Structural search for dense packing of concave and convex shapes in two dimensions

Nabiha T. Elias, Toby S. Hudson
E-mail: Toby.Hudson@sydney.edu.au
School of Chemistry, The University of Sydney, NSW 2006, Australia.

Abstract. We describe and implement a structural search methodology which divides the global configuration space into enumerable subspaces, termed isopointal sets, and then searches within them using simulated annealing. This division allows us to follow a natural order for searching the infinite global configuration space, by respecting the observed tendency for molecular or colloidal particles to form crystals with a very limited set of local environments and symmetries. The method also produces near-optimal structural candidates with each symmetry searched, a useful feature for systems exhibiting polymorphic crystallization. We have applied this to dense packing simply-connected hard shapes in two dimensions, but the same method is applicable to optimizing arbitrary Hamiltonians, or ab-initio structural searches in higher dimensions.

1. Introduction
When we say that natural materials crystallize, or self-assemble, we are actually summarising the result of a momentous structural search process. There are infinitely many ways of arranging particles into a crystal, and the nucleation of a crystal from a melt or solution involves stochastically testing many different configurations, in search of those that lower the free energy of the system. Throughout the fluid, there are of the order of $10^{23}$ particles, continually forming and reforming local clusters searching for the rare arrangement of a supercritical nucleus. Even with this inbuilt search-parallelism, there is no guarantee that the globally optimal configuration will be found. Crystallization is an intrinsically kinetic phenomenon. Indeed many molecular crystals exhibit polymorphism, where a number of crystalline arrangements are observed as metastable products of this process. What hope then do we have of predicting these arrangements with the humble computer?

The good news is that in most cases, crystallization is a reliable process, and even if not thermodynamically optimal, the polymorphs are very good, and limited in number. Additionally, the structure formed is usually simple, having a low number of different particle environments (distinct occupied Wyckoff sites). The equivalent observation in the mathematics of tiling is that tile-transitive, or isohedral tilings (where a generating tile can be mapped onto any other tile by the symmetry operations) are quite frequent, but tiles that require anisohedral tiling are rarer. This means that a comprehensive search of structures with a small number of unrelated local environments should usually be sufficient to find the thermodynamically optimal structure.

One conventional option is to approximate the small number of distinct particle environments by a small, flexible unit cell.[1, 2, 3] This is reasonable for crystals with low overall symmetry,
but becomes wasteful in highly symmetric situations. For example, the dense packing of binary spheres seen in the super-lattice analogous to the atomic NaZn$_{13}$ structure, formed by some colloidal opals, has only three inequivalent sites (and two of those are on fixed high-symmetry points). Simulating the structure as a single flexible unit cell would require 112 particles, with a corresponding search space with 338 degrees of freedom. Making use of the high symmetry allowed an exhaustive search of the relevant isopointal set, with only three degrees of freedom, obtaining structures with a density higher than that of segregated uniform spheres.[4]

This paper reports the implementation of a proposed methodology which divides the global search domain into enumerated sub-domains, ordered by a gradual increase in the number of distinct particle sites.[5] We obtain three potential advantages over conventional structural search methodologies. The first is that a local optimum is found within each sub-domain searched, thus generating a range of good candidate structures with different symmetries. This is useful for the study of molecules exhibiting polymorphism, because a single exploration of the global search domain can locate multiple discrete polymorphs. The second advantage is the ability to search within only specific sub-domains when there is reason to believe they will contain the solution. Thirdly, because each subdomain is defined by the symmetry of the configurations it explores, each sub-search uses the minimal set of degrees of freedom, obtaining the efficiency advantages discussed in the previous paragraph.

As a proof-of-concept, we apply our search to the problem of packing simply-connected hard shapes in two dimensions (2-d), including non-convex shapes. We have chosen this limited testing domain for a few reasons. There are only seventeen distinct types of crystalline symmetry in 2-d, called the wallpaper groups. The equivalent number of space groups in three-dimensional (3-d) space is 230, which makes the 2-d case considerably cheaper computationally. Many conjectured dense packings and theorems are available for comparison and validation in 2-d, because of the overlap with the mathematical packing and tiling community. A final benefit is that solutions are easily visualized. Despite the limited testing domain used in this paper, the same method is applicable to optimizing arbitrary Hamiltonians or _ab-initio_ systems in higher dimensions. We have previously had success when searching specific isopointal sets for binary hard sphere packings in 3-d,[6, 4, 5, 7] and intend to apply the method to other particle shapes and interactions.

The packing of shapes in space is a fundamental constraint and ordering principle with widespread applications throughout the natural sciences. This kind of problem has been studied since the time of Aristotle,[8] but is sufficiently difficult that the best packing of even one of the simplest shapes, identical spheres, was unproven until around 2005.[9] Although the consequences of packing are evident at many size scales, including atomic compounds,[10] precious opals,[11], biological systems, and even everyday fruit packing, one of the most ubiquitous are the self-assembly of molecules or nano-particles into crystalline materials. In fact, the modern mathematical field of packings and coverings had its origins in chemical crystallography.[12]

Even 2-d packing problems have some scientific relevance. Molecular packing appears to be the primary determinant of the structure of monolayers formed at interfaces, across a wide range of molecules. This hypothesis is based on the observation that the symmetry groups which allow an arbitrary shape to contact the largest number of neighbouring objects (6-fold coordination), the wallpaper groups $p2$, $p1$, $p2gg$, and $pg$, are much more frequently observed in molecular monolayers than other wallpaper groups.[13] A similar observation has been made for 3-d molecular crystal structures.[14] Although the optimality of these actual molecular packings has not been tested, this is convincing evidence of the general importance of packing, even in systems where specific interactions of functional groups also play a role. Despite this understanding, the prediction of crystal structure based on the composition of the individual building blocks is still in its infancy.[13] This is in part due to the algorithmic paralysis caused
by the infinite search space which presents itself whenever a large collection of shapes is to be arranged optimally.

To the best of our knowledge, few systematic studies of congruent-shape packing principles have been undertaken, but some results are known or conjectured. Any convex shape can be packed in a dual lattice (with a wallpaper group with at least the symmetries of \( p2 \)) on the plane with a density of at least \( \approx 0.866025 \). Many specific shapes, including parallelograms, hexagons, triangles, and a range of specific irregular pentagons, can tile the plane (achieve a perfect density of 1) using only congruent copies of a single shape.\(^{[15]}\) Circles have long been known to have the fairly poor optimal density of \( \pi/\sqrt{12} \approx 0.9069 \). A handful of other convex shapes are thought to have a lower optimal packing, including regular pentagons (\((5 - \sqrt{5})/3 \approx 0.92131)\(^{[15]}\), regular heptagons (\( \approx 0.8926)\(^{[15]}\), and the smoothed-octagon (\( \approx 0.902414)\). The best-known packing of the smoothed octagon has been shown to be optimal amongst a specific subset of packings, but is conjectured to be the worst best-packing of all centrally-symmetric convex shapes.\(^{[17]}\)

2. Algorithm design

The key difference between our program and other methods of structural search is that we search only within one isopointal set at a time. Because the structures in an isopointal set must have specific symmetries, the number of free parameters is reduced relative to the general search problem. If a particular Wyckoff site is occupied, it may have zero, one or two free parameters that specify the exact positioning of that site, but all of the particles specified by that site are symmetry-related, and thus their orientations and positions are governed by those same few variables. To complete the global search, the possible isopointal sets must be generated and searched individually. This is possible because isopointal sets are enumerable. The enumeration proceeds as follows.

Firstly, we iterate the maximum number of distinct Wyckoff sites to be occupied. Since most molecular packings on surfaces, and known or conjectured optimal congruent-shape packings have all particles of one type in the same symmetry-orbit, corresponding to a single occupied Wyckoff-site (also called isohedral or tile-transitive), we have concentrated our efforts on searching for structures with only one occupied Wyckoff site. We have also briefly tested our program on some shapes that are known to only pack optimally in structures which are not isohedral, such as the pentagonal shapes discussed by Schatschneider.\(^{[16]}\) The combinatorial explosion quickly takes hold, so even a comprehensive study of ternary systems would be computationally expensive.

The second level of the enumeration is to loop through all 17 wallpaper groups. These set the underlying minimum symmetry of the packing. Again arguments from experiments can be made that some of these are much more commonly observed than others. This is rationalized by the fact that these common wallpaper groups are those that allow an arbitrary particle shape to pack such that all particles are in contact with six neighbours. In this paper we have chosen not to limit the search in this way. Instead we will search for optimal configurations within every wallpaper group, to establish the feasibility of generating multiple candidates for polymorphism.

Within a wallpaper group, there are a number of different Wyckoff sites. One of these is always the general site, a location with no special symmetry requirements, and as such, any type of particle could occupy this general site. The others are locations on particular symmetry elements of the group, such as mirror planes, or points of rotation. These locations impose requirements about what kind of particle can be placed on such a site, and so, depending on the shape we are studying, they sometimes cannot be occupied. Each allowed combinations of Wyckoff sites defines a so called isopointal set. So, for example, an isopointal set labelled \( p3(\text{ad}) \) would indicate the search space defined by the wallpaper set \( p3 \) with one particle on Wyckoff site \( a \) and another on site \( d \). This is a space of structures rather than a particular structure, because
Wyckoff site \(d\) in wallpaper group \(p3\) has variable coordinates, \(x\) and \(y\), a particle rotation, \(\phi\), and the (hexagonal) cell parameter can also be varied. In some cases, these parameters can be varied until the structure actually obtains a higher overall symmetry, but for search purposes, we consider all accessible structures to be within the isopointal set.

In principle any ergodic search algorithm could be used to search within the isopointal set. In our case the core algorithm we have chosen is a standard simulated annealing Monte Carlo optimization algorithm, which varies the free parameters associated with the packing structures, including the unit cell angle and side lengths, along with the positions and orientations of the particles. The function it seeks to optimize is simply the density of the structure - the inverse of its packing fraction. This density is determined entirely by the shape and number of the particles, and the cell volume, provided the particles do not overlap. In the case of overlap, the cell is expanded until the overlap is removed.

3. Simulations

We ran structural searches to optimize the packing of shapes with a range of different symmetries, including some known for their packing properties. The shapes we report here are the unit circle, the regular polygons with 5, 6, 7, 9, and 12 sides, the smoothed octagon, a pointy-egg shape with mirror symmetry, and a (non-convex) comma.

The egg and comma are defined parametrically, with an angle-dependent radius of the form given in equations 1 and 2. These specific shapes have no particular importance, they were simply selected to increase the diversity of the test set. Our current implementation allows for the study of any shape that can be specified parametrically as a function of the angle, including non-differentiable or non-convex shapes like these. In the results we report here, the shapes were constructed using 3600 radial spokes, which were tested for intersections to determine overlaps with neighbouring shapes. The area of each shape was also numerically determined from these segments. Thus the calculated packing values have a small numerical error, of the order of \(\pm 0.001\). The numerical results of our simulations are shown in Table 1.

\[
r_{\text{egg}}(\theta) = \frac{1}{2\sin^2((3\pi - \theta)/4)}
\]

\[
r_{\text{comma}}(\theta) = 0.4 + 0.6 \times (2(\sqrt{\theta/2\pi} - 0.5))^4
\]

It is immediately notable that many of the simulations for a given shape have converged on the same dense-packed structure even though they were exploring isopointal sets in different wallpaper groups. Nevertheless, depending on the symmetries of the shape to be packed, simulations in some wallpaper groups are unable to symmetrize like this. Thus there are often a range of diverse dense-packed structures, with different symmetry, which could be used as candidates in free energy calculations to predict potential polymorphism. This is particularly true of the less symmetrical shapes, so for example, Figure 1 shows five different structures all with high packing fractions.

We also note that the wallpaper groups most frequently observed in molecular monolayers, \(p2\), \(p1\), \(p2gg\), and \(pg\),[13], are consistently good performers across the range of shapes we have tested. This suggests that searches limited to these groups may be sufficient in systems or situations where an exhaustive global search is not required or proves too expensive computationally.

Detailed inspection of the structural and numerical outcomes for each of the well-known shapes gives maximum-packing results generally in line with literature expectations, as well as a range of high-density polymorphs. For regular pentagons and heptagons, the densest structures found both have symmetry \(p2mg\), and are identical to those produced by Kuperberg’s constructive method,[15] which are the densest known packings of these shapes and appear in many different contexts.[18, 19, 20] (Our result in the \(pg\) set appears slightly better for the
Table 1: Maximum packing fractions found for searches with a particular wallpaper group. In each isopointal set, only one Wyckoff site was occupied. In cases where more than one isopointal set is possible within the same wallpaper group, the best performing search is reported here. The full structures for all isopointal sets are available in the supplementary material.

| Wallpaper group | comma | egg | pentagon | heptagon | smoothed octagon | nonagon | hexagon | dodecagon | circle |
|-----------------|-------|-----|----------|----------|-----------------|--------|---------|----------|-------|
| p1              | 0.939 | 0.912 | 0.819 | 0.864 | 0.902 | 0.892 | 1.000 | 0.928 | 0.907 |
| p2              | 0.950 | 0.881 | 0.897 | 0.890 | 0.902 | 0.894 | 1.000 | 0.928 | 0.907 |
| pm              | 0.881 | 0.807 | 0.691 | 0.741 | 0.825 | 0.760 | 0.751 | 0.804 | 0.785 |
| pg              | 0.941 | 0.912 | 0.922 | 0.893 | 0.901 | 0.897 | 1.000 | 0.928 | 0.907 |
| cm              | 0.853 | 0.911 | 0.854 | 0.862 | 0.902 | 0.888 | 1.000 | 0.928 | 0.907 |
| p2mm            | 0.756 | 0.794 | 0.691 | 0.739 | 0.825 | 0.757 | 0.750 | 0.804 | 0.785 |
| p2mg            | 0.889 | 0.919 | 0.921 | 0.893 | 0.902 | 0.869 | 1.000 | 0.928 | 0.907 |
| p2gg            | 0.923 | 0.916 | 0.870 | 0.885 | 0.903 | 0.893 | 1.000 | 0.928 | 0.907 |
| c2mm            | 0.817 | 0.859 | 0.854 | 0.842 | 0.902 | 0.827 | 1.000 | 0.928 | 0.907 |
| p4              | 0.818 | 0.833 | 0.843 | 0.844 | 0.851 | 0.831 | 0.841 | 0.862 | 0.842 |
| p4mm            | 0.493 | 0.794 | 0.674 | 0.729 | 0.825 | 0.776 | 0.804 | 0.804 | 0.785 |
| p4gm            | 0.820 | 0.816 | 0.823 | 0.844 | 0.850 | 0.827 | 0.833 | 0.862 | 0.842 |
| p3              | 0.903 | 0.897 | 0.859 | 0.883 | 0.893 | 0.895 | 1.000 | 0.928 | 0.907 |
| p3m             | 0.646 | 0.639 | 0.651 | 0.638 | 0.670 | 0.884 | 0.933 | 0.928 | 0.907 |
| p3lm            | 0.696 | 0.897 | 0.871 | 0.883 | 0.890 | 0.888 | 1.000 | 0.928 | 0.907 |
| p6              | 0.689 | 0.790 | 0.736 | 0.761 | 0.776 | 0.784 | 1.000 | 0.928 | 0.907 |
| p6mm            | 0.473 | 0.778 | 0.706 | 0.572 | 0.718 | 0.734 | 1.000 | 0.928 | 0.907 |

pentagons, but converges structurally and numerically as the shape resolution is increased.) For smoothed octagons, we have also converged on the known optimal density of $\approx 0.9024$,[17] and we realise this with two distinct packings, one in the $c2mm(a)$ isopointal set, and one in the $p2mg(c)$ set. For the hexagon, dodecagon and circle, the high symmetry of these particles means that the known optimal packings of these shapes can easily be obtained from many different isopointal sets. The one known failure in these simulations occurs for the nonagons, for which a $p2$ structure with a density of 0.9010 was recently reported.[3] Our optimal solution has a very close density of 0.897, but is structurally different. Our corresponding $p2$ search appears to have been trapped in a local optima in the six dimensional search space, but we still consider it beneficial to have reduced it from the nine dimensions required for a global search of the two particle unit cell.

4. Conclusions
We have successfully demonstrated a proof-of-concept of the enumerated isopointal set structural search, treating the example problem of searching for homogeneous packings of congruent hard shapes in 2-d. Overall densest packings were almost always in line with the best known prior results, apart from one case which fell slightly short due to the simulated annealing simulation getting trapped in a local optima. These simulations also reproduced the tendency observed for molecular or colloidal particles to favour crystals with a very limited set of local environments and symmetries. The division of the global search into subspaces ensured diverse competitive structural outcomes, especially in the cases with less symmetric particles. This aspect bodes well for the use of this search methodology in the study of polymorphic systems.

The step up from 2-d to 3-d structure search will result in an order of magnitude expansion in the number of isopointal sets that must be searched. As noted earlier, there are...
Figure 1: The five densest packings found for the comma shape, each with a different wallpaper group.

230 space groups as compared to 17 wallpaper groups, and space groups often have more Wyckoff sites. Thankfully in many applications this can be scaled back by prioritizing the space groups known to dominate the known crystal structures. But even when an exhaustive search is desired, it will be feasible given that each result reported here required a few hours on a single CPU core, and different isopointal searches can be run on different cores. The speed of an individual test for overlapping particles is directly related to the efficiency. We have not sought to optimize this, but many algorithms are available, due to its importance in computer graphics. Our shapes were defined with 3600 segments, which should be easily sufficient to accurately specify most 3-d shapes.

Acknowledgments
TSH wishes to acknowledge helpful discussions with Prof. Peter Harrowell.

Supporting information
Supporting information is available, which gives optimized packing structures for each isopointal set searched. This material is available under Supplementary Data.

References
[1] Torquato S and Jiao Y 2009 Nature 460 876-879
[2] Filion L, Marechal M, van Oorschot B, Pelt D, Smallenburg F and Dijkstra M 2009 Physical Review Letters 103(18) 188302
[3] de Graaf J, van Roij R and Dijkstra M 2011 Physical Review Letters 107(15) 155501
[4] Hudson T S 2010 The Journal of Physical Chemistry C 114 14013–14017
[5] Hudson T S and Harrowell P 2011 Journal of Physics: Condensed Matter 23 194103
[6] Kummerfeld J K, Hudson T S and Harrowell P 2008 The Journal of Physical Chemistry B 112(35) 10773–10776
[7] O’Toole P I and Hudson T S 2011 The Journal of Physical Chemistry C 115 19037–19040
[8] Aristotle 350 BC On the Heavens vol III translated by J. L. Stocks.
[9] Hales T C 2006 Discrete & Computational Geometry 36 5–20
[10] Parthé E 1961 Zeitschrift für Kristallographie 115 52–79
[11] Sanders J V 1980 Philosophical Magazine A 42 705–720
[12] Fejes Tóth G 2004 Handbook of discrete and computational geometry ed Goodman J E and O’Rourke J (CRC Press) chap 2
[13] Plass K E, Grzesiak A L and Matzger A J 2007 Accounts of Chemical Research 40 287–293
[14] Mighell A D, Himes V L and Rodgers J R 1983 Acta Crystallographica Section A 39 737–740
[15] Kuperberg G and Kuperberg W 1990 Discrete & Computational Geometry 5(1) 389–397
[16] Schattschneider D 1978 Mathematics Magazine 51(1) 29–44
[17] Hales T C 2011 ArXiv e-prints ArXiv:1103.4518v1
[18] Henley C L 1986 Physical Review B 34(2) 797–816
[19] Baker T S, Caspar D L D and Murakami W T 1983 Nature 303 446–448
[20] Schilling T, Pronk S, Mulder B and Frenkel D 2005 Physical Review E 71(3) 036138