Introducing Alpha Shapes for the Analysis of Path Integral Monte Carlo Results

Patrick J. Moran, Department of Computer Science
Marcus Wagner, Department of Physics
National Center for Supercomputing Applications
University of Illinois at Urbana-Champaign, 61801, USA

Abstract

We present a new technique for the visualization and analysis of the results from Monte Carlo simulations based on α-complexes and α-shapes. The specific application we present in this article is the analysis of the quantum-mechanical behavior of hydrogen molecules and helium atoms on a surface at very low temperatures. Our technique is an improvement over existing techniques in two respects. First, our approach allows one to visualize the points on a random walk at varying levels of detail and interactively select the level of detail that is most appropriate. Second, using α-shapes one can obtain quantitative measures of spatial properties of the system, such as the boundary length and interior area of clusters, that would be difficult to obtain otherwise.

1 Introduction

Path Integral Monte Carlo (PIMC) [1] is an exact computational method based on Feynman's [7] Path Integral formalism for modeling many-body quantum mechanics at finite temperatures. In PIMC, the simulated particles are moved with probabilities depending on their quantum mechanical expectation values to be found in their old and new positions, i.e. before and after a move. As the system of physical particles evolves, sets of particle coordinates are generated from which energy and other physical quantities are computed. One can use the results from such simulations to address questions such as "Does a liquid adsorbed on a solid surface spread out into a film and 'wet' (coat) the substrate, or does it form droplets instead?", "If the adsorbate forms droplets, then what is their size distribution?" and "Do the droplets percolate into a network of rivulets, or is there a phase separation into large unconnected adsorbate puddles and patches of dry surface area?" To answer such questions, one typically visualizes the PIMC results using a few standard techniques. For example, Figure 1 illustrates a typical visualization. There are a couple of observations that we can make immediately about the figure. First, in the case of particles on relatively flat surfaces, two-dimensional visualizations are sufficient even though in the simulation the particles are in 3-d. Second, the visualization needs to handle the fact that the simulations are often done using periodic boundary conditions. A closer look at the figure will reveal that the data are replicated four times, with the contents of the dashed box representing the fundamental domain. Using visualizations such as that in Figure 1, one can answer the questions posed above, at least qualitatively.

In this article we present a new technique for analyzing PIMC results based on α-complexes and α-shapes. Alpha complexes are a geometric technique for describing shape and space [3, 5]. Given a point set S, one can define a family of shapes that capture the intuitive notion of fine (S alone) to coarse (the convex hull of S) shape. Alpha shapes are based on the Delaunay triangulation of the point set. The Delaunay triangulation may be familiar to some readers, but the definition of the Delaunay triangulation as a simplicial complex may be less familiar. Simplicial complexes are standard concepts in algebraic topology and are a general, mathematically rigorous way of describing regions in space in terms of discrete "building blocks" such as edges and triangles. One advantage of developing our definitions in terms of simplicial complexes is that the definitions can be generalized to higher dimensional spaces [6].

The input to α-shapes is a finite set of points and, if the data were computed using periodic boundary conditions, the length of one periodic cycle for each dimension. Information about the path followed by each particle (as displayed by the edges in Figure 1) is not part of the input. All the points in the input set are treated equally.

In the following section, we define the terminology

1070-2385/94 $04.00 © 1994 IEEE

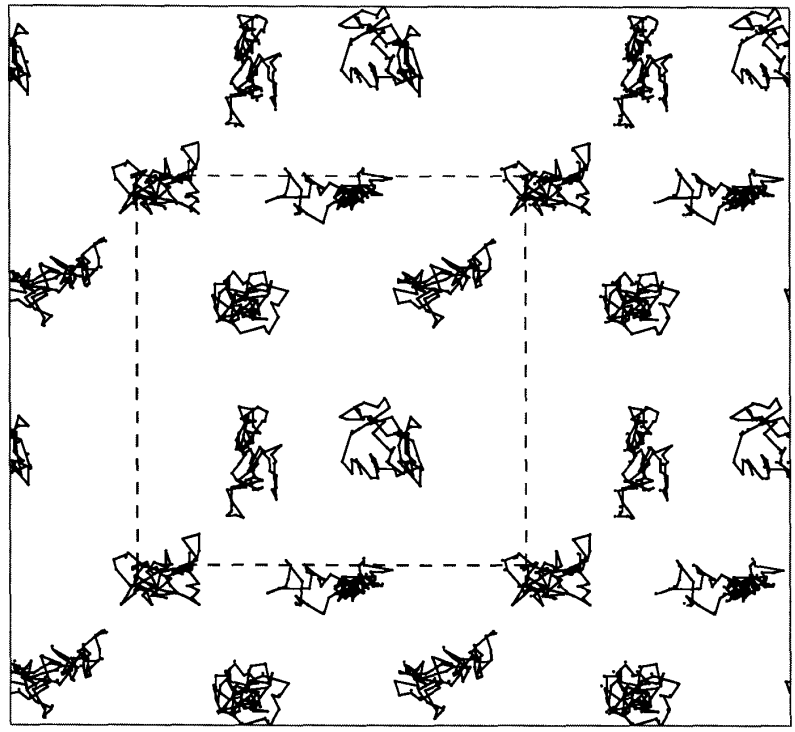

Figure 1: A typical visualization technique for Path Integral Monte Carlo data.

and concepts used in this article. In Section 3 we discuss some of the implementation issues. Section 4 presents results with two example data sets. Finally, in Section 5 we provide some concluding thoughts.

2 Definitions

We present the basic definitions needed to understand α-complexes and shapes in $\mathbf{R}^2$. References to more detailed presentations of the concepts that follow are also included.

2.1 General Position

In the following definitions we assume that no 3 points are collinear and no 4 points are cocircular. The exclusion of these degenerate cases is known as *general position*. Degenerate cases do occur in practice, and our implementation can handle them, but the definitions are clearer if we do not cover the special cases here.

2.2 Simplices and Complexes

In $\mathbf{R}^2$ we consider a *simplex* to be the convex hull defined by 1, 2 or 3 points. Such simplices are known as vertices, edges and triangles, respectively. A subset of the set of points T defining a simplex σ defines another simplex, a *face* of σ. A simplex defined by a proper subset of T is a *proper face* of σ. Triangles have edge and vertex faces, edges have two vertex faces. We exclude the case of an empty subset of T for these definitions.

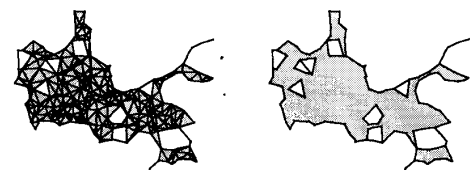

Figure 2: An α-complex and an α-shape.

A *simplicial complex* $\mathcal{C}$ is a collection of simplices

satisfying two conditions:

1. If σ is in $\mathcal{C}$, then every face of σ is in $\mathcal{C}$.
2. The intersection of $\sigma_1, \sigma_2 \in \mathcal{C}$ is either empty or a face of both σ_1 and σ_2.

If $\mathcal{C}$ is a simplicial complex, then $\mathcal{C}' \subseteq \mathcal{C}$ is a *subcomplex* if it also satisfies the simplicial complex conditions. The *shape* of a simplicial complex $\mathcal{C}$ is the union of the the simplices in $\mathcal{C}$. The shape of $\mathcal{C}$ is sometimes written as $|\mathcal{C}|$. Figure 2 illustrates a simplicial complex and its shape. Notice that the simplicial complex is not necessarily connected, and it may contain holes inside regions.

A *triangulation* of a finite point set S is a simplicial complex $\mathcal{C}$ such that each simplex in $\mathcal{C}$ is defined by points in S and $|\mathcal{C}| = \text{Conv}(S)$ is the convex hull of S.

2.3 The Delaunay Triangulation

For a simplex σ defined by the point set T, let B_σ specify the smallest open disk where the points in T are on the disk boundary, and let ρ_σ be the radius of B_σ. For a vertex, we let $\rho_\sigma = 0$.

Let ab be an edge of a triangulation in $\mathbf{R}^2$ and the face of two triangles abc and abd. Edge ab is *locally Delaunay* if d is not inside B_{abc} (or equivalently, c is not inside B_{abd}). Edges on the boundary of a triangulation are always considered to be locally Delaunay. The *Delaunay triangulation* of a finite point set is the triangulation where all the edges are locally Delaunay. The Delaunay triangulation has many interesting and useful properties. Preparata and Shamos [11] describe the triangulation and a closely related structure, the Voronoi diagram, in detail.

2.4 Alpha Complexes and Shapes .

For a triangulation of a finite point set S, we classify each edge as one of two types. An edge σ is *attached* if $B_\sigma \cap S \neq \emptyset$ and *unattached* otherwise. When testing an edge in the Delaunay triangulation it is sufficient to consider the vertices of the two triangles which an edge adjoins. An edge ab which is the face of triangles abc and abd is attached if $B_{ab} \cap \{c, d\} \neq \emptyset$. Vertices and triangles are always considered to be unattached.

Given the Delaunay triangulation $\mathcal{D}$ defined by a finite point set S, an *α-complex* $\mathcal{C}_\alpha \subseteq \mathcal{D}$ is a subcomplex where each simplex $\sigma \in \mathcal{C}_\alpha$ is either

- unattached and $\rho_\sigma \leq \alpha$
- the face of another simplex in $\mathcal{C}_\alpha$

An α-shape is the shape of an α-complex.

Figure 3 illustrates an example point set ($\alpha = 0$) and three additional α-shapes. If the shapes are considered in clockwise order starting at the upper left, then the shapes are in order of increasing α.

2.5 The Union of Disks

An alternate technique for visualizing a path point set S is to draw a disk of radius α centered at each point in S. As with α-shapes, we would like to use this technique not only to display the data but also to make quantitative statements. Fortunately, there is a close connection between an α-complexes and the union of a set of disks, Edelsbrunner [6] describes this connection in detail, including formulae for computing boundary lengths and area. The right hand side of Figure 4 provides some intuition to how the α-complex encodes the relationships of the disks. For example, notice that if there is an edge between two vertices, then the corresponding disks intersect. Likewise, if a triangle is part of the α-complex then the three corresponding disks have a common intersection.

2.6 Betti Numbers

Betti numbers are a concept from algebraic topology. The kth Betti number, β_k, is the rank of the kth homology group of a topological space. Munkres [10] formally defines Betti numbers and many associated concepts in detail. In this article we are particularly interested in β_0 and β_1. β_0 provides a count of the number of connected components of a shape. Informally, β_1 is a count of the number of holes in a shape embedded in $\mathbf{R}^2$. In cases where there are periodic boundary conditions, as for the torus in Figure 6, this intuitive interpretation is almost correct but not quite. Rather than introduce all the formalism necessary to rigorously define β_1 we encourage the reader to refer to [10], while for this article we continue to use the rough notion of "hole count."

2.7 Signatures

A *signature function* is a function $f : \mathbf{R} \rightarrow R$ that maps a real value to a value in range R. We use signature functions that map α to quantities such as the interior area and boundary length of $\mathcal{C}_\alpha$. We can also compute signatures measuring topological quantities such as the number of connected components. Since many signatures only change when $\mathcal{C}_\alpha$ changes, it is convenient to define signatures that map an integer index to a range R. For an integer s, let $[s]$ represent the set $\{1, 2, \ldots, s\}$ of indices. An index from the set $[s]$ is also known as a *rank*. A *discrete signature function* is a function $f : [s] \rightarrow R$ that maps each index

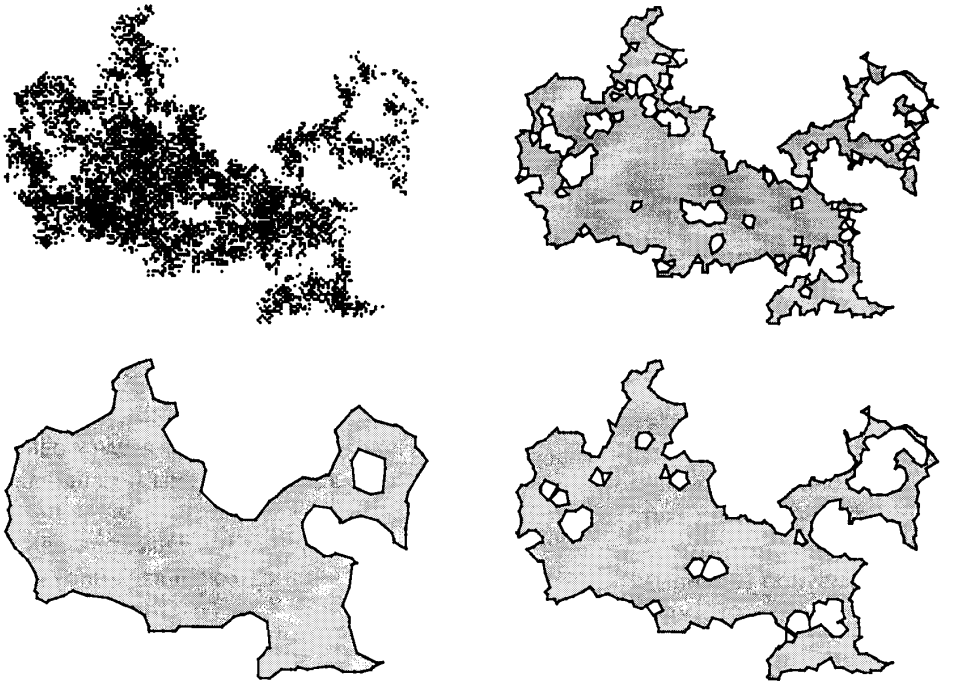

Figure 3: A point set based on the fractal motion of one atom shown at four α values.

$i \in [s]$ to a value $f(i)$ in range R. We will drop the adjective discrete if it is clear from context.

Signature functions provide information about how the α-shape or α-complex evolves as α grows. They thus hint at shapes with certain properties and can be used in choosing interesting shapes.

3 Implementation

3.1 General Position

The assumption of general position in the previous section allowed us to present the definitions in a simpler form, without the need to cover many special cases. We would like our implementation to also be free of such special cases, both for better readability and reliability. To achieve this goal, we use "Simulation of Simplicity" [4], or SoS for short. SoS is a module that provides primitives to answer queries where a degenerate case may occur. In $\mathbf{R}^2$, SoS provides two primitives. `Left_turn` takes three point arguments and answers whether the third point is to the left or right of a line through the first two. `In_circle` takes four points as arguments and answers whether the fourth point is inside a circle circumscribing the first three. To resolve a degeneracy, SoS simulates a globally consistent infinitesimal perturbation and makes decisions based on the thus obtained non-degenerate data. The fact that degenerate cases occur is hidden from the rest of the implementation. We encourage the reader to see [4] for details.

3.2 Selecting Simplices

Given a value for α by the user, we select simplices from the Delaunay simplicial complex to get the α-subcomplex. To accomplish this, we compute a threshold ϱ for each simplex as a preprocessing step. A simplex σ is part of $\mathcal{C}_\alpha$ if its threshold $\varrho \leq \alpha$. If σ is unattached, then we define $\varrho = \rho_\sigma$. Otherwise, if σ is an attached edge, then

$$\varrho = \text{Min}\{\rho_\tau \mid \tau \in \text{Up}(\sigma, \mathcal{D})\}$$

where $\text{Up}(\sigma, \mathcal{D})$ returns the set of simplices that σ is the face of. In this case $\text{Up}(\sigma, \mathcal{D})$ returns the triangles that edge σ adjoins. Notice the difference between ρ_σ

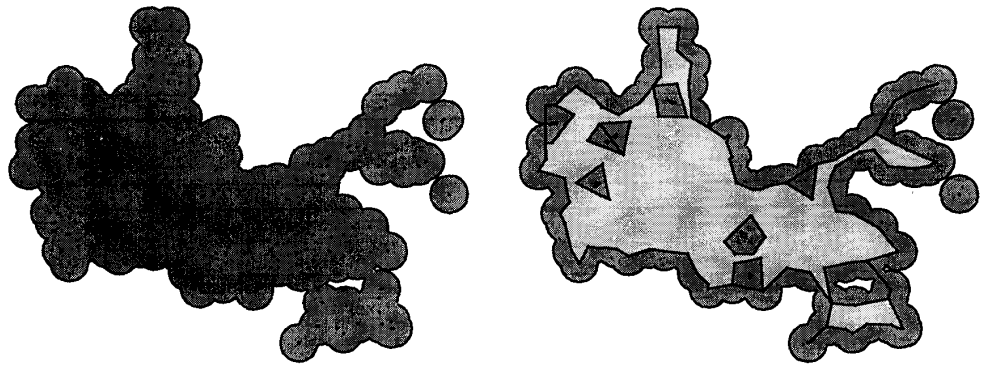

Figure 4: The union of α radius disks and the same plot with the α-shape superimposed.

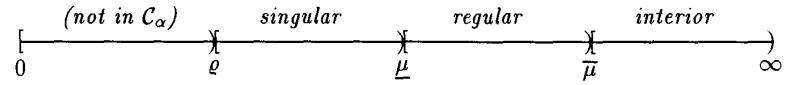

Figure 5: The life of a simplex in $\mathcal{C}_\alpha$ for $0 \leq \alpha < \infty$.

and ϱ. For a simplex σ, ρ_σ specifies the radius of σ, regardless of whether it is unattached or attached. On the other hand, ϱ is a threshold specifying when σ first becomes part of $\mathcal{C}_\alpha$.

We further classify each vertex and edge $\sigma \in \mathcal{C}_\alpha$ as:

singular	if $\sigma \in \text{Bd}(\mathcal{C}_\alpha)$ and σ is not a proper face
regular	if $\sigma \in \text{Bd}(\mathcal{C}_\alpha)$ and σ is a proper face
interior	if $\sigma \notin \text{Bd}(\mathcal{C}_\alpha)$

where $\text{Bd}(\mathcal{C}_\alpha)$ returns the set of simplices on the boundary of $\mathcal{C}_\alpha$. These three classifications correspond to three (possibly empty) subintervals of the interval $[\varrho, \infty)$, and we can define two thresholds $\underline{\mu}$ and $\overline{\mu}$ to delimit the subintervals. Figure 5 illustrates how a simplex is classified by ϱ, $\underline{\mu}$ and $\overline{\mu}$. To make it unnecessary to treat the simplices on the boundary of $\mathcal{D}$ as special cases, we consider each boundary simplex to be the face of a triangle τ where $\rho_\tau = \infty$. With that small modification, we define $\underline{\mu}$ and $\overline{\mu}$ for each vertex and edge in $\mathcal{D}$ to be:

$$\underline{\mu} = \text{Min}\{\rho_\tau \mid \tau \in \text{Up}(\sigma, \mathcal{D}),\ \tau \text{ unattached}\}$$

$$\overline{\mu} = \text{Max}\{\rho_\tau \mid \tau \in \text{Up}(\sigma, \mathcal{D})\}$$

As α advances from 0 towards infinity, a simplex σ becomes regular when it first becomes the proper face of another simplex in $\mathcal{C}_\alpha$, and σ becomes interior when every simplex that it is the face of in $\mathcal{D}$ is part of $\mathcal{C}_\alpha$.

3.3 Algorithms and Complexity

There are many algorithms to choose from for computing Delaunay triangulations, see for example [11]. Our implementation uses the randomized incremental algorithm presented by Guibas, Knuth and Sharir in [8]. This algorithm has $O(n^3)$ worst case time for n points, but performs in $O(n \log n)$ expected time. We refer the reader to [8] for more details. In practice constructing the Delaunay triangulation is the longest step in terms of time.

The number of simplices in the Delaunay triangulation $\mathcal{D}$ in $\mathbf{R}^2$ can be computed easily based on Euler's relation. Let F_k be the number of k-simplices in $\mathcal{D}$. If $F_0 = n$, then $F_1 \leq 3n - 6$ and $F_2 \leq 2n - 4$. Let $m = F_0 + F_1 + F_2$ represent the number of simplices in $\mathcal{D}$, then $m \leq 6n - 10$. The maximum number of α-shapes defined by a point set is equal to the number of unique radii for unattached simplices. The maximum value for this number s would occur if all the simplices were unattached and had unique radii, therefore $s \leq 5n - 9$. Computing ϱ, $\underline{\mu}$ and $\overline{\mu}$ for each simplex can be accomplished in time proportionate to the

number of simplices, thus the threshold calculations for the simplices can be completed in linear time.

Given an α-complex $\mathcal{C}_\alpha$ with the edges labeled as singular, regular or interior, computing basic measures such as interior area and boundary length is straightforward. The area is the sum of the area of each triangle in $\mathcal{C}_\alpha$. The boundary length is the sum of the regular edge lengths plus twice the sum of the singular edge lengths. One could compute a signature such as the interior area signature simply by computing the area for each complex in the α-shape family. Delfinado and Edelsbrunner [2] describe a more efficient approach for computing the area, boundary length, β_0 and β_1 signatures for a shape family based on incremental techniques.

S	Card(S)	time (sec)
Figure 2	257	2
Figure 6	800	8
Figure 3	8193	58

Table 1: Example initial computation times.

3.4 Performance

The computations carried out by our implementation can be put into two categories: those that take place initially and those that occur each time one selects a new α value. For the largest data set described here, the time to compute and display an α-shape given α is less than one second, so we concentrate on providing a few example times for the initial computation step. Table 1 lists timings based on some of the data sets used for our illustrations, the computations were done on an SGI Indigo Elan workstation. These timings reflect an implementation that has not been extensively optimized.

4 Results

We present two examples illustrating the use of α-shapes for visualization. In the first example we consider a point set where the configuration suggests a distinct "shoreline", in the second we consider the random walk of a single particle.

4.1 A "Shoreline" Example

Our first example is from a PIMC simulation of a vacuum interface of solid molecular hydrogen, H_2, with a partially filled top layer [12]. The simulation is set up in a rectangular box with periodic boundary conditions. In this particular simulation, a snapshot of which is shown in Figure 6, T= $0.5K$ and the partially filled top layer has 1/3 as many H_2 molecules as a layer with the same surface normal direction would have, deep inside the bulk. In this context an important question that can be addressed using α-shapes is whether or not the partially filled H_2 top layer spreads out over the surface, or whether it forms a step or droplet, the latter depending on whether it is solid or liquid. We observe surface wetting when a partially filled liquid top layer spreads out over the substrate surface, which can occur only if the energy penalty per particle caused by generating additional droplet shoreline length does not exceed the particles' kinetic energy. To quantify this and predict the existence or absence of a wetting transition we need the numerical value of the line tension of a droplet formed by a partially filled surface top layer. Note that a melted surface step or terrace amounts to a droplet which may or may not spread out over the substrate.

It is very common that the surface of a solid is not completely flat in the sense that the atoms or molecules at the surface usually do not exhibit the same regular crystalline arrangement as deep inside the bulk material. The particles in the surface top layer may be able to lower their total free energy by rearranging themselves, and during the material growth there may not have been enough particles nearby to completely fill the surface top layer, so that terraces formed.

Let N_P denote the number of particles in such a partially filled surface top layer, atoms or molecules. The line tension γ of a droplet or a terrace can then be computed as

$$\gamma = N_P \, (e_P - e_F) \, / \, L_P$$

where e_P and e_F are the binding energies per particle to the bulk substrate in a partially filled and a completely filled *top* layer, respectively, both readily obtained with PIMC. The total shoreline length L_P, however, which measures the combined circumference of all droplets or terraces in the partial top layer, must be calculated with α-shapes because a consistent, repeatable way to define the size of any droplets on the surface is needed, and specifying this "by hand" is not reliable.

Furthermore, the density of particles per area A in a droplet or terrace is important because if that value differs from the corresponding density in a bulk layer, then there must be a structure misalignment somewhere between the bulk interior and its surface which favors surface melting. Similar to the case of the free random walker depicted in Figure 3, Figure 6

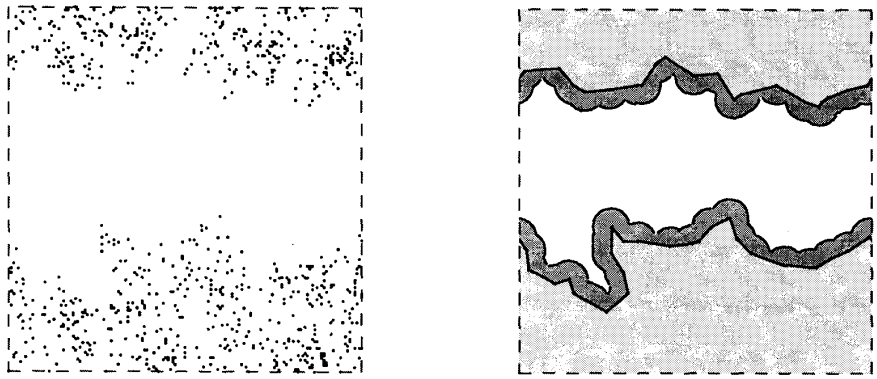

Figure 6: The point set within periodic bounds for the first example, an α-shape and the disk union.

emphasizes that neither N_P nor A are always easy to obtain.

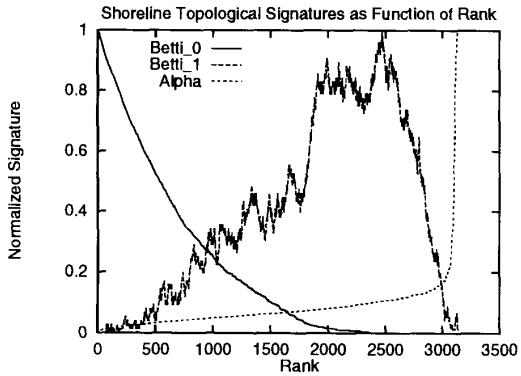

Figure 7: Topological signatures for first example.

Figures 7 and 8 illustrate basic topological and metric signatures for the shoreline example. Betti_0 (β_0) in Figure 7 corresponds to the number of connected components for each α-shape. Betti_1 (β_1) corresponds to the first Betti number for each α-complex. β_1 is roughly the number of "holes" in each complex. The graph in Figure 7 also includes a plot of the minimum α contained in the interval associated with each rank. This is not a topological signature, but it is included to illustrate how the intervals associated with each rank are distributed over the range of α. Typically the intervals are not evenly distributed, most are in the low part of the α range.

In some cases, particularly when the user is interested in metric signatures, it is constructive to plot signatures as a function of α. For example, Figure 8

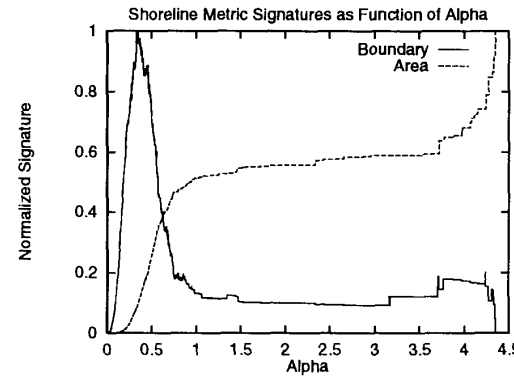

Figure 8: Metric signatures for first example.

shows the boundary length and area signatures plotted in this manner. Note how both signatures are relatively flat over a wide range of α, suggesting that shapes in this range may be natural to the point set. Indeed, the shapes correspond to the two shoreline configuration shown in Figure 6.

4.2 A Single Particle Random Walk

The random walk of a free particle is equivalent to a colloid particle diffusing in a fluid, or to an atom migrating on a solid surface after adsorption. Consider points along the path traced by such a random walker, shown in the upper left of Figure 3, which is well known to be a fractal object[9], i.e. the traveled distance scales as a non-integer power of the linear length scale, and the path looks similar regardless of the magnification. For a physicist or engineer it is of interest how "spread out" the diffusing particle is,

i.e. how much area it has swept over, but also the overall shape of the visited area, because both allow conclusions regarding the roughness of that surface, or analogously, the effective mass of the diffusing particle. The area needs to be given numerically but its shape is most readily understood by way of visualization. Alpha shapes provide a unique definition of the surface area along with the numerically exact computation of its measure which would otherwise be hard to estimate from the set of particle coordinates shown in the upper left of Figure 3.

5 Conclusion

We have presented a new technique for studying both the qualitative and quantitative nature of the shape of point sets generated by Path Integral Monte Carlo. While we focused on a fairly specific application in this article, our technique may be applicable to many scientific disciplines, in particular those where there is a need to visualize and measure the often vaguely defined notion of shape.

Acknowledgements

We would like to thank Professor David Ceperley for providing some of the initial inspiration for this work and for providing the data sets used in figures 1, 2, 3 and 4. We would also like to thank the α-shapes research group headed by Professor Herbert Edelsbrunner for advice and encouragement.

References

[1] D. M. Ceperley and E. L. Pollock, Path-integral computation techniques for superfluid ^{4}He. In: S. Caracciolo and A. Fabrocini (eds.), *Proceedings of the Elba Conference on Monte Carlo Methods in Theoretical Physics, 1990*, ETS Editrice, Pisa, (1992), 35-71.

[2] C. Delfinado and H. Edelsbrunner. An incremental algorithm for betti numbers of simplicial complexes. *ACM 9th Annual Symposium on Computational Geometry.* (1993), 232-239.

[3] H. Edelsbrunner, D. G. Kirkpatrick and R. Seidel. On the shape of a set of points in the plane. *IEEE Trans. Inform. Theory.* **IT-29**, (1983), 551-559.

[4] H. Edelsbrunner and E. Mücke. Simulation of Simplicity: a technique to cope with degenerate cases in geometric algorithms. *ACM Trans. Graphics* **9**, (1990), 66-104.

[5] H. Edelsbrunner and E. Mücke. Three-dimensional alpha shapes. *ACM Trans. Graphics* **13**, (1994), 43-72.

[6] H. Edelsbrunner. The union of balls and its dual shape. *ACM 9th Annual Symposium on Computational Geometry.* (1993), 218-231.

[7] R. P. Feynman, *Statistical Mechanics*, Addison-Wesley, 1972.

[8] L. J. Guibas, D. E. Knuth and M. Sharir. Randomized incremental construction of Delaunay and Voronoi diagrams. *Proc. 17th Ann. International Coll. Automata, Lang., Progr. 1990.* Lecture Notes in Computer Sci., Springer-Verlag, **443**, 414-431.

[9] B. Mandelbrot, *The Fractal Geometry of Nature*, W. H. Freeman and Company, New York, 1977.

[10] J. R. Munkres. *Elements of Algebraic Topology.* Addison-Wesley, Redwood City, California, 1984.

[11] F. P. Preparata and M. I. Shamos. *Computational Geometry – an Introduction.* Springer-Verlag, New York, 1985.

[12] M. Wagner and D. M. Ceperley, Path integral monte carlo simulations of H_2 surfaces. *J. Low Temp. Phys.* **94**, (1994), 161-183.