A COMPUTATIONAL TOPOLOGY APPROACH TO HARD SPHERES IN A BOX

JACKSON GORHAM

1. An Introduction to the "Hard Disks" in a Box Model

As an attempt to better understand the physical properties of atoms and how they interact together on a quantum level, physicists have looked toward models based upon hard spheres for insight. Similar in spirit to the underpinnings of the Boltzmann model, the hard spheres model sets a foundation upon which to frame this topic and begin to tackle interesting problems. In order to simplify matters, we can study this model in two-dimensional space; accordingly, we can now speak of these hard spheres as hard disks. As a consequence of its rich, mathematical structure, this hard disks model has recently attracted both statistical physicists and mathematicians. In its simplest form, the hard disks model studies the nature of packing some number n of identical disks each with an arbitrary radius r into a unit square under the constraint that no two disks overlap. This idea is very old and dates back to the original models by the first atomists.

While Graham and others have done work to find the greatest possible radius for a various number of disks [6] [1], this paper instead focuses on the topological properties of the configuration space. By identifying each disk by its center's coordinates in $(0,1)^2$ and juxtaposing these n ordered pairs to create a $2n$-tuple, each possible configuration of n disks corresponds to a point in the Euclidean space $\mathbb{R}^{2n}$. With this embedding into Euclidean space, we can adjoin this set with a topology. The result is commonly called the configuration space.

Recent research has examined implications of this model from many different angles. For instance, work by Werner Krauth has shown that when there are many disks and approximately 72 percent of the unit square can be covered by those disks, a phase transition occurs as has been confirmed by empirical experiments [9]. On the other hand, some research has approached this model from a statistical mechanics angle [8]. Among researchers in that camp, it is widely believed that the configuration space is connected, for if not, the Metropolis Markov process is not ergodic. Surprisingly, we will show that this is not true. This is one of the major facets of this paper; even in the case with just 5 disks, the configuration space is complex enough to evade many of the simple properties that these spaces were assumed to have.

The hard disks model has practical applications as well. Because of the strong relationship between the ability of molecules to freely move and the phase of the matter (solid,

liquid, gas) these molecules induce, there is much concern to understand how these disks can move about as their radii change. One way to approach the issue of the disks' mobility is by determining how the number of disjoint components varies with the radius r. However, as of now, those researching this problem understand "very, very little about the topology of the set," including the connectedness of this space [3]. Hence the structure of π_0—the set of path components for a given radius and number of disks—is the primary concern of this paper. In this way we will be able to show the configuration space is not connected.

One strategy is to work under the framework of hierarchical clustering. After defining a metric between points in the configuration space, one can study how the connectedness of the space changes as the maximum allowable distance between points is increased. This paper's idea resembles the agglomerative approach, i.e., increasing an energy threshold E until each pair of clusters finally becomes one. Using terminology from the literature, we are "softening" the disks by relaxing the constraint no two disks can overlap. The interesting caveat of this strategy is our choice of metric: given two different points in the configuration space, the distance between them is defined as the minimum (over all paths between the two points) of the maximum energy required to move from one point to another. However, while this metric is much better suited to study this problem in hard disks than say the Euclidean metric, it is also much harder to calculate. This is the need for our computational algorithm.

This algorithm is the main feature of this paper: with a synthesis of pre-existing algorithms, we have been able to describe the persistent homology π_0 for $2, 3, 4$ and 5 disks as the radius r varies. More importantly, the advantages of using this algorithm over others is underscored by its ability to detect components in the case of 5 disks that otherwise would go unnoticed.

2. Choosing an Energy Function with the Configuration Space

From now on, given some number of disks n and radius r of these disks, let us define the configuration space $Config(n; r)$. For any $x \in [0, 1]^{2n}$, let us write x in the form $(x_1, \ldots, x_n)$ so that $x_k \in \mathbb{R}^2$ represents the center of a particular disk for each $1 \leq k \leq n$. Then we define the configuration space to be the set

$$(1) \qquad \{x = (x_1, \ldots, x_n) \in [0, 1]^{2n} \mid d(x_i, x_j) \geq 2r \text{ for any } i \neq j\}$$

where d is the Euclidean distance for $\mathbb{R}^2$. Note there is a subtle detail in this definition of the configuration space; if there were n hard disks with radius r embedded in the unit square, the coordinates of the center of each disk would have to be between r and $1 - r$. Later in this paper when we work out calculations for explicit n, we will introduce the notation of $ConfigBox(n; r)$ to address this subtlety. However, in order to alleviate the annoying technicality of having the boundary of this space vary with r when we only care about the connectedness of this space, we instead let the centers' coordinates vary from 0 to 1. Because π_0 is invariant under any scaling of the entire space, we can scale down these coordinates appropriately later to embed them in $[r, 1 - r]^{2n}$ without altering the

connectedness. Also, on another note, by granting $Config(n; r)$ the subspace topology from the bigger Euclidean space, we can talk about the topological structure of the space.

In order to work with hierarchical clustering, we next need some energy function $E : [0, 1]^{2n} \to \mathbb{R}$. So given n disks each with radius r, one possible candidate is

$$(2) \qquad E(x) = \begin{cases} 0 & : x \in Config(n; r) \\ 1 & : x \notin Config(n; r) \end{cases}$$

The problem though is that this energy function depends on r and worse yet, it is not differentiable on $[0, 1]^{2n}$. However, by softening the disks—that is, allowing the disks now to overlap—while we do lose some information about the space information we can utilize a family of functions that are differentiable:

$$(3) \qquad E_m = \sum_{i<j} \frac{1}{d(x_i, x_j)^{2m}} \text{ for any positive integer } m.$$

We claim these functions behave like the candidate $E(x)$ in (2) as $m \to \infty$. Suppose $y, z \in [0, 1]^{2n}$. Let us define $r_y = max\{r \in \mathbb{R} | y \in Config(n; r)\}$ and similarly for r_z. Note that for any $m \in \mathbb{N}$,

$$(4) \qquad E_m(y) = \sum_{i<j} \frac{1}{d(y_i, y_j)^{2m}} \geq \left(\frac{1}{2r_y} \right)^{2m}$$

and similarly

$$(5) \qquad E_m(z) = \sum_{i<j} \frac{1}{d(z_i, z_j)^{2m}} \leq \binom{n}{2} \left(\frac{1}{2r_y} \right)^{2m}$$

Using these inequalities, we note that if $0 < r_y < r_z$, then we must have

$$(6) \qquad \frac{E_m(y)}{E_m(z)} = \frac{\sum_{i<j} d(y_i, y_j)^{-2m}}{\sum_{i<j} d(z_i, z_j)^{-2m}} \geq \frac{(2r_y)^{-2m}}{\binom{n}{2}(2r_z)^{-2m}} = \binom{n}{2}^{-1} \left(\frac{r_z}{r_y} \right)^{2m}$$

which tends to infinity as $m \to \infty$. Thus once m is large enough, we see that $E_m(y) > E_m(z)$ if and only if $r_y < r_z$, and moreover, configurations that have a smaller minimum distance between disks are penalized increasingly more as m increases. Practically speaking, this means that as we increase m, we are hardening the disks. The greatest advantage of using these functions is that they are differentiable whenever defined and now permit flows along the gradient.

3. Identifying Stable Equilibria (Vertices)

Now that we have specified our particular family of energy functions, we can transform the problem of determining the structure of this amorphous, subspace of $\mathbb{R}^{2n}$ into a discrete problem. Note that as r decreases and new components emerge, they will actually be local minimums for the above energy functions provided m is large enough. Hence we really only need to focus on the various local minima.

In order to attain these, we use an algorithm denoted flowGradient. We start by picking a point p_0 uniformly randomly from the set $[0, 1]^{2n}$. After specifying some small step size $\epsilon > 0$ and energy function E_m, we recursively generate a sequence of points $p_k \in [0, 1]^{2n}$ by

$$(7) \qquad\qquad p_{k+1} = p_k - \frac{\bigtriangledown E_m(p_k)}{\mid \bigtriangledown E_m(p_k)\mid} \cdot \epsilon.$$

By jumping the distance ϵ in the direction opposite the gradient $\bigtriangledown E_m$ recursively, eventually a local minima around p_0 will be attained. It is worth saying here that if one of the coordinates of a point p_k does not lie in the interval $[0, 1]$, we simply force this coordinate to be 0 if its negative and 1 otherwise. This projects any point outside of $[0, 1]^{2n}$ directly back into this space. If this process eventually becomes constant, we call this limit point an equilibrium. However, if the equilibrium is an isolated point in $Config(n; r)$ for some $r > 0$, we call this point a stable equilibrium. These will be the equilibria at immediate interest. In fact, while a local minimum is always an equilibrium, an equilibrium may not always be a local minimum.

One example of this is the formation with five disks introduced later called ThreeCorner-Pents. Though it is an equilibrium, it is not a local minimum. However, this need not be a bad thing. Including these types of formations does not ever hurt our process, and in fact sometimes they improve the path-detecting process. Regardless, by flowing along these gradients, we are able to turn the problem of examining $Config(n; r)$ for all r into a finite problem of finding the paths connecting these equilibriums with the lowest energy threshold.

Finally, there is one huge benefit of using this flowGradient method instead of just trying to sample randomly from the configuration space. When r approaches the values where $Config(n; r)$ becomes empty, the probability of obtaining a point in $Config(n; r)$ by sampling uniformly randomly from $[0, 1]^{2n}$ is nearly zero. Given that at these larger radii r most of the important phase transitions are taking place, it is crucial to utilize algorithms that capture these unlikely points in the space. Thus while there is no guarantee that by using the flowGradient method all these equilibria will be found, it is by far a better option in order to understand the connectedness of these spaces.

4. Algorithm to Detect Low Energy Paths

The ultimate goal of the *hybrid algorithm* is to find the path between two points P_0 and P_1 in the configuration space which requires the smallest maximum energy over all points along the path. That is, amongst the set of all continuous parameterizations $P =$
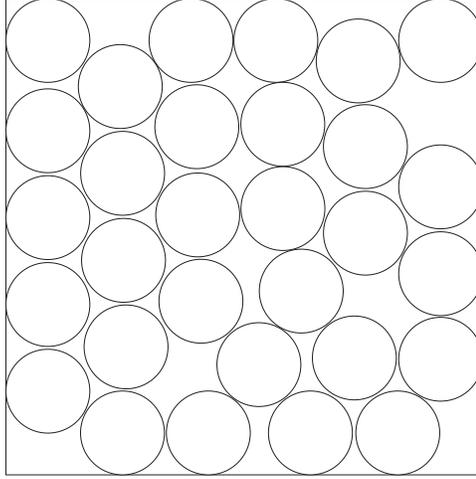
FIGURE 1. This is an example of an equilibrium for 30 disks found after running the gradientFlow 2000 times with a jump size of .005. Here the radius of each disk is .088.

$\{f : [0,1] \rightarrow Config(n;r) \mid f(0) = P_0 \text{ and } f(1) = P_1\}$, we wish to find $\gamma \in P$ such that $sup\{E(\gamma(t)) : 0 \le t \le 1\} = inf\{sup\{E(f(t)) : 0 \le t \le 1\} : f \in P\}$.

At first glance, this is exactly what an algorithm by the name of the Nudged Elastic Band Method appears to do. Given some type of differentiable energy function E, the algorithm begins with the path as a linear interpolation of N points between P_i and P_f. Much like Newton's Method, the algorithm then tries to adjust the path recusively so in the end for any i

$$(8) \qquad \bigtriangledown E(P_i) \cdot t_i)t_i = \bigtriangledown E(P_i) \text{ where } t_i \text{ is the unit tangent vector at } P_i.$$

If this holds for all points along the path, then the path is at least in a local minimum energy threshold. So to accomplish this computionally, at each step in the iteration, the algorithm proceeds as follows: given a set of N points along the path $\{P_1, P_2, \ldots, P_N\}$, beginning with $i = 2$, we calculate the direction vector

$$(9) \qquad D_i = (\bigtriangledown E(P_i) \cdot t_i)t_i - \bigtriangledown E(P_i).$$

Here the tangent unit vector t_i is calculated by normalizing the vector $P_{i+1} - P_{i-1}$ to have length 1. Thus by calculating D_i for each i and then adding these changes to each of corresponding points $P_2, P_3, \ldots, P_{N-1}$, we can move the path toward the optimal one. This process continues until (7) is reached.

There are two problems with solely using this method. First, if there is some symmetry along the trajectory of the path, the Nudged Elastic Band Method may actually rip the path apart rather than optimizing it. Though this is not obvious, this can happen for paths which are symmetrical with respect to their endpoints. Second and more importantly, this

method only returns the closest local minimum path. Because the ideal path between two points could be quite long and complicated, we need something that is better at capturing the absolute minimum.

Both these problems are handled quite well by simulated annealing. Simulated annealing is a recursive random process that is very good at closing in on the absolute minimum without getting trapped in local mins. [3] Simulated annealing is actually just a special case of the Metropolis algorithm. The general idea is to first define some objective function F that is trying to be optimized and a desired maximum length L you would like to move in a single iteration. Then given some point p, we pick some random point p' in the open disk $D(p, L)$, and if $F(p') < F(p)$ then we replace p with p' and repeat. If not, then we flip a weighted coin that comes up heads with probability $\frac{F(p)}{F(p')+F(p)}$, and if the coin does show up heads, we still replace p by p'. The strength of this algorithm is its ability to jump out of local minimums and continue flowing toward a global minimum.

In order to adapt this into our situation, we have made some slight changes. First rather than defining the objective function to the maximal energy among all the points on the path, we have instead defined the objective function to be an approximation of the line integral I:

$$(10) \qquad I = \sum_i |P_{i+1} - P_i| \cdot E_m\left(\frac{P_i + P_{i+1}}{2}\right) \approx \int_P E_m(t)\, dt.$$

The reason for using the line integral is that if we only used $I = sup\{E(P_i))\}$, we would frequently be adding noise to parts of the path that wouldn't improve anything. Hence in order to only reward energy lowering deviations, we used the line integral as the objective function. As an added bonus, it also penalizes moves that stretch the path out too much that would potential rip the continuity apart.

Second, our choice of p' is made by randomly picking one of the points P_k on the path, and then moving that the center of that disk by adding some randomly-chosen vector $v \in [0, 1]^2$ normalized to have some randomly-chosen length from $[0, L]$. In order to promote continuity, each point along the path is also changed by a scalar multiple of v, with the points closer to the initial point being moved more than others. This prevents the path from becoming too disconnected. Though rare, it is worth noting here that whenever we get a path which contains points whose coordinates lie outside of $[0, 1]$, we simply restart the algorithm from the beginning.

Third, throughout the entire process, we keep track of the best found path from all iterations, and we use this path after running the desired number of iterations. This allows us to use the best found path, and not just the final result from the algorithm.

Accordingly, the hybrid algorithm is a combination of these two methods to generate the optimal path: by first picking two points and defining the initial path to be the linear interpolation of N points between them, we run the Metropolis algorithm to get us close to the optimal path, and then zero into the optimal path via the Nudged Elastic Band. The result of combining these two methods is surprisingly powerful. As a case study, we

present the case of determining $|\pi_0|$ for 5 disks in a box. This was also accomplished for 3 and 4 disks, but the case with 5 disks is the most interesting.

5. CASE STUDY: 5 DISKS

As outlined by Gunnar Carlsson, there are algorithms that have proven to be quite powerful to determine the persistent homology of numerous topological spaces [2]. By using randomly selected data points from a manifold to build a simplicial complex that emulates the topological structure of the space, many topological properties from the data can be extrapolated. In fact, using one computer automated version of this algorithm—denoted JPLEX—it was possible accurately illustrate the persistent homology of $Config(n; r)$ when $n = 2, 3$, and 4.

However, as for the case of $n = 5$ disks, the algorithm began producing some very irregular results. In order to utilize JPLEX, first $10,000$ points were uniformly randomly sampled from the space $Config(n; r)$ for some r, and then the first two Betti numbers (β_0, β_1) were calculated for various values including the neighboring radius paramater R_{max} and number of landmark points. Landmark points were a set of points chosen inductively, where the first one was randomly chosen from all available points, and the kth landmark point was chosen to maximize the distance to the next closest landmark point. The values for R_{max}, listed in the first column, were obtained by taking fractions of the distance R which measured how finely the landmark points covered the dataset. The values in the first row are the number of landmark points chosen to represent the data.

		Landmark Points				
		50	100	150	200	250
	$R/30$	$(1, 0)$	$(1, 2)$	$(1, 18)$	$(1, 74)$	$(1, 113)$
	$R/20$	$(1, 0)$	$(1, 0)$	$(1, 15)$	$(1, 52)$	$(1, 64)$
R_{max}	$R/15$	$(1, 0)$	$(1, 0)$	$(1, 2)$	$(1, 13)$	$(1, 44)$
	$R/12$	$(1, 0)$	$(1, 0)$	$(1, 1)$	$(1, 9)$	$(1, 43)$
	$R/8$	$(1, 0)$	$(1, 0)$	$(1, 0)$	$(1, 0)$	$(1, 4)$

Note here $\beta_0 = |\pi_0|$ was always 1, meaning that the space was supposedly always connected, which will be shown later to be false when $r = .18$. Unfortunately, JPLEX here mistook the fractured structure of this space as porous rather than splintered into numerous little clusters. One likely explanation for this discrepancy is that by randomly sampling from this space, many of the small disconnected components were simply not observed. The data points were obtained randomly picking 5 points in the unit square and finding the minimum distance amongst all pairs of points. If this distance was greater than $2r$, the point was made into a 10-tuple and recorded; otherwise the point was dropped. The process continued until $10,000$ points were choosen. Because of this naive approach of sampling, only roughly one out of every million iterations was successful. Hence not only was sampling randomly becoming futile, but JPLEX was having a tough time distinguishing the very close, though disconnected components. This inspired a new approach.
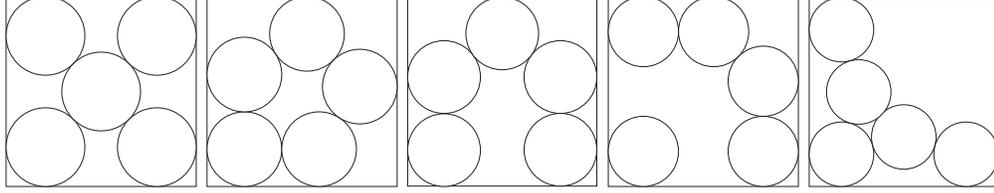
FIGURE 2.

Configuration	Energy threshold	radius
X formation	$E_X \approx 132$	$\dfrac{1}{2 + 2\sqrt{2}} \approx .207$
OneCornerPent	$E_{OCP} \approx 373$	$r_{OCP} = \dfrac{6 + \sqrt{2} - \sqrt{2 + 4\sqrt{2}}}{18 + 4\sqrt{2}} \approx .196$
House	$E_H \approx 436$	$r_H = \dfrac{5}{26} \approx .192$
ThreeCornerPent	$E_{TCP} \approx 626$	$r_{TCP} = \dfrac{1}{4 + \sqrt{2}} \approx .185$
Lambda	$E_\lambda \approx 3577$	$r_\lambda = \dfrac{1 + \sqrt{2} - \sqrt{3}}{4} \approx .171$

Before jumping into the substance of the case of 5 disks, for the simplicity of notation, let us define $ConfigBox(n; r)$ in the same manner the configuration space was originally defined, but this time notice the coordinates must lie between r and $1 - r$:

$$(11) \qquad ConfigBox(n; r) = \{(x_1, \ldots, x_n) \in [r, 1 - r]^{2n} | d(x_i, x_j) \geq 2r \ \forall i \neq j\}$$

Note now $x_i \in [r, 1-r]^2$ for each i. In this way, while the computations will be run with the ideas presented in the theoretically friendly $Config(n; r)$, by scaling down these final results we can extend these outcomes to the unit box.

Now in order to attain a set of stable equilibria, the flowGradient method was run $25,000$ times with a jump size of .0125, 12000 iterations per outcome, and an energy exponent of both $m = 5$ and $m = 25$. From these $50,000$ points in $\mathbb{R}^{10}$, removing all duplicates and symmetries yielded 5 distinct stable equilibria. The five sets of equilibra are shown in Figure 1 with their corresponding names and energy levels when $m = 5$. As mentioned above, note that the Lambda formation is a very unlikely candidate and would be very difficult to capture by simply randomly sampling from the topological space.

The next step was to determine the paths between these equilibria and their symmetric counterparts that actually affected the connectedness of the space. For each of these equilibria, a corresponding matrix was created that contained all that equilbrium's symmetries and permutations. Thus the matrix containing all the permutations and symmetries of the X formation denoted AllXs was only a 120×10 matrix, while the four other matrices were 480×10. Finally using the Hybrid algorithm outlined in the past section, we calculated the minimum energy threshold needed for each path between each pair of the five equilibria and their permutations.
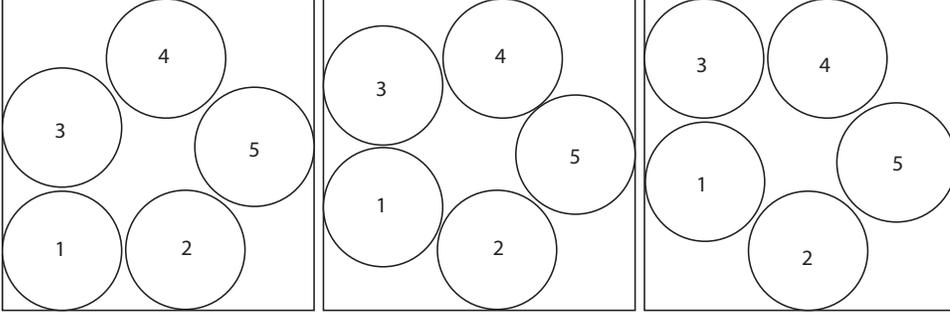
FIGURE 3. **Path 1** The least-demanding energy path between two disjoint components, as it only requires $E_{Path1} \approx 419$ as an energy threshold. The maximum radius for which this path exists is $r_{Path1} \approx .1913$.

Given that only a few of the paths actually would have small enough energy thresholds to make an impact, we only focused on paths that returned an energy threshold less than $10^3 = 1000$. While this was far below the energy level needed for the Lambda formations, these formations were isolated enough from the other equilibria that they really comprised their own separate case, which we will examine later. After moding out the paths that were equivalent up to permutation or symmetry, there were only 9 distinct paths.

First note decreasing the radius r toward zero and increasing the maximum energy threshold are analogous for the purpose of understanding the evolution of $ConfigBox(n; r)$. Thus to simplify how the evolution of $|\pi_0|$ unfolds, we will only be concerned with raising the energy threshold E from 0 until all points and paths have been accounted for. We will partition the interval $[0, \infty)$ into other intervals so that as E passes through these intervals, the value of $|\pi_0|$ may be affected.

Interval 1: $E \in [0, E_X)$. Since no points have an energy less than E_λ, the space is empty here.

Interval 2: $E \in [E_X, E_{OCP})$. At this point, the X formation finally emerges. Since $E < E_{OCP}$, there are only the X formation and its permutations. Moreover, as no paths have yet come into play, they are all disjoint. Hence $|\pi_0| = 5! = 120$.

Interval 3: $E \in [E_{OCP}, E_{Path1})$. Now the OneCornerPents have emerged, and because no paths between any points have not been established, these add another $4 \times 5! = 480$ new components. Thence we have $|\pi_0| = 120 + 480 = 600$.

Interval 4: $E \in [E_{Path1}, E_{Path2})$. At this point, the first edges finally come into play, as illustrated in Figure 3. Surprisingly, if this rotation of the five disks is iterated multiple times, we see that not only can we connect the two components where the path starts and ends, but all the OneCornerPents which maintain the same cyclical ordering of disks are connected. To be specific, suppose we label all the disks 1 through 5. Then two components are unconnected if and only if the 5-cycles in S_5 formed by reading off the disks' numbers clockwise are not equal. Accordingly, these 480 disjoint components immediately collapse into $4! = 24$ unconnected ones. So we see $|\pi_0| = 120 + 24 = 144$.
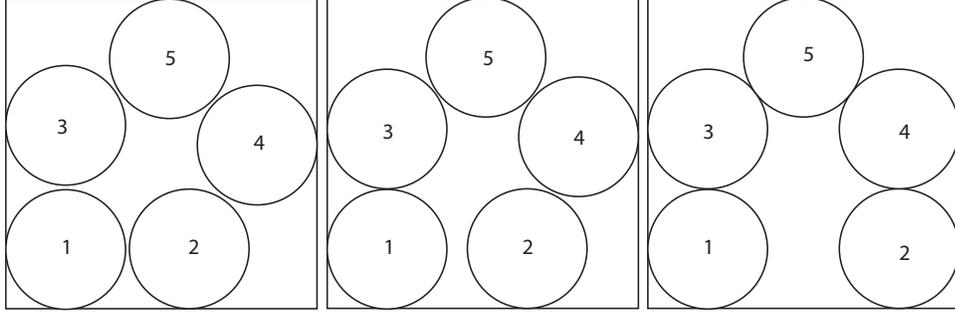
FIGURE 4. **Path 2** This path connects each House formation to a nearby OneCornerPent formation. Note the path is monotonic with respect to the energy function, and thus this path exists whenever $r \leq r_H$.

Interval 5: $E \in [E_{Path2}, E_{Path3})$. Interestingly enough, the emergence of the next groups of formations and paths do not change the picture at all. We will handle this in steps. Note that as soon as E reaches E_H and admits the House formations, Path 2 simultaneously emerges connecting these formations to the nearby OneCornerPents. For any House formation, the Hybrid Algorithm detected a monotone, energy increasing path to that configuration from a nearby OneCornerPent. Hence $E_H = E_{Path2}$, and $|\pi_0|$ remains unchanged.

The same is true for the ThreeCornerPents. As soon as E reaches E_{TCP} and these formations appear, they are connected by monotone energy-increasing paths from a nearby OneConerPent. Again, the beauty of the path being monotonic is that as once this configuration becomes possible for a given radius, it is simultaneously connected to one of the 24 circles for that exact same radius.

So we have seen that for $E \in [E_{Path1}, E_{Path3})$, while new formations emerge, the number of unconnected components $|\pi_0|$ remains at 144. In fact, while the Hybrid Algorithm returned numerous paths with energy thresholds less than 1000, all of those with thresholds less than 900 were some composition of $Path1, Path2$, and paths connecting the ThreeCornerPents. Hence even with our energy threshold E up to 900, there is no change in the number $|\pi_0|$.

Interval 6: $E \in [E_{Path3}, E_\lambda)$. Here the picture finally changes. At this energy, there is a viable path from a OneCornerPent formation to an X formation, as shown in Figure 5. Hence any disk can be shuffled into the center from a OneCornerPent formation, and accordingly we end up connecting the entire space! So for the first time, the entire space is connected, and $|\pi_0| = 1$. However, this is not the end of the story.

Interval 7: $E \in [E_\lambda, E_\lambda + \epsilon)$. At the energy level of $E = E_\lambda$, the Lambda formations finally appear and are actually unconnected from themselves and the other massive connected component. This will imply for a brief instant $|\pi_0| = 481$, and then $|\pi_0|$ will return to 1 for all greater E. Surprisingly, we see $|\pi_0|$ is not monotonic E increases, even if E becomes large!
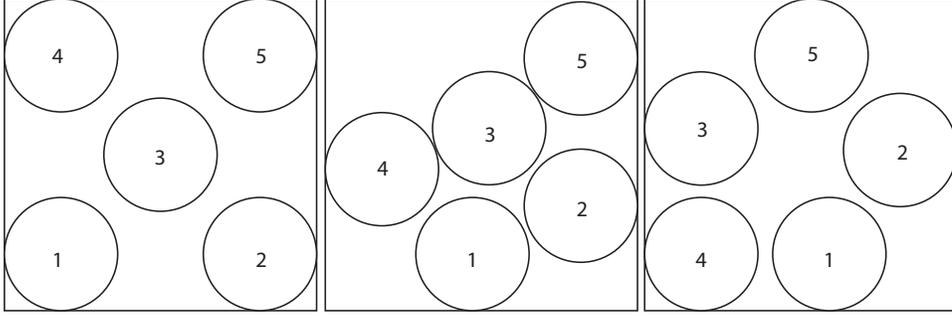
FIGURE 5. **Path 3** This path finally comes into existence when $r = r_{Path3} \approx .1815$ and $E \approx 914$. Note that this connects the entire space once it is admitted.

In order to prove the Lambda formation adds 480 disjoint components for a small interval of E, we proceed in two steps. First we will show that for the maximal radius r_λ, the Lambda formation is indeed rigid. By rigidity, we mean by interpreting each Lambda formation as a point in $ConfigBox(5, r_\lambda)$, that point is an isolated point. This is the subject of our first theorem:

Theorem 1. *The Lambda formation is rigid; that is, for some sufficiently small $\epsilon > 0$, if we denote P as one of the 480 distinct Lambda formations, then $ConfigBox(5, r_\lambda) \cap D(P, \epsilon) = \{P\}$.*

Proof: It suffices to show that if we appropriately stretch and translate P so that its exterior coordinates are 0 and 1 instead of r_λ and $1 - r_\lambda$, the new point we call P_0 is rigid in the original configuration space $Config(5; r)$. Let us call the new radius after stretching and translating r_λ to be R, and let us label the five circles' centers such that the three outer centers now lie at the points $p_1 = (1, 0), p_2 = (0, 0)$ and $p_3 = (0, 1)$, while the other two centers p_4 and p_5 are the centers lying inside the triangle $\triangle p_1 p_2 p_3$ as shown in Figure 6. By convention, $P_0 = (p_1, p_2, \ldots, p_5)$.

Suppose that we move each of the p_i to another point $q_i \in [0, 1]^2$, and in this way P_0 is moved to $Q = (q_1, \ldots, q_5) \in [0, 1]^{10}$. We show that for a small enough ϵ, the only possible choice for $Q \in D(P_0, \epsilon) \cap Config(5, R)$ is P_0 itself. This will be done in two steps. First we will show that if $Q \neq P_0$, then either q_4 or q_5 must have changed.

To see why, consider the open disc $D(p_4, \delta)$ for some very small $\delta > 0$. As $\delta \to 0$, the set $D(p_4, \delta) \cap D(p_1, 2R)$ nearly becomes the entire half of the disk $D(p_4, \delta)$ that is closest to p_1. This same relationship holds if we replace p_1 with p_2. Now note that the angles $\angle p_1 p_4 p_2$ and $\angle p_1 p_4 p_5$ are actually $\frac{5\pi}{6}$ radians, and so by picking a small enough ϵ, we can ensure

$$(12) \qquad \overline{D(p_4, \epsilon)} - [D(p_1, 2R) \cup D(p_2, 2R) \cup D(p_5, 2R)] = \{p_4\}$$
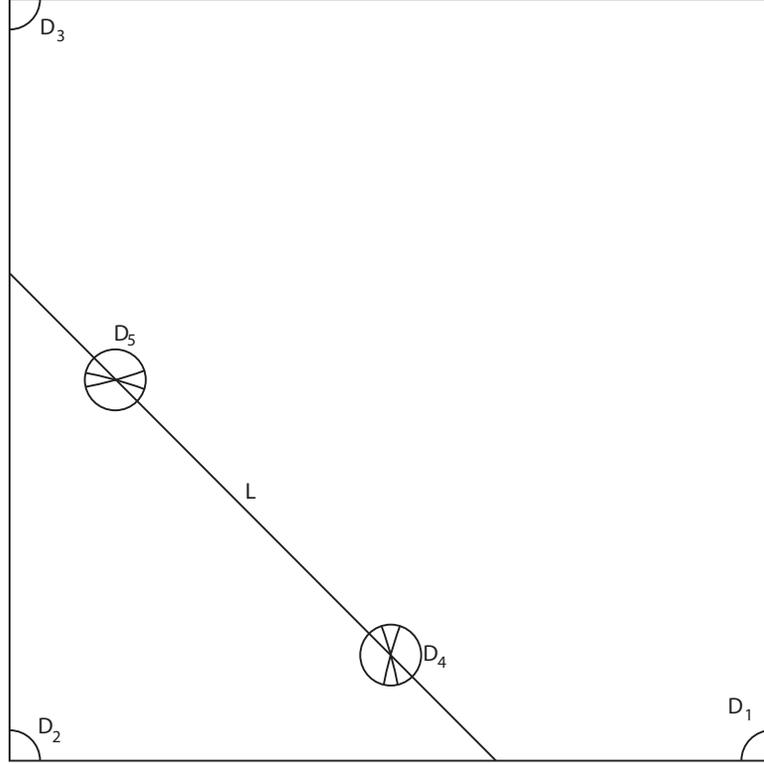
FIGURE 6. Illustration for the following proof.

By exchanging p_4 with p_5 and using p_2, p_3, and p_4 instead of p_1, p_2, and p_4, we can make the same conclusions. Most importantly, for this same ϵ, we have

$$(13) \qquad \overline{D(p_5, \epsilon)} - [D(p_2, 2R) \cup D(p_3, 2R) \cup D(p_4, 2R)] = \{p_5\}$$

Now define regions $D_1, D_2, \ldots, D_5$ such that $D_i = D(p_i, \epsilon) \cap [0, 1]^2$ for all i. We claim moving just one of the five centers q_i from q_i to another point in the region D_i means that for some $i \neq j$, the distance between q_i and q_j is less than $2R$. Note that if $Q \neq P_0$ and neither q_4 nor q_5 has moved then least one of q_1, q_2, or q_3 has moved within its respective region. But as $D_1 - D(p_4, 2R) = \{p_1\}, D_2 - D(p_4, 2R) = \{p_2\}$, and $D_3 - D(p_5, 2R) = \{p_3\}$, moving q_1, q_2, or $, q_3$ would force $Q \notin Config(5, R)$.

Thus at least q_4 or q_5 has moved also. Next we prove our second step: we show that regardless of where q_1, q_2, and $, q_3$ move within their regions, the only places q_4 and q_5 can move without getting within $2R$ to these exterior centers is toward one another. This will conclude the proof that P is rigid.

Note that for any $q_1 \in D_1$, we have $D(p_1, 2R) \cap D(p_4, \epsilon) \subseteq D(q_1, 2R) \cap D(p_4, \epsilon)$. To see why, note that $D_1 - D(p_4, 2R) = \{p_1\}$. Hence by symmetry, we see that no matter

where the centers q_1, q_2, and q_3 move, the centers of the inner disks q_4 and q_5 must lie in the regions $S = D_4 - [D(p_1, 2R) \cup D(p_2, 2R)]$ and $T = D_5 - [D(p_2, 2R) \cup D(p_3, 2R)]$ respectively.

Let L be the line through p_4 and p_5. We claim both S and T lie on the same side of L. Because the angle between L and the line through p_4 and p_2 is $\frac{\pi}{3}$ radians, by choice of our small ϵ, we can be sure $\partial D(p_4, \epsilon) \cap S$ does not intersect L. By symmetry, the same argument holds for T, and thus we know both S and T lie above L.

Now we claim using this fact, we can show for any $q_4 \in S$ and any $q_5 \in T$, we have $d(q_4, q_5) \le 2R$. First note that if q_4, q_5, and p_5 were collinear, then because $d(q_4, p_5) \le 2R$ and q_5 must lie between q_4 and q_5, we would have $d(q_4, q_5) < 2R$. But if these three points were not collinear, then the angle $\angle q_4 p_5 q_5$ must be acute. Here we use the fact that both S and T lie on the same side of L, and $S \subseteq D(p_5, 2R)$ and $T \subseteq D(p_4, 2R)$. Remember, by choosing ϵ sufficiently small, we know $\angle p_5 q_4 q_5$ is also acute. Thus as these two angles are acute, we have that $d(q_4, q_5) \le d(p_5, q_4) \le 2R$. But note if we have equality then either $q_4 = p_4$ or $q_5 = p_5$. Thus w.l.o.g. suppose $q_5 = p_5$. Then $S - D(q_5, 2R) = \{p_4\}$, and so $q_4 = p_4$. Thus we see the only possible for $q_4 = p_4$ and $q_5 = p_5$, and hence $Q = P_0$ as desired.

$\square$

Thus we know that indeed, the Lambda formation is rigid exactly when the radius is r_λ. However, we can assert something a bit stronger. We claim there is some $r^* < r_\lambda$ such that for all $r \in (r^*, r_\lambda]$ that the space $ConfigBox(5; r)$ has no less than 480 disjoint components.

Theorem 2. *There is some positive real $r^* < r_\lambda$ such that for any $r^* < r \le r_\lambda$, the space $ConfigBox(5; r)$ retains the 480 unconnected components present in $ConfigBox(5, r_\lambda)$.*

Proof: Let $X = [0, 1]^{2n}$, and let $B(r, \epsilon) \subseteq X$ such that

$$(14) \qquad B(r, \epsilon) = \{x \in X | \text{ for some } y \in ConfigBox(n, r) \text{ we have } d(x, y) < \epsilon\}$$

Essentially, $B(r, \epsilon)$ is an open subset of X such that a point is included in $B(r, \epsilon)$ if and only if it is within ϵ of some point in the configuration space associated with r. Now, given the definition of $ConfigBox(n, r)$, we know that for any n and r that this set is semialgebraic. Accordingly, invoking Hironaka's triangulation theorem [7], we know that this space is homeomorphic to some simplicial complex $\mathcal{C}$. Moreover, note that $ConfigBox(n, r)$ is a closed subset of X, and as X is compact by the Heine-Borel theorem, we know $ConfigBox(n, r)$ must be compact as well.

However, note that this implies $\mathcal{C}$ must also be compact, as homeomorphisms preserve compactness. But we know each component of a simplicial complex is both an open and closed subset of the simplicial complex, and thus $\mathcal{C}$ must have a finite number of components. For the set of disjoint components form an open cover of $\mathcal{C}$, and because we know there must be a finite subcover, the number of components of $\mathcal{C}$ must be finite. But this implies the number of components of $ConfigBox(n, r)$ must be finite too.

Thus given $n = 5$ and $r = r_\lambda$, because there are only a finite number of components, we can pick some $\delta > 0$ small enough so that $B(r_\lambda, \delta)$ is still composed of 480 disjoint components. Generally, suppose we identified the T disjoint components of $ConfigBox(n, r)$ as $c_1, c_2, \ldots, c_T$, and let

$$(15) \qquad \epsilon_{i,j} = \inf\{d(x, y) | x \in c_i \text{ and } y \in c_j\}$$

We know $\epsilon_{i,j} \geq 0$, and moreover, we claim $\epsilon_{i,j} > 0$. If not, then we would have some sequence of $\{(x_k)\} \subseteq c_i$ and $\{(y_k)\} \subseteq c_j$ such that $d(x_m, y_m) \to 0$. But as each component is compact, we know that c_i would have to intersect c_j, and this is a contradiction. Thus $\epsilon_{i,j} > 0$. Hence to find the appropriate δ for our case, we simply take $\delta = \min\{\epsilon_{i,j} | 1 \leq i < j \leq 480\}$, which is guaranteed to be greater than 0.

Finally, consider the function $f : X \to \mathbb{R}$ which for each $x \in X$, $x \mapsto r_x = \sup\{r \geq 0 | x \in ConfigBox(n, r)\}$. Note that f is continuous. Moreover, we know that $B(r_\lambda, \delta)$ is an open subset of the closed set X, and hence $X \setminus B(r_\lambda, \delta)$ is a closed set. Thus it is also compact. This implies that f must achieve its maximum on this set, let us denote this value attained as r^*. We know that $r^* < r_\lambda$, for otherwise the point $p \in X \setminus B(r_\lambda, \delta)$ giving $f(p) = r^*$ would also be in $B(r_\lambda, \delta)$.

By construction of $B(r_\lambda, \delta)$, we see that any path connecting two components must intersect $X \setminus B(r_\lambda, \delta)$, and for this interval of the path f must attain the value r^*. Thus for any $r \in (r^*, r_\lambda]$ we know all the original paths must remain disjoint, as desired.

$\square$

As an aside, one can prove an even stronger result. In fact, given any positive integer n and r, for some interval $(r^*, r_\lambda]$ the homotopy type of $ConfigBox(n, r)$ does not change at all, and accordingly the number of components must remain exactly the same. The proof is technical and much more involved, and relies on approaches requiring either Stratified Morse Theory or Min-type Morse Theory [5] [4]. Using this piece of knowledge, we know for a short interval above E_λ, $|\pi_0|$ jumps up to 481, and then quickly returns back to 1. As there are no more formations that have yet to be added, we see that the space then remains connected for all larger E. Hence we have a complete understanding of the evolution of $|\pi_0|$ over all positive E, even though we do not know the exact values r where π_0 changes.

The dendrogram in Figure 7 illustrates the evolution graphically. Interestingly enough, we see that $|\pi_0|$ is not unimodal with respect to E. The second observation is that even if the space $ConfigBox(n; r)$ becomes completely connected, there are possibly smaller r for which the space becomes disconnected again.

6. Concluding Thoughts

As demonstrated in the previous analysis, utilizing an algorithm combining methods in random processes and gradient flows was not only powerful at dispelling the topological features of the space, but both its two pieces were crucial for the success of the algorithm. Without using stimulated annealing first, the nudged elastic band method sometimes ripped paths apart when the endpoint could be reached by a symmetrical reflection of the starting
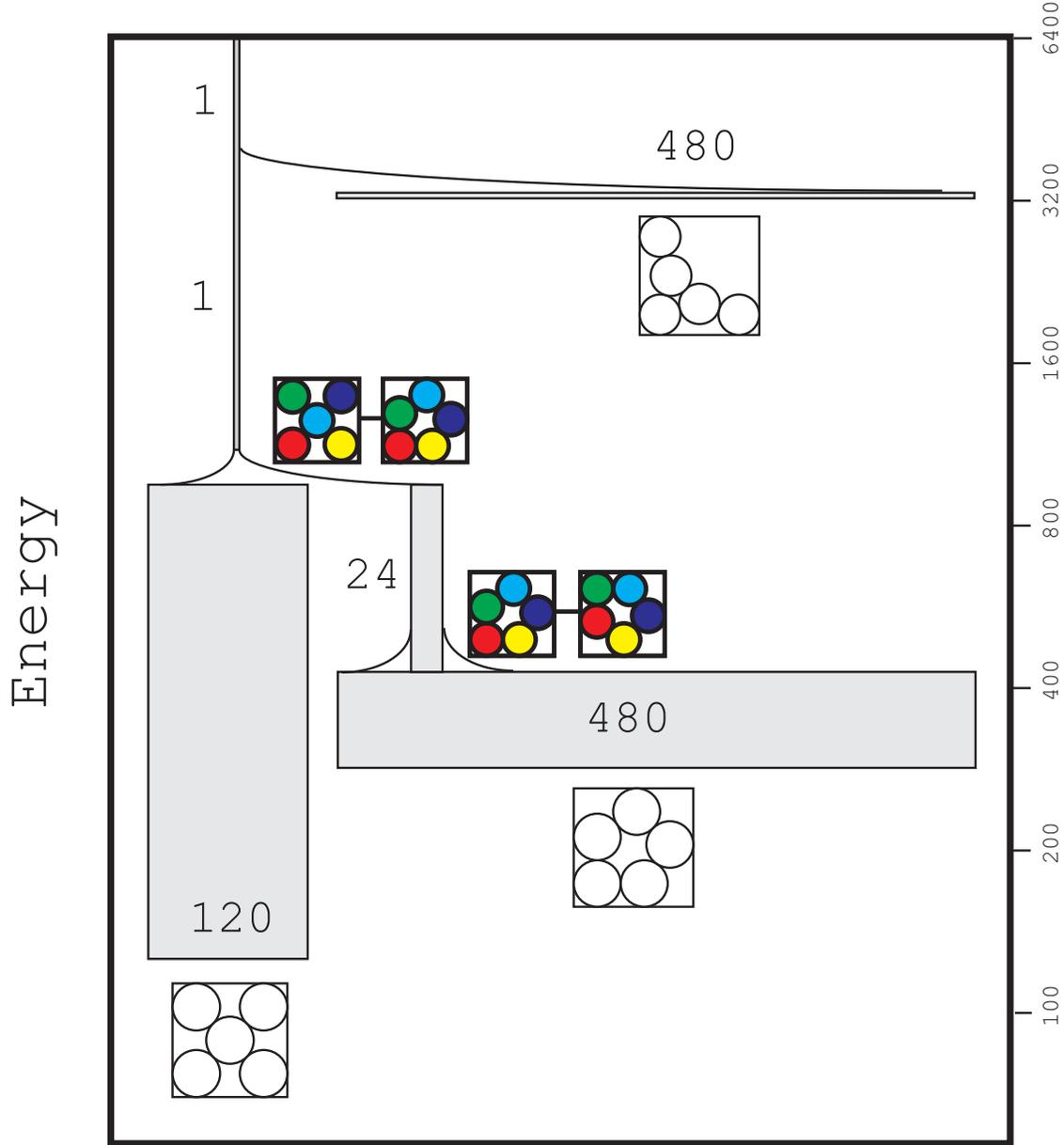
FIGURE 7. The above dendrogram illustrates the evolution of π_0 as E increases from 0. The shaded boxes represent actual components, while the curved pieces represent the edges adjoining components.

point. Moreover, because the elastic band method zeroed in on the closing minimum to the linear interpolation between the starting point and end point, there was no guarantee that the produced path would necessarily be optimal. For these two reasons, stimulated annealing was crucial in breaking the symmetries and moving the path closer to a global minimum.

The elastic band method was also essential. Though simulated annealing is great at moving in toward a global minimum, the path is slightly perturbed from the actual minimum path it was near. Thus because knowing the maximum energy threshold for each path was necessary to understand the order at which disconnected components became connected, zeroing in these approximations was key. Because of the differing strengths in each of these methods, the combination of the two reinforced each other's efficacy.

As was seen in the case of 5 disks, the evolution of these topological spaces is extremely complex. In fact, while this algorithm does offer some numerical approaches to understand the topological connectedness of the space, relatively little is known on the theoretical side of these spaces. One possible avenue to build upon this work is to apply this approach to situations with more disks and also different boundary conditions (say a torus instead of a unit square). While perfectly understanding the evolution of π_0 made be unattainable, uncovering rough estimates for a large number of disks could provide valuable qualitative information relating to phase changes. Another possible avenue of research would be to apply this hierarchical clustering to other areas of statistical topology and see if this process is fruitful in other settings.

References

[1] David W. Boll, Jerry Donovan, Ronald L. Graham, and Boris D. Lubachevsky. Improving dense packings of equal disks in a square. *Electron. J. Combin.*, 7:Research Paper 46, 9 pp. (electronic), 2000.

[2] Gunnar Carlsson. Topology and data. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):255–308, 2009.

[3] Persi Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009.

[4] V. Gershkovich and H. Rubinstein. Morse theory for Min-type functions. *Asian J. Math.*, 1(4):696–715, 1997.

[5] Mark Goresky and Robert MacPherson. *Stratified Morse theory*, volume 14 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1988.

[6] R. L. Graham and B. D. Lubachevsky. Repeated patterns of dense packings of equal disks in a square. *Electron. J. Combin.*, 3(1):Research Paper 16, approx. 17 pp. (electronic), 1996.

[7] Heisuke Hironaka. Triangulations of algebraic sets. In *Algebraic geometry (Proc. Sympos. Pure Math., Vol. 29, Humboldt State Univ., Arcata, Calif., 1974)*, pages 165–185. Amer. Math. Soc., Providence, R.I., 1975.

[8] Matthew Kahle. An application of disc packing to statistical mechanics. *Arxiv*, arXiv:0908.1830v2, 2009.

[9] Werner Krauth. *Statistical mechanics*. Oxford Master Series in Physics. Oxford University Press, Oxford, 2006. Algorithms and computations, Oxford Master Series in Statistical Computational, and Theoretical Physics.