattler identification, the jammed packing has a contact number of $Z = 3.956$ and has not reached the expected isostatic value of $Z = 4$. As δ is increased, we do not see $Z = 4$ until $\delta = 2.5 \times 10^{-5}$, well above the rattler cutoff.

The first striking feature of this result is that for a system with a high

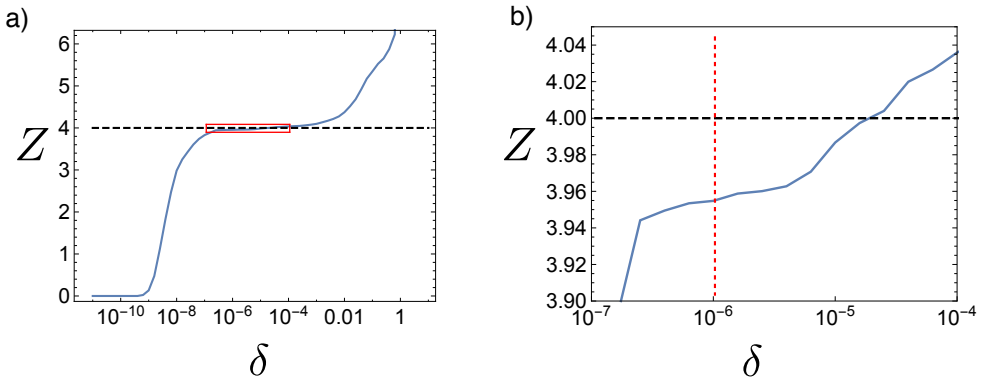


Figure 6.6: a) Plot of Z versus δ for a packing of $n = 800$ with rattlers excluded from contact counting. b) A close up of the highlighted region of (a). The vertical dashed line indicates the value of the imposed rattler cutoff. At this value, the packing has a value of $Z < 4$.

degree of hexatic order, one might expect an average contact number near $Z = 6$. This is not the case due to the fact that the curvature of the surface induces strain in the crystalline regions of the packing[8], combined with the hard contact constraints: because there can be no overlap, the obliqueness of the packing results in each particle having only 4 contacts in the crystalline regions of the packing.

The second, more significant feature of this result is that the packing appears to be hypostatic with a contact number of $Z < 4$. Packings in 2D require $Z = 4$ in order to be stable, so a value of $Z < 4$ appears to be at odds with the fact that these packings are metric jammed.

To further explore this apparent deficit in required contacts, we plot the number of missing contacts, $(4 - Z)n_j + 2$ (where, as earlier, n_j is the number of non-rattlers that make up the jammed sub-packing) as a function of n_j in fig. 6.7. The reason for the $+2$ comes from a careful degree of freedom counting argument and is especially relevant for low particle numbers. For a packing of n particles, there are $2n$ particle coordinates. The ellipsoidal

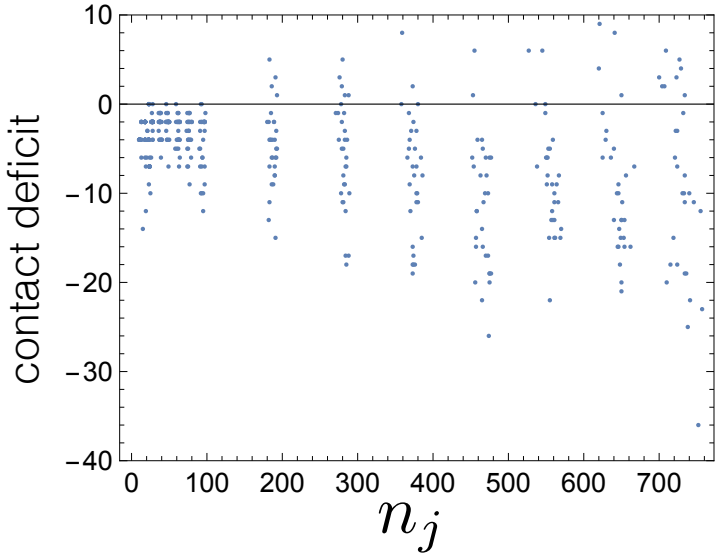


Figure 6.7: The number of missing contacts as a function of n_j (the number of non-rattlers), for 30 packings at each n. Note that the magnitude of the lower bound of this dataset increases roughly linearly with n_j.

surfaces studied here have a trivial rotational symmetry, reducing the number of relevant degrees of freedom to $2n-1$. The surface evolution motions that we allow for introduce an addition degree of freedom, bringing the number back to $2n$. Because the impenetrability constraints are inequality constraints, to constrain a point in an N dimensional configuration space, $N+1$ contacts are needed (e.g. 3 contacts are needed to hold a disk in place in $2D$.) Thus, a total of $2n+1$ contacts are needed, and because each contact is shared between two particles we expect to count $4n+2$ contacts. In fig. 6.7 we focus on the lower bound of this data as it is the fewest number of contacts that is most interesting, while anything with more contacts may be over-constrained due to states of self-stress[42]. We see that the deficit in the number of contacts grows roughly linearly with n_j, with a slope of 0.033.

Because we have confirmed that these packings are stable, there must be some additional constraints not accounted for solely by contact counting. To investigate the nature of these constraints, we focus on a simple ideal packing

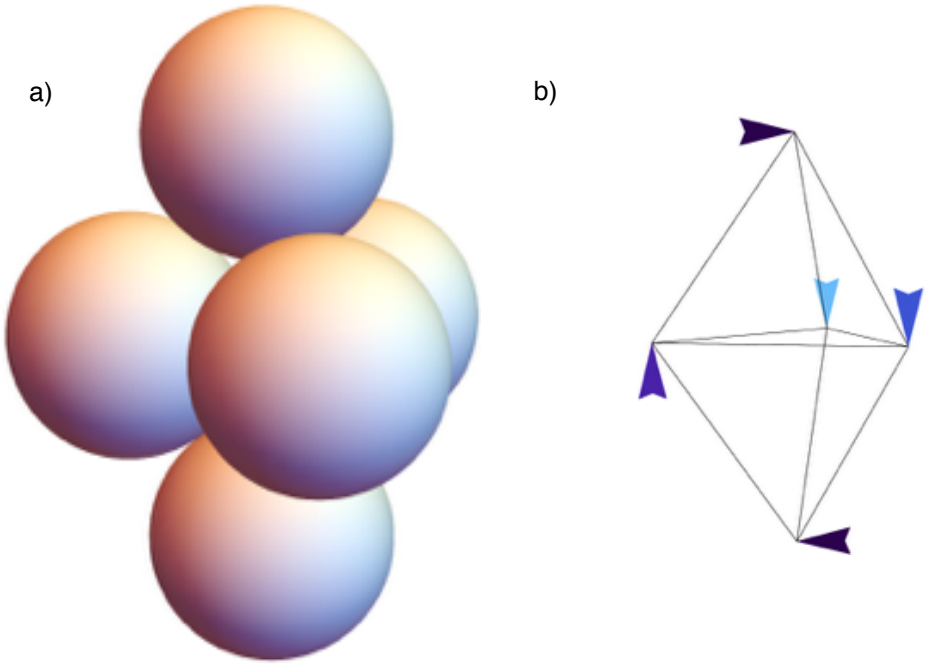


Figure 6.8: A simple jammed packing of five particles. a) The particle configuration, with five particles packed on a surface with an aspect ratio of $\sqrt{2}$. b) An unjamming motion found by the linear program for this packing, which appears allowed to first order, but any finite displacement in these directions is prevented by the surface curvature.

of five particles on a commensurate ellipsoid, pictured in fig. 6.8. The particle coordinates lie on the vertices of a triangular bipyramid with equilateral faces (i.e. two tetrahedra on top of one another) and the surface aspect ratio is $\sqrt{2}$. If we consider a fixed surface, this packing requires ten contacts for collective jamming, using the contact requirements outlined above. However, two particles have three neighbors and three particles have four neighbors, giving a total of nine interparticle contacts, falling short of the ten constraints required. Applying the unjamming linear program to this configuration uncovers an apparent unjamming motion which attempts to rotate the configuration in a uniform circular motion about an axis lying in the ellipsoid's equatorial plane. We find, however, that motion along this unjamming direction is impossible. The reason for this is that although this motion appears possible to first order, the surface imposes high order constraints which are not accounted for in the linearization of the system's constraints. As the particles try to move, the curvature of the surface causes them to change direction and collide with their neighbors, preventing motion. Thus we see that the curvature of the surface is the source of these additional constraints, resulting in the apparent hypostaticity.

Similar nonlinear constraints have been seen in other particle packings. For packings of ellipsoids and other non-spherical particles in Euclidean space, it is well understood that the non-spherical nature of the particles imposes higher order constraints on particle motions[22]. The reason for this is that to first order, the surface of these particles at a contact point is indistinguishable from a sphere and thus an infinitesimal rotation will not result in a collision. However, to higher order, a particle's surface at a contact point can look flatter, and the flattened surface will tend to collide with neighbors as it rotates (illustrated in fig. 6.9). In these systems, the linearization of constraints does

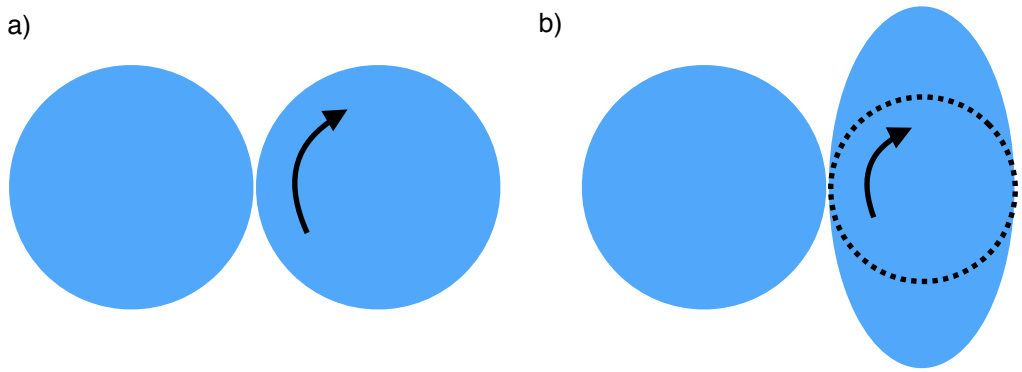


Figure 6.9: a) A spherical particle is rotationally symmetric, and so rotational degrees of freedom do not matter. b) Non-spherical particles can appear underconstrained to first order, in that infinitesimal rotations do not immediately result in overlap. However, finite rotations will cause overlap due to the relative flatness of the particle compared to a sphere.

not capture the variation in curvature of the particles' surfaces. For packings in curved space, it fails to capture the curvature of the space because locally the space looks flat.

6.4 Soft particles

As discussed in the *Background* chapter, packings of particles with soft, finite range potentials can be compressed beyond the point at which they become stable, i.e. the jamming point, with packing fraction ϕ_c at zero temperature and applied stress. For disordered packings (typically bidispersed packings in 2D or 3D, or monodipsersed packings which are able to avoid crystallization in 3D), as the density is increased above the jamming point, a number of specific mechanical properties are observed which distinguish them from crystalline packings. They show an excess of low-frequency vibrational modes (in contrast to the Debye law for ordered solids[]), as seen in a number of glassy and disordered systems[]. They also exhibit critical scaling laws as the density is increased above the jamming point. This is true for the contact number Z, as

well as the bulk modulus B and the shear modulus G. These scaling properties reveal the nonlinear nature of packings near the jamming point.

To generate packings of soft particles, we again employ the dynamic packing algorithm, replacing the hard particle interactions with a compact Hertzian interaction,

$$V_{ij} = \frac{\epsilon}{2/5}\left(1 - \frac{r_{ij}}{\sigma_{ij}}\right)^{2/5} \Theta\left(\frac{\sigma_{ij}}{r_{ij}} - 1\right)$$

where ϵ determines the energy scale, r_{ij} is the distance between the centroids of particles i and j, σ_{ij} is the sum of the radii of the particles, and Θ is the Heaviside step function enforcing a finite interaction range. Both monodispersed packings and bidipsersed packings with a radius ratio of 1.4 are produced. The surface relaxation proceeds past the point where all particles are overlapping, creating overjammed configurations. We apply an energy minimization to snapshots of the configuration taken throughout the relaxation process, fixing surface geometry and using a conjugate gradient method[]. From these energy minimized configurations, we expand the packing quasi-statically (i.e. minimizing the energy after each small expansion step) at fixed aspect ratio. This allows us to find the jamming point packing fraction ϕ_c corresponding to the initial energy minimized configuration (as ϕ_c is a property specific to a given packing, not a universal value[]), while also generating packing fractions in the intermediate range to study the mechanical behavior of the packings as a function of $\phi - \phi_c$.

First we investigate the vibrational modes of the packings. To calculate vibrational modes along the surface, we impose a harmonic energy penalty for particle motions normal to the surface, with an energy scale much larger than the particle interaction energy scale. We then calculate the Hessian matrix of the packing energy with respect to the particle coordinates in 3D and diagonalize it to find eigenfrequencies and eigenmodes, and ignore modes

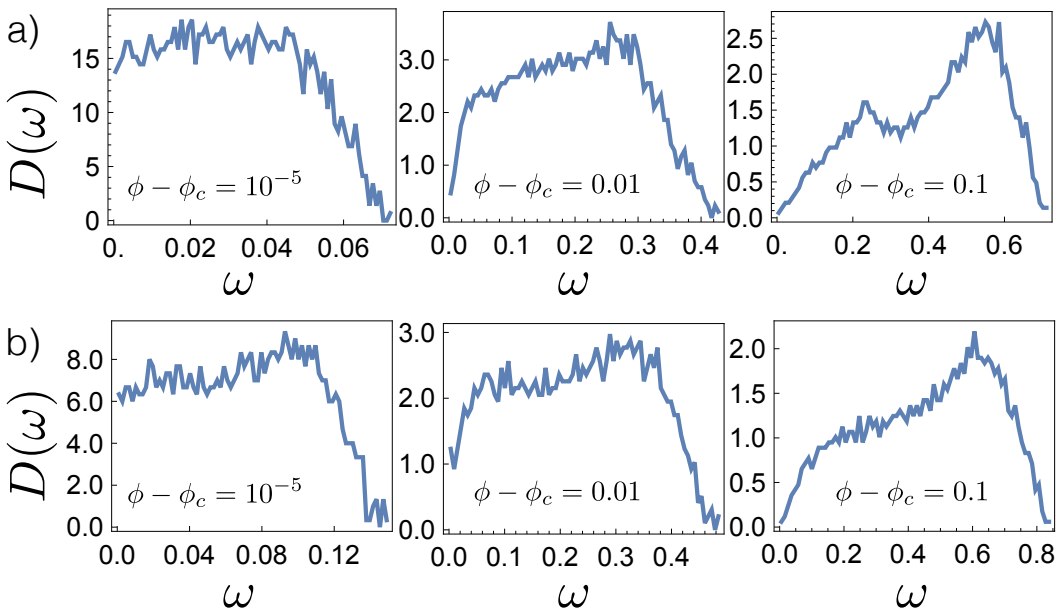


Figure 6.10: The density of states of vibrational modes at various packing fractions, for (a) monodispersed and (b) bidispersed packings. As the packings approach the jamming point, there is an abundance of low frequency modes.

normal to the surface which are easily identifiable due to their much larger magnitude.

Figure 6.10 shows the density of vibrational frequencies for packings at various values of $\phi - \phi_c$. We see that as ϕ approaches ϕ_c, the so-called boson peak[**?**], an excess of low frequency modes, shifts towards $\omega = 0$, and at very low packings fractions ($\phi - \phi_c = 10^{-5}$), there is a significant density of states extending down to $\omega = 0$.. This is a signature of marginal stability: below ϕ_c, the particles are not in contact and all vibrational modes are zero-frequency modes. As a packing approaches the jamming point from above, it develops an abundance of very low frequency modes. Both monodispersed and bidispersed packings show similar behavior.

Next we look at the scaling of $Z - Z_c$ with respect to $\phi - \phi_c$, where Z_c is the contact number at which packings are marginally stable. For bidipsersed packings, we see a power law behavior with exponent 0.50 ± 0.03 (mean and

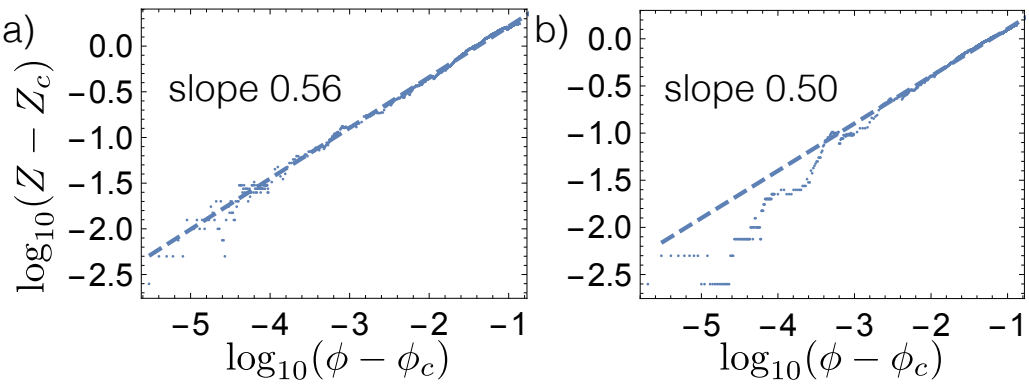


Figure 6.11: Scaling of contact number above the jamming point for (a) a monodispersed packing and (b) a bidispersed packing, with power law fits displayed. For the bidisperse packing, the power law fit is for data above $(\phi - \phi_c) = 10^{-3}$.

standard deviation, based on 9 packings), at least at higher packing fractions ($\phi - \phi_c > 10^{-3}$), consistent to the value of 0.5 seen for bidisperse packings in Euclidean space[**?**]. Bidispersed packings deviate from this behavior at lower packing fraction. For monodisperse packings, we see a slightly higher exponent of 0.58 ± 0.04 (based on 10 packings). The changing contact number is an indication of nonlinear behavior: as the packing is compressed, packings that were not in contact start interacting, changing the structure of the energy landscape.

Finally, we investigate the scaling of the elastic moduli of packings near jamming. Disordered jammed systems exhibit a number of nonlinear elastic behaviors. These can be seen by comparing the instantaneous response and infinite-time response of their bulk and shear moduli. The instantaneous moduli are calculated by applying a uniform compression or shear and then calculating the response of the system pressure. The infinite-time moduli are calculated by applying the same deformation, but then minimizing the configuration energy before calculating the system response. If the system is linear, then the deformation will scale or shear the configuration's local energy land-

scape but will not change its structure and the configuration will still be at a local energy minimum. In disordered jammed materials, however, a difference is observed between the instantaneous and infinite-time response. For the bulk modulus, both the instantaneous response B_0 and the infinite-time response B_∞ show a power law which scales as $(\phi - \phi_c)^{\alpha-2}$, where α is the exponent of the interaction potential (e.g. $5/2$ for Hertzian interactions.) Despite having the same power law exponent, the power laws have different coefficients such that $B_\infty < B_0$. The difference is more extreme for the shear moduli G_0 and G_∞ which actually have different power law exponents: $G_0 \propto (\phi - \phi_c)^{\alpha-2}$ and $G_\infty \propto (\phi - \phi_c)^{\alpha-1.5}$.

The bulk and shear moduli can be derived from the pressure tensor. The pressure tensor of the packing can be calculated by[]

$$p_{\alpha\beta} = A^{-1} \sum r_{ij\alpha} \frac{r_{ij\beta}}{r_{ij}} \frac{dV}{dr_{ij}}$$

where A is the surface area, V is the full configuration energy, and $r_{ij\alpha}$ is the component of $\vec{r}_{ij}$ along the surface coordinate alpha (we take $\vec{r}_{ij}$ in 3D and take the projection along the surface tangent vectors $\vec{t}_\theta$ and $\vec{t}_\phi$ in the polar and azimuthal directions, respectively, at both positions $\vec{r}_i$ and $\vec{r}_j$ and use the average between the two points.) From this, the bulk modulus can be calculated from the pressure, $B = \phi \frac{dp}{d\phi}$, where $p = \frac{1}{2} \sum_\alpha p_{\alpha\alpha}$. The shear modulus is given by $G = \frac{d\Sigma}{d\gamma}$ where $\Sigma = p_{\theta\phi}$ and γ is the applied shear. The shear is applied by twisting the configuration around the ellipsoid symmetry axis such that $\frac{ds_\phi}{ds_\theta}$ (where s_θ and s_ϕ are arclengths along the polar and azimuthal directions) is constant, i.e. there is a uniform shear rate across the surface. After applying a shear we fix the positions of several particles near the poles of the surface to ensure that the packing will not relax back completely after

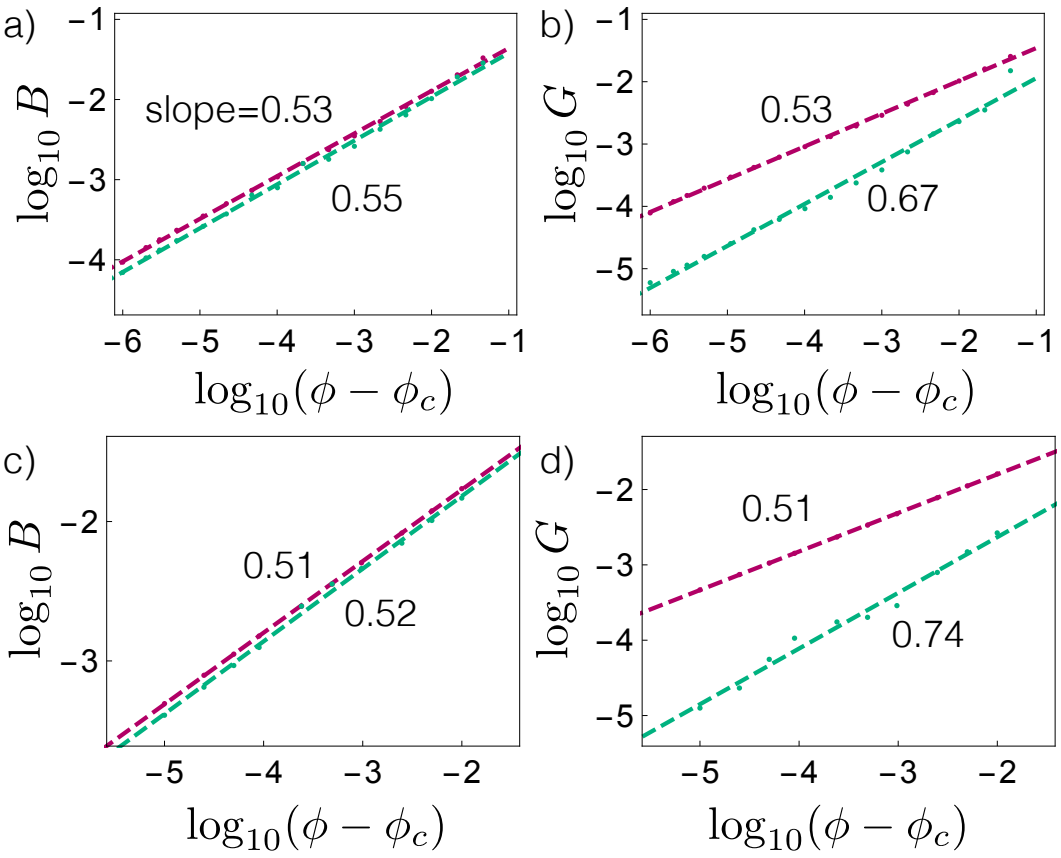


Figure 6.12: (a) Bulk and (b) shear modulus for a monodisperse packing. (c) Bulk and (d) shear modulus for a bidispersed packing. In all cases, we see a downward shift between the instantaneous and infinite-time moduli, indicating nonlinear behavior. For the shear moduli, we also see a change in the scaling exponent. The change is not as extreme for the monodispersed packing as in the bidispersed packing, indicating behavior intermediate between that seen for crystalline and random jammed packings.

the energy minimization, and we exclude the fixed particles (and the area they cover) from the pressure tensor calculation.

For the bulk modulus, we see the same behavior in monodispersed packings as that expected for disordered packings: we see an exponent of about 0.5 (as expected for a Hertzian potential) for both B_0 and B_∞, and we see that $B_\infty < B_0$. For the shear modulus, we do see a change in the exponent, but not as large of a change as is expected for disordered packings. We see power laws of $G_0 \propto (\phi - \phi_c)^{0.53}$ and $G_\infty \propto (\phi - \phi_c)^{0.67}$, while for a disordered

packing one would expect a power law of $G_\infty \propto (\phi - \phi_c)^1$. In the bispersed case, interestingly, we see the exponent increase to 0.74 rather than 1. This is higher than the monodispersed case, but lower than usually seen in disordered packings.

It is rather surprising that monodipsersed packings on a 2D surface share so many properties with disordered packings, given that the monodisperse packings are relatively well ordered. There are two effects, both stemming from geometric frustration, which lead to the packings exhibiting these properties near the jamming point. First, the surface curvature necessitates topological defects in the packing. These defects correspond to localized regions of disorder. Second, the curvature causes strain in the nearly-hexagonal packing. Thus, instead of the surface being covered by a perfect hexagonal lattice with each particle in contact with six neighbors, the lattice is slightly oblique and most particles have four contacts — allowing for the average contact number $Z \approx 4$. The fact that the lattice is still largely well ordered, but with localized regions of disorder, may explain why the shear modulus scaling laws appear to be intermediate between what one would expect for a fully ordered and a fully disordered system.

Chapter 7

Percolation transition in the packing of bidispersed particles on curved surfaces

BIDISPERSED hard spheres provide a simple model of a glass transition[], at high temperature they flow freely, but as the temperature is reduced they become kinetically arrested and form rigid but highly disordered structures[28]. At zero temperature and stress, a similar *jamming* transition to rigidity occurs as a function of density[20], in 2D this tends to occur near a packing fraction of $\phi = 0.84$[43].

Sphere packings, the high density and zero temperature limit of these processes, have been extensively studied in both 2D and 3D Euclidean space[20, 21, 28, 44] revealing strong dimensional dependence: 2D monodispersed spheres tend to crystallize readily, because the locally dense hexagonal packing fills space; in 3D the locally dense tetrahedral packing cannot fill space, permitting a random close packed structure that is the subject of much debate[]. Even in 2D, however, disorder can be induced in bidispersed systems. Molec-

ular dynamics simulations have shown that there is a transition from order to disorder as the degree of bidispersity is increased[45, 46, 47, 48], and statistical models of bidispersed particle packings have been used to predict the local features of disordered bidispersed packings[49, 50]. The degree of order or disorder can be measured by an order parameter such as the hexatic bond orientational order[4].

As discussed in the *Background* chapter, monodispersed packings on spheres undergo a scar transition. They are mainly crystalline with a transition between isolated defects for small particle number and chains of defects or *scars* akin to grain boundaries in bulk systems that occur above a critical number of particles $N_c \approx 110$ and grow with system size[10, 9]. The scars may join in a asterisk-like motifs[**?**] and are aligned by anisotropic curvature[51].

In this chapter, we investigate the packing of bidispersed particles on a spherical surface as a simple model of how glasses interact with curvature. We determine the packing fraction as a function of particle number N, bidispersity $b = (r_1 - r_2)/(r_1 + r_2)$ where $r_1 \geq r_2$ and fraction of large particles $\chi = N_1/N$. We determine the packing fraction and connectivity as a function of these parameters, and by identifying topological defects from the neighbor graph show that variation in these parameters is explained by a percolation transition due to growth of the scars.

7.1 Simulation parameters

Packings with high coverage packings were created using a modified version of the dynamic packing algorithm described in subsection 3.1.2. We use the same particle dynamics (including only diffusion), but generate the packing on a shrinking spherical surface, whose radius starts at $R = 1$ and decreases

at a constant rate. The surface relaxation timescale τ_r is given by the time it takes the radius to shrink to $R = 0$. Particles are randomly assigned to two categories corresponding to the larger and smaller radii respectively. The larger particle radius is set at $r_1 = 0.03$. The radius of the smaller particles is calculated from $r_2 = r_1(1 - b)/(1 + b)$. The simulation proceeds until the surface can no longer shrink without creating particle overlaps as outlined in 3.1.2.

As in chapter 7, configurations produced by this procedure are referred to as *arrested*, because they remain metastable if the simulation is restarted, however eventually a stochastic diffusion move will unjam the arrested configuration, potentially facilitating further inflation and a consequent increase in the packing fraction. This process occurs in real glasses and is known as *aging*. We employ the unjamming linear program outlined in subsection 3.2.4 to artificially age the arrested structures, by finding and executing an *unjamming* motion of the particles and further inflating them. Iterative unjamming and inflation guides the packing toward a state that is collectively jammed with respect to movement of the particles and further inflation. Each arrested structure was subjected to this artificial aging process to produce a corresponding ensemble of *jammed* structures.

7.2 Packing fraction

An ensemble of simulations was run for values of bidispersity on the interval $b \in [0, 1]$ with $N = 800$ particles and $\chi = 1/2$; the packing fraction Φ, i.e. the ratio of surface area of the sphere enclosed by the particles to the total area, was calculated for each configuration and the a fraction of the best were selected for each value of b. The mean of this restricted set is shown as a function of

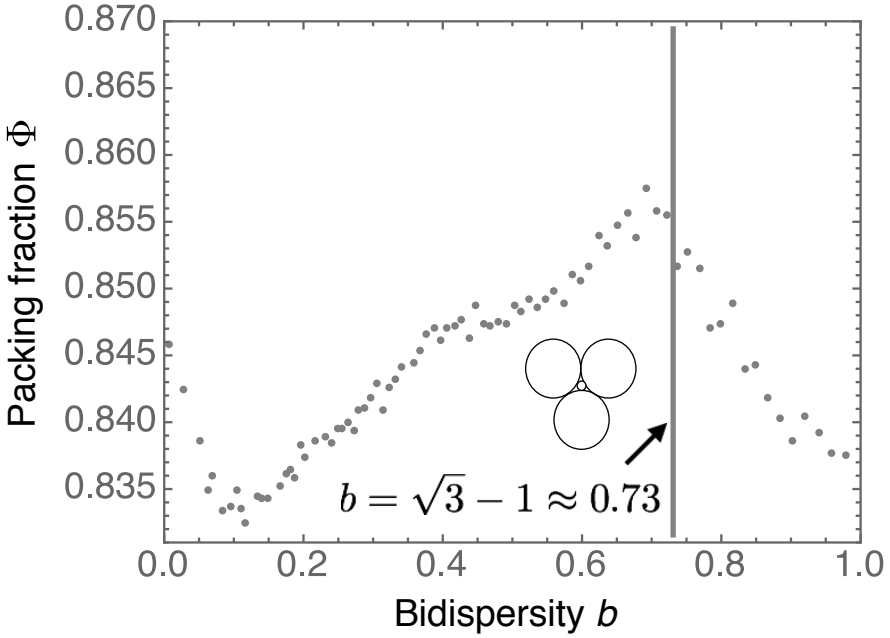


Figure 7.1: Packing fraction as a function of bidispersity $b = (r_1 - r_2)\,/\,(r_1 + r_2)$ where $r_1 > r_2$ for several values of particle number N. The maximum near $b = 0.73$ occurs for an Apollonian packing, i.e. where smaller particles fit in the interstices of the larger particles as depicted in the inset.

b in fig. 7.1. From the monodispersed case, a small amount of bidispersity immediately introduces disorder and reduces the packing fraction as expected. Above a critical value of bidispersity $b_c \sim 0.2$, however, this trend reverses and Φ increases again up to a maximum value of 0.858 at $b = b_A \sim 0.7$ and then decreases as $b \to 1$. The maximum at $b = b_A$ is immediately explicable: it corresponds to the special ratio of radii at which the smaller particles fit exactly in the interstices between the larger particles [fig. 7.1 inset]; we call this the *Apollonian point* in reference to the tiling. The packing fraction at $b = 1$ corresponds exactly with that for $N/2$ particles. No such immediate explanation is obvious for the low bidispersity results, which appear to be well mixed; we therefore seek a more detailed understanding of the structure.

7.3 Scar Percolation

For monodispersed particles, neighbors are assigned by constructing a Voronoi tessellation from the particle centers of mass, partitioning the surface into N polygonal regions closest to a particular particle[3]. Two particles are considered to be neighbors if they share an adjacent edge on the Voronoi tessellation. Generalizing this construction to bidispersed particles using a weighted distance fails to uniquely assign all points on the surface to a particle; two alternatives that have been proposed[**?**] are the r*adical tessellation* and the *navigation map*, both of which recover the Voronoi tessellation in the limit of monodispersed spheres. The radical tessellation utilizes the radical plane as a separatrix between each pair of particles; the navigation map partitions the surface into regions closest to the surface of the particles rather than their center of mass. We found little difference between quantities calculated from these constructions and use the radical tessellation exclusively in the remainder of

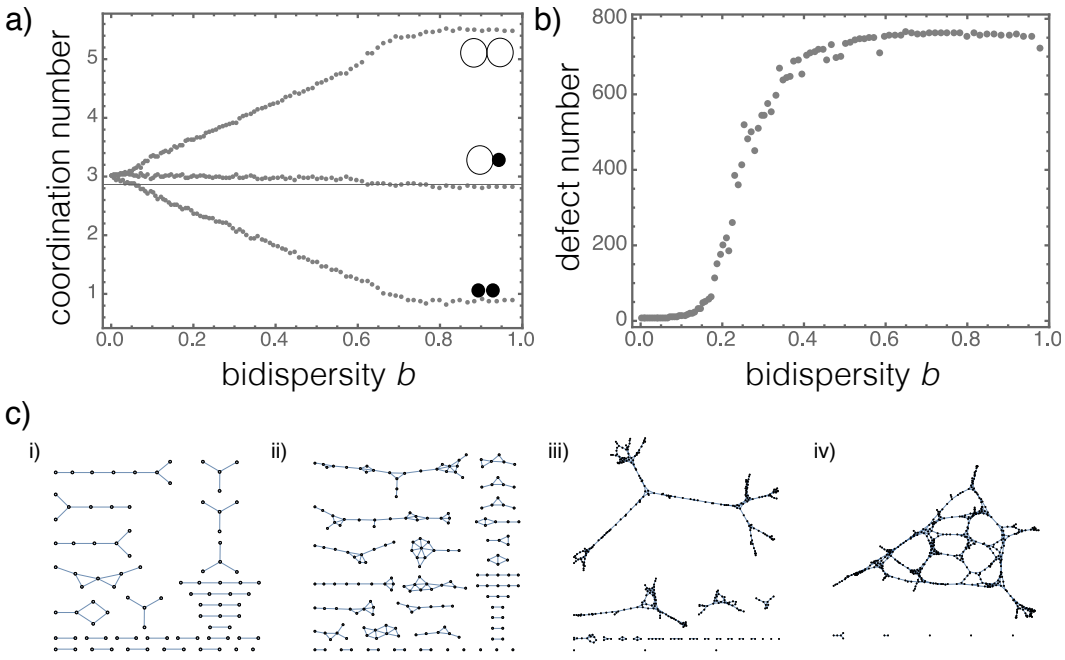


Figure 7.2: a) Average coordination number for large-large, large-small and small-small inter-particle contacts respectively; b) Number of particles with coordination number $C \neq 6$ as a function of bidispersity; c) Representative defect subgraphs for i) $b = 0$; ii) $b = 0.1$, iii) $b = 0.2$ and iv) $b = 0.4$ illustrating growth of the scar network.

the chapter.

From the radical tessellation, the adjoint neighbor graph was constructed for each packing and the coordination number determined for each particle. In fig. 7.2(a) we plot the average coordination number per particle, separated into large-large, large-small and small-small contacts. At infinitesimal b, each particle has six neighbors, three smaller and three larger on average. With increasing b, the number of large-small contacts per particle remains a constant value of three; larger particles gain more large neighbors while smaller particles lose small contacts. At the Apollonian point, the smaller particles are surrounded by three larger neighbors, while the larger particles are on average surrounded by six large neighbors and three smaller neighbors. For $b > b_A$, the coordination numbers remain constant, consistent with the discussion above whereby the smaller particles simply are caged within the interstices of the larger particles.

In fig. 7.2(b), we plot the number of particles that possess a non-hexagonal coordination number as a function of bidispersity. This reveals that a transition is occurring: As b increases from zero, this number rises slowly until a point $b \approx 0.1$, when it increases much more rapidly. This point coincides with the position of the minimum in the packing fraction in fig. 7.1. Above bidispersity $b \approx 0.5$, essentially none of the particles possess six neighbors. To determine the nature of this transition, it is necessary to examine the detailed evolution of the structure. A useful tool to do this is to take the defect subgraph of the connectivity graph by deleting all vertices that have six neighbors. Illustrative examples of these subgraphs for different values of b are depicted in fig. 7.2(c). At $b = 0$, [fig. 7.2(c)(ii)] the subgraph is disconnected; the small connected subgraphs correspond to the previously-studied scars, and are essentially linear in morphology, with a small number of branches. As bidispersity

increases to $b = 0.1$ [fig. 7.2(c)(ii)] the connected subgraphs are still recognizably scar-like in nature, though many have a more branching morphology, but substantially longer. By $b = 0.2$. [fig. 7.2(c)(iii)], the defect subgraph remains disconnected, but is now dominated by a few largely connected graphs that are mostly linear with branches. For $b = 0.4$ [fig. 7.2(c)(iv)] the defect subgraph is mostly a single connected structure with a small number of additional isolated defects; it is no longer branching, but consists of linear elements that link into a foam-like structure. For $b > 0.4$, the defect subgraph retains this structure, but becomes more dense.

Chapter 8

Experiments

THE simulations studied in this thesis model the phenomenon of arrest relaxation in Pickering emulsions. As discussed in section 2.2, a Pickering emulsion is a mixture of two immiscible fluids plus colloidal particles. Arrested emulsion droplets are created by a process of deformation, adsorption, relaxation, and arrest: some deformation is applied to an emulsion droplet driving it into a non-spherical shape with increased surface area; because of the increased surface area, more colloidal particles can adsorb at the droplet interface; the deforming force is removed and the droplet begins to relax towards an area-minimizing spherical shape; finally, if the surface coverage of the colloidal particles is large enough, the particles become densely packed as the surface area is reduced and elastic contact forces between particles prevent the surface from relaxing any further. This results in a droplet held in a non-spherical shape which is covered in a dense particle monolayer.

In this chapter we describe the creation of arrested emulsion droplets, as done by Patrick Spicer's group in the Department of Chemical Engineering at the University of New South Wales. The system components are described as

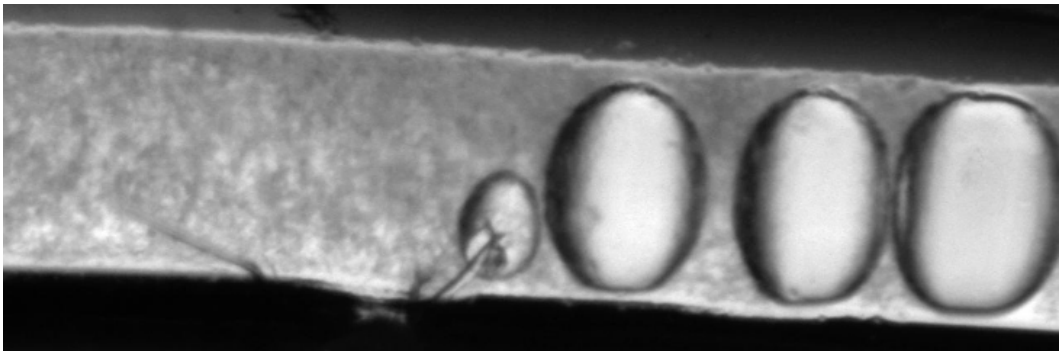


Figure 8.1: Elongated pickering emulsion droplets being created by a microfluidics device. Image courtesy Patrick Spicer.

well as the specific mechanisms by which droplets are deformed and relaxed. We also discuss various physical effects which influence droplet formation as well as challenges in creating the droplets. Finally, we present an image analysis applied to a few example droplets, to demonstrate how these experiments can be compared to the simulations presented in the rest of this thesis.

8.1 Pickering emulsion preparation

Pickering emulsions are composed of two immiscible fluids (typically an aqueous and oil-based phase) and colloidal particles. The Spicer group uses hexadecane as the oil phase, carbopol (a yield stress fluid) as the aqueous phase, and silica colloids, slightly larger than 1 micron in diameter. The colloids are first dispersed within the oil phase and allowed to equilibrate. A coaxial-flow microfluidic setup[52] is then used to extrude oil droplets into a surrounding carbopol phase. Oil is ejected from a central syringe, while carbopol is flowed past this syringe. A Plateau–Rayleigh instability causes the stream of oil to separate into distinct droplets; however, they are held in an elongated shape by the yield stress of the carbopol. The colloids which are initially held internally in the oil phase can then adsorb at the droplet interface. An image of

elongated droplets in production is given in fig. 8.1.

8.2 Droplet relaxation

Once elongated droplets have been formed, they can be relaxed by altering the carbopol so that the droplet surface tension overcomes the carbopol's yield stress. Carbopol is an aqueous suspension of high molecular weight cross-linked poly(acrylic acid) polymers[53]. By changing the chemical environment, the polymer network can swell or deswell, changing the rheological properties of the carbopol fluid[54]. Thus, by adding salt to the carbopol matrix surrounding an elongated droplet, the yield stress is reduced and the droplet can then relax towards a spherical shape. By lowering the concentration of added salt (which slows its diffusion diffusion into the carbopol), the relaxation rate of the droplet can be slowed.

8.3 Other physical effects and complications

As discussed in chapter 5, a number of phenomena affect the Pickering emulsion systems and may influence the ordering of the particles on the final arrested droplet. The most prominent effects observed in the experimental system are interparticle attraction and gravity.

Interparticle attractions are clearly present in the system, as evidenced by particles aggregating into rafts before arrest. An example of these aggregates is shown in fig. 8.2 for a sparsely covered droplet. Interparticle attractions are most likely due to capillary effects[55]. If a particle induces a deformation in the fluid-fluid interface around it, this deformation will increase the surface area of the interface. If two nearby particles do this, the area added by the induced deformations can be reduced by moving the particles closer together.

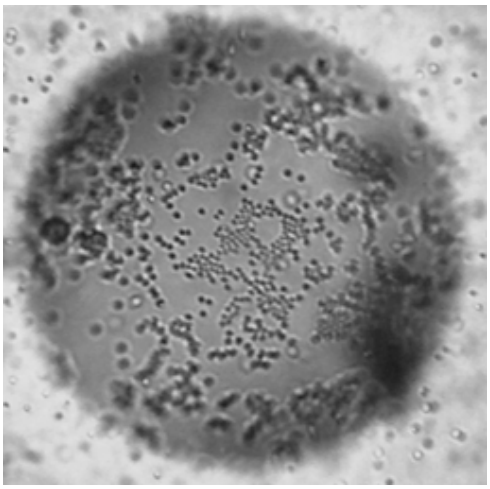


Figure 8.2: Aggregates of particles form due to attractive interactions. Image courtesy Patrick Spicer.

Thus, surface tension drives attractive forces between particles.

The question then arises of what causes the particles to deform the surface. Particles on a spherical interface, in the absence of gravity, will not deform the interface[55]. However, when gravity is significant, this will pull particles downward and induce a surface deformation. This would cause particles on the top and bottom of the surface to be attracted to one another, but would not affect particles on the side of the droplets. Due to constraints on imaging, it is more difficult to view the sides of the droplets than the tops and bottoms, so this behavior has not been confirmed. On a surface with anisotropic curvature, quadrupolar interactions will be induced[56], although this effect is likely small for the system at hand. Quadrupolar interactions can lead to square-lattice ordering being induced in packings on curved surfaces, but this requires careful tuning[**?**].

Gravity is observed to cause particles to sink towards the bottom of a droplet before arrest. While thermal effects in colloidal suspensions often overcome gravitational effects, the density and size of the silica particles used prevents this.

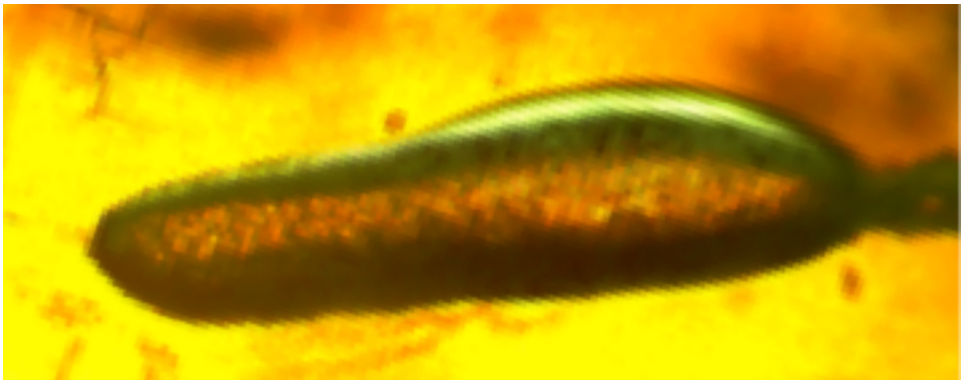


Figure 8.3: An asymmetrically relaxed droplet. Image courtesy Patrick Spicer.

One challenge to creating symmetric ellipsoidal droplets is achieving uniform addition of salt to the surrounding carbopol matrix. If the salt is not added uniformly, it may cause one part of the droplet interface to yield sooner, leading to asymmetric relaxation as pictured in fig. 8.3.

8.4 Image analysis

Given a bright-field microscope image of an arrested droplet, an image analysis can be performed to identify defects in the packing. This is done by identifying particles based on brightness maxima in the image. Once particle locations have been determined, a Delaunay triangulation is performed to determine coordination numbers and thus defect charges for each particle.

Figure 8.4 shows three experimental images, as well as the same images with particles colored by coordination number. Note that coordination numbers for particles near the edge of the image are not shown, due to spurious defects being detected at the boundary of the packing, as well as the fact that near the edges, the droplet surface is being viewed at an angle. It is clear in each image that there is a high degree of hexagonal ordering. However, there appear to be more defects in these packings that observed in the static simula-

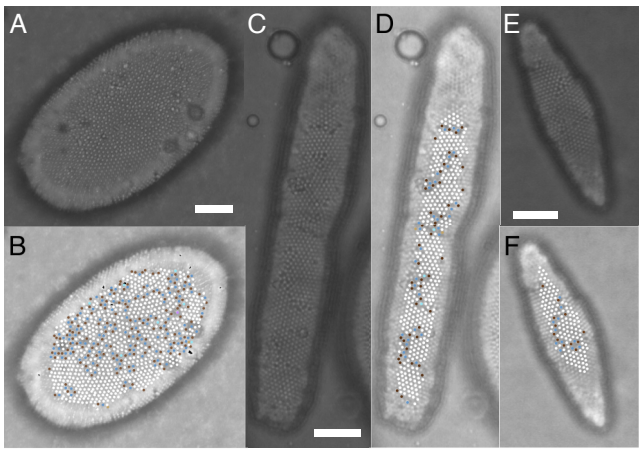


Figure 8.4: Experimental data for particle-stabilized droplets of aspect ratio (A, B) 1.6, (C, D) 5.1,and (E, F) 3.0. Scale bars represent 15 mm. (A, C, E) Microscope images; (B, D, F) reconstructed particle positions, colored by coordination number as determined by Delaunay triangulation of the particle centroids—4: light brown, 5: dark brown, 6: white, 7: dark blue, 8: light blue, 9: purple. In general, defects appear to be more common and are more likely to be found at low- curvature regions of the droplet in the experiments than in static simulations. Microscopy images courtesy Patrick Spicer.

tions. It may be possible that this is due to the dynamical effects discussed in chapter 5. A large scale comparison of theoretical and experimental droplets is underway to determine which effects dominate in the experiments.

Chapter 9

Conclusion

9.1 Summary of results

In chapter 4, we studied the role of geometry on hard sphere packings by using an inflation algorithm to generate packings of spheres on ellipsoidal surfaces with fixed aspect ratio, and investigated the structure of topological defects in these packings. We found, as expected, that the defect charged density is predicted by Gaussian curvature. However, finite size effects can be observed which cause deviations at fine length scales. For packings on highly prolate ellipsoids (aspect ratios above 2.6), there is a secondary peak in the defect number density, between the low-curvature center and the high-curvature ends. This excess in the number of defects does not correspond to any obvious feature of the surface geometry. It may, however, be due to the hard-particle nature of these packings, as packings of soft-particles do not exhibit a clear secondary peak.

We see a scar transition on spherical surface, as expected, although we see that for hard particles the scar transition occurs at a lower particle and is much sharper than for soft particles. A heuristic model is posited which accounts for

"core" defect charges, favored in regions of high curvature, plus scars attached to these core defects, whose lengths grow as the inverse of curvature. This model does well at capturing the the defect number distribution for low aspect ratio surfaces, and roughly captures the defect number distribution for highly prolate ellipsoids while not accounting for the secondary peak. The model correctly predicts a softening of the scar transition as the aspect ratio increases.

We find that scars have a directional coupling to curvature, although this coupling is highly nontrivial. On prolate ellipsoids, coupling is observed to directions of both high and low principal curvatures at different aspect ratios and surface regions. For very high aspect ratio (greater than 5), we observe quadrupolar order, in which scars tend to lie along directions between the directions of principal curvature. On oblate ellipsoids, the coupling is simpler, as scars tend to form chiefly along the high-curvature central rim.

Finally in chapter 4, we perform a thorough search of our aspect ratio and particle number parameter space at low particle number in order to build a catalog of symmetric packings. We find a number of patters: packings of n-fold rotational symmetry occur in bands along increasing particle number and aspect ratio, as extra rows of n particles are added and the surface elongates; low particle numbers are more likely to be achiral; and oblate packings are capable of high symmetry at higher particle number than are prolate packings.

In chapter 5, we study packings created using a dynamic surface evolution algorithm, developed for this work, which incorporates additional physical forces, namely interparticle attraction and gravity. We see qualitatively different results for varying relaxation rates (i.e. by varying the diffusion timescale relative to the relaxation timescale.) For slow relaxation, higher quality packings are produced, as measured through both packing fraction and defect number. Slower relaxation rates result in a secondary peak in the defect density,

although this peak is much more pronounced than for packings on static surface, indicating that the surface evolution plays a critical role in ordering. For fast relaxation, the defect density instead shows a secondary peak at the middle of the droplet rather than between the middle and the ends. We find that this is due to surface evolution causing particles to densify towards the ends of the droplet due to surface deletation, causing a front of order that moves in from the ends to the center.

For attractive particles, a well-ordered packing with gaps and few defects forms early on and persists throughout droplet evolution, resulting in more uniform ordering but lower order on average than for the purely diffusive slow relaxation case. For significant gravity, we again see more uniform ordering along the axis of surface in the arrested state due to gravity promoting early ordering, though we see a significant variation in ordering around between the top and bottom of the surface.

The generic trends we see with different dynamical effects are that: 1) these effects suppress diffusion, resulting in lower order on average; 2) they favor densification according to some spatial distribution early in the surface evolution, which translates into order which is quenched to some extent into the final packing.

Finally, in chapter 7 we examine the rigidity of packings produced by arrested relaxation and study the concept of jamming on curved surfaces. Because these packings are produced on an evolving surface with no boundary, we introduce the concept of metric jamming to describe packings of hard spheres which are stable to both collective particle motions combined with surface evolution. By adapting Donev's linear program[**?**] for unjamming to work on curved surfaces, we are able to show that generically the arrested packings produced by the relaxation algorithm are not collectively jammed, but a com-

bination of unjamming and further surface relaxation can be used to age them into a metric jammed state. The metric jammed states appear to be hypostatic based on their average contact number; this is due to the surface curvature imposing non-linear constraints on the packing. The deficit in the number of contacts grows linearly with the number of particles; the lower bound in contact number of shared between monodispersed and bidispersed packings. It is especially interesting that monodisperse packings can be marginally stable, as monodisperse packings in flat space are typically crystalline and overconstrained. While the monodispersed packings we see on curved surfaces are largely crystalline, geometric frustration causes strain within the crystalline regions, allowing for a low contact number.

We also investigate packings of particles with soft Hertzian interactions in order to study the behavior of packings near the jamming point. In a number of ways, the monodispersed packings resemble disordered jamming packings much more than crystalline packings, despite their high degree of order. We see that for for monodisperse packings, the spectra of dynamical modes resembles those seen in packings in flat space near the jamming point: packings just above the jamming point have an excess of low frequency modes, which reduces as the packing is compressed above the jamming point. We also find that the elastic properties exhibit nonlinear behavior: the infinite-time bulk and shear moduli are lower than the instantaneous bulk modulus, and the infinite-time shear modulus has a different scaling exponent, although the exponent is between that expected for crystalline and disordered materials.

9.2 Open questions and future work

Regarding the geometric effects in packings on static surfaces, two major questions remain. One is the origin of the excess defects seen on highly prolate ellipsoids. This effect is not explained by Gaussian curvature, and there is evidence that it is due to hard particle interactions. Why this should be the case is not clear. The highly complex directional coupling of scars to curvature anisotropy is also intriguing, and a simpler analytical model which predicts this would be satisfying.

The dynamic simulations make a number of very clear predictions, as outlined above. However, many of these are yet to be verified. As our experimental collaborators are able to produce and analyze larger numbers of arrested droplets, and better control the relative strengths of different dynamic effects, it will be interesting to see how the experiments compare to the simulations presented here.

In terms of jamming, it is unclear why the contact deficit scales with particle number. There are clear connections between contact deficit and surface topology: the curved surfaces appear to allow for larger arching structures within the contact network, but the distribution of the size of these arches (i.e. the number of sides of the polygons within the contact network) has not been quantitatively predicted.

An important new direction that this work could be taken involves modeling the deformation of the surface on which the packing lies, for example due to feedback from particles. We have begun work on this by studying packings in which particles are held on the surface by a harmonic potential. Preliminary findings are presented in appendix A.

Appendix A

Buckling with soft surface constraints

All the work in this thesis has dealt with particles constrained exactly to lie on a specified surface, either a surface that is fixed or evolves through some predetermined family of shapes. While this model helps us understand the effects of curvature on particle packings, a more robust model would allow the surface to deform more freely. The surface might evolve based on surface tension, and particles could introduce feedback on the surface. One simple step in this direction is to still specify the shape of the surface, but to relax the constraint holding particle on the surface. This can be acheived, for example by using a surface-particle potential which is harmonic in the particle's distance from the surface.

The dynamic relaxation algorithm can be easily modified to accomodate soft surface constraints. We have done so and run preliminary simulations to see what effect these soft particle constraints have. Fig. A.1 shows an example of one result. This simulation includes Hertzian inter-particle interactions, diffusion, and a harmonic surface constraint. It is apparent that packing is

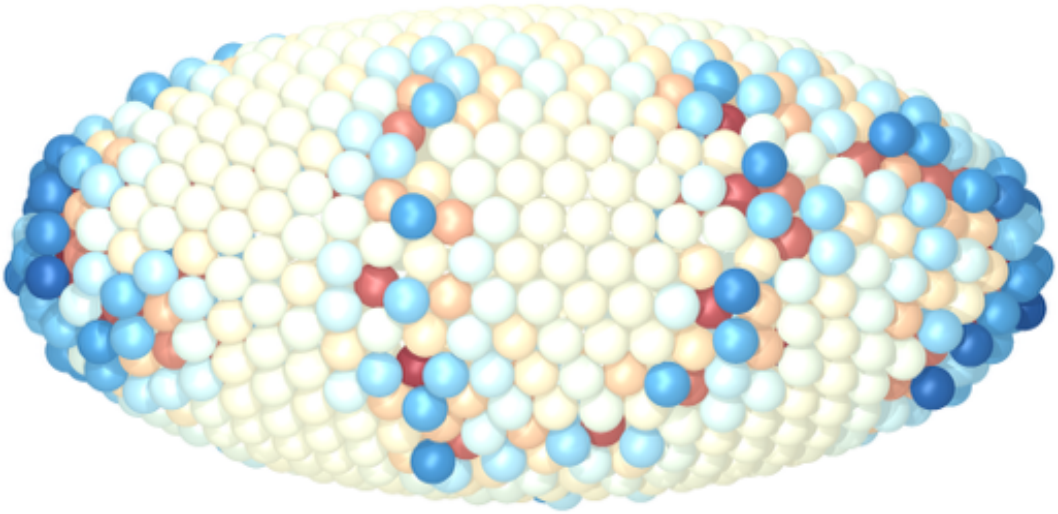


Figure A.1: A packing in which the particles are held to the surface by a harmonic constraint acting normal to the surface. Color indicates displacement from the surface: red particle have been pushed towards the interior of the droplet and blue particles have been pushed outwards.

buckling near defects and scars in the packing. In future work, we plan on exploring this idea more deeply and quantifying the correlation between defects and buckling. This is an important step in understanding the connection between the ordering of the packing on an arrested emulsion drop and its structural stability.

Appendix B

Optimizing the second hyperpolarizability with minimally parametrized potentials

In this appendix and the following two appendices, I include papers which I have published or have submitted for publication for the sake of completeness, as a record of the other contributions I have made over the course of my Ph.D.

This paper, published in the Journal of the Optical Society of America B, is an attempt to understand what features of a molecule lead to optimal nonlinear optical properties. Specifically, we use a simple parametrization of a 1D quantum mechanical potential to model molecules and attempt to optimize the second hyperpolarizability of this potential. We find that it takes a small number of parameters to create optimal potentials, and that the parameters chosen are well aligned with the structure of the objective function, providing design advice for tuning molecules to achieve optimal nonlinear properties.

Optimizing the second hyperpolarizability with minimally-parametrized potentials

C. J. Burke and T. J. Atherton
Department of Physics and Astronomy, Center for Nanoscopic Physics, Tufts University, 4 Colby Street, Medford, MA. 02155

J. Lesnefsky
Department of Physics, University of Illinois at Chicago, 845 W. Taylor St., Chicago, IL 60607-7059

R. G. Petschek
Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio, USA 44106

The dimensionless zero-frequency intrinsic second hyperpolarizability $\gamma_{int} = \gamma/4E_{10}^{-5}m^{-2}(e\hbar)^4$ was optimized for a single electron in a 1D well by adjusting the shape of the potential. Optimized potentials were found to have hyperpolarizabilities in the range $-0.15 \lessapprox \gamma_{int} \lessapprox 0.60$; potentials optimizing gamma were arbitrarily close to the lower bound and were within $\sim 0.5\%$ of the upper bound. All optimal potentials possess parity symmetry. Analysis of the Hessian of γ_{int} around the maximum reveals that effectively only a single parameter, one of those chosen in the piecewise linear representation adopted, is important to obtaining an extremum. Prospects for designing new chromophores based on the design principle here elucidated are discussed.

PACS numbers: 190.0190, 160.0160, 160.4330.

I. INTRODUCTION

Developing materials with high electronic nonlinear susceptibilities is of fundamental importance for a wide variety of applications such as optical solitons, phase conjugate mirrors and optical self-modulation[1, 2]. These susceptibilities are defined by considering a material in the presence of an electric field E and expanding the induced polarization in a Maclaurin series,

$$P = \alpha E + \beta EE + \gamma EEE + O(E^4), \tag{1}$$

where α is the linear susceptibility encountered in simple dielectrics and β and γ are the nonlinear susceptibilities, referred to as the first and second hyperpolarizabilities respectively. Each of these quantities is generally a tensor whose components depend on the frequency of the incident electromagnetic radiation. Here, we consider the zero-frequency limit of these quantities, and neglect the motion of the nuclei due to changes in the electronic wavefunction as a result of the applied field. The coefficients α, β and γ are material properties which depend on the electronic structure of the constituent molecules, the degree to which they are ordered, and symmetries present either in the molecules or in their arrangement. Due to these many factors, the synthesis of new materials with high values of β and γ as demanded by applications is highly nontrivial. A remarkable result due to Kuzyk [3, 4] is that quantum mechanics requires that the first and second hyperpolarizabilities β and γ are bounded: specifically, γ obeys the inequality,

$$-\left(\frac{e\hbar}{\sqrt{m}}\right)^4 \frac{N^2}{E_{10}^5} \leq \gamma \leq 4\left(\frac{e\hbar}{\sqrt{m}}\right)^4 \frac{N^2}{E_{10}^5} \equiv \gamma_0^{\max}, \tag{2}$$

where N is the number of electrons, E_{10} is the energy difference between the ground and first excited states and m is the electron mass. It is natural to define the intrinsic hyperpolarizability as a figure of merit to characterize the proximity of a particular system to this limit,

$$\gamma_{int} = \gamma/\gamma_0^{\max} \tag{3}$$

and to ask: how to create materials that achieve optimal γ_{int}? The discovery of the bounds (2) has motivated a number of experimental studies that have demonstrated that carefully tuning the electronic states and geometry of chromophores can lead to higher second hyperpolarizabilities[5–8]. Generic design principles motivated by fundamental theory would therefore be desirable. Unfortunately, the procedure used to derive the bounds (2) cannot directly provide these; they were obtained by optimizing γ for a three-level ansatz with respect to the dipole matrix elements and energy level spacings $E = E_{10}/E_{20}$ and *not* by constructing an explicit potential. Indeed, the assumptions behind the derivation have been questioned[9, 10] and the limits need not be achievable with a local potential; it has been speculated recently these may require exotic Hamiltonians[11].

Subsequent work, following approaches developed in earlier studies of β[12, 13], has attempted to address this in two ways: First, by identifying universal features of Hamiltonians with γ near $\gamma_0^{\max}$ by a Monte Carlo search[14] and, secondly, by numerically optimizing γ_{int} with respect to the shape of a local potential[11]. This latter work found potentials which have second hyperpolarizabilities in the range $-0.15 \leq \gamma_{int} \leq 0.60$, which represents an *apparent* bound that is more restrictive than the bound of (2). Moreover it was demonstrated that the optimized potentials spectra and dipole moments were broadly consistent with those identified in

the earlier Monte Carlo study.

While these strategies provide useful goals for chemists attempting to design new nonlinear chromophores, they do not provide insight into which features of the potential are necessary to optimize γ_{int}, or how many free parameters should be necessary to achieve optimal or near optimal γ. In a previous paper[15], we developed a technique to examine the analogous question for β: by optimizing potentials described by increasing numbers of free parameters and examining the eigenvalues of the Hessian matrix at each maximum, we identified the combinations of parameters most important to the optimization. The analysis revealed that effectively only two parameters were necessary to maximize β, and hence that a surprisingly broad range of potentials with high β exists around each maximum.

In this work, we apply the same technique to the problem of optimizing γ. At first sight, the problem appears to be more difficult than that for β since the expression for γ is much more complicated and the bounds for positive and negative γ are different. Remarkably, however, we will show that effectively only *one* parameter is necessary to optimize γ in either direction and, moreover, that in each case it is one of the parameters utilized in our representation of the potential. In this sense, we are able to suggest much more clearly a possible design strategy for materials with high γ than for β within the limitations of the model. At least within our representation of the potentials, we find that the potential that maximizes γ is rapidly varying, while for negative γ, as for β, quite generic, slowly varying potentials are adequate. The paper is organized as follows: in section II the calculations performed are described; the results are presented and discussed in section III; conclusions are drawn in section IV.

II. MODEL

It is first necessary to generalize the method described our previous paper on optimizing the intrinsic first hyperpolarizability β_{int} [15]: in the present work, γ_{int} is to be optimized by adjusting the shape of a one-dimensional piecewise-linear potential. Such a potential with $N+1$ segments may be represented,

$$V(x) = \begin{cases} A_0 x + B_0 & x < x_0 \\ A_n x + B_n & x_{n-1} < x < x_n,\ n \in \{1, ...N-1\} \\ A_N x + B_N & x > x_{N-1}, \end{cases} \tag{4}$$

with the positions x_n and slopes A_n as the adjustable parameters and where the B_n are chosen to enforce continuity. Because γ_{int} is invariant under trivial translations and rescalings of the potential, some of these parameters were fixed $x_0 = 0$, $B_0 = B_1 = 0$, and $A_1 = \pm 1$. These choices, together with a change of origin and rescaling allow for any potential. Thus maximizing with these constraints allows faster optimization. Furthermore, the left- and right-most slopes are required to be negative and positive, respectively, ensuring only bound electron states. Finally, for technical reasons, having to do with the asymptotic behavior of the Airy functions introduced in eq. 7 below, it is difficult to allow the sign of any slope to change during an optimization. In consequence, we have chosen to restrict $\mid A_i \mid > .005$, and, as appropriate to do separate optimizations for each interesting sign of each slope.

A second representation for the potential was also considered where parity symmetry was specifically enforced. This was motivated by previous work[16] which identifies parity as important for optimizing γ_{int}, particularly for the lower bound. The potentials with enforced $\mathcal{P}$ symmetry were constructed on the half line $x \geq 0$ with N segments,

$$V(x) = \begin{cases} A_n x + B_n & x_{n-1} < x < x_n,\ n \in \{1, ...N-1\} \\ A_N x + B_N & x > x_{N-1} \end{cases} \tag{5}$$

with $x_0 = 0$ and requiring $V(-x) = V(x)$. Again x_n and A_n are adjustable parameters and $x_0 = 0$, $B_0 = B_1 = 0$ are fixed. The parameter A_1 was set to either -1 or $+1$ to study the consequences of both cases.

For such a potential with a uniform applied electric field of strength ϵ, the wavefunction obeys the Schrödinger Equation in each segment,

$$\left[-\frac{1}{2}\frac{\mathrm{d}^2}{\mathrm{d}x^2} + (A_n + \epsilon)x + B_n \right] \psi_n = E\psi_n, \tag{6}$$

in units such that $e = 1$, $\hbar = 1$, and $m_e = 1$. The solution in each segment is written in terms of the well-known Airy functions,

$$\begin{aligned} \psi_n(x) \ = \ & C_n \mathrm{Ai}\left[\frac{\sqrt[3]{2}(B_n - E + x(A_n + \epsilon))}{(A_n + \epsilon)^{2/3}} \right] + \\ & + D_n \mathrm{Bi}\left[\frac{\sqrt[3]{2}(B_n - E + x(A_n + \epsilon))}{(A_n + \epsilon)^{2/3}} \right]. \end{aligned} \tag{7}$$

To solve for the coefficients C_n and D_n the usual boundary conditions are imposed, i.e. that the wavefunction $\psi(x)$ and its derivative $\psi'(x)$ are continuous at the boundary between segments. Additionally, in the end segments, the wavefunction must vanish as x goes to $\pm\infty$ fixing $D_N = 0$. There are a total of $2N$ linear equations in the coefficients for the arbitrary case and $4N-2$ equations and coefficients for the $\mathcal{P}$-symmetric case, which can be written in matrix form

$$W \cdot u = 0 \tag{8}$$

where u is a vector comprised of the C_n and D_n coefficients and W is a matrix which depends on E, ϵ and the parameters A_n and x_n.

The allowed energy levels are found by numerically finding the roots of

$$\det W = 0 \tag{9}$$

with $\epsilon = 0$. It is readily possible, as previously done for β, to obtain from (8) an expression for the second hyperpolarizability,

$$\gamma \equiv \frac{1}{6} \left. \frac{\mathrm{d}^4 E_0}{\mathrm{d}\epsilon^4} \right|_{\epsilon=0}; \tag{10}$$

this is achieved by repeatedly differentiating the matrix W using the Jacobi formula,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon} \det W = \mathrm{Tr}\left(\mathrm{adj}W \cdot \frac{\mathrm{d}W}{\mathrm{d}\epsilon} \right), \tag{11}$$

where adj$W$ is the adjugate of W (since W is singular), and applying the chain rule

$$\frac{\mathrm{d}W}{\mathrm{d}\epsilon} = \frac{\partial W}{\partial \epsilon} + \frac{\partial W}{\partial E}\frac{\mathrm{d}E}{\mathrm{d}\epsilon}. \tag{12}$$

Having performed similar calculations to those in [15], we arrive at the expression

$$\frac{\mathrm{d}^4 E}{\mathrm{d}\epsilon^4} = -\frac{\mathrm{Tr}\left[\left(\frac{\mathrm{d}^3}{\mathrm{d}\epsilon^3}\mathrm{adj}W \right) \cdot \frac{\mathrm{d}W}{\mathrm{d}\epsilon} + 3\left(\frac{\mathrm{d}^2}{\mathrm{d}\epsilon^2}\mathrm{adj}W \right) \cdot \frac{\mathrm{d}^2 W}{\mathrm{d}\epsilon^2} + 3\left(\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathrm{adj}W \right) \cdot \frac{\mathrm{d}^3 W}{\mathrm{d}\epsilon^3} + \mathrm{adj}W \cdot W''' \right]}{\mathrm{Tr}\left(\mathrm{adj}W \cdot \frac{\partial W}{\partial E} \right)}, \tag{13}$$

where

$$W''' = \frac{d^4 W}{d\epsilon^4} - \frac{\partial W}{\partial E}\frac{d^4 E}{d\epsilon^4}. \tag{14}$$

From this, γ is readily obtained and the intrinsic second hyperpolarizability $\gamma_{int} = \gamma/\gamma_0^{\max}$ calculated for a given set of parameters.

The quantity γ_{int} was optimized numerically for both arbitrary and $\mathcal{P}$-symmetric potentials with varying numbers of segments using the `FindMaximum` function of *Mathematica* 8, an implementation of the Interior Point method for constrained optimization. Both maxima and minima of γ_{int} were obtained from a large number of randomly generated starting points, and also manually chosen starting points with large values for γ_{int}. Once an optimum γ_{int} was found, the extent to which each of the parameters was important to the extremum was characterized by calculating the Hessian matrix,

$$H_{ij} = \frac{\partial^2}{\partial P_i \partial P_j}\gamma_{int},$$

where the P_i are the parameters, and calculating its eigenvalues and eigenvectors. Since the Hessian matrix characterizes the local curvature of the objective function in the parameter space around the extremum, these quantities give the magnitudes and directions of the principal curvatures. As stressed in previous work[15] these curvatures implicitly depend on a measure implied by this equation that is peculiar to our numerical parameterization of the problem. More physically relevant measures can also be used to calculate eigenvalues. While these make quantitative changes in the eigenvalues and vectors, they do not make qualitative changes, and we give results for this "numerically natural" measure below.

III. RESULTS AND DISCUSSION

Our optimized potentials, together with the ground and first excited state wavefunctions, are displayed in Fig. 1 for both arbitrary (eq. 4) and $\mathcal{P}$-symmetric (eq. 5) parametrizations. The associated parameter values are listed in Table I. The potentials and wavefunctions are displayed on a transformed position and energy scale,

$$\begin{aligned} \bar{x} &= (x - \langle x \rangle)/(E_1 - E_0)^{1/2} \\ \bar{V}(\bar{x}, \{P\}) &= (V(\bar{x}, \{P\}) - E_0)/(E_1 - E_0) \end{aligned} \tag{15}$$

such that the ground state energy is $E_0 = 0$, the difference between the ground and first excited state energy is $E_1 - E_0 = 1$ and the position expectation value for the ground state is $\langle x_0 \rangle = 0$. This rescaling does not change γ_{int} and permits convenient comparison of the results of each optimization. To identify the relative importance of each of the parameters to the optimization, the results of the eigenanalysis of the Hessian matrix are also displayed for selected potentials in Fig. 1; the j-th eigenvalue of the Hessian, h^j, is listed alongside a plot of the variation in the potential $\Delta V^j(x)$ in the direction of the associated eigenvector,

$$\Delta V^j(x) = \left. \frac{\partial \bar{V}(\bar{x}, \{P_i + \alpha v_i^j\})}{\partial \alpha} \right|_{\alpha=0}, \tag{16}$$

where v_i^j is the i^{th} component of the j-th eigenvector. Note that the values of V and x in the right hand side of (16) are renormalized as a function of α using (15) so that the variations presented automatically preserve the properties $\langle x_0 \rangle = 0$ and $E_1 - E_0 = 1$. In order to assist in interpreting these plots, we include a detailed example of how these quantities are calculated in the appendix.

The first set of results displayed in Fig. 1(a) are optimized potentials with no enforced symmetry and specified by 3 or 5 free parameters. The optimized γ_{int} for

Bound	Description	Ref.	γ_{int}	A_0	A_2	A_3	x_1	x_2
Upper	3 param. arb.	1(a)(i)	0.43277	2.33932	0.10189	—	1.04960	—
	5 param. arb.	1(a)(ii)	0.43602	4.28494	0.08829	0.37688	1.42131	3.95728
	2 param. $\mathcal{P}$	1(b)(i)	0.58220	—	0.00550	—	0.57426	—
	3 param. $\mathcal{P}$	1(b)(ii)	0.59707	—	0.00500	—	0.88533	7.95755
Lower	3 param. arb.	2(a)	−0.11409	82.471	237.57	—	1.25227	—
	2 param. $\mathcal{P}$	2(b)	−0.14996	—	31.6157	—	1.48005	—

Table I: List of parameters of optimized potentials. Here the Ref. column refers to the subfigures in the present paper where the respective potentials are displayed.

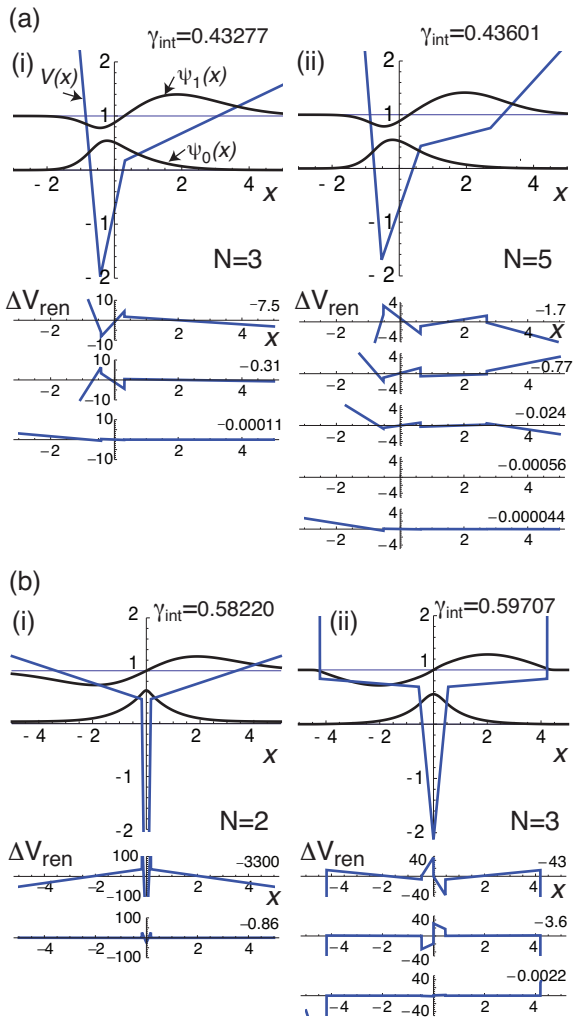


Figure 1: Optimized potentials for the *upper* bound of γ_{int} with (a) no enforced symmetry and (b) $\mathcal{P}$-symmetry. The energies of the ground and first excited state are indicated by horizontal lines; the corresponding wavefunctions are also displayed. The plots have been rescaled to facilitate comparison by ensuring $E_1 - E_0 = 1$ and $\langle x_0 \rangle = 0$ while preserving γ_{int}.

both the lower and upper bounds of γ_{int} potentials fall somewhat below the apparent bounds observed in [14]. For the upper bound, the best results assuming all slopes except the first are positive are what appears to be a local maximum value of $\gamma_{int} \backsimeq 0.43$ [Fig. 1(a)(ii)]. A

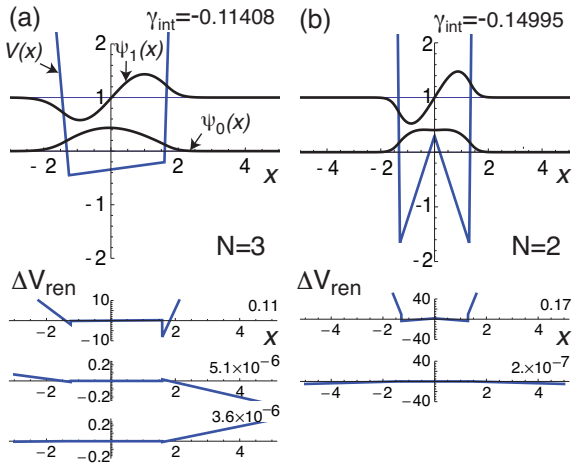


Figure 2: Optimized potentials for the *lower* bound of γ_{int} with (a) no enforced symmetry and (b) $\mathcal{P}$-symmetry.

similar value was also found in [11] for potentials with no constrained symmetry, but the potentials found here do not closely resemble those found in that work. The eigenvalues and eigenvectors displayed below the potential in [Fig. 1(a)(ii)] show that only two of the eigenvalues are significant in magnitude and are associated with the shape of the potential in the middle while the small eigenvalues are associated with the outer slopes. These results are reminiscent of those found for the first hyperpolarizability, where β_{int} was found to approach its maximum value for the same class of potentials with a similarly small number of parameters and analysis of the Hessian revealed that only effectively two parameters were important to the maximization. For the lower bound, the 3 parameter system [Fig. 2(a)] converges on a shape approaching a square well. The 5 parameter system also converges on a $\mathcal{P}$-symmetric potential and, because of this, further discussion of the lower bound is deferred to a subsequent paragraph.

A possible explanation for the fact that the optimized positive γ_{int} for arbitrary potentials in Fig. 1 falls short of the bounds established in [3, 11, 14] is that the hyperpolarizability is sensitive only to some features of the potential and that many local extrema exist. Rather than making extensive runs starting from a variety of potentials, we chose to use the more constrained subset of $\mathcal{P}$-symmetric potentials (5) and found that, indeed,

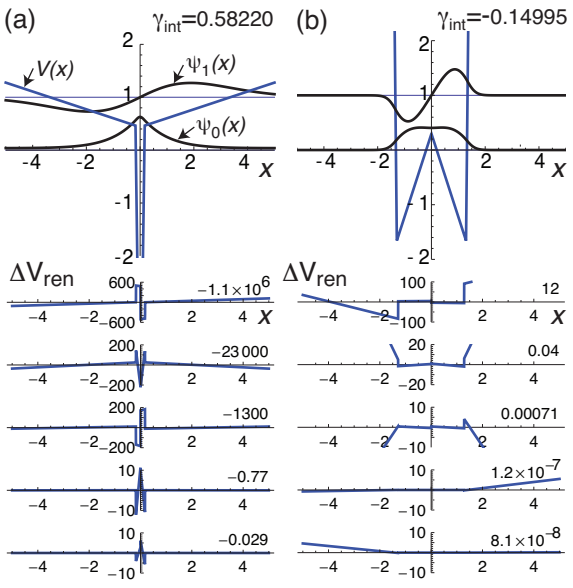


Figure 3: Five parameter potentials with no enforced symmetry optimized using two-parameter $\mathcal{P}$-symmetric optima as the starting point. (a) Upper and (b) lower bounds.

even with only 2-3 parameters, much higher values of γ_{int} could be obtained as shown in Fig. 1(b). This is reminiscent of the observation in [11] that these bounds could only be reached if a $\mathcal{P}$-symmetric starting point was used for the search; in this work, we not only enforce the symmetry of the starting point but at all times in the optimization.

For the upper bound, we found a two-parameter potential close to the apparent maximum but below the theoretical maximum [Fig. 1(b)(i)]. The shape is characterized by shallow outer slopes with a divot in the center; the ground state wave function is localized to the divot with the highly delocalized first excited state above the divot. It was found to be necessary to constrain the slope $A_2 > 0.005$ since the method of calculating γ_{int} presented in section II fails for shallow slopes. To avoid the unphysical feature of delocalized higher excited states, a distant wall was added to construct a three-parameter potential [Fig. 1(b)(ii)]. Such a change is expected *a priori* from previous work[17] to make no significant difference to γ_{int} as it is far from the region where ψ_0 and ψ_1 are large. For this well-like potential, we performed a maximization, adjusting A_2, x_1, and x_2 while fixing the outer walls to have a large slope ($A_3 = 100$). The best potential found possesses $\gamma_{int} = 0.59707$, and has $A_2 = 0.005$ which is the shallowest slope allowed by the constraint. To facilitate comparison with optimization methods that study the energy spectrum e.g. [18], we include the rescaled energy levels of this potential: $E_2 = 1.144$, $E_3 = 1.775$, and $E_4 = 2.179$. The eigenvectors of the Hessian, calculated for the subspace excluding A_2, are well aligned with the parameters: the most significant eigenvector corresponds to x_1, the outer boundary of the divot. The other eigenvector corresponds to x_2, the position of the outer walls; this would be expected to have relatively little influence on γ_{int} since it controls a feature where the ground state wavefunction is small. Fixing the outer slopes at $A_3 = 100$ as above, we also attempted to optimize γ_{int} with A_2 constrained to be negative. It was found that γ_{int} increased as the slope approached zero, until a point was reached where the calculations became numerically unstable due to the asymptotic properties of the Airy functions. The highest value of γ_{int} which was found within a region of parameter space for which the calculation was still stable was $\gamma_{int} = 0.5915$, lower than the current maximum. We then performed similar optimizations on potentials with strict hard wall boundary conditions. The calculations for these potentials *were* numerically stable in all regions of parameter space which were explored: For $A_2 > 0$, a maximum of $\gamma_{int} = 0.5959$ was found. For $A_2 < 0$ a higher maximum of $\gamma_{int} = 0.5968$ was found, though this is still lower than the current maximum of $\gamma_{int} = 0.5971$. Since a higher γ_{int} is found in potentials with $A_2 < 0$ than for potentials with positive A_2 in cases with strict hard wall boundary conditions, we speculate that a value of γ_{int} higher than the current maximum found might be found for the finite A_3 case. Nonetheless, we do not expect to see a significant improvement as the current maximum is already within $\sim 0.5\%$ of the maximum value found in previous studies.

For the lower bound, a potential with the best value of $\gamma_{int} = -0.1500$ was found using only two parameters [Fig. 2(b)]. This potential is characterized by steep outer walls and a "bump" in the middle: the ground state and first excited state wavefunctions cover the same spatial extent, but the bump causes the ground state to become spread out and relatively flat. The rescaled energy levels of this potential are $E_2 = 3.502$, $E_3 = 6.660$, and $E_4 = 10.65$. Eigenanalysis of the Hessian of γ_{int} about this solution shows that one eigenvalue is significantly larger than the other, indicating that only one of the parameters is physically relevant. Moreover, the eigenvectors of the Hessian for this potential are aligned with the parameter space chosen to represent the potential. The higher eigenvalue is associated with an eigenvector along the x_1 direction, which determines the position of the outer walls; the smaller eigenvalue is associated with the parameter that controls the slope of the outer walls. Because γ_{int} is invariant under rescalings of the form (15), a potential with identical γ_{int} can be constructed for a well of arbitrary width by tuning the slope of the bump.

The two parameter $\mathcal{P}$-symmetric potentials identified as the satisfying the apparent lower bound can be equivalently represented by a five parameter arbitrary potential. Optimization of the five parameter potential indeed finds this potential as the apparent global maximum. There are therefore no directions in this new parameter space which lead to a higher γ_{int}, despite relaxing the requirement that $x_{-n} = -x_n$ and $A_{-n} = -A_n$. While the existence of asymmetric potentials with more negative

γ_{int} cannot be ruled out, our analysis confirms that a $\mathcal{P}$-symmetric potential satisfies the apparent lower bound. We repeated this procedure for the upper bound using the two parameter $\mathcal{P}$-symmetric potential displayed in Fig. 2(b) as the starting point for optimization. It was found that despite relaxing the symmetry constraint no further improvement could be made so that the potential of Fig. 2(b) is also a local optimum with respect to the expanded parameter space. Analysis of the hessian was also performed on both of these five parameter optima. The eigenvalues and associated eigenvectors, shown in Fig. 3, reveal that for both upper and lower bounds the variation in the potential associated with the largest eigenvalue is indeed asymmetric.

The fact that we are able to achieve optimized γ_{int} within 1% of previous limits with far fewer parameters than in other representations of the potential[11], and also fewer parameters than required to numerically optimize β_{int}, is surprising since the calculated expressions for γ_{int} are far more complicated than those for β_{int}. Since the dimensionality of this parameter space is so small, it is instructive to visualize it directly: For our two-parameter optimizations, the value of γ_{int} is plotted over a portion of the parameter space [Fig. 4(a) and (b)]. The plot for the lower bound merely illustrates the results obtained from analysis of the Hessian, i.e. that the minimum is strongly curved about the optimal value of x_1 but shallow with respect to A_2. The plot for the upper bound is more interesting: the region of parameter space for which γ_{int} is within 2% of the maximum value obtained is highlighted showing a clear ridge.

By calculating values of $X = x_{01}/x_{01}^{\max}$ and $E = E_{01}/E_{02}$, the natural parameters of the three-level ansatz[4], as a function of (A_2, x_1), we are able to display γ_{int} re-parametrized in (E, X) space [Fig. 4(c) and (d)]. Here, $x_{01}^{max} = 1/\sqrt{2E_{10}}$ in our units. Notice that the entire region explored in the numerical parametrization (A_2, x_1) collapses onto a narrow, elongated region in (E, X) space for the upper bound [Fig. 4(c)] and a complicated curved line for the lower bound[Fig. 4(d)]. These plots confirm the results of eigenanalysis of the Hessian: that essentially only a single parameter (or 2 for the upper bound) characterizes optimal γ_{int}.

The results of the three-parameter optimization and the plot in fig. (4) both suggests that the truly optimal $\mathcal{P}$-symmetric potential for the upper bound has shallow outer slopes $A_2 \to 0$. Such a potential can be transformed, using (15), to a potential of equivalent γ_{int} but where the outer slope is unity and the central well is far narrower and sharper. Since the central divot for the transformed potential resembles a Dirac delta function, we studied the second hyperpolarizability of the family of potentials

$$V(x) = |x| - \alpha\delta(x) \tag{17}$$

where α is the single adjustable parameter. Values of γ_{int} as a function of α are displayed in fig. 5 and the maximum value is found to be $\gamma_{int} = 0.58194$ which occurs

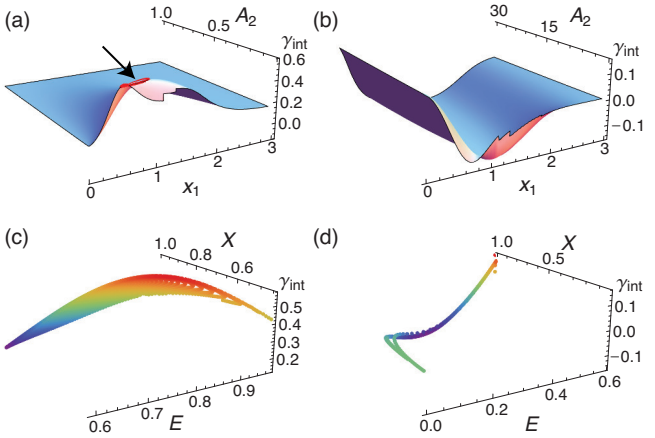


Figure 4: Visualization of the variation of γ_{int} as a function of the parameters x_1 and A_2 of a two-parameter $\mathcal{P}$-symmetric potential for the (a) upper bound and (b) lower bound. The region of parameter space where γ_{int} is within 2% of the maximum value is highlighted in (a) and indicated by an arrow. γ_{int} is also shown in the (E, X) parameter space for the (c) upper bound and (d) lower bound.

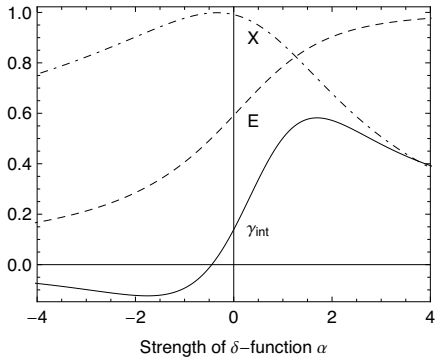


Figure 5: Plot of γ_{int}, $E = E_{10}/E_{20}$ and $X = x_{10}/x_{10}^{max}$ as a function of α for the for the potential $V = |x| - \alpha\delta(x)$.

when $\alpha = 1.69552$. Despite the simplicity of this one-parameter potential, the result is only 3% smaller than the best reported so far and only fractionally smaller than those found with the two and three parameter potentials above. This family of potentials is less satisfactory for optimizing the negative γ_{int}, achieving $\gamma_{int} = -0.1236$ which is $\sim$ 18% short of the best potential. Another interesting feature of this parametrization is that by adjusting α, one can obtain an arbitrary value of the energy ratio E for the first three energy levels. The energy levels of the first and second excited states are fixed, and adjusting α changes only the ground state. This is illustrated by the plot of E in fig. 5: its range contains all values from 0 to 1 as α varies from $-\infty$ to $+\infty$.

We now turn to the question of whether the simple, but optimal or nearly so, potentials that were identified above resemble either those previously found [11] or possess the universal features identified in [14]. Comparing our best optimized potentials and wavefunctions to those in [11],

Bound	Potential	γ_{int}	E	$f(E)$	X
Upper	3-level ansatz[4]	1	0	1	0
	Best from [11]	0.5993	0.5767	0.7009	0.5150
	Arb. 5 param.	0.43601	0.5765	0.7012	0.6512
	$\mathcal{P}$ 3 param.	0.59707	0.8741	0.2504	0.7468
	$\lvert x\rvert - \alpha\delta$	0.58194	0.6842	0.5354	0.7385
	Triangle well $\lvert x\rvert$	0.1392	0.5918	0.6836	0.9911
Lower	3-level ansatz[4]	-0.25	0	1	± 1
	Best from[11]	-0.1500	0.1493	0.982447	0.6658
	$\mathcal{P}$ 2 param.	-0.1500	0.2855	0.93298	0.9416
	$\lvert x\rvert - \alpha\delta$	-0.1236	0.2438	0.951765	0.9222
	Infinite square well	-0.1262	0.3750	0.881103	0.9801

Table II: Second hyperpolarizabilities and physical parameters $X = x_{01}/x_{01}^{\max}$ and $E = E_{01}/E_{02}$ for the optimized potentials obtained in this work.

some qualitative similarities are apparent. For the upper bound, the potentials in Watkins et al. are roughly symmetric near their lowest point. They feature a central divot within a wider well where the ground state wave function is localized within the central divot and the first excited wave function is relatively delocalized compared to the ground state. For the lower bound, the potentials feature a steep well within which both the ground state and first excited state are localized, and a central bump which causes the ground state to be spread out within the well. These qualitative features are shared by the potentials obtained in this work, but many extraneous details are removed by the highly constrained, judiciously chosen representation.

In table II we display values of X and E, together with the values of these parameters that extremize γ_{int} for the three-level ansatz[4] and those possessed by the best previously found potentials[11]. We also display the value of the energy function $f_\gamma(E)$ defined in eq. 26 of reference [18] which has been shown to characterize some features of the energy spectrum that are relevant to high γ_{int}. Values are also shown for two elementary potentials, the triangle well and infinite square well. The results for the upper bound are quite consistent with those of Watkins *et al.* if the breadth in the range of these values identified by the Monte Carlo study[14] is taken into account. The values of $f(E)$ obtained are consistent with the results of [18] that suggest a value of $f_\gamma(E)$ approaching 1 typically occurs for potentials associated with large γ_{int}; our result for the 3 parameter $\mathcal{P}$ symmetric potential seems to be an interesting exception. The values of E and X are also displayed in Fig. 5 as a function of α, the strength of the δ function, in the potential (17); E is a monotonically increasing function of α while X is a monotonically decreasing function. Crudely, these explain the existence of a maximum γ_{int} as representing the trade-off between increasing the motion of the electron (associated with high X) versus enhancing transitions to other states (associated with low E).

IV. CONCLUSION

We have optimized the intrinsic second hyperpolarizability γ_{int} of a piecewise linear potential well with respect to parameters that control the shape of the potential. We found solutions that lie within the range $-0.15 \leq \gamma_{int} \lessapprox 0.60$ in agreement with the apparent bounds established in previous numerical optimizations[11]; these both fall short of the Kuzyk limits[4]. By using two types of potential, one where all slopes were allowed to vary and another with explicitly enforced symmetry, we demonstrated that $\mathcal{P}$-symmetric potentials satisfy the apparent lower bound for γ_{int} and come within $\sim 0.5\%$ of the apparent upper bound. The parametrization used constrains the potential to be smooth, preventing the occurrence of rapid oscillations which do not affect γ_{int} [17]. Because of this and the strong symmetry constraint, the optimal $\mathcal{P}$-symmetric potentials found were characterized by only $2-3$ parameters. Of these, *a posteriori* analysis of the Hessian revealed that effectively only one or two, for the lower and upper bound respectively, were important

These results are reminiscent of those obtained earlier for β_{int}[15], yet the number of parameters required to optimize γ_{int} appears to be smaller even though it is a more complex object, containing more terms and involving higher derivatives. At least part of the reason for this is that for γ_{int} there exists a "compatible" symmetry operator, $\mathcal{P}$, which can be used to constrain the shape of the potentials; this was not possible for β_{int} where $\mathcal{P}$-symmetric potentials automatically have $\beta = 0$. However, even though we have shown that $\mathcal{P}$-symmetric potentials can have optimal or near-optimal second hyperpolarizabilities, it is not clear whether the apparent upper bound $\gamma_{int} \sim 0.6$ achieved by Watkins and Kuzyk can be achieved with a $\mathcal{P}$-symmetric potential or whether a small amount of asymmetry is necessary. Moreover, the reason why local potentials fall short of the Kuzyk bounds remains opaque.

The small parameter space allows us to propose a clear design paradigm for new chromophores, within the limitations of the model one-electron 1D system studied. Neglected here, for example, are multi-electron interactions, molecular ordering and inter-molecular electron hopping. Nonetheless, the potentials obtained could be realized, for example, by a centrosymmetric molecule with a central attractive or repulsive group—for positive or negative γ_{int} respectively. The strength/electronegativity of the central group and the ratio of the length of the central and peripheral groups can then be tuned to give high γ_{int}. Unfortunately, most practical chromophores are π conjugated systems in which there are approximately as many electrons as there are sites on the molecule, thus the consequences for them from this single electron calculation are clearly very speculative. Nevertheless, homologous sequences already studied for high γ as a function of chain length, e.g. [7] could likely be enhanced by including such a central group with different electroneg-

ativity. Any other means of achieving a potential well or inducing a significant phase shift in wavefunctions passing through the center of the molecule is also likely, even in multi-electron systems, to offer a route for achieving larger γ_{int}. The present analysis provides other important insights: first, that since the "true" parameter space for γ_{int} is so small, only rough tuning of the molecular design ought to be necessary. Secondly, this work again confirms that there are a large set of modifications to optimized potentials, e.g. rapid oscillations, that will not change γ_{int} and need not be considered in planning what molecules to synthesize. As has been previously noted, materials with high γ_{int} could also be realized more directly in other ways, such as through composite materials.

Appendix

To illustrate how the Hessian reveals which features of the potential that are important to optimal γ_{int}, and to assist in interpreting the variations of the potential depicted in figures 1-3, we develop a simple example for a given point in parameter space. Consider the two-parameter $\mathcal{P}$-symmetric potential at minimum γ_{int} [Fig. 2(b)]. The Hessian of γ_{int} at the minimum is

$$H = \begin{pmatrix} \frac{\partial^2}{\partial A_2^2} & \frac{\partial^2}{\partial A_2 \partial X_1} \\ \frac{\partial^2}{\partial X_1 \partial A_2} & \frac{\partial^2}{\partial X_1^2} \end{pmatrix}$$

$$\gamma_{int} = \begin{pmatrix} 1.54 \times 10^{-6} & -4.70 \times 10^{-4} \\ -4.70 \times 10^{-4} & 0.165 \end{pmatrix}.$$

The eigenvalues of this matrix are 0.165 and 2.01×10^{-7} and their respective eigenvectors are

$$\begin{pmatrix} -0.002843 \\ 0.999996 \end{pmatrix}, \begin{pmatrix} -0.999996 \\ -0.002843 \end{pmatrix}.$$

The larger eigenvalue corresponds to an eigenvector pointing almost exactly along the X_1 direction in parameter space, and the smaller one points along the A_2 direction. This is clearly illustrated in fig. 4(b), where there is a valley along the A_2 direction, while moving along the X_1 direction leads to a less negative value of γ_{int}.

It should be noted that during optimization, in order to fix the scaling(15) which is irrelevant to the hyperpolarizability, the arbitrary constraint was imposed that $A_1 = -1$. In plotting the local variation in the potential associated with the eigenvectors described by (16) and displayed in the bottom of fig. 2(b), we ease this constraint and instead apply a *post-hoc* scaling such that $E_{10} = 1$ and require that E_{10} remains equal to 1 as the potential is varied away from the extremum. This new constraint is more natural because E_{10} defines the characteristic energy scale of problem as seen from (2).

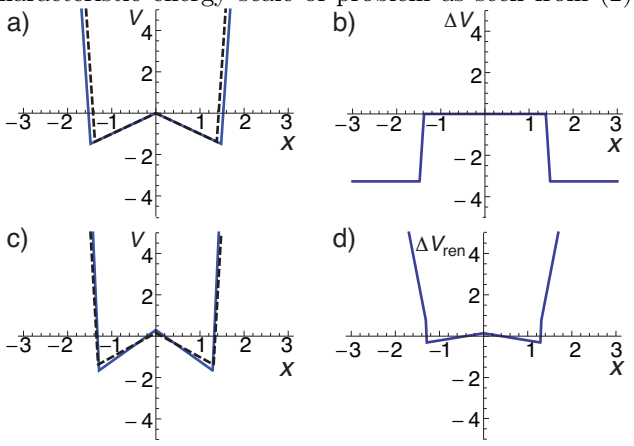


Figure 6: Examples of constructing variations in the potential away from a given point in parameter space. (a) Original unrescaled potential at minimum γ_{int} (solid blue line) and a perturbation (in the $-X_1$ direction) of this potential (dashed black line). (b) The difference between the optimum potential and the perturbation. (c) Rescaled versions of the potentials in (a) along with (d) their difference.

This scaling is a unique, one-to-one mapping between the two spaces of potentials with the different constraints imposed; it allows one to identify the features of the potential which are most relevant to optimizing γ_{int} unencumbered with features associated with the irrelevant scaling.

Fig. 6 shows an example variation of the potential in both the original parametrization and after rescaling to fix E_{10}. The variation is in the $-X_1$ direction and is exaggerated (i.e. it is not a small local variation as in (16)) so that it is clearly visible. In fig. 6(b) there is a sharp jump between the inner and outer segments which is indicative of a change in the X_1 parameter. Note that in general, the size of the jump depends not just on the variation in X_n but also on the surrounding slopes. In fig. 6(d) we see that upon rescaling, the slopes of the piecewise segments also contribute to the change in γ_{int}.

[1] R. W. Boyd, *Nonlinear Optics, 3rd ed.* (Academic Press, 2008).

[2] Y. R. Shen, *The Principles of Nonlinear Optics* (John Wiley & Sons, 2003).

[3] M. Kuzyk, "Physical limits on electronic nonlinear molecular susceptibilities," Phys. Rev. Lett. **85**, 1218–1221 (2000).

[4] M. Kuzyk, "Fundamental limits on third-order molecular

susceptibilities," Opt. Lett. **25**, 1183–1185 (2000).

[5] M. Fujiwara, K. Yamauchi, M. Sugisaki, K. Yanagi, A. Gall, B. Robert, R. Cogdell, and H. Hashimoto, "Large third-order optical nonlinearity realized in symmetric nonpolar carotenoids," Phys. Rev. B **78**, 161101 (2008).

[6] J. M. Hales, J. Matichak, S. Barlow, S. Ohira, K. Yesudas, J.-L. Bredas, J. W. Perry, and S. R. Marder, "Design of polymethine dyes with large third-order optical nonlinearities and loss figures of merit," Science **327**, 1485–1488 (2010).

[7] T. Luu, E. Elliott, A. Slepkov, S. Eisler, R. McDonald, F. Hegmann, and R. Tykwinski, "Synthesis, structure, and nonlinear optical properties of diarylpolyynes," Org. Lett. **7**, 51–54 (2005).

[8] J. C. May, I. Biaggio, F. Bures, and F. Diederich, "Extended conjugation and donor-acceptor substitution to improve the third-order optical nonlinearity of small molecules," Appl. Phys. Lett. **90**, 251106 (2007).

[9] B. Champagne and B. Kirtman, "Comment on "physical limits on electronic nonlinear molecular susceptibilities"," Phys. Rev. Lett. **95**, 109401 (2005).

[10] B. Champagne and B. Kirtman, "Evaluation of alternative sum-over-states expressions for the first hyperpolarizability of push-pull pi-conjugated systems," J. Chem. Phys. **125**, 024101 (2006).

[11] D. Watkins and M. Kuzyk, "Universal properties of the optimized off-resonant intrinsic second hyperpolarizability," J. Opt. Soc. Am. B **29**, 1661–1671 (2012).

[12] K. Tripathy, J. Moreno, M. Kuzyk, B. Coe, K. Clays, and A. Kelley, "Why hyperpolarizabilities fall short of the fundamental quantum limits," J. Chem. Phys. **121**, 7932 (2004).

[13] J. Zhou, U. Szafruga, D. Watkins, and M. Kuzyk, "Optimizing potential energy functions for maximal intrinsic hyperpolarizability," Phys. Rev. A **76**, 053831 (2007).

[14] S. Shafei, M. Kuzyk, and M. Kuzyk, "Monte carlo studies of the intrinsic second hyperpolarizability," J. Opt. Soc. Am. B **27**, 1849–1856 (2010).

[15] T. J. Atherton, J. Lesnefsky, G. A. Wiggers, and R. G. Petschek, "Maximizing the hyperpolarizability poorly determines the potential," J. Opt. Soc. Am. B pp. 1–8 (2012).

[16] M. Kuzyk and C. Dirk, "Effects of centrosymmetry on the nonresonant electronic third-order nonlinear optical susceptibility," Phys. Rev. A **41**, 5098 (1990).

[17] G. Wiggers and R. Petschek, "Comment on" pushing the hyperpolarizability to the limit"," Opt. Lett. **32**, 942–943 (2007).

[18] S. Shafei and M. Kuzyk, "Critical role of the energy spectrum in determining the nonlinear-optical response of a quantum system," J. Opt. Soc. Am. B **28**, 882–891 (2011).

Appendix C

Maximizing the hyperpolarizability of 1D potentials with multiple electrons

This paper, published in the Journal of the Optical Society of America B, is a follow up to the paper in the previous appendix. Here, we explore potentials with multiple elections. We study both the first and second hyperpolarizability. We again find that very simple potentials are able to give high non-linear responses, and we correlate properties of the wavefunction to a high nonlinear response, which is useful information for chemists trying to design molecules.

Maximizing the hyperpolarizability of 1D potentials with multiple electrons

Christopher J. Burke and Timothy J. Atherton*
Department of Physics and Astronomy, Tufts University, 574 Boston Avenue, Medford, Massachusetts 02155, USA

Joseph Lesnefsky and Rolfe G. Petschek
Department of Physics, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106, USA

We optimize the first and second intrinsic hyperpolarizabilities for a 1D piecewise linear potential dressed with Dirac delta functions for N non-interacting electrons. The optimized values fall rapidly for $N > 1$, but approach constant values of $\beta_{int} = 0.40$, $\gamma_{int}^{+} = 0.16$ and $\gamma_{int}^{-} = -0.061$ above $N \gtrsim 8$. These apparent bounds are achieved with only 2 parameters with more general potentials achieving no better value. In contrast to previous studies, analysis of the hessian matrices of β_{int} and γ_{int} taken with respect to these parameters shows that the eigenvectors are well aligned with the basis vectors of the parameter space, indicating that the parametrization was well-chosen. The physical significance of the important parameters is also discussed.

I. INTRODUCTION

Nonlinear optical materials are the active constituent for many applications such as light modulators, contrast agents for medical imaging and therapy, optical solitons, phase conjugation mirrors and optical self-modulation. In each of these, the performance of the system is improved by using a material with a stronger nonlinear response, quantified by various nonlinear susceptibilities defined by expanding the induced polarization P in a power series in the applied electric field,

$$P = \alpha E + \beta EE + \gamma EEE + O(\epsilon^4). \tag{1}$$

Here, α is the linear susceptibility familiar from dielectric materials; β and γ are the nonlinear susceptibilities and are referred to as the first and second hyperpolarizability respectively. These quantities are in general frequency-dependent tensors that depend on the electronic structure of the constituent molecules, their symmetry, ordering and the material in which they are embedded. In the present work we focus on the off-resonant molecular contribution. Much effort has been expended over the years in synthesizing new molecules with higher β or γ. Comparisons between materials must be made carefully however, because these quantities increase trivially with the size of the molecule.

Important progress on developing suitable figures-of-merit for comparison was made by Kuzyk, who showed that fundamental quantum mechanics requires that β and γ for the off-resonant case are bounded by the inequalities,

$$|\beta| \leq \sqrt[3]{4}\left(\frac{e\hbar}{\sqrt{m}}\right)^3 \frac{N^{3/2}}{E_{10}^{7/2}} \equiv \beta_0^{max}, \tag{2}$$

$$\left(\frac{e\hbar}{\sqrt{m}}\right)^4 \frac{N^2}{E_{10}^5} \leq \gamma \leq 4\left(\frac{e\hbar}{\sqrt{m}}\right)^4 \frac{N^2}{E_{10}^5} \equiv \gamma_0^{max} \tag{3}$$

where E_{10} is the difference between the ground and first excited state and N is the number of electrons participating. From the maximum values β_0^{max} and γ_0^{max}, one defines intrinsic hyperpolarizabilities,

$$\beta_{int} = \beta/\beta_0^{max},\ \gamma_{int} = \gamma/\gamma_0^{max}. \tag{4}$$

The intrinsic quantities have the property that they remain invariant under a simultaneous rescaling of energy and length,

$$x \to x' E^{1/2},\ V(x) \to V'(x')E \tag{5}$$

and hence are useful quantities for comparing materials because they remove the irrelevant scaling with size. Analysis of extant materials following the derivation of the bounds in eq. (3) revealed that all of them fell short of the fundamental limits by more than an order of magnitude, an observation that has catalyzed a great deal of research over the past decade on how to create materials that approach these fundamental limits. The derivation of the bounds provides some guidance—for optimal β and γ only three states are assumed to significantly contribute to the hyperpolarizabilities and the optimum can be achieved by tuning the dipole transition matrix elements and energy level spacings—but does not construct an explicit molecule or potential that has these properties.

Subsequent work, thoroughly reviewed in [1] has attempted to determine whether these predictions are universal and to explicitly construct potentials that approach them. One approach has been to conduct Monte Carlo searches of Hamiltonians with arbitrary spectra and dipole transition elements to identify those with large β_{int} and γ_{int}. The results support the three-state hypothesis, though the optima found in such calculations need not correspond to a local potential. To address this, several authors have numerically optimized the intrinsic hyperpolarizabilities with respect to the potential function for 1 electron, using various representations of the potential, including power laws[2], elementary functions

* timothy.atherton@tufts.edu

with a superimposed Fourier series[3], piecewise linear potentials[4, 5] and quantum graphs[6–9]. The best potentials from these different studies possess hyperpolarizabilities within the bounds of eq. (3) with the best known values found of $\beta_{int} \sim 0.71$ and $\gamma_{int} \sim 0.60$ achieved in several studies with qualitatively different potentials. It has therefore been speculated that the fundamental limits may require exotic potentials and not be achievable with local potential functions. The effect of including multiple electrons on the optimized potentials has, however, received relatively little attention. Watkins and coworkers [10] found for $N = 2$ electrons that the best intrinsic hyperpolarizabilities are somewhat lower than for the $N = 1$ electron case, but even with electron-electron interactions included, the universal features identified in other studies remained the same.

In this paper, we apply the potential optimization technique to potentials with $N > 2$ electrons that interact only through Pauli exclusion. This is a key step towards simulating realistic molecules. We find that the best values of β_{int} and γ_{int} fall off with increasing N from the $N = 1$ electron case, but rapidly converge to a universal value. The small number of parameters in our potentials allows a detailed exploration of the "landscape" of β_{int} and γ_{int} around the maximum. As in previous work, the hyperpolarizabilities are more sensitive to one parameter than the other. Dimensional and approximate analytical arguments allow us to provide physical interpretations of these two parameters in terms of the wavefunctions of the highest occupied molecular orbital.

The paper is organized as follows: in section II the choice of potential, calculation and optimization techniques are briefly reviewed; in section III we present results for β_{int} and γ_{int} separately together with some discussion of the implications of our results for identifying the features of potentials most important to the hyperpolarizabilities; brief conclusions are presented in section IV.

II. MODEL

Following a similar approach to that established in our earlier papers[4, 5], we optimize β_{int} and γ_{int} with respect to the shape of a piecewise linear potential dressed with Dirac delta functions for N electrons interacting only through Pauli exclusion. We perform this optimization for two carefully chosen potentials depicted in Fig. 1 as well as a more general potential. The first type is an asymmetric triangular well with a delta function at the center [fig. 1(a)],

$$V_1(x) = -\alpha\delta(x) + \begin{cases} -A_0 x, & x < 0 \\ x & x \geq 0 \end{cases}, \qquad (6)$$

parametrized by the left hand slope $A_0 > 0$ and the strength of the δ-function α. The effect of the δ function is to introduce a sudden change in the phase of the wavefunction. Lytel *et al.* [9] recently showed that the addition of a δ function to a 1D potential has an equivalent effect on the wavefunction to adding a side chain on a quantum graph. This correspondence suggests that 1D potentials dressed with δ-functions could be engineered in molecules by the addition of appropriate side groups.

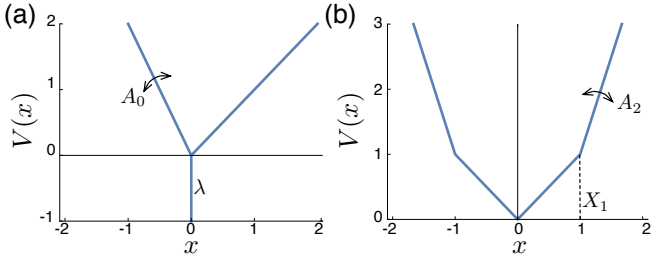


Figure 1. Schematics of the potential classes in which $\beta_{\rm int}$ and $\gamma_{\rm int}$ are optimized. (a) Asymmetric triangle will with a delta function at the center. The slope on the right side is fixed at 1 while the slope of the side A_0 and the strength of the delta function α can be varied. A_0 can be fixed at 1 to study a symmetric 1-parameter potential. (b) a 2 parameter symmetric linear piecewise potential. The first slope to the right of the origin is fixed at 1. The position of the boundary between the first and second elements X_1 and the second slope A_2 can be varied. The left side of the potential is constrained to be the reflection of the right side.

The second type of potential we consider, depicted in fig. 1(b), contains only linear elements defined for $x > 0$,

$$V_2(x) = \begin{cases} x & 0 < x < X_1 \\ A_2(x - X_1) + X_1 & x \geq X_1 \end{cases}, \qquad (7)$$

and for $x < 0$ defined by enforcing $\mathcal{P}$ symmetry, i.e. $V(-x) = V(x)$; this potential is specified by two parameters X_1 the position of the boundary between the two elements and A_2 the slope of the outermost element. These potentials were motivated by our results in [4, 5] that only 2 parameters at most were important to the optimization of both β_{int} and γ_{int}; they have been designed to achieve the known limits for $N = 1$ electrons with residual flexibility. For example, we showed in [5] that a triangular well with a δ-function of variable strength at the center, i.e. taking the potential (6) and fixing $A_0 = 1$, was able to reach within 3% of the upper bound for γ_{int} with λ as the only free parameter. The parametrization has also been chosen to eliminate variables irrelevant to β_{int} and γ_{int} associated with translations of the potential and rescalings of the form (5).

We also minimized β_{int} for a piecewise linear potential with m elements,

$$V(x) = \begin{cases} A_0 x + B_0 & x < x_0 \\ A_n x + B_n & x_{n-1} < x < x_n \\ A_m x + B_n & x > x_{m-1} \end{cases}, \qquad (8)$$

with positions x_n and slopes A_n as the parameter set describing the potential. We used this potential in our earlier paper on maximizing β_{int} for one electron[4]. There

are some necessary constraints on the parameters: the x_n are strictly ascending; $x_0 = 0$ and $B_0 = 0$ with no loss of generality and $B_1 = B_0$ with the remaining constants B_n given by,

$$B_n = \sum_{m=1}^{n-1} (A_m - A_{m+1}) x_m. \tag{9}$$

The energy scale associated with the potential is also fixed, as in the previous paper, by setting $A_1 = 1$. Having imposed these constraints, there remain $2N - 1$ free parameters.

For each of these potentials, β_{int} and γ_{int} were calculated for N electrons as follows: first, the Schrödinger equation is written for each segment as,

$$\left[-\frac{1}{2}\frac{d}{dx^2} + (A_n + \epsilon)x + B_n \right] \psi_n = E\psi_n \tag{10}$$

where A_n and B_n are the slope and offset in the nth segment and ϵ is the applied electric field. This can be solved analytically using the well-known Airy functions,

$$\psi_n(x) = C_n \mathrm{Ai}\left[\frac{\sqrt[3]{2}\left(B_n - E + x\left(A_n + \epsilon\right)\right)}{\left(A_n + \epsilon\right)^{2/3}} \right] + D_n \mathrm{Bi}\left[\frac{\sqrt[3]{2}\left(B_n - E + x\left(A_n + \epsilon\right)\right)}{\left(A_n + \epsilon\right)^{2/3}} \right]. \tag{11}$$

To solve for the coefficients C_n and D_n in each element, a set of boundary conditions are assembled at the edge of each element from the usual conditions, i.e.,

$$\psi_{n+1}(X_n) - \psi_n(X_n) = 0 \tag{12}$$

$$\psi'_{n+1}(X_n) - \psi'_n(X_n) = \alpha_n \psi_n(X_n) \tag{13}$$

where α_n is the strength of the delta function centered at X_n. The requirement that $\psi \to 0$ as $x \to \pm\infty$ eliminates two coefficients. The boundary conditions can be written as a set of linear equations,

$$W \cdot u = 0 \tag{14}$$

where u is a vector comprised of the C_n and D_n coefficients and W is a matrix that depends on E, ϵ and the parameters A_n and X_n. The single electron energy levels λ_i for the potential are determined by numerically finding the roots of,

$$\det W = 0, \tag{15}$$

setting $\epsilon = 0$. Having determined these, we construct the non-interacting N electron ground state from the single electron states by successively filling the energy levels using the *aufbau* principle; we similarly determine the first excited state by promoting an electron from the highest occupied orbital to the lowest unoccupied orbital. In this work we focus only on even values of N, as this simplifies determining which electron to promote. The N-electron ground state energy E_0, and that of the first excited state E_1, are then determined by summing over the energies of the individual single electron energies,

$$E_n = \sum_i \nu_i^n \lambda_i, \tag{16}$$

where ν_i^n is the occupation number for the i-th single electron level in the n-th multi-electron state.

The hyperpolarizabilities β and γ are obtained by differentiating the ground state energy as a function of ϵ,

$$\beta \equiv \frac{1}{2}\frac{d^3 E_0}{d\epsilon^3},\ \gamma \equiv \frac{1}{6}\frac{d^4 E_0}{d\epsilon^4}. \tag{17}$$

The necessary derivatives can be related to the single-electron energy levels by differentiating (16) with respect to ϵ. An important advantage of the piecewise linear potential is that the necessary derivatives of the single electron energy levels λ_i are conveniently obtained by repeated differentiation of the determinant eq. (15) using the Jacobi formula,

$$\frac{d}{d\epsilon}\det W = \mathrm{Tr}\left(\mathrm{adj}W \cdot \frac{dW}{d\epsilon}\right),$$

where adj$W$ is the adjugate matrix of W, together with the chain rule,

$$\frac{dW}{d\epsilon} = \frac{\partial W}{\partial \epsilon} + \frac{\partial W}{\partial \lambda_i}\frac{d\lambda_i}{d\epsilon}. \tag{18}$$

To avoid repetition, formulae are available for these derivatives in references [4] and [5]. Having evaluated these derivatives, the intrinsic hyperpolarizabilities are easily calculated numerically and a program to do so was implemented in *Mathematica 10*. In subsequent sections, we present the optimized results as a function of N, together with the corresponding best potentials.

III. RESULTS

A. First hyperpolarizability

We optimized β_{int} with respect to the parameters of the asymmetric δ-function potential of eq. (6) as well as the more general piecewise linear potential with of eq. (8) with four linear elements and 6 parameters. The number of linear elements was chosen because earlier work showed no improvement in β_{int} after 6 parameters[4]. The highest values found for increasing N are displayed in fig. 2(a) for both of these. It is apparent that the best achievable result diminishes for $N > 2$ but rapidly reaches a plateau of $\beta_{int} \sim 0.4$ and that both potentials give consistent results.

For $N \leq 8$ electrons, the two choices of potential give very consistent results, suggesting that we have indeed found a likely global optimum. For $N > 8$ electrons, however, the optimization procedure failed to find a maximum of β_{int} for the general linear potential that approached that of the asymmetric δ-function potential.

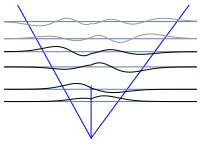
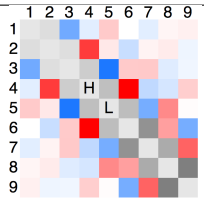
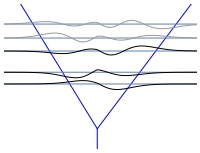
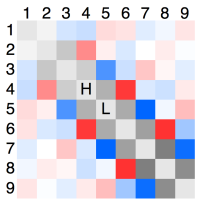
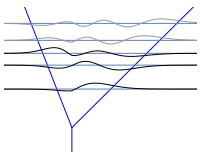
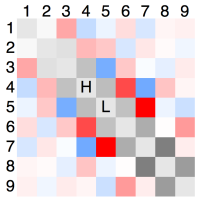
Value of β_{int}	Potential and wavefunctions	A_0	α	Hessian Eigenvalues	Hessian Principal Eigenvector	x_{nm}	X	E
0.404 (100%)		1.334	-2.229	-6.84, -0.099	$\begin{pmatrix} 0.999 \\ 0.035 \end{pmatrix}$		0.573	0.500
-0.372 (96%)		1.227	1.928	7.70, 0.130	$\begin{pmatrix} 0.999 \\ 0.027 \end{pmatrix}$		0.570	0.500
0.370 (92%)		2.568	2.307	-1.24, -0.014	$\begin{pmatrix} 0.986 \\ -0.162 \end{pmatrix}$		0.569	0.523

Table I. Asymmetric δ-function potentials with locally optimal β_{int}. Results for each potential are shown columnwise: the value of β_{int} , and the fraction of the globally optimum value; the potential and wavefunctions with black lines indicating states that are occupied and light gray lines indicated unoccupied states in the ground state configuration; optimized values of the parameters A_0 and δ; the eigenvalues and eigenvectors of the Hessian matrix of β_{int} with respect to the parameters; a plot of the dipole transition matrix $x_{nm} = \langle n\,|x|\,m\rangle$ with the Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) indicated by H and L respectively, coloring described in the text; the energy ratio E and dipole transition moment X for the Kuzyk three-state model.

We speculate that this is because the number of local maxima increases with N; we shall show that this is true later for the more carefully chosen potentials at least. Because of this, we did not consider the general piecewise linear potential further but display the optimized potential and wavefunctions for $N = 2$ and $N = 8$ respectively in fig. 2(b) and (c). That these potentials achieve similar values of β_{int} to the asymmetric δ-function potentials, despite visually appearing very different, supports our conclusion in previous work[4] that β_{int} is poorly determined in potential space with many irrelevant directions.

Due to the small parameter space of the asymmetric δ-function potential, it is possible to directly visualize the objective function; this is displayed in fig. 2(d) for $N = 1$ or 2 electrons (the magnitude is scaled by $2^{-1/2}$ for the $N = 2$ electron case), with corresponding optimized potential and wavefunctions shown in fig. 2(e). The objective function for $N = 8$ electrons and the optimized potential and wavefunctions are shown in fig. 2(f) and (g) respectively. It is immediately apparent that as N increases, the objective function acquires additional local extrema.

For the $N = 1$ case, the global optimum found is $\beta_{int} = 0.701632$ at $A_0 = 35.283$ and $\alpha = 1.1189$, which is only marginally short of the best known value of $\beta_{int} = 0.708951$ found from optimizing many different classes of potential [4]. The global maximum lies at the top of the long, narrow ridge viewed in fig. 2(d), a feature of the objective function that was also seen in earlier studies of more complicated potentials [4]. Its presence implies that β_{int} is much less sensitive to one of the parameters than the other, which can be quantified by computing the hessian matrix of β_{int} with respect to the parameters,

$$h = \begin{pmatrix} \frac{\partial^2}{\partial^2 A_0} & \frac{\partial^2}{\partial A_0 \partial \alpha} \\ \frac{\partial^2}{\partial A_0 \partial \alpha} & \frac{\partial^2}{\partial^2 \alpha} \end{pmatrix},$$

and finding the eigenvalues and eigenvectors. These quantities respectively measure the curvature and principal directions of the objective function around the maximum. We previously used this technique in [4] to show that that while the best known value of $\beta_{int} = 0.708951$ was obtained by optimizing a piecewise linear potential with 6 free parameters, in fact β_{int} was effectively only sensitive to 2-3 parameters at the optimum. Here, the eigenvalues of the hessian evaluated at the global maximum of β_{int} are -0.32277 and -0.00066 and the associated eigenvectors are $(0.00965, -0.99995)$ and $(0.99995, 0.00965)$. Hence, β_{int} is very sensitive to the value of α, the second parameter and much less sensitive to the value of A_0; Since the eigenvectors are nearly

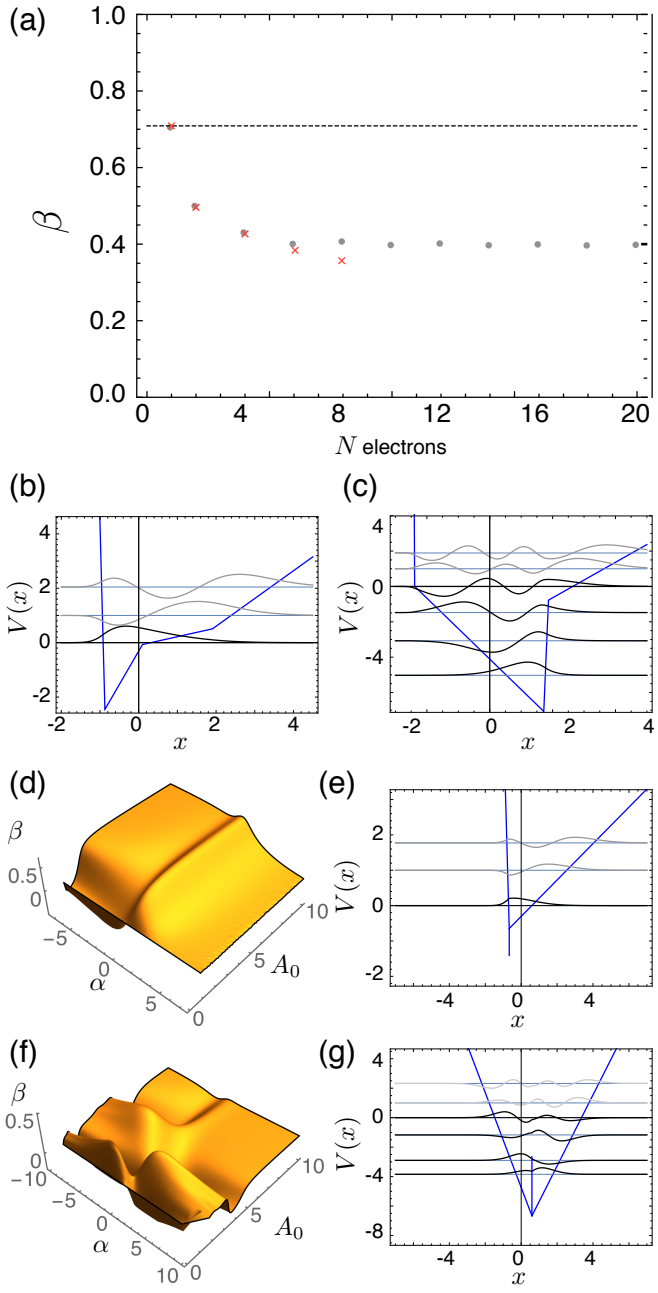


Figure 2. (a) Maximum β_{int} achieved in 2-parameter asymmetric δ-function potential. Results for the asymmetric delta function potential are shown as grey circles; results for an arbitrary linear potential are shown as red crosses. The dashed line represents the largest values of β_{int} for one electron found to date. Optimized potential and wavefunctions for the arbitrary linear potential with (b) $N = 2$ electrons and (c) $N = 8$ electrons. For the asymmetric δ-function potential and $N = 2$, (d) the objective functions β_{int} plotted versus the δ-function potential shape parameters A_0 and α and (e) the optimized potential and wavefunctions. Corresponding plots for $N = 8$ electrons are shown in (f) and (g). For all subfigures, wavefunctions plotted in black are occupied in the ground state; those in grey correspond to the two lowest unoccupied states.

parallel to basis vectors $(1, 0)$ and $(0, 1)$ in parameter space—as is visible from the orientation of the ridge in fig. 2(d)—it is clear that β_{int} is sensitive to these features of the potential specifically, and not some combination of them. In contrast, the eigenvectors in [4] were not well aligned with the parameter space and so it was not possible to ascribe high β_{int} to particular features of the potential. The design advice from this study is much clearer: to optimize β_{int}, create an asymmetric potential well with a steep wall on one side, i.e. set $A_0 \gg 1$; then add an attractive group in the center and tune the strength of attraction, i.e. carefully adjust α as this largely determines β_{int}.

A similar analysis was applied to the multi-electron case. In table I, we display the three extrema with largest β_{int} for $N = 8$, together with a plot of the potential and wavefunctions; parameter values of A_0 and δ at the optimum and the results of the eigenanalysis. Unlike the $N = 1$ case, the global optimum has a repulsive δ-function; the next two solutions have attractive δ-functions. The existence of both attractive and repulsive extrema supports the paradigm proposed by Lytel *et al.* [9] in their work on optimization of quantum graphs. They suggest that large β_{int} is achieved by introducing a disturbance at some point in the π-electron chain of molecule, e.g the addition of a side group. The disturbance then induces a phase shift in the wavefunction, producing a change in dipole moments sufficient to achieve large β_{int}. In our work, the δ-function serves to provide the disturbance; the insight of Lytel *et al.* is that it is the overall phase shift at the disturbance that is the relevant parameter, not its detailed nature. Hence both attractive and repulsive features can provide an appropriate phase shift.

Just as for the $N = 1$ case above, eigenanalysis of the hessian for the multi-electron case shows that the eigenvectors remain well aligned with the parameter basis vectors for the multi-electron case. Surprisingly, while α was found to be the most important parameter for $N = 1$, it is A_0 that appears to be most significant for the $N = 8$ case. The ratio of eigenvalues is also less extreme, around 10^{-2} rather than 10^{-4} as before. These results suggest that when applying the design approach herein proposed, i.e. an asymmetric well with a phase shift-inducing feature, to real systems, tuning both parameters may be important to achieve high β_{int}.

We also performed eigenanalysis for the more general linear potential; as a prototypical example for $N = 8$ electrons the eigenvalues were $\left(-458, -0.4, -0.1, -3 \times 10^{-3}, 1 \times 10^{-4}, 4 \times 10^{-6}\right)$ indicating that only one parameter is important. Interestingly, with increasing N the lowest 7 eigenvalues remained roughly constant while the largest eigenvalue strongly increased: For $N = 2$; the principal eigenvalue was found to be -2.7, for $N = 4$ it was -196 and for $N = 6$ it was -320. This progression is interesting because it suggests that for large N, the problem in some sense becomes simpler as the important parameter dom-

inates the others to an ever increasing extent. Unfortunately, as with previous work[4], the eigenvectors are not clearly aligned with the parameter space, so the general linear potential provides less useful information than the highly constrained asymmetric δ-function potential.

We also display for each potential a visualization of the first few position matrix elements $x_{nm} = \langle n|x|m\rangle$. These are important because, as discussed more fully in the appendix below, the hyperpolarizabilities can be expressed as a sum over states involving x_{nm} as well as the energy-level differences $E_{nm} = E_n - E_m$. It is therefore natural to examine this matrix to determine which transitions contribute most to the hyperpolarizabilities. The interpretation of this matrix is, however, complicated by the fact that many combinations of these parameters are individually irrelevant to the hyperpolarizabilities. As is well-known, for example, the three state model[3, 11, 12] achieves the bounds quoted in equation (3), and only requires two parameters $E = \frac{E_{10}}{E_{20}}$ and $X = \frac{|x_{01}|}{|x_{01}^{MAX}|}$ with $|x_{01}^{MAX}| = \sqrt{\frac{\hbar N}{2mE_{10}}}$.

The x_{nm} matrices are dominated by the tridiagonal terms, and the diagonal elements can be eliminated from the expressions for the hyperpolarizabilities, e.g. by using the dipole-free sum over states (DFSOS) formula. To aid inspection, we have therefore omitted the diagonal terms and plotted the first off-diagonal terms, i.e. those with $|n-m| = 1$, in greyscale. The remaining terms are plotted in a scheme where intensity corresponds to magnitude and red or blue refers to the positive or negative sign of the term respectively. Reflecting the potential in space and changing the signs of odd-indexed wavefuctions changes such plots only by changing the signs, and for $n - m \neq 0$ the colors of x_{nm} for $n - m$ even. To avoid confusion, we have chosen the potential or its mirror image in such a way that x_{nm} is positive for n the HOMO and m the LUMO+1. The off-tridiagonal terms are important to β_{int}, because a tridiagonal matrix would yield $\beta_{int} = 0$ as for the Harmonic Oscillator. Clearly, however, for all the local optima displayed in table I, the $|n-m| = 2$ terms are much larger than the remaining $|n-m| > 2$ terms.

For each local maximum in table I, we display calculated values of the E and X parameters. For the three state model, these parameters yield optimal $\beta_{int} = 1$ for values of $E = 0$ and $X = 3^{-1/4} = 0.760$. However, past studies[13] of optimized potential functions for $N = 1$ electrons find values of $E \approx 1/2$ and $X \approx 0.789$ regardless of the starting potential. Optimization of Quantum graphs[7] produces mildly different values of $E \approx 0.4$ and $X \approx 0.79$. For the optima presented here, we also find $E = 1/2$, but the results seem to favor a value of $X = 0.57$. This result is consistent with the results of eigenanalysis of the hessian, which suggests only one of the parameters E and X can be important.

The dipole matrix plots, together with the value of X or $X' = x_{HOMO,LUMO+1}/x_{HOMO,LUMO}$ allows us to appreciate the compromises made in this optimization. All contributions to β in the dipole-free SOS formula involve three states, the product of the transition moments between them and a function of the energies. From the x-matrix plots it is clear that matrix elements become smaller very quickly moving away from the diagonal. Thus the SOS for β is expected to be dominated by contributions from terms that involve only one off-tridiagonal element. There are exactly two sets of three states involving only the one off-tridiagonal matrix element: the ground state, the excited state in which one electron has been excited from the HOMO to the LUMO and one of two doubly excited states: the state in which one electron has been excited from the HOMO to the LUMO+1 or that in which one electron has been excited from the HOMO-1 to the LUMO. As the HOMO to LUMO+2 and HOMO-1 to LUMO transition matrix elements have opposite signs, these contributions have opposite signs. Generically the transition moments are larger between higher energy states to the HOMO to LUMO+1 and so are expected not to dominate.

The smaller result for many electrons seems, at least in part, to be explained by the negative contribution of the HOMO-1 to LUMO contribution. This contribution is lacking for 1 or two electrons when there is no HOMO-1. Other than this, the pattern of matrix elements looks substantially similar for all the maxima—including ones we have not displayed. Thus it seems that the smaller results for β_{int} for more than two electrons may be explained by the need to include four states rather than two, given the near degeneracy of the two doubly excited states.

B. Second hyperpolarizability

The symmetric triangular well with a δ-function, i.e. eq. (6) with $A_0 = 1$, is known to achieve near-optimal results for γ_{int} for $N = 1$ despite only containing one free parameter. Due to the simplicity of this potential, it is possible to directly visualize γ_{int} as a function of this parameter, the strength of the delta function α, for different values of N. The results are displayed in fig. 3(a). Immediately apparent is an alternation of the sign of the γ_{int} curves as N is increased by 2: for $N = 2, 6, 10...$ the maximum value of γ_{int} occurs for positive α—corresponding to a δ-function wall in the middle of the well, while for $N = 4, 8...$ the maximum value of γ_{int} occurs for negative α. The magnitude of α for which the maxima occur increases only slightly with N, however. The behavior of the minimum value of γ_{int} is similar, but shifts even less.

Some insight into these results is obtained by examining the effect of the δ-function on the single electron energy level spectrum for the potential, shown in fig. 3(b). Clearly, only the even wavefunctions are affected by the δ-function due to the $\mathcal{P}$-symmetry of the potential. While the placement of λ_0 can be freely adjusted by changing α, the higher even energy levels are bounded by their invariant odd neighbors, e.g. $\lambda_1 < \lambda_2 < \lambda_3$. The freedom

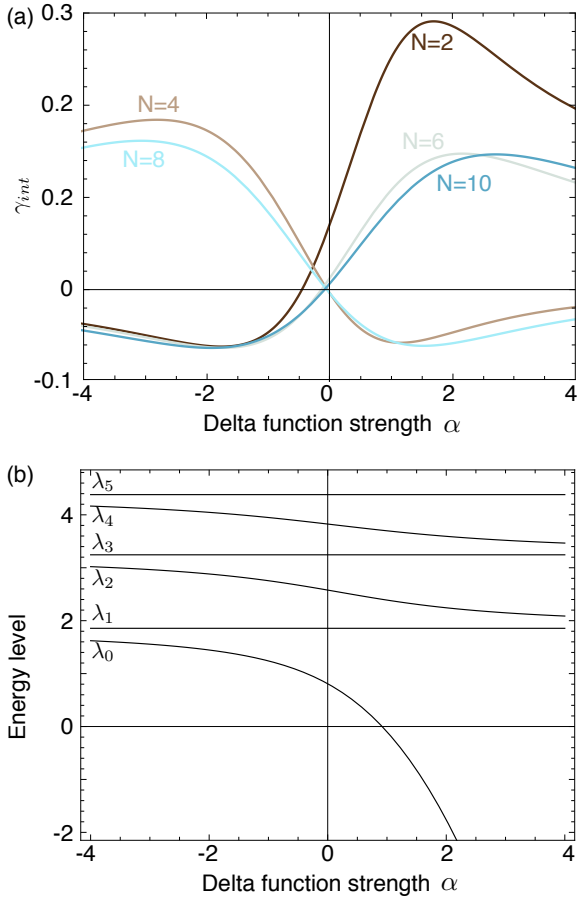


Figure 3. (a) $\gamma_{\rm int}$ versus δ function strength α for varying electron number N in the symmetric δ function potential. Each time N is increased by 2, the sign of α for which the maximum or minimum occurs changes sign. (b) The energy level structure of the symmetric δ function potential as a function of δ function strength α.

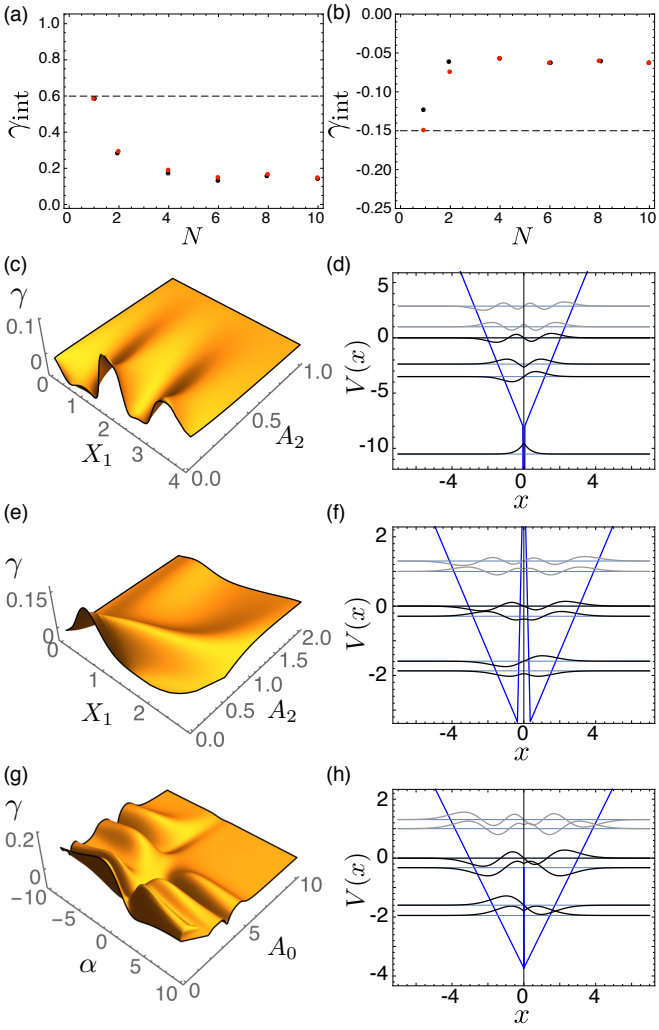


Figure 4. Maximum (a) and minimum (b) $\gamma_{\rm int}$ achieved for varying increasing numbers of electrons; results for the symmetric linear piecewise potential are shown in red and for the symmetric δ function potential in black. (c) For $N = 8$, the objective function $\gamma_{\rm int}$ plotted versus the linear piecewise potential shape parameters A_2 and X_1 for positive A_1 and (d) the best potential and wavefunctions obtained. (e) Results for negative A_1 and (f) the corresponding potential and wavefunctions. (g) Results for the δ function potential shape parameters α and A_2 and (h) optimized wavefunctions and potential.

of selecting λ_0 relative to λ_1 allows the $N = 2$ case to achieve γ a high fraction of the Kuzyk maximum, while the restricted higher levels only permit a lower value of γ_{int} to be achieved. The reason for the alternation is also apparent. For $N = 2$, the Highest Occupied Molecular Orbital (HOMO) in the ground state is λ_0 while the Lowest Unoccupied Molecular Orbital (LUMO) is λ_1; increasing α serves to *widen* the HOMO-LUMO gap. On the other hand, for $N = 4$ the HOMO is λ_1 and the LUMO is λ_2; increasing α for this case serves to narrow the HOMO-LUMO gap. This trend continues with α widening the HOMO-LUMO gap for $N = 2, 6, 10...$ and narrowing the HOMO-LUMO gap for $N = 4, 8$; the alternating effect of α explains the different signs of γ_{int} with N.

To determine whether these results are universal, we also optimized γ_{int} for the asymmetric δ-function potential eq. (6) as well as the linear piecewise potential given by eq. (7). Shown in fig. 4(a) and (b) are the best maximum and minimum γ_{int} obtained as a function of N. Although there are small differences between results obtained with different potentials, the same trend is clear: that the best γ_{int} falls off with increasing N but rapidly reaches a constant value, yielding an apparent feasible range of $-0.05 < \gamma_{int}^{max} < 0.2$. These apparent bounds are shared by all three potentials.

We plot the objective function results for $N = 8$ for several different scenarios: For the linear piecewise potential, the sign of A_1 must be chosen to be positive or negative prior to optimization. The objective function is

shown for $A_1 = +1$ in fig. 4(c) revealing several local maxima; the corresponding optimized potential, which maximizes γ_{int} , and wavefunctions are shown in fig. 4(d). If alternatively, $A_1 = -1$, a minimum of γ_{int} is obtained; the objective function is rather simpler as apparent in fig. 4(e) and the corresponding potential and wavefunctions are shown in fig. 4(f). Note that despite the arbitrariness of the linear potential, the optimized potentials strongly resemble the symmetric δ-function potential, validating its use as *ansatz* earlier.

The asymmetric δ-function potential has a much more complicated landscape for γ_{int} as is evident in fig. 4(g). Many local minima and maxima exist. The best three for both positive and negative γ are displayed in table II, together with the parameter values and results of the hessian eigenanalysis. Many of the results above for β are also seen for γ: both attractive and repulsive δ-functions lead to large γ, in agreement with the phase interruption paradigm. Eigenvalues of the hessian differ by a similar ratio of $\sim 10^{-2}$ and again the eigenvectors are well aligned with the basis vectors of the parameter space. Just as for the first hyperpolarizability with $N = 8$ electrons, it appears that A_0 is the most important parameter.

Intriguingly, the best negative γ_{int} obtained does not possess $\mathcal{P}$ symmetry in sharp contrast to the $N = 1$ case, although a $\mathcal{P}$-symmetric solution within 1% of this value also exists. It is possible this is due to the freedom of choosing the relative position of the lowest energy level λ_0 relative to λ_1; less freedom exists for the higher single particle energy levels.

Position matrix elements x_{nm} are shown in II for each potential. Note that these matrices for $\mathcal{P}$ symmetric potentials have a characteristic checkerboard structure. Elements immediately off the diagonal $|n - m| = 1$ are plotted on a greyscale while non-tridiagonal elements are plotted in red and blue and it is evident, as for β_{int}, that the matrices are diagonally dominant. For positive γ_{int} , the $\mathcal{P}$-symmetric global optimum has large terms on the $|n - m| = 3$ off diagonal, which is consistent with a few-state hypothesis; here the HOMO→LUMO+2 and HOMO-1→LUMO+1 terms make the strongest contribution. The transition matrices for the non $\mathcal{P}$-symmetric solutions, however, seem to have little in common with the global optimum, which is perhaps to be expected since the values of γ_{int} for these secondary maxima fall some 20% below it.

For negative γ_{int}, where the two best solutions have nearly identical values, the transition matrix for the $\mathcal{P}$-symmetric optimum resembles the solution for positive γ_{int}, i.e. HOMO→LUMO+2 and HOMO-1→LUMO+1 are the dominant off-tridiagonal terms. The non $\mathcal{P}$-symmetric solution has an interesting structure: the HOMO→LUMO element is large, while the HOMO→LUMO+1 and HOMO-1→LUMO elements are significantly weaker than the other $|n - m| = 2$ elements. A similar pattern, where off-tridiagonal elements are suppressed around the Fermi surface, is also seen in the third highest optimum that is also non $\mathcal{P}$-symmetric. It is for negative γ_{int}, therefore, that we find the clearest correspondence with the transition matrix elements of the three-state ansatz.

We now connect these results to the Sum-Over States Formula for γ, which includes two types of terms: There are both terms involving the ground state and three other states, m, n, and l, energy factors and the matrix element product $x_{0l}x_{0n}x_{lm}x_{nm}$ or $x_{0m}x_{ml}x_{ln}$ which we shall hereafter call "four state" terms. There are also terms involving the ground state and two excited states, n, m, referred to hence as "three state" terms. A three state term proportional to $(x_{0n}x_{nm})^2$ makes a positive contribution to γ, one proportional to $(x_{0n}x_{0m})^2$ will make a negative contribution. This is more easily seen from eq. (73) in [1], if m is considered the unconstrained summation variable, n and l are (as before) both considered to be different from m, and the energy factors are made symmetric in n and l. Moreover, only three state terms involving at least one off-tri-diagonal matrix element can contribute. If there are no off-tri-diagonal matrix elements, the matrix elements and energies are constrained to be those for the harmonic oscillator, for which $\gamma = 0$. The largest terms in γ from four-state terms involves three excited states, three-tridiagonal matrix elements and one matrix element for which $|n - m| = 3$. The largest contributions to γ from three state terms involves the square of a tri-diagonal matrix element and the square of an $|n - m| = 2$ matrix element. These last matrix elements, and so also these terms, are forbidden by $\mathcal{P}$ symmetry.

Now looking at the x matrices, we see that the the $\mathcal{P}$-symmetric potentials have three large contributions, with the signs of these three contributions oscillating with period 2 in the empty state in the highest or lowest singly occupied state. The resultant cancellations may again partly explain the smaller γ_{int} achieved for these potentials. The difference between positive and negative γ is not the kinds of terms involved (as suggested in [14]) but rather in the signs of the off-tridiagonal matrix elements. The three-state terms that are allowed for non $\mathcal{P}$-symmetric potentials are more similar to those considered in [14], except that these can have either sign.

To facilitate comparisons with other work, we again computed values of the Kuzyk three-state model parameters E and X. For the lower bound, all three extrema favor $E = 0.5$ and $X \approx 0.65$. An earlier study by Watkins *et al.*[3], who optimized γ_{int} for different potentials and $N = 1$, found for negative γ_{int} $X \approx 0.66$ and two clusters of solutions with $E \approx 0.15$ or $E \approx 0.53$. Our results are therefore in excellent agreement. For the upper bound, our results display more variation with $E = 0.765$ and $X = 0.615$ for the best potential found. Watkins *et al.* similarly found more variation in positive γ_{int}, obtaining results around $E \approx 0.57$ with $X \approx 0.5$. As for β, the discrepancy is explicable because eigenanalysis reveals that effectively only one parameter is truly important, seemingly the slope of the left hand boundary.

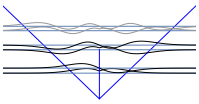
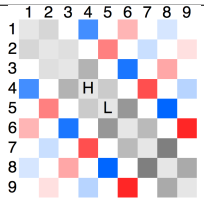
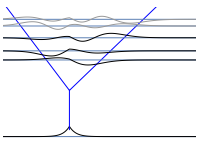
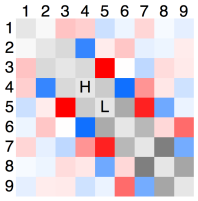
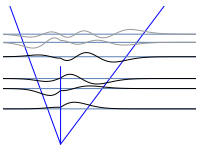
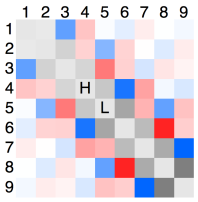
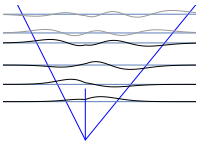
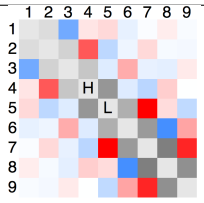
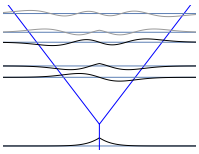
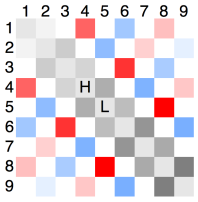
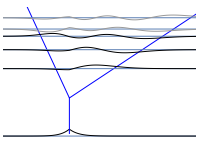
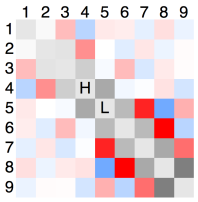
Value of γ_{int}	Potential and Wavefunctions	A_0	α	Hessian eigenvalues	Hessian eigenvectors	x_{nm}	X	E
0.161 (100%)		1.000	-2.963	-0.597, -0.048	$\begin{pmatrix} -0.999 \\ 0.053 \end{pmatrix}$		0.615	0.765
0.134 (83%)		1.375	2.534	-1.926, -0.039	$\begin{pmatrix} -0.999 \\ -0.004 \end{pmatrix}$		0.516	0.638
0.127 (79%)		2.043	-3.765	-1.597, -0.039	$\begin{pmatrix} -0.999 \\ 0.024 \end{pmatrix}$		0.511	0.638
-0.0613 (100%)		1.630	-2.107	6.970, 0.0187	$\begin{pmatrix} 0.999 \\ -0.016 \end{pmatrix}$		0.655	0.500
-0.0609 (99%)		1.000	1.524	13.78, 0.0311	$\begin{pmatrix} 0.999 \\ -0.0006 \end{pmatrix}$		0.654	0.500
-0.0502 (82%)		3.260	2.467	1.162, 0.0123,	$\begin{pmatrix} 0.997 \\ -0.069 \end{pmatrix}$		0.639	0.500

Table II. Asymmetric δ-function potentials with locally optimal γ_{int}. Results for each potential are shown as follows, columnwise: the value of γ_{int} , and the fraction of the globally optimum value; the potential and wavefunctions; optimized values of the parameters A_0 and δ; the eigenvalues and eigenvectors of the Hessian matrix of γ_{int} with respect to the parameters; a plot of the dipole transition matrix $x_{nm} = \langle n|x|m\rangle$ with the HOMO and LUMO indicated by H and L respectively and coloring as described in the text; the energy ratio E and dipole transition moment X for the Kuzyk three-state model.

C. Discussion

The above subsections have presented optimized potentials for β_{int} and γ_{int} with increasing N, expressed in a parameter space that *a posteriori* is found to coincide with what is important to the hyperpolarizabilities. As discussed, this is a significant development over previous work, where the parameter space was larger the optimized potentials gave a less clear sense of which features are important. Nonetheless, the parameters A_0 and α are an artifact of the parameterization chosen, and it is therefore desirable to identify quantities that are invariant under reparametrization and rescaling of the form in eq. (5). This is particularly important because, if we generalize to the space of arbitrary potentials, there are many other solutions with identical values of the hyperpolarizability. For the $N = 1$ case, we know that the hy-

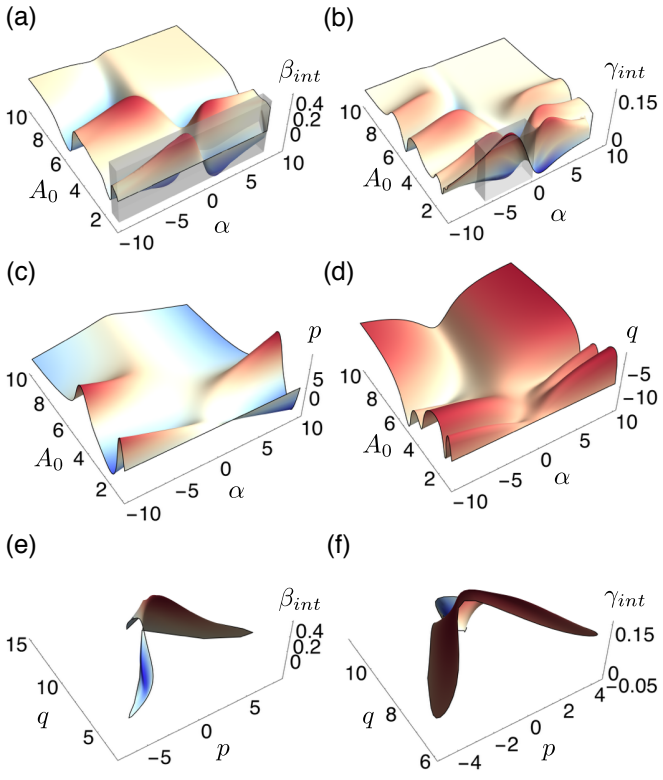


Figure 5. Objective functions (a) β_{int} and (b) γ_{int} for the asymmetric δ-function potential shown in the natural parameter space A_0 and α; wavefunction parameters at the classical turning points (c) $p = \psi_l'/\psi_r' - \psi_r'/\psi_l'$ and (d) $q = -\Delta x^2 \left|\psi_l'\psi_r'\right|^{2/3}$ plotted with respect to A_0 and α. A selected region of the objective function, enclosed in the grey volume in (a) and (b), is show replotted with respect to p and q in (e) and (f) respectively.

perpolarizabilities can be completely reconstructed from the ground state wavefunction[15]. Extending this to the $N > 1$ case, it is easy to see that the hyperpolarizablities can be constructed for this problem as appropriate integrals of the occupied single particle wavefunctions alone. As the largest term is expected to the HOMO, the important parameters ought to be expressible in terms of parameters crucial to the HOMO.

Moreover, there is a need to connect the dipole-free sum-over-states view, which prescribes certain values of the dipole transition matrix elements, and the potential view that we pursued above. We therefore performed an approximate analysis of the problem, inserting WKB ansatz wavefunctions into expressions for E_{nm} and x_{nm} to determine what details of these wavefunctions most significantly affect these parameters. The full calculations presented in the Appendix below support this argument and suggest that the three parameters introduced in the next paragraph are the crucial ones..

On the basis of approximate analytic arguments made in the appendix, and a numerical experiment that follows, we believe that there are three length scales that are important to the hyperpolarizabilities around the maximum. One is Δx, the separation between the turning points of the HOMO. The other two are length scales L_k for k and an element of $\{r, l\}$ that characterizes the rate at which the wavefunction varies near the right and left turning points of the HOMO. A variety of nearly equivalent forms can be given for these lengths and dimensionless ratios containing them, including $L = (V')^{-1/3}$, and more global formulae related the derivatives of the wavefunction at the turning points or to the normalization of the HOMO. From these three lengths, it is possible to construct two dimensionless parameters that are arguably important to the hyperpolarizability, for example,

$$\begin{aligned} p &= \psi_l'/\psi_r' - \psi_r'/\psi_l' \\ q &= -\Delta x^2 \left|\psi_l'\psi_r'\right|^{2/3} \end{aligned} \tag{19}$$

are invariant under reparameterization and rescaling. Here Δx is the distance between the turning points, ψ_l' and ψ_r' are the slopes of the wavefunctions where the subscripts l and r refer to the left and right turning point respectively. As some of the definitions of L are zero when V' is infinite at the turning points, and it is still possible for the hyperpolarizabilities to approach the maximum in this case, appropriate dimensionless combinations should not be infinite in this limit. As β_{int} requires asymmetry in the potential, it is expected that a parameter odd under $\mathcal{P}$ symmetry, such as p, mostly controls β while one even under $\mathcal{P}$ symmetry, such as q, mostly controls γ_{int}.

In fig. 5 we show the objective functions (a) β_{int} and (b) γ_{int} plotted in the parameter space of the asymmetric δ-function potential, i.e. A_0 and α. In the same parameter space, we show in 5(c) and (d) the values of the wavefunction parameters p and q from eq. (19). From these plots, we see that much of the structure in the objective functions β_{int} and γ_{int} is attributable to these new parameters: notice that the position of the ridges in 5(c) corresponds to the ridges in 5(a), while those in 5(d) corresponds, more roughly, those in 5(b). In figs. 5(e) and (f), the portion of the objective function in figs. 5(a) and (b) that is enclosed within the grey cuboid is reprojected into the new (p, q) parameter space. Plots of the objective functions—not shown here—in (p, q) space for the full range of (A_0, α) from fig. 5(a) and (b) closely resemble the structure observed in this reduced reduced region.

From these results, we conclude that β_{int} is largely determined by p while γ_{int} is largely determined by q, and the remaining parameter in each case must be tuned less precisely to achieve the optimum. It is also now clear that the asymmetric δ-function performs so well due to a fortuitous correspondence: the A_0 parameter directly controls the ratio of the slope of the potentials at the turning points, which is readily related to the slope of the wavefunctions at the classical turning points, i.e A_0 and p are simply related.

We display in table III the values of p and q calculated for each of the optima of β_{int} and γ_{int} shown in

	Optimum value	p	q
β_{int}	0.403	1.69	−8.99
	−0.372	−1.28	−7.66
	0.370	2.11	−9.26
$\gamma_{int} > 0$	0.161	0	−9.30
	0.134	−1.85	−6.40
	0.126	−1.88	−6.95
$\gamma_{int} < 0$	−0.0613	0.456	−10.54
	−0.0610	-1.02×10^{-5}	−9.30
	−0.0503	1.214	−10.54

Table III. Values of dimensionless parameters p and q constructed from the HOMO wavefunction for each of the optimized $N = 8$ solutions displayed in tables I and II

		N				
		2	4	6	8	10
β_{int}	p	−4.17	1.94	−2.52	1.69	−1.96
	q	−1.37	−4.00	−7.11	−8.99	−12.6
$\gamma_{int} > 0$	p		5.15×10^{-3}	2.08	0	1.92
	q		−3.64	−4.97	−9.30	−9.62
$\gamma_{int} < 0$	p		0	0	0.456	0
	q		−3.65	−7.35	−10.5	−13.2

Table IV. Dimensionless parameters p and q calculated from the best solution found as a function of N.

table I and II. For each of the optimization problems, i.e. $|\beta_{int}|$, $\gamma_{int} > 0$ and $\gamma_{int} < 0$ the values of at least one of these parameters are internally quite consistent with each other. This is particularly so for the negative γ_{int} results where the secondary optima are close to the global optimum. Values of p and q for the global optimum as a function of N are shown in table IV. These reveal several trends: First $\mathrm{sgn}(\beta_{int}) = \mathrm{sgn}(p)(-1)^{n_{nh}}$ where n_{nh} is the number of nodes in the HOMO. Otherwise, it is expected that $\mathcal{P}$ symmetry just changes the sign of β_{int} and p. Also, a sign alternation occurs in p for β_{int}; q seems to increase with N for γ_{int}, while p is found alternately 0 or some small value. The *a posteriori* consistency of these parameters supports the argument above that these are the "real" parameters of the optimization problem.

IV. CONCLUSION

We have optimized the intrinsic hyperpolarizabilities β_{int} and γ_{int} for non-interacting multi-electron systems with respect to the shape of several classes of potential: a piecewise linear potential, and an asymmetric triangular well with a δ-function. The best values obtained for β_{int} and γ_{int} drop from the $N = 1$ case and approach an apparent feasible range of $|\beta_{int}| < 0.4$ and $-0.05 < \gamma_{int}^{max} < 0.2$ for N larger than around eight electrons. The asymmetric δ-function potential achieves these bounds and, due to the small number of parameters, and *a posteriori* verification that the parameters are indeed relevant, provides a design prototype for synthesis of new chromophores. For β_{int}, a molecule should have asymmetric walls and possess an attractive or repulsive feature in the middle—a main chain functional group or side-chain—that promotes a rapid change in the phase of the wavefunction. The asymmetry of the boundary and the strength of the attraction or repulsion should then be tuned to achieve high β_{int}. For γ_{int}, the molecule should be essentially $\mathcal{P}$-symmetric with a central attractive or repulsive feature that should similarly be tuned.

By approximate analysis, we also determined that the *ad hoc* parameters of our potential can be re-expressed in terms of the dimensionless, scale invariant parameters derived from the shape of the HOMO wavefunction at the classical turning points. These new parameters are important both because they explain the success of our original parametrization and because they provide a new wavefunction-centered approach to screen potential chromophores for large hyperpolarizabilities.

The results also provide important information on how well the many theoretical studies of single electron systems might describe real molecules that posses multiple electrons. While the apparent bounds quoted above are more restrictive than the $N = 1$ case, it seems, encouragingly, that the overall design paradigms described above apply equally well to both cases. Information from other studies on target values of the E and X parameters of the three state model for $N = 1$ also seems to remain valid with increasing N. The insights of [4, 5] that only a very small number of parameters are required remain valid, and happily, this appears to be increasingly so with large N. Because of this and because the objective function for multiple electrons acquires many more local extrema, it should in principle be *easier* to tune multi-electron systems.

Appendix A: Approximate Analysis

In this appendix we use approximate techniques to argue that the dimensionless parameters identified in the main text constructed from the separation between the turning points of the HOMO and the slopes of the potential at these turning points ought to mostly explain the hyperpolarizability. The simple, zeroth order argument is given in the next two paragraphs. A more detailed but still very approximate calculation follows.

For all potentials with large hyperpolarizabilities, the transition matrix elements between the frontier wavefunctions—those that are close to the Fermi level—must be large. When these matrix elements are large, there are very strong constraints on the energy differences between the various states, and hence we can regard these parameters are fixed by the matrix elements and we need only calculate the transition matrix elements to predict the hyperpolarizabilities.

In order to calculate these transition matrix elements approximately, we note that wavefunctions fall rapidly in the classically forbidden region and oscillate in the classically allowed region. Thus most of the transition ma-

trix integral between two states necessarily comes from the region between or very close to the turning point of the lower energy state. Moreover, far from the turning points, each wavefunction becomes relatively small, in effect because the particles are (classically) moving relatively quickly. Thus most of the contribution to these integrals comes from the region close to the turning point of the lower energy wavefunction. From this it is clear that the distance between the turning points of the HOMO is an important parameter.

The wavefunction near a turning point is constrained by the slope of the potential at the turning point. Moreover, the ratio of the amplitudes of the wavefunction at the two turning points is limited by general principles, described below. While the wavefunction of the higher energy state near the turning points of the lower energy states is more free to vary, provided the potential is not too far from linear, they are still largely constrained by the slope of the potential. Thus the slopes of the potential at the turning points of the HOMO also seems to be a very important parameter. From these parameters, it is possible to construct two dimensionless parameters, and it is known from numerical experiments that only two parameters seem to be important to maximizing hyperpolarizabilities for model potentials.

We now provide a more detailed, but nonetheless approximate, analysis. We begin with the sum-over-states formulae for the hyperpolarizabilities. For instance, the off-resonant expression for the second hyperpolarizability is,

$$\gamma_{xxxx} = 2e^4 \left(2 {\sum_{lmn}}' \frac{x_{0l}\bar{x}_{lm}\bar{x}_{mn}x_{n0}}{E_{l0}E_{m0}E_{n0}} - {\sum_{mn}}' x_{0m}^2 x_{0n}^2 \left\{ \frac{1}{E_{m0}^2 E_{n0}} + \frac{1}{E_{n0}^2 E_{m0}} \right\} \right), \tag{A1}$$

which contains three kinds of quantity: the energy level differences $E_{nm} = E_n - E_m$, matrix elements $x_{nm} = \langle n \left| x \right| m \rangle$ and barred quantities $\bar{x}_{nm}$ that contain dipole terms,

$$\bar{x}_{nm} = \begin{cases} \Delta x_{n0} = x_{nn} - x_{00} & , \; n = m \\ x_{nm} & , \; n \neq m \end{cases}. \tag{A2}$$

As is well known, the dipole terms can be eliminated from these expressions using the sum rules[16, 17], leaving only the transition elements x_{nm} and energy level differences E_{nm}. Numerous previous studies, as well as the results above, have shown that only a few states—in fact 2-3—near the Fermi energy contribute significantly to the hyperpolarizabilities. Hence, the question of what is important to the hyperpolarizability maybe be addressed by understanding how the potential and wavefunctions affect these quantities. In the remainder of the appendix, we will answer this question by constructing WKB ansatz wavefunctions, inserting them into expressions for these quantities and examining the form of the results.

In order to proceed, we make some simplifying assumptions. First, we assume that the true optimum potential is sufficiently smooth that the WKB approximation yields good approximations, at least near the turning points, to the relevant wavefunctions, i.e. those near the Fermi surface. This is justified by previous studies[15] that have shown the addition of small rapidly varying perturbations to the potential doesn't affect the hyperpolarizabilities. The ansatz potentials studied in this work, with delta functions and changes in slope at isolated points, all satisfy this criterion.

Second, we shall assume that for wavefunctions near the Fermi energy there are only two classical turning points, except possibly for isolated delta functions in the potential that may violate this rule. We justify this because the presence of multiple turning points would result, at least approximately, in roughly independent particles in the separate classically allowed regions. This would lead to hyperpolarizabilities that grow $\propto N$ rather than $\propto N^{\frac{3}{2}}$ (for β) or $\propto N^2$ (for γ) as the Kuzyk bounds imply. The turning points shall be denoted x_n^l and x_n^r, corresponding to the left and right turning point respectively, and are found by solving $V(x_n^l) = E_n$. In this expression, and hereafter, l indexes the left or right turning point.

Within the above assumptions it is possible to write an ansatz wavefunction for the nth state that is valid except for the region near the classical turning points,

$$\psi_n(x) = \begin{cases} A_n(x)\cos\left[\phi_n(x) - \Delta\phi_n(x)\right], & E > V \\ A_n(x)\exp\left[-\gamma_n^l(x)\right], & E < V \end{cases} \tag{A3}$$

where $E > V$ and $E < V$ refer to the classically allowed and forbidden regions respectively. The functions γ_n^l are defined by,

$$\gamma_n^l(x) = \left| \int_{x_n^l}^{x} \left(V(x') - E_n\right)^{\frac{1}{2}} dx' \right|, \tag{A4}$$

and the function $\phi_n(x)$ is given by,

$$\phi_n(x) = \frac{\pi}{4} + \int_{x_n^<}^{x} \left(E_n - V(x')\right)^{\frac{1}{2}} dx'. \tag{A5}$$

The remaining functions $A_n(x)$ and $\Delta\phi_n(x)$ are smooth and slowly varying, except where the potential has delta functions or sharp changes; nonetheless they are necessarily smoother than the potential. For the smooth potentials considered here, $\Delta\phi(x)$ is of order π or smaller.

$A(x)$ is relatively close to the semiclassical result $A(x) \approx a(x)v(x)^{\frac{1}{2}}$ where $v(x)$ is the classical velocity and $a(x)$ changes only by amounts that are asymptotically small where the WKB approximation is valid, except very close to the turning points. The dependence of $a(x)$ can easily be made more precise by using the uniform asymptotic WKB approximation to the wavefunction[18] though this does little to illuminate the discussion. Moreover, even if WKB is invalid somewhere between the turning points, the magnitude, $|a(x)|$ is, roughly speaking, the rate at which the electron "turns" at the turning point and so is expected to be the same at both turning points. In WKB there is no reflection in classically allowed regions and hence it is obvious that $|a(x)|$ cannot change. More generally, however, any structure between the turning points must have equal magnitude for the transmission and reflection from each side. This implies that $|a(x)|$ must have the same values at both turning points. Moreover, $a(x)$ can be chosen to be real near both turning points, and the ratio of a near the two turning points alternates sign as we increase the energy of the wavefunctions.

The approximate energies are found, as usual, by solving

$$n + \frac{1}{2} = \phi(x_n^>) = \int_{x_{n^l}}^{x_n^r} (E_n - V(x'))^{\frac{1}{2}}\, dx'. \tag{A6}$$

Combining copies of the WKB equation (A6) for two different states n and m, we obtain,

$$n - m = \int_{x_{n^l}}^{x_n^r} (E_n - V(x'))^{\frac{1}{2}}\, dx' - \int_{x_m^l}^{x_m^r} (E_m - V(x'))^{\frac{1}{2}}\, dx'. \tag{A7}$$

These integrals may be combined by making a linear change of variable $x' \to x_p^l + y\Delta x_p$ where $y \in [0,1]$ and $\Delta x_p = x_p^r - x_p^l$.

$$n - m = \int_0^1 dy \left[\Delta x_p \left[E_n - V\left(x_n^l + y\Delta x_p\right)\right]^{\frac{1}{2}} - \Delta x_p \left[E_m - V\left(x_m^l + y\right)\right]^{\frac{1}{2}}\right]. \tag{A8}$$

The square roots in the integrand can be combined by completing the square,

$$n - m = \int_0^1 dy \frac{\Delta x_n^2 \left[E_n - V(\Delta x_n y + x_n^l)\right] - \Delta x_{t,m}^2 \left[E_m - V(\Delta x_m y + x_m^l\right]}{\Delta x_n \left[E_n - V(\Delta x_n y + x_n^l)\right] + \Delta x_{t,m} \left[E_m - V(\Delta x_m y + x_m^l)\right]}. \tag{A9}$$

The form of the integrand in (A9) is instructive: near the turning points, i.e. as $y \to 0$ and $y \to 1$, the numerator goes to zero linearly in y while the denominator goes to zero as a square root; the integrand therefore vanishes like $y^{1/2}$ near the turning points. Hence, the majority contribution to this integral comes from the spatial region far from the classical turning points, particularly if the two energies are similar and the potential near the turning points is slowly varying. Moreover, this integral tends to smooth out small, high frequency variations in the potential; it follows that if $|n - m|$ is small, then $E_n - E_m$ is largely determined by the form of the potential far from the turning points.

We now turn to the dipole matrix elements x_{nm}, which can be computed using the position formula,

$$x_{nm} = \int_{-\infty}^{\infty} dx x \psi_n \psi_m. \tag{A10}$$

We shall restrict our analysis to small values of $|n - m|$, because only these states contribute significantly to the hyperpolarizability. Moreover, approximate analysis of these integrals is complicated for large $|n - m|$ because, while small, they are strongly dependent on the detailed analytic behavior of the potential and wavefunctions.

Substituting the ansatz wavefunction (A3) into the position formula (A10),

$$x_{nm} \approx \int_{x_m^<}^{x_m^>} dx \frac{1}{2} A_n A_m \left[x - \frac{1}{2}(x_{t,m}^r + x_{t,n}^l)\right] \cos(\phi_n - \phi_m + \Delta\phi_m - \Delta\phi_m), \tag{A11}$$

. This integral is then the integral of a function that has $n-m-1$ nodes, roughly evenly spaced through the interval, and has a somewhat (algebraically) larger magnitude near the turning points. From this formula, it is apparent that the transition matrix elements depend primarily on the nature of the wavefunctions near the classical

turning points, i.e. separation between the two turning points and as well as the shape of the wavefunctions in their vicinity. As the shape of the wavefunctions are most dependent on the slopes of the potential at the turning points and there must be a connection between the energy differences and the transition matrix elements, this strongly argues that the separation between the turning points of the HOMO and the slopes of the potential at these turning points should be among the most important heuristic parameters that determine the hyperpolarizabilities.

[1] M. G. Kuzyk, J. Perez-Moreno, and S. Shafei, Physics Reports **529**, 297 (2013).

[2] S. M. Mossman and M. G. Kuzyk, "Optimizing hyperpolarizability through the configuration space of energy spectrum and transition strength spanned by power law potentials," (2013).

[3] D. S. Watkins and M. G. Kuzyk, J. Opt. Soc. Am. B **29**, 1661 (2012).

[4] T. Atherton, J. Lesnefsky, G. Wiggers, and R. Petschek, Journal of the Optical Society of America B: Optical Physics **29**, 513 (2012).

[5] C. J. Burke, T. Atherton, J. Lesnefsky, and R. Petschek, JOSA B **30**, 1438 (2013).

[6] R. Lytel, S. Shafei, J. H. Smith, and M. G. Kuzyk, Phys. Rev. A **87**, 043824 (2013).

[7] R. Lytel and M. G. Kuzyk, Journal of Nonlinear Optical Physics & Materials **22**, 1350041 (2013).

[8] R. Lytel, S. Shafei, and M. G. Kuzyk, Journal of Nonlinear Optical Physics & Materials **23**, 1450025 (2014).

[9] R. Lytel, S. M. Mossman, and M. G. Kuzyk, Journal of Nonlinear Optical Physics and Materials **24**, 1550018 (2015).

[10] D. S. Watkins and M. G. Kuzyk, The Journal of chemical physics **134**, 094109 (2011).

[11] M. G. Kuzyk, Physical review letters **85**, 1218 (2000).

[12] K. Tripathy, J. P. Moreno, M. G. Kuzyk, B. J. Coe, K. Clays, and A. M. Kelley, The Journal of Chemical Physics **121**, 7932 (2004).

[13] J. Zhou, U. B. Szafruga, D. S. Watkins, and M. G. Kuzyk, Phys. Rev. A **76**, 053831 (2007).

[14] M. G. Kuzyk, Opt. Lett. **25**, 1183 (2000).

[15] G. A. Wiggers and R. G. Petschek, Opt. Lett. **32**, 942 (2007).

[16] J. Pérez-Moreno, K. Clays, and M. G. Kuzyk, The Journal of Chemical Physics **128**, 084109 (2008).

[17] M. G. Kuzyk, Phys. Rev. A **72**, 053819 (2005).

[18] C. Bender and S. Orszag, *Advanced Mathematical Methods for Scientists and Engineers I: Asymptotic Methods and Perturbation Theory*, Advanced Mathematical Methods for Scientists and Engineers (Springer, 1999).

Appendix D

Developing a project-based computational physics course grounded in expert practice

In this paper, submitted to the American Journal of Physics, we present a computational physics course which we designed and taught. We explain the process by which we designed the course, using interviews with computational experts to guide us. We outline the structure of the course, present examples of student work and feedback, and perform a study using independent analysis by computational experts to assess student progress throughout the course.

Developing a project-based computational physics course grounded in expert practice

Christopher Burke and Timothy J. Atherton*
Department of Physics and Astronomy, Tufts University, 574 Boston Avenue, Medford, MA 02155

We describe a project-based computational physics course developed using a backwards course design approach. From an initial competency-based model of problem solving in computational physics, we interviewed faculty who use these tools in their own research to determine indicators of expert practice. From these, a rubric was formulated that enabled us to design a course intended to allow students to learn these skills. We also report an initial implementation of the course and, by having the interviewees regrade student work, show that students acquired many of the expert practices identified.

I. INTRODUCTION

Computers have been used to solve physics problems since they were first created[1], and with their ever increasing ubiquity, computing has become a fundamental scientific practice in Physics. Some brief examples[2]: in High Energy Physics, computers automatically select detected events for recording, store them and allow processing of the results on large-scale compute Grids; moreover, theoretical predictions are made by simulations. In Astronomy, formation processes are simulated at multiple length scales from star formation to galaxy evolution. In Condensed Matter Physics, many-body quantum mechanics simulations predict the electronic structure of crystals and molecules, bulk properties from molecular dynamics and continuum behavior by solving PDEs. Moreover, almost since they became available, computers have been used in Physics Education for visualization and simulation[3], and their immense potential as educational tools was recognized early on[4, 5]. This tradition is continued today in projects such as the PhET[6] simulations, designed for classroom use.

Given the central importance of computation to physics research, and hence the need for appropriately trained students and professionals to perform it, it is perhaps surprising that computing is often only weakly integrated into the physics curriculum. Several approaches have been tried to remedy this[7]: At the largest scale, whole majors on Computational Science have been created[8]; courses[9] or sequences of courses[10, 11] have been created within departments; computational projects and homework problems have also been integrated into existing courses and laboratories, both for majors[12] and at the introductory level[13, 14]. Encouragingly, a large-scale survey performed about a decade ago[15], showed widespread support for these efforts. A valuable starting point for those seeking to implement one of these approaches is a special double volume[16] of the American Journal of Physics on Computation published in 2008 and including several of the articles cited above. Of particular practical help is the extensive Resource Letter[17], an updated version of[18].

* timothy.atherton@tufts.edu

In this work, we discuss a Computational Physics class recently created at Tufts University. Central to our vision was a desire to ground the course in actual professional practice, following a movement to refocus science education around real scientific practices[19]. We therefore conducted interviews of scientists, mathematicians and computer scientists to determine these practices. The results of the these interviews were distilled into a rubric that aims to describe indicators of expert practice in computational physics. From the rubric, we used a backwards course design approach to create a sequence of projects that allow students to acquire these practices in a structured manner. We gave the course for the first time in the Spring 2015 semester, and conducted an after-delivery assessment exercise to determine whether student work did indeed contain evidence of expert practice. We report in detail the design process in the hope it might be of use to others planning project-based courses with other subject matter.

The structure of the paper follows our design and assessment exercise: the rubric and its development is described in section II; the course design process is explained in section III; brief comments on the initial implementation are given in section IV; our *a posteriori* analysis of the success is presented in section V. We finish with discussion and conclusions in section VI.

II. RUBRIC DEVELOPMENT

The initial phase of our design process was to develop a rubric for the course, capturing in a single document indicators of expert practice in Computational Physics. The purpose of this was threefold: First, and most obviously, the rubric would be a tool to help assess the quality of student work submitted in the course. Second, by making this document available to students during the course, we wished to convey these excellent practices as a norm and engender a climate of professionalism in the classroom. Finally, we anticipated that by distilling the practices into a rubric, the resulting document would be a powerful tool to assist in the design of the course, ensuring that the content and assessment would be mutually aligned, and that both would prepare students to solve computational problems in actual physics research.

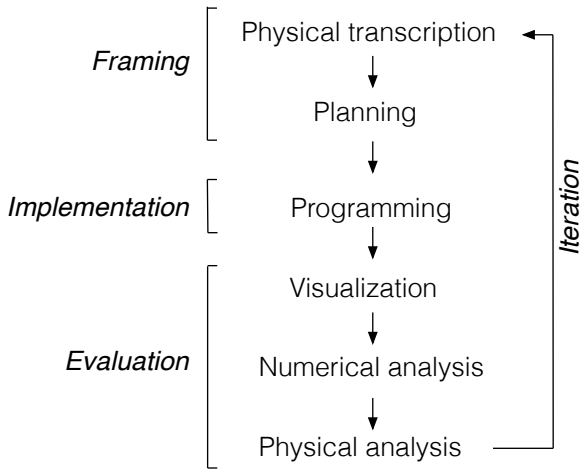


Figure 1. Initial model of expert problem solving in Computational Physics.

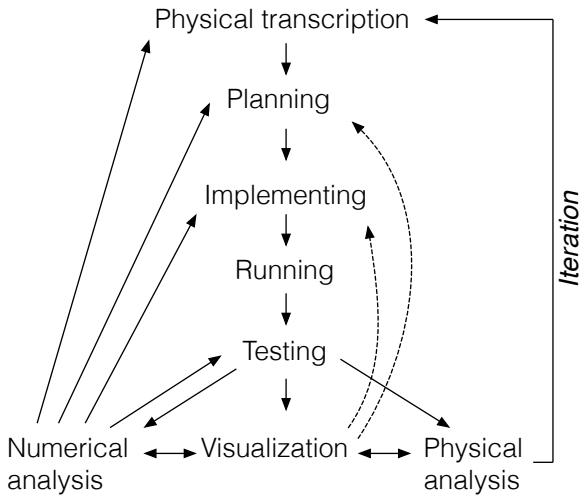


Figure 2. Evidence-based model of expert problem solving in Computational Physics.

We began by determining the typical content of Computational Physics courses. We examined course syllabi from a range of institutions, Boston University, University of Connecticut, East Tennessee State University, Northeastern University and Oregon State University. From these syllabi, and our own experiences as professionals, we proposed an initial set of competencies that seem essential to the practice of computational physics: *Physical transcription*, the formulation of the physical system of interest into a mathematical problem and creation or selection of algorithm(s) to solve it; *Planning*, organizing the overall program into modules and selecting appropriate data structures; *Programming*, implementation of the approach; *Visualization*, post-processing of the results and selection of the appropriate representations to interpret them; *Numerical analysis*, assessing error and stability of the program; *Physical analysis*, relating the results of the program back to the initial problem and determining whether the approach has provided the required insight.

Our initial model of how these competencies were used by experts to solve problems in physics research is shown in figure 1. We anticipated a formulation phase, in which the physical system and research questions of interest were transcribed into a math problem and then an algorithm developed or selected to solve the problem. Following this, the program is implemented. Having run the program, the results are evaluated both to assess whether the results were numerically meaningful, i.e. whether the program gave stable, repeatable results, and physically meaningful, i.e. whether they answered the original research questions. We envisioned that these activities would be repeated iteratively several times, as the initial product may not fully address the original questions, or the results may lead to new questions. This sequence mirrors famous work on problem solving[20] as well as physics-specific research[21]. We particularly note the parallel with the Integrated Problem-Solving model[22, 23], which divides the problem-solving process into *framing, implementation and evaluation*, and was recently adapted to assess student learning in Computational Materials Science[24].

From our initial list of competencies and problem solving model, we devised interview questions (listed in Appendix I) to determine actual professional practices associated with these competencies. We selected five faculty and postdocs from Physics, Computer Science and Mathematics who actively use computation in their research and have published refereed papers involving computation. These five were interviewed by one of the authors (CJB) and these interviews transcribed. The two authors then coded the transcripts separately according to the prompt "identifiable items we could look for in student work". Following this step, the coded items were compared and only common items retained. We then categorized the coded items according to our initial competency model.

In performing this exercise, we identified additional competencies that we had not proposed in our initial model: *Running*, incorporating the ability to create scripts and file structures to execute the program and collect the data, and *Testing*, incorporating comparison of the output of the program against known solutions as well as profiling the program to assess performance. We renamed the competency *Programming* in the initial model to *Implementation* to reflect our inclusion of important non-programming practices associated with implementation, e.g. creation of external documentation and use of version control.

Furthermore, the evidence of the interviews forced us to abandon our initial cyclic model of how Computational Physics programs are created and instead create a more highly connected and iterative model, depicted in fig. 2. Of particular note, we found evidence that ideas from numerical analysis inform almost every as-

pect of computational physics problem solving, and that experts typically use these ideas as a sub-step in each of the Physical Transcription, Planning, Implementing, Testing and Visualization steps. Testing and Numerical Analysis proceed in a closely iterative fashion. The Numerical analysis, Visualization and Physical Analysis steps of evaluating the program output are also highly interconnected and mutually inform the selection of representations to plot and how to interpret the results. Perhaps our most surprising result is that Visualization was shown not to be limited to assessing the program output, but also a key tool to plan and implement effective programs. Overall, the evidence suggests that while experts do indeed solve Computational Physics problems iteratively, the execution of their strategy does not involve a clearly defined sequence of steps but rather invocation of different competencies as required. While the number of interviewees in this study is small, their nonlinear approach to problem solving parallels that seen in other domains[25].

A draft rubric was prepared from the coded items, organized along the new competency model, which seeks to define expert practice through the indicators identified from the interviews. The participants interviewed were invited to give feedback on the rubric draft at a panel, together with representatives from the Tufts Center for the Enhancement of Learning and Teaching (CELT) and Tufts Technology Services Educational Technology group who both possessed relevant expertise in computing education. Following the recommendations of the group, we constructed the final version of the rubric shown in table II.

III. COURSE DESIGN

Having created the rubric described in section II, we set out to design a course in which students could learn the expert practices identified in a structured manner. Following other studies[9], we decided to adopt a project-based approach. Our overall vision for the course was that students would complete a structured sequence of projects to acquire the practices from the rubric in a scaffolded manner. We proposed to spend the majority of the class time—$2\frac{1}{2}$ hours per week—with the students working in groups, guided by the instructor and TA; we would introduce each class with a short "micro lecture" introducing each project and its associated concepts and algorithms. We also anticipated that the students would create a final term project where they would select and formulate the physics problem to be solved themselves.

First, we sought to distill the essence of the rubric further into a small number of overall course goals for students to learn in the course:—

1. Develop mathematical and computational models of physical systems.

2. Design, implement and validate a functioning code for such a model.

3. Understand the role of numerical analysis in formulating, designing and interpreting computer models.

4. Use visualization and physical analysis to test hypotheses.

5. Connect theory, experiment and simulation.

To identify possible projects, we surveyed 17 faculty from the Astronomy, High Energy Physics, Cosmology and Condensed Matter Physics research groups in the Tufts Physics and Astronomy department. For each of the proposed projects—which we supplemented with some of our own—we selected practices from the rubric that might be particularly important. While good programming is a ubiquitous requirement, for example, the Ising model is a suitable problem to study random numbers and issues of repeatability. We then ordered this collection of projects to determine those that could be completed with little specialist knowledge to those requiring more, and identified dependencies between projects.

Following the backward course design approach[26], we selected a subset of projects that allowed us to span the full list of practices and constructed the course plan shown in table IV. Reading column-wise for each project is an estimate of the time required—1 or two weeks—the practices we intended to focus on in the project, the project title, the content of the project as viewed from a traditional course, and any additional skills that may be required for successful completion of the project. From this overall plan, we developed project descriptions, available as supplementary material[27], as well as brief "micro-lectures" to communicate both the content (i.e. algorithms and concepts) and the Learning Objectives (i.e. the professional practices). The problems themselves will be described in more detail in the following section concerning our initial implementation.

IV. IMPLEMENTATION

A. Student backgrounds

1. *Demographics*

An initial implementation of the course ran in the Spring semester of 2015 and was ultimately completed by $n = 22$ students out of 23 registrants. The course contained 7 women and 14 men. Some 13 were Physics or Astronomy majors, 8 were Computer Science majors with some double majoring in both Physics and Computer Science, 3 were graduate students in Physics and Education.

Competency	Indicators
Physical Transcription	Analytical methods are first employed to understand the problem to the greatest extent possible, including identification of symmetries, length scales and timescales. An appropriate mathematical model and computational representation are chosen including choice of algorithm and discretization. Multiple methods are considered and their strengths, limitations and complexity are evaluated using literature where appropriate. Important physical constraints and conserved quantities are identified, and the approximations made are clearly stated. The purpose of the calculation and desired results are clearly articulated.
Planning	The program to be written is broken into modules and functions that can be designed, tested and debugged independently. A suitable and efficient representation of the data, such as classes and data structures, is chosen appropriately for the algorithm. Relevant libraries, software packages and existing code are identified. The need for parallelization is considered and incorporated into the design if necessary. An appropriate language is chosen based on needs of the problem, ease of implementation, maintainability and performance.
Implementation	The code can be easily understood and convinces the reader it works through careful commenting, descriptive variable and function names and validation of input. Coding standards are developed and obeyed amongst the implementation team. Comments document the physics, are in proportion to the complexity of the section, and identify input and output to functions. Outside documentation thoroughly describes how the program works and how to run it such that future users and maintainers can easily work with the program. The code is efficient but not at the expense of readability and maintainability. Professional programming tools, e.g. version control and debuggers are used where available.
Testing	Each element of the program is carefully tested separately and together at each step. The program is verified on test cases with known solutions identified in the planning process. Appropriate metrics are used to analyze performance and identify need for optimization. Visualization is used to provide insight into whether the algorithm is working.
Running	Initial conditions are chosen judiciously. Output is organized and labeled and input parameters used in each run are recorded. Multiple runs, if necessary, are automated efficiently through scripts.
Visualization	Visualization is used to gain intuition regarding the output and to present final results in a compelling way. Important results are visualized so that they require little effort to decipher and are appropriate for the given data. Relevant tools are used to make the visualization. Strengths and limitations of the choice of representation and alternatives are discussed. Visualization, including physical objects, may also be used to display program and data structures if necessary.
Numerical Analysis	The source and nature of all approximations made are identified and their impact on the result discussed. The most significant sources of error are carefully analyzed and estimates of the error are given; ideally these are used to guide the algorithm, e.g. in refining the discrete representation. Conditions for stability are identified and quoted. The scaling of computational time with problem size is understood. Reproducibility is verified for non-deterministic algorithms. Floating point arithmetic is used carefully.
Physical Analysis	Results are compared to hypotheses. Adherence to physical constraints (e.g. energy conservation) is verified. Possible improvements or alternate implementations are identified. The results are cross checked with alternative simulation strategies.

Table II. Rubric based on expert practice.

2. Prior experience

In this section, we discuss the prior experience of students in the class. We first address the issue of programming language. The course design presented above is intentionally language independent because evidence from our interviews indicates that good programming practices are largely independent of the language chosen. Moreover, language selection for a given problem is an important skill. The language chosen should be one used by professional physicists; within the general family of languages familiar to the students; have good quality development tools and have libraries or intrinsic functions to support useful algorithms and visualization.

To aid this decision, and generally guide our preparation, we circulated a brief survey before the course to assess student course preparation in physics, computer science and math as well as programming language familiarity and confidence in programming. The questions are listed in Appendix II. From the 14 responses, we coded student preparation in each subject as Introductory, Intermediate, Advanced or Graduate depending on the highest level course taken by the student. The courses General Physics 1 and 2, Calculus 2, Multivariable Calculus and Introduction to Programming were classified as Introductory. Modern Physics, Linear Algebra, Differential Equations and Data Structures were classified as Intermediate. Any courses with higher numbers were classified as Advanced.

Week	Learning Objectives	Assignments	Content	Additional skills
1-2	Documenting code Testing each component separately Version control	Double pendulum	Discretization — ODEs — Integration	Group work
3-4	Dividing code into functions Identifying sources of error	1D quantum mechanics	Eigensystems — Root finding — Shooting	Reading papers
5	Selecting a model or algorithm Identifying improvements	Model fitting	Optimization — Statistics	Consulting textbook/internet resources
6	Identifying sources of approximation Using visualization to analyze results	Linear combination of atomic orbitals	Linear Algebra	
7-8	Scripting Reproducibility	Ising model of magnetism	Monte Carlo — Random Numbers	
9-10	Identify symmetries / conserved quantities Test physical acceptability of sol'n	Time dependent Schrödinger equation	PDEs	
11-12	Identify existing packages Choose representation of data structures Analyze stability/scaling Analyze and optimize performance	Laplace's equation	Relaxation — Interpolation/Extrapolation Finite Differences Finite Elements	
ongoing	Develop a hypothesis Choose language Use visualization to present a compelling case	Final project		

Table IV. Course plan

	None	Intro.	Inter.	Adv.	Grad.
Physics	0	1	7	4	2
CS	4	6	3	1	0
Math	0	3	7	4	0

Table V. Subject preparation.

		Physics		
		Intro.	Inter.	Adv.
CS	None	0	1	3
	Intro.	1	3	2
	Inter	0	2	1
	Adv..	0	1	0

Table VI. Correlations between subject preparation in Physics and Computer Science.

The results for subject preparation are shown in table V. Most obviously, students were best prepared in Physics, but only slightly less well prepared in Math. While most students had some familiarity with Computer Science, it is very clear that preparation in this subject was weaker than the other two subjects. A table of student numbers sorted by both Physics and Computer Science preparation, displayed in table VI, shows that the majority of students had an intermediate or better preparation in Physics and only introductory or no preparation in computer science. A small minority, however, had an intermediate or better preparation in in both subjects. No students had taken advanced or graduate classes in both subjects.

Language	Familiar	Confident
C++	10	8
Mathematica	9	4
Python	9	4
Java	6	3
MATLAB	5	3
FORTRAN	1	0

Table VII. Language familiarity.

Languages with which students expressed familiarity or confidence are shown in table VII. The list is somewhat abbreviated: exactly one student each expressed familiarity with Javascript, Lua, Objective-C, Ruby, Scheme, SQL, Swift and Visual Basic. All respondents self-assessed their own experience with programming as familiar or better. Interestingly, *all* students provided concrete examples of programming tasks they had completed outside Introductory Computer Science classes, referring to data analysis, simulations and programs written for research purposes.

With this information, we decided to use both *Mathematica* and Python for the course as these were both near the top of languages students were already familiar with. Both of these languages are high level, easy to

Preferred group size	No. respondents
2	4
3	6
4	2
5	0

Table VIII. Preferred group size.

learn and provide rich visualization tools; *Mathematica* also provides access to many useful algorithms. Because *Mathematica* is better suited to short programs and has the shallower learning curve, it was used in the first half of the class; we then switched to Python for the Ising model and subsequent projects, except for the project on Laplace's equation. Additionally, we allowed students to use other languages if they wished, though this option was only utilized for the final projects. By giving students the experience of using more than one language, we were able to explicitly compare and contrast their strengths and weaknesses.

B. Logistics

In this section, we describe our approach to organizational matters. The class met twice a week for $1\frac{1}{4}$ hour sessions. As noted above, the typical class structure involved short presentations by the instructor on the project physics background, a numerical analysis topic, or one of the competencies identified in the rubric. *Numerical Recipes*[28] was recommended as a textbook, and was used by students as a reference for standard algorithms. Support was provided through portions of class time dedicated to student group work during which the instructor and TA were available to help troubleshoot issues, as well as traditional office hours. A *Piazza* forum was also provided for the course which received 2-3 posts per week. Questions on the forum were typically answered quickly, on average within half an hour. Students tended to use the forum to share links to helpful resources, organize groups for the final project, and pose syntax questions.

A key issue in designing the class was the choice of group size. Following generic advice, the first two projects listed below were executed in teams of 4 or 5; project 3 was performed in pairs; project 4 was performed in teams of 4. At this point in the semester, we circulated a mid-semester survey with one question that asked students what they thought the optimal group size was based on their experiences. We had 12 responses, which are summarized in table VIII. Students were also asked to explain the reasons for their selection, which revealed two opposing effects: logistics, such as finding convenient times to meet physically and synchronizing efforts favor smaller group sizes while division of labor tends to favor larger group sizes. However, students reported that in larger groups, there tended to be people who didn't contribute significantly. We therefore concluded that groups of 3 were preferable for this course and used this size for the remainder of the semester with occasional groups of 4 where required to accommodate the whole class.

Another important matter is the issue of grading in a team-based class. As described above in section III, the instructors gradually scaffolded in the competencies during the course and this was reflected in the grading. For example, the first project was graded primarily on the quality of the code as well as successful completion of the project. A key mechanism we established to provide accountability was to require short individual self-assessments to be completed after each project online through Tufts' Course Management System (Sakai). The questions asked are listed in appendix III. These submissions allowed the instructors to understand who had done what as well as where students felt their own work stood in comparison to the grading criteria. Project submissions were graded as a team, but each student received an individual grade and feedback form—collated and emailed automatically by a *Mathematica* script—including their own self-assessment as a component. The instructor occasionally modified grades if there were sufficient evidence that an individual had failed to contribute meaningfully to a project, though this was used very sparingly.

An overview of the projects, as well as the project descriptions themselves, is provided in supplementary material[27]. A final project, performed in parallel with the last third of the class, was intended as the summative assessment, allowing students to attempt a problem of their own choosing and replacing the final exam of a typical class. A list of project ideas was provided, largely connecting with TJAs research in Condensed Matter Physics, but included the option to formulate a problem that connected with student's research field. Some students chose to work individually, others in teams of up to four. The class concluded with a mini research symposium where each group gave a short five minute presentation showcasing their findings. Each student was required to fill out a self-assessment different from the other projects, asking instead how the project demonstrated their professional practice in each competency.

C. Student work

We present two brief case studies of student work that illustrate some of the ways groups creatively engaged with the projects and used visualization to analyze the performance of their program. The first example comes from project 1, the Double Pendulum. In this project, each group implemented a different algorithm, and the project was performed in two stages: first, students used their assigned algorithm to solve the simple harmonic oscillator equations and presented their findings to the class, afterwards they extended their code to solve the double pendulum. The numerical analysis concepts of Error, Order and Stability were introduced in

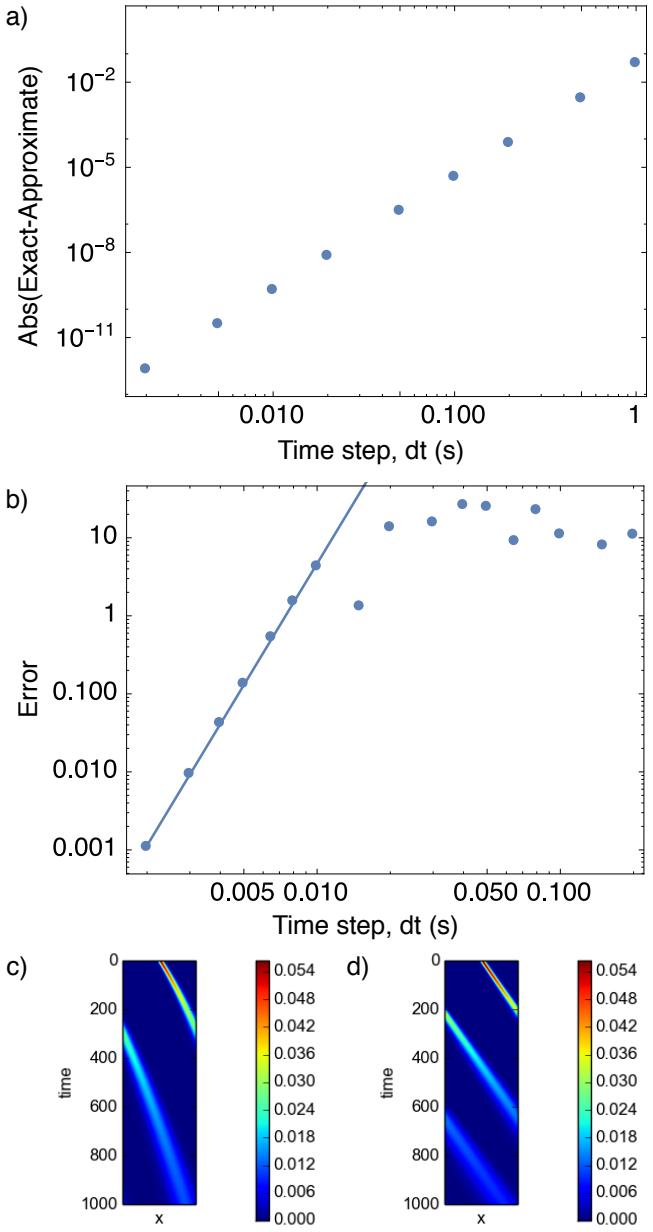


Figure 3. Examples of student-generated visualizations. (a) Finding the dependence of the error in a Runge-Kutta integration of the simple pendulum; (b) a similar plot for the double pendulum. (c) and (d) Comparing the performance of two different algorithms for integrating the TISE.

class, but only the performance of the Euler algorithm was discussed in detail. The group that used the Runge-Kutta algorithm spontaneously conducted an analysis of the discrepancy between their numerical solution and the analytical solution as a function of time-step. They presented the graph shown in fig. 3(a) to the class after the first part of the project, and used the `FindFit` function of *Mathematica* to verify that the error was $\propto (\Delta t)^4$ as expected. In the second part of the project, they created a similar plot [fig. 3(b)], comparing the solution at a specified Δt to that with much smaller $\Delta t = 0.0005$. The graph was annotated in their notebook by the comment,

> *"show max error and the convergence fit for small dt, which is actually of order 5 in this case. Note that the error is considerably larger, even at small dts, than the more well-behaved single pendulum, and also that the chaos in the system is very evident at larger dt"*

indicating that this group had appreciated that the Runge-Kutta algorithm was performing as expected in some regime of Δt, but failing in others, and that the highly nonlinear double pendulum equations were more challenging to solve.

The second example comes from project 6, the Time Dependent Schrödinger Equation, in which students solved a number of classic 1D problems. For their report, most groups presented snapshots of the wavefunctions at different times. One group, however, utilized a kymograph representation to show both spatial and temporal evolution. Two examples taken from their report are shown in fig. 3(c) and (d), both showing the propagation of a gaussian wavepacket as a function of time. Periodic boundary conditions are enforced, so the wavepacket exits from the right and re-enters from the left during the simulation. The two figures represent different algorithms, fig. 3(c) is a non-unitary cell-centered finite difference scheme; fig. 3(d) is a unitary scheme. Paired with these figures was the text,

> *"By applying periodic boundary conditions, we see the particle exit the frame on the right and reappear on the left, as expected. The results from the two different algorithms are shown in Figure X. Notice how the first algorithm gives undesirable results, such as dissipation and a non-constant velocity of the wavepacket. The second algorithm, however, gives the expected results."*

showing that the students selected this representation to make such comparisons straightforward. The wavepacket in fig. 3(c) clearly changes velocity as the slope of the line changes. The students went on to use this representation to verify time-invariant behavior, i.e. eigenfunctions, as well as reflection from a barrier and crystal.

V. ANALYSIS

In this section, we present feedback on the initial implementation of the course from several sources. An informal feedback session was held during class and students completed the usual course evaluation required by the university. Furthermore, we asked the experts interviewed to develop the rubric in section II to regrade student work to determine whether it aligned with their conception of professional practice.

A. Informal feedback

Students in the final class were asked to collaboratively construct ideas for things to do differently in future classes. First among these were the course requirements, which were kept deliberately minimal in the first iteration. Students thought that the requirements should be Modern Physics due to the presence of Quantum Mechanics projects, an intermediate math class such as Linear Algebra and the second Computer Science class, i.e. Data Structures. We note that many, but not all, students would have met these requirements given the backgrounds presented in section IV A. A second important issue was the level of course credit and class time. The course was initially offered as a Special Topics class worth 1 course credit (equivalent to 3 Semester Credit Hours); students felt that $1\frac{1}{3}$ or $1\frac{1}{2}$ might be more appropriate given the workload, similar to a lab course, and that additional class time e.g. through an associated recitation would be valuable. When prompted by the instructor, students preferred to increase the course credit rather than delete projects. It was felt that an opportunity to cross-examine other groups work would be helpful, first by having students upload their work to a repository, and second by incorporating an after project de-brief. Finally, students requested more structure and scaffolding in the initial projects, including internal deadlines and clearer stages. Following the class, several students noted that they had continued to work on projects initiated in the class; we are aware of at least one project that resulted in a journal publication.

B. Course evaluations

The course evaluation was completed by 15 students. Tufts course evaluation questions use a 5 point scale (Very poor—Less than satisfactory—Satisfactory—Very good—Excellent). Students' overall evaluation of the course was a mean of 4.47 corresponding to Excellent and compared to a departmental average for undergraduate major courses of 4.1, weighted by enrollment. The mean response to the question "How would you rate the success of the course in accomplishing its objectives as stated on the course syllabus" was 4.40 and "How would you rate the use of out-of-class activities to promote your learning" yielded a mean of 4.20, indicating that the project-based approach was well received. Of mild concern is that students' responses to a question on the use of in-class time received a lower mean score of 3.73, suggesting that some reorganization of this aspect might be necessary.

Written response questions provided more detail to these views. In response to the question, "In what ways has this course made you think differently or more deeply?", most students commented on the subject matter, for example,

> *"this class gave me a glimpse of many different types of interesting physics problems. The range of problems tackled during the class gave me a sense of the range of applications of computation in physics"*
>
> *"The projects were felt 'real'. They were non-trivial and felt like the kinds of problems a real computational physicist might work on"*

or the problem-solving aspects,

> *"It was really amazing to get to learn about this approach to solving problems—how to approximate solutions, how to tell how accurate the approximation is, and how to make physics so much more accessible with computers"*
>
> *"The class made me think more deeply about computation. Instead of being content to write code that gets something done, I think about how the code is getting something done and what methods will be best to reduce error, runtime, etc. This is an extremely useful and marketable class. I could have used this ages ago."*

but also group work,

> *"Combining coding skills with communication skills was a new and different task for me."*
>
> *"I am better at working in groups and I now have an idea of what it is like to do computational physics as a profession."*

In response to "What aspects of this course worked best to facilitate your learning?", students universally mentioned the projects, which were *"super interesting and well thought out"*. However, positive experiences with group work was also a theme, in spite of the challenges,

> *"Group work was great. I say this in spite of my complaints about certain group members. Students were forced, in these projects, to go out and discover how to make a certain approach work. This led to strong engagement and lasting impression."*
>
> *"Different groups worked together in different ways, and not all of them were great. But I feel like I really know and like a ton of the people in the class because we were together through the hardest of times. I also feel like I was able to learn a lot of different things from everyone I worked with."*

The final question on how to improve the course provoked a great deal of responses. The vast majority of these repeated topics in the in-class discussion, including the need for prerequisite courses, additional class time and managing group size. Many of the remainder touched on more delicate themes about group work. Challenges

identified included that *"there was a TON of variance in ability"* as well as others referring to differing levels of effort put in by some students. One student felt that,

> *"[The instructor should] more clearly adjust grades of students who don't participate equally in groups. I have a sense that it was 2-3 students who were consistently bad group members, but it was a definite problem. Without the grade incentive, I honestly think that there's a portion of students who will just not do the work and coast along."*

Another student suggested additional structure for group work,

> *"[People] need to be taught time management if they are going to be working in groups. In order to do this, I think that you should have intermediate deadlines for the first few projects (not just the first one), and you should require that students turn in a work plan (including internal deadlines) before they do anything else for each project. Without an authority endorsing this sort of behavior, responsible group members are unlikely to be able to hold other group members accountable."*

Several students also mentioned the need for additional supporting resources,

> *"I would make sure those of us with less experience have clear resources for information. It seemed like I had no way of figuring out what to do or where to turn sometimes."*
>
> *"[I'd like to see] Reading recommendations to get necessary mathematical background"*
>
> *"I would provide some more background information for people who may not have extensive math or programming background. The* Mathematica *and python tutorials that were posted were really helpful, so more online resources or 'crash course' type things for some of the computational and mathematical methods we used would have been great to get everyone on the same page."*

Overall, feedback from students in both formal and informal contexts suggests that the course was a success, that this form of pedagogy is effective, if intensive, but needs additional support to accommodate students with different course backgrounds as well as more structure in managing group dynamics.

C. Post course grading study

From the initial implementation of the course, we wished to determine whether student work produced in the class satisfied the views on professional practice held by the experts whom we interviewed to create the rubric. Moreover, we wished to establish whether there was evidence that these were acquired during the class itself.

	Submission A					Submission B				
	1	2	3	4	5	1	2	3	4	5
Phys. Trans.	F	F	F	P	P	G	G	G	G	G
Planning	—	G	G	G	F	G	G	G	G	G
Implementation	—	F	F	G	F	G	F	G	G	G
Testing	F	—	—	G	—	G	—	—	G	G
Running	F	P	—	P	F	G	G	G	G	G
Visualization	P	F	F	—	—	G	G	G	G	G
Num. Ana.	P	—	P	P	P	F	G	G	G	G
Phy. Ana.	P	P	P	P	—	G	G	G	G	G

Table IX. Comparison of expert grading of two projects.

To test inter-grader consistency, we selected two submissions from project 5 that the we felt displayed a large difference in quality, receiving a C grade and an A grade respectively. We asked the experts to grade these two projects, for each competency assigning a label from the categorizations Good—Fair—Poor as well as Insufficient Evidence. The results are shown in table IX. Submissions A and B were clearly distinguished from one another in quality by the graders, but the degree of consistency is different between the two. To find out why, we held a panel with four of the experts in attendance and invited them to discuss their rationale where there seemed to be discrepancies. In discussion, it seemed that most of the difference between graders could be attributed to whether or not they looked at the code to make their distinction or just looked at the report. For example, grader 4 felt that the quality of the implementation for submission A was good, having looked at the code and noted extensive commenting. The other graders had given more weight to the quality of the report. The experts all agreed that submission B represented work of a high quality. In the panel, we asked all the graders whether they felt that submission B accorded with their own vision of expert practice and they answered affirmatively. We therefore established that at least some students displayed evidence of expert practice as articulated by our rubric, and determined additional guidance necessary to promote inter-grader consistency.

To track the work longitudinally, we amended our grading instructions to require examination of the code and asked the graders to grade other projects from the course. Project 3 involving data fitting was not graded because it was substantially shorter and more prescriptive than the other projects. The projects were stripped of information that identified their ordering, and randomly assigned to different graders; each submission was graded by a single grader, and each competency was graded on the same Good—Fair—Poor scale above. In total, 24 submissions from possible total of 49 were re-graded.

From these, a mean grade was calculated for each

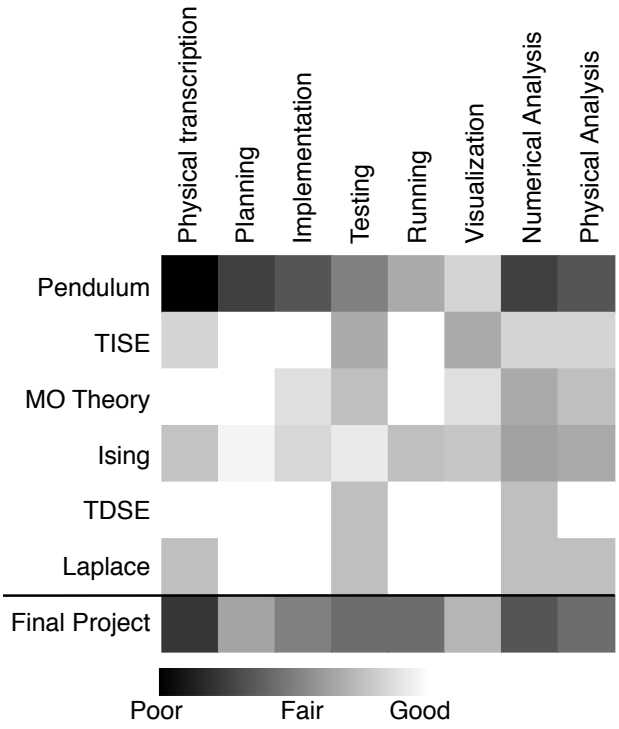


Figure 4. Post-course grading study. Project submissions were regraded after the class by the expert participants

project that is displayed in the visualization in fig. 4. This is intended to provide a snapshot of how well students acquired the competencies during the class. Some trends are immediately clear: happily, it appears that after the initial project, the quality of work rapidly improved, and continued to improve up to the last class project (Laplace's equation). The very swift improvement after the first project suggests that the students already possessed relevant abilities—whatever they were—to do this kind of work, but needed to find out how to apply them. It is possible that the presence of the rubric helped this.

There are differences between the skills: some such as Planning, Implementation and Running seem to have been picked up without trouble. Visualization improved substantially later in the class. Two skills, however, Testing and Numerical Analysis seem to have presented more difficulty. It is important to be careful in interpreting this: it is not necessarily the case that students failed to test their code adequately, though this may be true, but rather that the evidence of testing was insufficiently detailed to the graders. Physical transcription and Physical analysis were also somewhat inconsistent, suggesting that the more abstract competencies are more challenging for students to acquire. In future iterations of the class, additional scaffolding and emphasis should be introduced to address these competencies. It is also clear that the final projects were substantially weaker, in the eyes of the graders, than most of the class projects. In part, this is because the write-ups for these projects were kept very short; it is also inevitable that formulating and executing a project is a much more complex task than solving a prescribed problem. While the instructors were very pleased with the results of the final projects, it appears that in future classes additional time, support and reporting requirements should be introduced to improve the final projects.

VI. CONCLUSION

We present a project-based Computational Physics Course that was designed around expert practice obtained from interviewing faculty and distilled into a rubric that summarizes identifiable features of this practice. A successful initial implementation of the course, together with extensive post course analysis, was also presented as well as matters to consider in future iterations of the course. By having the expert interviewees grade student work produced in the class, it is clear they found evidence of excellent practices that were acquired by students in the class. It is furthermore clear that the rubric can be used as a tool to separate a strong submission from a weak one.

Our work builds upon previous studies of Computing in Physics Undergraduate Education and proposed course structures by providing an concise version of what constitutes expert practice in this discipline, a tool to measure it and a reference implementation. Due in part to the small size of this study, there are remain numerous avenues for further investigation. For instance, a key limitation is that expert practices were self reported and not directly observed. Even so, our interviews complicate simplistic models of expert problem solving originally formulated for introductory classes. At the very least, such models may simply be inapplicable to complex research-oriented problems such as those attempted by students in this class. There is a clear need for more realistic models, of which that presented in fig. 2 is likely to prove a crude initial sketch. Of course, a scaled-up multi-class multi-university study on project-based approaches to Computational Physics is likely to prove deeply informative, particularly since similar techniques are now being tried in introductory classes. Given the central importance of group dynamics to this strategy, which we have shown to be both challenging and rewarding, micro-level approaches such as video observation and interviews with students might help untease how students participate in group work in this context. Our work also provides a valuable example of backwards course design that may be applicable in physics to domains other than computation, such as laboratories or graduate classes focussed on preparing students for research.

While our study has shown that students taking a single course on Computational Physics *can* acquire expert practice, the broader question of the role of Computation in Physics Education remains open. Is it best to have a small number of courses such as the one presented that fo-

cus primarily on Computation, is it better to intersperse Computation into many classes, or is a mixed approach needed? Identifying strategies that produce graduates and PhD students well prepared to perform computational work that meets the needs of future research, industry and other careers is surely of considerable benefit to the community of Physicists.

ACKNOWLEDGMENTS

The authors thank Prof. Michelle Wilkerson as well as Sheryl Barnes and Dona Qualters for much invaluable expertise in helping us design and conduct the study, and Profs. Roger Tobin and David Hammer as well as Anna Phillips for providing thoughtful advice on the manuscript. We are grateful to Tufts University for a Tufts Innovates! seed grant to create and implement the course. CJB acknowledges funding from the Graduate Institute For Teaching (GIFT) program at Tufts University that partially supported his participation. TJA is supported by a Cottrell Scholar Award from the Research Corporation for Science Advancement.

I. APPENDIX: EXPERT INTERVIEW QUESTIONS

The following sequence of questions was asked in the expert interviews.

1. Can you think of a problem in your research area that might be suitable for an undergraduate computational physics course?
2. Could you explain the physics of the problem?
3. What are the possible computational approaches to solve this problem?
4. What are the advantages and disadvantages of these approaches?
5. Can you tell me what you see as the issues involved in transforming a physics problem into an algorithm or set of equations to solve?
6. Thinking about the problem from earlier, how would you go about planning a computer program to solve this problem?
7. Could you sketch a flowchart for this program?
8. I'm going to show you a page of code. Tell me what you like about the programming, and what things would you change?
9. What do you think constitutes good programming?
10. What sort of visualization strategies do you think are useful? Which do you use and what challenges do you come across?
11. What sort of visualization might you use for the example problem?
12. Using numerical analysis strategies, for example thinking about error, order, and stability, how would you assess the numerical performance of your algorithm?
13. More generally, what types of numerical analysis do you use to analyze the algorithms you employ in your work?
14. How would you use the results of your hypothetical program to perform a physical analysis of the system you are modeling?
15. What else might you consider while trying to come to physical conclusions from a set of computational results?

II. APPENDIX: STUDENT BACKGROUND SURVEY QUESTIONS

This short survey was circulated before the course.

1. What Physics Courses have you taken?
2. What Computer Science or Applied/Numerical Math courses have you taken?
3. What Computer Languages have you used? Which do you feel confident in?
4. Describe your overall experience and familiarity with programming. Feel free to cite projects you've undertaken.

III. APPENDIX: INDIVIDUAL SELF ASSESSMENTS

Self assessments were required after each project.

1. Describe your contribution to the project. Identify things that you yourself did.
2. Overall, what grade would you give to your own contribution to the project? *(A—Mastery. I think I did this to a professional level; B—Solid understanding. I got this, though there may be still residual mistakes; C—Progress. I'm still working on learning this.)*
3. How well did your team achieve the goals of the project? Explain briefly each member's contribution. Identify any challenges your team faced and how you overcame them.
4. Overall, what grade to your team's project submission as a whole? *(A—Mastery. I think I did this to a professional level; B—Solid understanding. I got this, though there may be still residual mistakes; C—Progress. I'm still working on learning this.)*

5. Did your team do anything over and above that required in the project description?

6. If you have other comments on your group's project, please write them here.

---

[1] Thomas Haigh, Mark Priestley, and Crispin Rope. *ENIAC in Action: Making and Remaking the Modern Computer*. History of Computing. MIT Press, 2016.

[2] Bill McCurdy, David Ceperly, Steve Koonin, Jeremiah Ostriker, Dan Reed, NCSA John Wilkins, and Paul Woodward. Computation as a tool for discovery in physics. *National Science Foundation, Arlington, Virginia*, 2002.

[3] Abraham Goldberg, Harry M. Schey, and Judah L. Schwartz. Computer-generated motion pictures of one-dimensional quantum-mechanical transmission and reflection phenomena. *American Journal of Physics*, 35(3):177–186, 1967.

[4] Alfred Bork. Interactive learning: Millikan lecture, american association of physics teachers, london, ontario, june, 1978. *American Journal of Physics*, 47(1):5–10, 1979.

[5] Alfred Bork. Special feature learning with computer simulations. *Computer*, (10):75–84, 1979.

[6] Katherine Perkins, Wendy Adams, Michael Dubson, Noah Finkelstein, Sam Reid, Carl Wieman, and Ron LeMaster. Phet: Interactive simulations for teaching and learning physics. *The Physics Teacher*, 44(1):18–23, 2006.

[7] Norman Chonacky and David Winch. Integrating computation into the undergraduate curriculum: A vision and guidelines for future developments. *American Journal of Physics*, 76(4):327–333, 2008.

[8] Rubin Landau. Computational physics: A better model for physics education? *Computing in Science Engineering*, 8(5):22–30, 2006.

[9] Claudio Rebbi. A project-oriented course in computational physics: Algorithms, parallel computing, and graphics. *American Journal of Physics*, 76(4):314–320, 2008.

[10] Ross L Spencer. Teaching computational physics as a laboratory sequence. *American journal of physics*, 73(2):151–153, 2005.

[11] David M Cook. Computation in undergraduate physics: The lawrence approach. *American Journal of Physics*, 76(4):321–326, 2008.

[12] Marcos D Caballero and Steven J Pollock. A model for incorporating computation without changing the course: An example from middle-division classical mechanics. *American Journal of Physics*, 82(3):231–237, 2014.

[13] Ruxandra M Serbanescu, Paul J Kushner, and Sabine Stanley. Putting computation on a par with experiments and theory in the undergraduate physics curriculum. *American Journal of Physics*, 79(9):919–924, 2011.

[14] Marcos D Caballero, Matthew A Kohlmyer, and Michael F Schatz. Implementing and assessing computational modeling in introductory mechanics. *Physical Review Special Topics-Physics Education Research*, 8(2):020106, 2012.

[15] Robert G Fuller. Numerical computations in us undergraduate physics courses. *Computing in science & engineering*, 8(5):16–21, 2006.

[16] Wolfgang Christian, Bradley Ambrose, et al. An introduction to the theme double-issue. *American Journal of Physics*, 76(4):293–295, 2008.

[17] Rubin H Landau. Resource letter cp-2: Computational physics. *American Journal of Physics*, 76(4):296–306, 2008.

[18] Paul L DeVries. Resource letter cp-1: Computational physics. *American Journal of Physics*, 64(4):364–368, 1996.

[19] Melanie M Cooper, Marcos D Caballero, Diane Ebert-May, Cori L Fata-Hartley, Sarah E Jardeleza, Joseph S Krajcik, James T Laverty, Rebecca L Matz, Lynmarie A Posey, and Sonia M Underwood. Challenge faculty to transform stem learning. *Science*, 350(6258):281–282, 2015.

[20] G. Pólya. *How to Solve it: A New Aspect of Mathematical Method*. Anchor Books: Science. Doubleday, 1957.

[21] Leonardo Hsu, Eric Brewe, Thomas M Foster, and Kathleen A Harper. Resource letter rps-1: Research in problem solving. *American Journal of Physics*, 72(9):1147–1156, 2004.

[22] Jonna Kulikowich Thomas Litzinger, Monica Wright and Peggy Van Meter. A cognitive study of modeling during problem solving. In *2006 Annual Conference & Exposition*, Chicago, Illinois, June 2006. ASEE Conferences.

[23] Thomas A Litzinger, Peggy Van Meter, Carla M Firetto, Lucas J Passmore, Christine B Masters, Stephen R Turns, Gary L Gray, Francesco Costanzo, and Sarah E Zappe. A cognitive study of problem solving in statics. *Journal of Engineering Education*, 99(4):337–353, 2010.

[24] Camilo Vieira, Alejandra J. Magana, Anindya Roy, Michael L. Falk, and Jr. Reese, Michael J. Exploring undergraduate students' computational literacy in the context of problem solving. volume 122nd ASEE Annual Conference and Exposition: Making Value for Society, 2015.

[25] Committee on Developments in the Science of Learning with additional material from the Committee on Learning Research and Educational Practice and Board on Behavioral, Cognitive, and Sensory Sciences; Division of Behavioral and Social Sciences and Education; National Research Council. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition*. National Research Council. National Academies Press, 2000.

[26] Grant P. Wiggins and Jay McTighe. *Understanding by Design*. Gale Reference. Association for Supervision and Curriculum Development, 2005.

[27] Supplementary Material. [URL will be inserted].

[28] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes 3rd Edition: The Art of Scientific Computing*. Cambridge University Press, New York, NY, USA, 3 edition, 2007.

Bibliography

[1] Claas de Groot, Ronald Peikert, and Diethelm Würtz. The Optimal Packing of Ten Equal Circles in a Square. Technical Report IPS Research Report No. 90-12, ETH Zurich, 1990.

[2] A Pérez-Garrido, MJW Dodgson, and MA Moore. Influence of dislocations in Thomson's problem. *Physical Review B*, 56(7):3640–3643, 1997.

[3] Atsuyuki Okabe, Barry Boots, Kokichi Sugihara, and Sung Nok Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. Wiley, 2 edition, 2000.

[4] DR Nelson and BI Halperin. Dislocation-mediated melting in two dimensions. *Physical Review B*, 19(5), 1979.

[5] M. Kleman and Oleg D. Lavrentovich. Topological point defects in nematic liquid crystals. *Philosophical Magazine*, 86(25-26):4117–4137, 2006.

[6] Gregory M. Grason. Topological defects in twisted bundles of two-dimensionally ordered filaments. *Physical Review Letters*, 105(4):1–4, 2010.

[7] Luca Giomi, Mark J. Bowick, Xu Ma, and M. Cristina Marchetti. Defect annihilation and proliferation in active Nematics. *Physical Review Letters*, 110(22):1–5, 2013.

[8] HS Seung and DR Nelson. Defects in flexible membranes with crystalline order. *Physical Review A*, 38(2):1005–1018, 1988.

[9] Mark Bowick, David Nelson, and Alex Travesset. Interacting topological defects on frozen topographies. *Physical Review B*, 62(13):8738–8751, October 2000.

[10] 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 (New York, N.Y.)*, 299(5613):1716–1718, March 2003.

[11] Vincenzo Vitelli, JB Lucks, and DR Nelson. Crystallography on curved surfaces. *Proceedings of the National Academy of Sciences of the United States of America*, 103(33):12323–12328, August 2006.

[12] Luca Giomi and Mark Bowick. Crystalline order on Riemannian manifolds with variable Gaussian curvature and boundary. *Physical Review B*, 76(5):054106, August 2007.

[13] Luca Giomi and Mark Bowick. Defective ground states of toroidal crystals. *Physical Review E*, 78(1):010601, July 2008.

[14] William T M Irvine, Vincenzo Vitelli, and Paul M Chaikin. Pleats in crystals on curved surfaces. *Nature*, 468(7326):947–951, December 2010.

[15] Spencer Umfreville Pickering. Emulsions. *Journal of the Chemical Society*, 91:2001–2021, 1907.

[16] Dmitry Ershov, Joris Sprakel, Jeroen Appel, Martien A. Cohen Stuart, and Jasper van der Gucht. Capillarity-induced ordering of spherical colloids on an interface with anisotropic curvature. *Proceedings of the Na-*

tional Academy of Sciences of the United States of America, 110(23):9220–4, 2013.

[17] Hsin-Ling Cheng and Sachin S Velankar. Controlled jamming of particle-laden interfaces using a spinning drop tensiometer. *Langmuir : the ACS journal of surfaces and colloids*, 25(8):4412–20, April 2009.

[18] M Cui, T Emrick, and TP Russell. Stabilizing Liquid Drops in Nonequilibrium Shapes by the Interfacial Jamming of Nanoparticles. *Science*, 342(October):460–463, 2013.

[19] Amar B. Pawar, Marco Caggioni, Roja Ergun, Richard W. Hartel, and Patrick T. Spicer. Arrested coalescence in Pickering emulsions. *Soft Matter*, 7(17):7710–7716, 2011.

[20] Andrea J. Liu and Sidney R. Nagel. The Jamming Transition and the Marginally Jammed Solid. *Annual Review of Condensed Matter Physics*, 1(1):347–369, August 2010.

[21] S. Torquato and F. H. Stillinger. Jammed hard-particle packings: From Kepler to Bernal and beyond. *Reviews of Modern Physics*, 82(3):2633–2672, September 2010.

[22] Aleksandar Donev, Robert Connelly, Frank H. Stillinger, and Salvatore Torquato. Underconstrained jammed packings of nonspherical hard particles: Ellipses and ellipsoids. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 75(5):1–32, 2007.

[23] Corey S. O'Hern, Leonardo E. Silbert, Andrea J. Liu, and Sidney R. Nagel. Jamming at zero temperature and zero applied stress: The epitome of disorder. *Physical Review E*, 68(1):011306, jul 2003.

[24] S. Torquato and F. H. Stillinger. Multiplicity of Generation, Selection, and Classification Procedures for Jammed Hard-Particle Packings. *Journal of Physical Chemistry*, 105:11849–11853, 2001.

[25] Z. W. Salsburg and W. W. Wood. Equation of State of Classical Hard Spheres at High Density. *The Journal of Chemical Physics*, 37(4):798, 1962.

[26] Aleksandar Donev, Salvatore Torquato, Frank H. Stillinger, and Robert Connelly. A linear programming algorithm to test for jamming in hard-sphere packings. *Journal of Computational Physics*, 197(1):139–166, 2004.

[27] Thomas Hales. A proof of the Kepler conjecture. *Annals of Mathematics*, 162(3):1065–1185, nov 2005.

[28] Aleksandar Donev. Jamming in hard sphere and disk packings. *Journal of Applied Physics*, 95(3):989–999, 2004.

[29] Randall Kamien and Andrea Liu. Why is Random Close Packing Reproducible? *Physical Review Letters*, 99(15):155501, oct 2007.

[30] S. Torquato, T. M. Truskett, and P. G. Debenedetti. Is random close packing of spheres well defined? *Physical review letters*, 84(10):2064–7, mar 2000.

[31] DA Wood, CD Santangelo, and AD Dinsmore. Self-assembly on a cylinder: a model system for understanding the constraint of commensurability. *Soft Matter*, 9:10016–10024, 2013.

[32] BD Lubachevsky and FH Stillinger. Geometric properties of random disk packings. *Journal of Statistical Physics*, 60(12):561–583, 1990.

[33] S. Torquato and Y. Jiao. Robust algorithm to generate a diverse class of dense disordered and ordered sphere packings via linear programming. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 82(6):1–14, 2010.

[34] Brian P. Flannery, Saul Teukolsky, William H. Press, and William T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 2007.

[35] Fausto Bernardini, Joshua Mittleman, Holly Rushmeier, Claudio Silva, and Gabriel Taubin. The ball-pivoting algorithm for surface reconstruction. *IEEE Transactions on Visualization and Computer Graphics*, 5(4):349–359, 1999.

[36] David R. Nelson. Toward a Tetravalent Chemistry of Colloids. *Nano Letters*, 2(10):1125–1129, October 2002.

[37] Salvatore Torquato. *Random Heterogeneous Materials: Microstructure and Macroscopic Properties*. Springer, 2002.

[38] Guangnan Meng, Jayson Paulose, David R Nelson, and Vinothan N Manoharan. Elastic instability of a crystal growing on a curved surface. *Science (New York, N.Y.)*, 343(6171):634–7, feb 2014.

[39] Shlomo Alexander. Amorphous solids: their structure, lattice dynamics and elasticity. *Physics Reports*, 296(2-4):65–236, 1998.

[40] Steven Atkinson, Frank H Stillinger, and Salvatore Torquato. Existence of isostatic , maximally random jammed monodisperse hard-disk packings. *Proceedings of the National Academy of Sciences*, 111(52):18436–18441, 2014.

[41] Aleksandar Donev, Salvatore Torquato, and Frank Stillinger. Pair correlation function characteristics of nearly jammed disordered and ordered hard-sphere packings. *Physical Review E*, 71(1):011105, January 2005.

[42] Cristian F. Moukarzel. Isostatic Phase Transition and Instability in Stiff Granular Materials. *Physical Review Letters*, 81(8):1634–1637, aug 1998.

[43] Corey O'Hern, Stephen Langer, Andrea Liu, and Sidney Nagel. Random Packings of Frictionless Particles. *Physical Review Letters*, 88(7):075507, January 2002.

[44] BD Lubachevsky, FH Stillinger, and EN Pinson. Disks vs. spheres: Contrasting properties of random packings. *Journal of Statistical Physics*, 64:501–524, 1991.

[45] Toshiyuki Hamanaka and Akira Onuki. Transitions among crystal, glass, and liquid in a binary mixture with changing particle-size ratio and temperature. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 74(1):1–7, 2006.

[46] M. Reza Sadr-Lahijany, Purusattam Ray, and H. Eugene Stanley. Dispersity-Driven Melting Transition in Two Dimensional Solids. *Physical Review Letters*, 79(17):3206, 1997.

[47] Wolfgang Vermohlen and Nobuyasu Ito. State diagram of polydisperse elastic-disk systems. 51(5), 1995.

[48] Hiroshi Watanabe, Satoshi Yukawa, and Nobuyasu Ito. Size-dispersity effects in two-dimensional melting. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 71(1):1–4, 2005.

[49] Sascha Hilgenfeldt. Size-topology correlations in disk packings: terminal bidispersity in order-disorder transitions. *Philosophical Magazine*, 93(31-33):4018–4029, 2013.

[50] P. Richard, L. Oger, J. P. Troadec, and a. Gervois. A model of binary assemblies of spheres. *European Physical Journal E*, 6(4):295–303, 2001.

[51] Christopher Joseph Burke, Badel Landry Mbanga, Zengyi Wei, Patrick Spicer, and Tim Atherton. The role of curvature anisotropy in the ordering of spheres on an ellipsoid. *Soft Matter*, 2015.

[52] Shelley L. Anna, Nathalie Bontoux, and Howard A. Stone. Formation of dispersions using flow focusing in microchannels. *Applied Physics Letters*, 82(3):364, 2003.

[53] Jong Yun Kim, Jun Yeob Song, Eun Joo Lee, and Seung Kyu Park. Rheological properties and microstructures of Carbopol gel network system. *Colloid and Polymer Science*, 281(7):614–623, 2003.

[54] J. O. Carnali and M. S. Naser. The use of dilute solution viscometry to characterize the network properties of carbopol microgels. *Colloid & Polymer Science*, 270(2):183–193, 1992.

[55] Peter A. Kralchevsky and Kuniaki Nagayama. Capillary interactions between particles bound to interfaces, liquid films and biomembranes. *Advances in Colloid and Interface Science*, 85:145–192, 2000.

[56] Chuan Zeng, Fabian Brau, Benny Davidovitch, and Anthony D. Dinsmore. Capillary interactions among spherical particles at curved liquid interfaces. *Soft Matter*, 8(33):8582, 2012.