



Research Article

Effect of Physically Realistic Potential Energy Form on Spatial Pattern Complexity in a Collective Motion Model

Austin M. Marcus ^{1,2} and Hiroki Sayama ^{1,2}

¹Systems Science and Industrial Engineering Department, Binghamton University (State University of New York), Binghamton, NY 13902, USA

²Center for Collective Dynamics of Complex Systems (CoCo), Binghamton University (State University of New York), Binghamton, NY 13902, USA

Correspondence should be addressed to Austin M. Marcus; amarcus6@binghamton.edu

Received 11 January 2023; Revised 14 March 2023; Accepted 15 March 2023; Published 27 April 2023

Academic Editor: Siew Ann Cheong

Copyright © 2023 Austin M. Marcus and Hiroki Sayama. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Collective motion models most often use self-propelled particles, which are known to produce organized spatial patterns via their collective interactions. However, there is less work considering the possible organized spatial patterns achievable by non-self-propelled particles (nondriven), i.e., those obeying energy and momentum conservation. Moreover, it is not known how the potential energy interaction between the particles affects the complexity of the patterns. To address this, in this paper, a collective motion model with a pairwise potential energy function that conserved the total energy and momentum of the particles was implemented. The potential energy function was derived by generalizing the Lennard–Jones potential to reduce to gravity-like and billiard-ball-like potentials at the extremes of its parameter range. The particle model was simulated under a number of parameterizations of this generalized potential, and the average complexity of the spatial pattern produced by each was computed. Complexity was measured by tracking the information needed to describe the particle system at different scales (the complexity profile). It was found that the spatial patterns of the particles were the most complex around a specific ratio in the parameters. This parameter ratio described a characteristic shape of the potential energy function that is capable of producing complex spatial patterns. It is suggested that the characteristic shape of the potential energy produces complex behavior by balancing the likelihood for particles to bond. Furthermore, these results demonstrate that complex spatial patterns are possible even in an isolated system.

1. Introduction

The phenomenon of pattern formation or the emergence of complex organizations in a system has been studied in various types of systems (e.g., biological, economic, and social) and in various aspects (e.g., spatial and temporal) [1–3]. When the system is composed of discrete particles and their spatiotemporal patterns at various scales are of interest, models of collective motion are used [4–7]. These methods model discrete particles explicitly, which makes analytical representations challenging, and so they are often explored via simulation. Typically, these models are interested in the patterns formed by a medium number of particles and how

the pattern is formed through dynamics at and between different scales in the system. Collective motion models typically consider “self-propelled” particles, which means that particles effectively have a source of energy which they can use to affect their motion in ways that real physical particles cannot. Energy and momentum conservation is not of relevance in these models, as they are primarily interested in producing complex patterns, which are known to be more prominent in driven systems [8]. On the other hand, molecular dynamics (MD) models are largely concerned with creating physically realistic simulations of molecular behavior at the microscopic scale but are primarily concerned with different questions. For example, they may be used to

understand molecular processes relevant to biology, such as protein-folding [9, 10], but are not usually concerned with general principles of organization [11].

Thus, it can be asked what the range of collective patterns producible by a physically realistic system is and how it depends on the basic forces between particles. To investigate this, the present work considers a minimal system that still produces complex spatial patterns. Specifically, it considers a collective motion model with a conservative pair-wise interparticle potential energy function and analyzes how this affects the spatial complexity of the collective patterns produced. While some models have explored pattern formation in a similar sense in physical-like particle systems, these have not specifically investigated changing the conservative potential or quantified the spatial complexity. For instance, it has been examined how the introduction of nonconservative forces on a Lennard–Jones substance changes the equilibrium state spatial distribution [12]; however, it did not consider variations of the conservative potential form. Similarly, previous work has used MD to simulate Rayleigh–Bénard convection [13], which examined collective pattern formation, but only considered billiard-ball like interactions between particles (elastic collisions with no long-ranged force). Interestingly, other work did compute the complexity of temporal patterns of the particles in a MD simulation [14] but limited its investigation to the specific interactions of water. Additionally, numerous equations-of-state for a Lennard–Jones substance have been developed, and while some approaches do derive macroscopic behavior from microscopic rules, they often focus on a specific interaction potential and homogeneous macroscopic properties such as heat capacity, speed of sound, or isothermal compressibility [15].

Additionally, in collective motion studies, the collective patterns are often considered qualitatively; that is, by inspection, it is clear that some outcomes are more complex than others. There are many quantitative measures of complexity that have been employed to make this characterization more precise and principled, but there is no general consensus on how complexity should be measured [16, 17].

This paper thus implements a discrete particle dynamics model, where the form of the potential energy was varied across runs to observe differences in the macroscopic patterns formed. The model was conservative and isolated, in order to determine the effect on pattern formation by the potential alone. The potential energy function was derived by combining three physically relevant potential functions (gravity, billiard-ball, and Lennard–Jones) into a single, general potential function, where each behavior is recoverable via parameter changes. Each of the selected realistic potentials represents qualitatively different patterns: ordered, homogeneous, and complex, respectively. Thus, the generalized potential relates physically relevant forces to the extremes of spatial pattern complexity. Since there is no consensus on how to measure complexity, a suitable complexity measurement based on the complexity profile [1, 18] was developed to quantify the complexity of the formed patterns.

It is found that the model produces a range of spatial clustering behavior depending on the parameters, ranging from simple to complex structures. The complexity measure developed has a strong nonlinear correlation with the parameters, and the patterns it identifies as complex match with intuition. Lastly, the parameterizations of the potential that produce the maximally complex behavior are those with a specific ratio among the parameters, describing a characteristic shape of the potential function that produces complex patterns. This indicates that complex patterns are produced when a balance between the interactions in the system is met.

2. Methods

2.1. Particle Simulation Model. The motion of abstract particles under different interaction rules was numerically integrated and their motion visualized and analyzed. A 200 particle system was used as it achieved a balance between computational speed and the formation of collective behaviors. The microscopic rules defining how particles interacted with each other were defined by a pair-wise potential energy function. This defined the potential energy between any two particles, as a function of the distance between them, and thus also defined the pair-wise forces via $F = -\nabla V$. The space the particles moved in was 2D for simplicity and with periodic boundary conditions to avoid possible additional effects on collective behavior from a reflecting boundary. Initial conditions for the system were chosen randomly so that particle positions were evenly distributed across the space and their velocities were evenly distributed in all directions and between the magnitudes of 0 and 1. Arbitrary units were used and were chosen so that the system’s qualitative behavior could easily be explored. Different densities of the particles were simulated by adjusting the volume of the space, and it was observed that denser configurations produced more complex collective behavior: higher density caused a higher rate of particle interactions, which increased the probability of bonds (“Bonds” is meant not in a chemical bond sense but in whether the particle has enough energy to escape its partner) forming. Thus, only the most dense configuration tested of an average initial interparticle distance of ≈ 2 distance units, corresponding to a density of 0.0625 particles per unit area, is reported in this paper.

Energy and momentum were conserved to a sufficient degree to produce accurate qualitative behavior by tailoring the numerical integration timestep to the system’s dynamics, which was verified by computing the root-mean-square error from the initial condition at a subset of timesteps ($\overline{\Delta E} \approx -3\%$, $\overline{\Delta p} \approx 1.78 \times 10^{-12}\%$). The position Verlet method [19] was used for numerical integration. This specific method was chosen due to its computational speed and adequate energy conservation properties in this case. Each simulation was run for 25,000 timesteps with a $\Delta t = 4 \times 10^{-3}$. The system was kept isolated from an environment in this study in order to make clear how the potential energy function itself affected the collective behavior. All simulation code was written in the Julia

programming language (an open-source language under the MIT license) [20] and is available online (see Data Availability section).

The form of the potential energy function was varied across simulation runs in order to find relationships between microscopic rules and macroscopic behavior. Each parameter configuration was run on nine different, randomly generated initial conditions to smooth out statistical variation in behavior.

2.1.1. Interaction Potential. Complexity is sometimes defined as a mean between completely coherent and completely homogeneous behavior [18]. Thus, potential energy functions were chosen to produce behavior between these extremes. Gravity is an example of interaction rules that may produce completely coherent behavior, as it tends to bring all of the particles into one location. The “billiard-ball” potential, which produces purely elastic collisions with no long-range interactions, is an example of interaction rules that may produce completely random behavior, similar to the behavior of the particles in an ideal gas; particles will wander randomly until they bounce off of each other. This random behavior approximates homogeneous behavior. It was hypothesized that the mean behavior between these two extremes would be induced by bonding behavior between particles. Thus, the Lennard–Jones potential was selected.

Using these potentials as boundary behavior, they were then generalized into a single equation where each characteristic shape can be recovered via specific parameter settings. Thus, by varying the parameters of this generalized potential, specific aspects of the potential function can be associated with certain types of collective behavior. The generalized potential equation, obtained by generalizing the Lennard–Jones potential form, is as follows:

$$V(x) = \left[\text{sign}(d)s \left(\frac{|d|}{1-s} \right)^{(1-(1/s))} (px)^{-a} \right] - (px)^{-sa}, \quad (1)$$

$$p = \left[\left(\frac{1-s}{|d|} \right)^{(1/s)} s^{(1-s)^{-1}} \right]^{a^{-1}}, \quad (2)$$

where $s \in (0, 0.8)$, $d \in (-\infty, 0), (0, \infty)$, and $a \in (1, \infty)$. s , d , and a control, respectively, the range of the potential, the presence and depth of the potential well, and the magnitude of the forces generated, although the effects of each parameter are not entirely independent. The smaller s is, the more long-range the potential becomes and vice versa. When $d < 0$, there is no minimum in the potential, and so the force is only ever attractive (e.g., gravity) (generally, in physics, the convention is to give attractive forces a negative potential energy, where particles are said to be bound if their total energy is negative and they escape when it is positive. The selection of zero as the cutoff point is arbitrary; the actual value of the potential does not matter for many applications, it being the differences in potential that drive motion), and the more negative d is, the sooner the potential becomes deeper. When d is very small but

positive, the minimum is very shallow, and thus it approximates the billiard-ball potential. As $d > 0$ increases, the minimum becomes deeper, representing stronger bonds between particles. The larger a is, the sharper the potential well (if $d > 0$) or the faster the potential drops off (if $d < 0$); in both cases, the magnitude of the force increases. Scaling the x -axis by p , $V(1) = 0$ when $d > 0$, making the effective “particle size” constant—though the location of the minimum shifts slightly with different s —making comparison of behavior across configurations more meaningful, as the magnitude of the distances between particles matters for visual inspection. Figure 1 shows equation (1) for different values of s and d , which demonstrates the full range of its relevant behavior. In all simulations, $a = 10$. Lastly, equation (1) recovers the Lennard–Jones potential at $a = 12, s = 0.5, d = \epsilon > 0$. Equation (1) recovers only Lennard–Jones potentials with $\sigma = 1$, due to p , which fixes the “particle size.” Note that for $d < 0$, when interparticle distance is less than 1, the potential energy was artificially truncated in the code to become constant. This was performed to define the behavior of collisions among the particles under this parameter value, choosing to make the particles pass through each other. Without modification, particles that collided would experience a very large force (singularity in the potential), making energy-conserving simulation intractable.

2.2. Quantification of Spatial Patterns. To quantify the complexity of the spatial patterns, a complexity measure called the “complexity profile” [1] was used. This method tracks the information needed to describe a system as a function of *observation scale*, producing not a single number but a curve that monotonically decreases with scale. The shape of this curve corresponds to the system’s complexity. Complex systems can be conceived as those systems that spread out the information it takes to describe them across scales [1]. The rationale for this is that simple systems either have no correlation between their parts (homogeneous) or complete correlation between their parts (ordered), allowing their behavior to be encapsulated by a few high-level variables. However, when the correlation between the parts of the system is in between these extremes, the system is complex and encapsulation by a few variables is difficult. Scale is considered because it accounts for the irrelevance of some parts for the description of the system. Intuitively, scale has the effect of blurring; on a larger scale, coarse descriptions of a system are sufficient. For example, describing the Earth on a large scale would only need to describe the general regions of land and water. However, describing it on a smaller scale would require more detail about the terrain, weather, geology, and forms of life present. Thus, in this sense, smaller scale descriptions require more information. Complex systems, since they spread information out across scales, are those that resist easy description even as irrelevant details are “blurred out.” While previous work [1] provides this conceptual framing of complexity, it does not provide a concrete measure

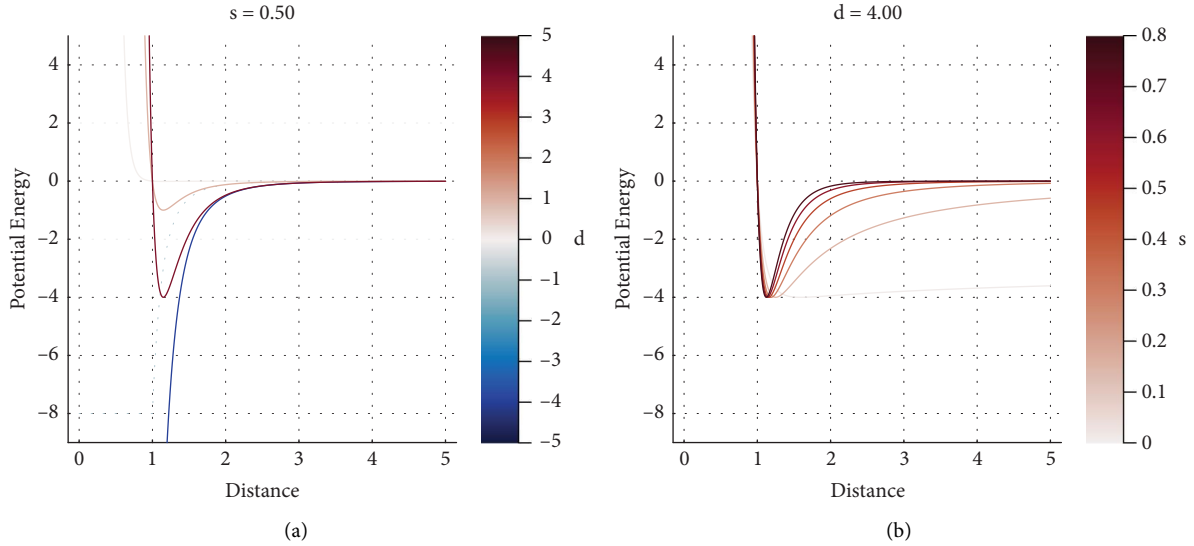


FIGURE 1: The generalized potential energy function (equation (1)): (a) d is varied, while s is held constant and (b) s is varied as d is held constant. The x -axis is the distance between two particles. The y -axis is the potential energy that exists between the two particles. In all cases, $a = 10$.

applicable here. The remainder of this section describes a complexity measure developed here inspired by this approach.

2.2.1. Spatial Information Measure. By plotting the information as a function of scale, the complexity of a system is contained in the shape of the resulting curve. Thus, the nature of the information measure used determines precisely which aspect of the system is being analyzed. The effect of increasing the scale of observation makes behavior below a certain level of coherence unobservable. In terms of a spatial distribution, this is similar to lossy image compression: if there is correlation between the color of nearby pixels, less information needs to be stored to reproduce a picture that has preserved the most coarse structures. The compressed image is different, pixel to pixel, but the essential parts of the image are still there; the large scale structure of the image is the same. As the amount of compression is increased, it is blurring the image more, which effectively says that only the largest scale information is relevant, and the amount of information needed is lower.

Thus, in the particle model considered here, the information needed to describe the spatial arrangement at a given scale was approximated by the number of clusters present, $L(x)$, and scale was defined as the minimum number of particles defining a cluster, x . This definition was chosen because, when particles are dispersed, it is necessary to specify each of their locations, but when they are clustered, only the locations of each cluster need to be specified. Thus, the effect of scale is emulated by increasing the minimum number of particles that define a cluster.

2.2.2. Spatial Complexity Measure. The complexity of a profile's curve can be computed by considering how it

changes over scale. Specifically, complex systems will have higher information content at all scales except for the smallest and largest. So, two curves can be compared by considering which one has more at intermediate scales. This means that both the information measure and the scale should grow logarithmically because this de-emphasizes effects at large values. It does so by reducing the area contributed by larger scales and the initial maximum of information present in all systems at small scales. Thus, the complexity of a given curve on the complexity profile was quantified using

$$C = \int I(\sigma) d\sigma, \quad (3)$$

where σ is scale and $I(\sigma)$ is the spatial information needed to describe the system at scale σ .

Applying this framework to the way information and scale were defined in this model gives $\sigma = \log_2 x$ and $I(\sigma) = \log_2 L(x) = \log_2 L(2^\sigma)$, and thus,

$$C \approx \sum_{x=1}^m \frac{\log_2 [L(x)]}{x \ln 2}, \quad (4)$$

where m is the maximum minimum cluster size considered (in this case, 100). In implementation, $\log_2 L(x) = 0$ was stipulated for $x \leq 1$. To compute C , clusters were identified using the DBSCAN algorithm [21] and implemented in the Julia package "Clustering.jl," part of the JuliaStats package (the software package is available from <https://github.com/JuliaStats/Clustering.jl>, under an MIT license). ϵ , which controls the density that defines a cluster, was set to 2, approximately twice the particle size. minPts , which defines essentially the minimum size of a cluster, was unconventionally set to 1 so that even isolated particles counted as a cluster rather than noise. For each parameterization and initial condition, 12 frames spread uniformly over the

second half of the simulation run (to ensure a dynamic equilibrium was reached) were selected. DBSCAN was applied to these frames individually, and then a complexity profile was computed for each by counting the number of clusters above a certain size, for all sizes (up to $m = 100$). These 12 profiles were then averaged together, and C was computed on this, defining the complexity of each parameterization and initial condition.

In order to calibrate C with respect to the least and most complex particle patterns, reference particle distributions (also with 200 particles in the same size space) were artificially generated and examined. Homogeneous and coherent distributions should be the least complex: these are patterns that are evenly spread out or concentrated in one clump, respectively.

Complex systems are those where the information it takes to describe them resists decreasing as the scale of observation increases. This corresponds to more area under the complexity profile curve and so should produce a higher C . Power-law behavior is generally associated with complexity, and so it was hypothesized that a power-law distribution in the size of the clusters would correspond with an intuitive evaluation of complexity. Specifically, clusters were randomly placed in the space, where their size was chosen according to the power-law probability density function as follows:

$$\rho(x) = \frac{1 - \gamma}{m^{1-\gamma} - 1} x^{-\gamma}, \quad (5)$$

where m is the maximum cluster size considered, x is the size of the cluster, and γ is the exponent of the power-law. In order to increase the area under the complexity profile curve, a small γ was used ($\gamma = 1.1$).

Twenty complex reference distributions were generated, and just one of each homogeneous and coherent was generated, as these latter ones do not vary. Representative images of the particle distributions are shown in Figure 2, and their C values are shown in Table 1.

3. Results

3.1. Particle Simulation. The particle model was simulated under different potential function parameter combinations, and videos of the particle’s movements were produced. Table 2 presents the parameter values used; all combinations of these parameters were simulated under nine initial conditions, resulting in 1,683 different simulations. Figure 5 shows representative frames from a subset of parameter configurations and initial conditions. There were basically three types of resultant collective behavior: the particles either clustered into one large mass (coherent regime), did not cluster at all (homogeneous regime), or they formed a number of smaller coexisting clusters (complex regime). Figure 6 depicts the potential energy function representative of each region. Note that the particles remained in constant motion in all cases, achieving a dynamic equilibrium in terms of the type of spatial patterns produced.

The coherent regime occurred when $s < 0.30$ and $d \neq 0.01$, making the interparticle potential long ranged, and it did not matter whether d was either negative or positive: in either case, gravity-like behavior was exhibited. This occurred because the long-ranged potential allowed particles far apart from each other to experience attraction, and $|d|$ was large enough for stable bonds to be formed between particles.

The homogeneous regime occurred when $d = 0.01$, across all values of s . This means that, when the potential was shallow, it did not matter how long-ranged the potential was and the particles did not form clusters. This makes sense, as this parameter region makes the attractive force between particles weak, and so the dominant mechanic is elastic collisions (or pass-through), recovering billiard-ball behavior.

The complex regime occurred when $s \geq 0.30$ and $d \neq 0.01$, forming a number of local clusters that became tighter and larger proportionally to $|d|$. This occurred because the larger s reduced the range of the potential so that distant particles were not drawn together, and the magnitude of d increased the strength of the potential between nearby particles. Thus, particles local to each other formed tight clumps and would interact with other clusters only if their incidental trajectories collided. Note that both negative and positive d allowed local clusters to form, but their clusters had different characteristics: $d < 0$ produced tighter clusters because the purely attractive force between them allowed smaller separations to occur, whereas $d > 0$ maintained a larger distance between particles via repulsion at the “particle-size” (Figure 1(b)).

3.2. Complexity Ranking. Figure 4(a) shows the complexity profile for a subset of all simulation runs (5% for visual clarity), constructed using the information measure described in Section 2.2.1 previously. Each line corresponds to a specific initial condition and parameterization of the potential energy. Curves that have a greater number of clusters at intermediate minimum cluster sizes correspond to more complex spatial arrangements of particles. Curves that correspond to less complex arrangements either start high and drop off at small scales or start low and remain low until large scales. C (equation (4)) was computed for each curve, and each curve was colored accordingly. Specifically, the upper and lower bounds of the color scale were defined by the maximum and minimum C observed across both the benchmark and simulated distributions.

Figures 4(b) and 4(c) show the average complexity profile for each regime in simulation and in the benchmark distribution. As expected, the power-law benchmark is the most complex according to C and has the highest values at intermediate scales. The complex regime in simulation is close to this curve, and it can be seen that as C increases, the curves approach the shape of the power-law benchmark curve. Note that the homogeneous and coherent regimes in simulation have much higher C than their benchmark counterparts. This shows that the ideal simple behavior is not achieved completely by the simulation, but this nevertheless

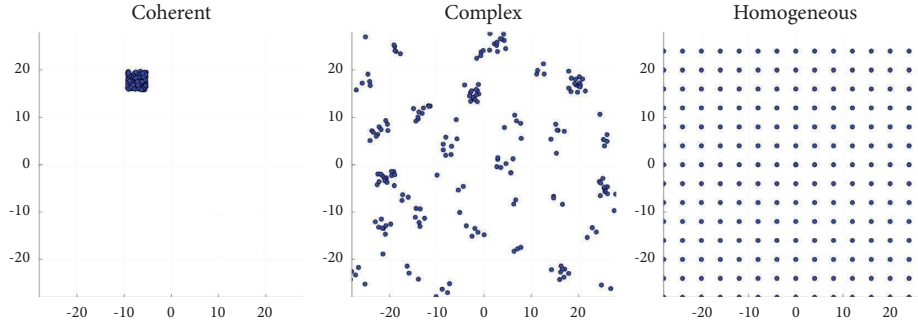


FIGURE 2: Particle distributions of the benchmark distributions used to calibrate C .

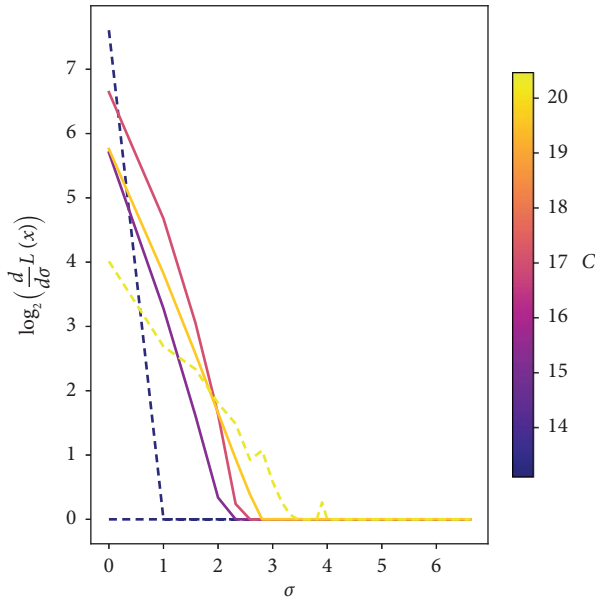


FIGURE 3: Discrete difference of the average complexity profile Figure 4(c), showing the number of clusters that have exactly the given size. The γ of a fitted power-law for each curve is shown in Table 1.

shows the meaning of changes in C on the spatial distribution.

Note that the power-law benchmark curve does not appear as a power-law in Figures 4(b) and 4(c). This is because at each point, the number of clusters *greater than* or equal to the scale is shown. The number of clusters having exactly the size of the scale at each point can be recovered by taking the discrete difference, in which a power-law is recovered. This is shown in Figure 3, and the γ of a fitted power-law for each of these average curves is given in Table 1.

The C values of each curve from simulation and the corresponding parameters are shown in Figure 7. The three regimes of behavior seen in Figure 5 are mirrored in Figure 7, and they occur in the same parameter regions. Both the coherent and homogeneous regimes have low complexity, while the complex regime has high complexity. Furthermore, Welch's unequal t test was conducted on the hypothesis that each pair of these three regimes actually comes from the same distribution, and it was found to strongly support the

alternative in all cases (see Table 3). The regimes' distribution over C are shown in Figure 8, depicting the clear separation between the complex and the simple (coherent and homogeneous) regimes. Thus, C aligns with intuitions about what constitutes a complex pattern and is well mapped into contiguous parameter regions.

Note that s and d have different impacts on C . There is a sharp change in C as s increases around $s = 0.225$, as well as when $|d|$ increases from its minimum value. Overall, C depends on both parameters, but s has a stronger individual effect on it. To quantify this, a linear regression model was fit between each parameter and C separately, where the parameter values were first normalized. The model found that s has $R^2 = 0.56$ with a slope of $0.05 (C/\%s)$ and d has $R^2 = 0.0001$ with a slope of $0.0008 (C/\%s)$. This means that s explains much more of the variance than d does (Figures 9(b) and 9(c)) and that on average a percent change in s changes C more than d does. However, the relationship is very nonlinear, and both variables are needed to explain the data well. Most notably, while s controls C more, the precise value of s that maximizes C depends on d , as can be seen in Figure 7.

Significantly, the most complex behavior occurs when a certain relationship between s and d is true. Namely, as s increases, $|d|$ must also increase in order to produce complex behavior. This can be seen in Figure 7 as a trend of lighter colored regions around a diagonal in s and d . This trend is shown more explicitly in Figure 9(a), which plots C against the ratio of the parameters. Here, it is clear that complexity is maximized around a very particular value of the ratio and drops off when moving away from this value. Note that the curve is not symmetric about $(s/d) = 0$. This is because positive values correspond to molecular-bonding-like behavior and negative values to gravity-like behavior, which produced slightly different spatial distributions of particles. The gravity-like parameter settings achieve a slightly higher maximum complexity than the molecular-bonding-like parameters, and their curve drops off quicker. This shows that complexity caused by molecular-bonding-like behavior is somewhat less sensitive to the parameter ratio: those patterns are slightly more robust.

The ratio that maximizes the complexity is found to be $|s/d| \approx 0.17$. This was computed as the mean of the top 10 most complex parameter settings for $d < 0$ and $d > 0$, which yields -0.18 and 0.16 , respectively. The dashed lines in

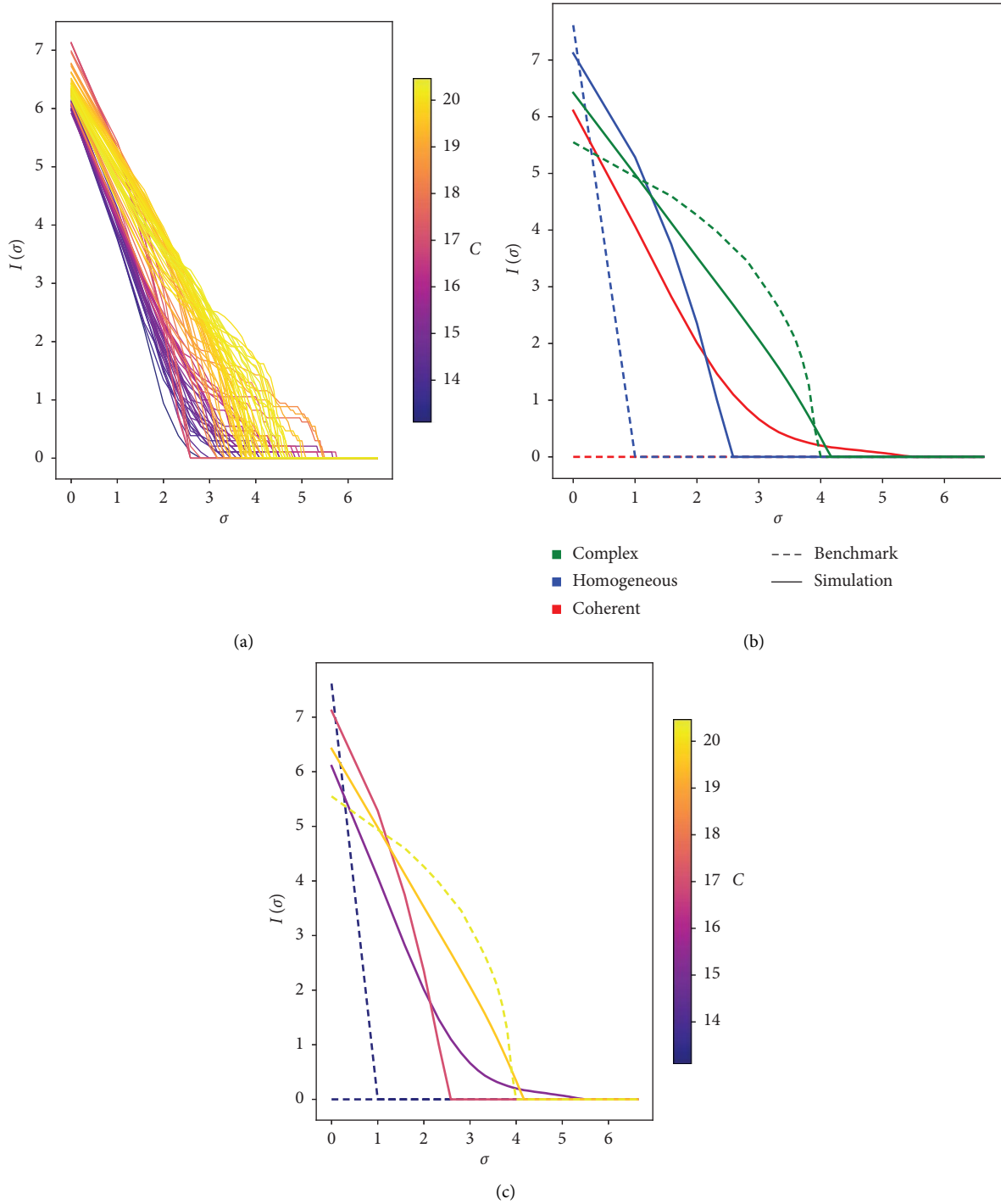


FIGURE 4: Complexity profiles. The horizontal axis is the scale of observation, and the vertical axis is information content. (a) Complexity profiles for a subset (5%), for visual clarity, of all simulation runs. Curves are colored by their C values. (b) Average complexity profile of the three observed regimes (solid) and the three benchmark distributions (dashed). Curves are colored by their regime. (c) Average complexity profile of the three observed regimes (solid) and the three benchmark distributions (dashed). Curves are colored by their C .

Figure 9(a) are drawn at these values, and they roughly approximate the center of the maximum clusters. If $|s/d| < 0.17$, the complexity is likely to be either very high or very low and then quickly drops off to always be low. If $|s/d| > 0.17$, the complexity slowly decays. Figure 10 plots the potential energy function for the s value that maximizes and

minimizes C at each d value. These potentials can be seen to have a characteristic shape that differs for positive versus negative d and which is described by the ratio 0.17. For $d < 0$, the most complex potentials are neither those with the sharpest drop off, nor the shallowest, but those in between, but with a bias towards the sharper curves. A curve is sharper

TABLE 1: Mean and standard deviation of C for each behavior regime in both the benchmark distributions and the simulations, as well as the γ (equation (5)) that best fits the average profile of each (Figure 3).

Regime	Benchmark (std)	Simulation (std)	Benchmark γ	Simulation γ
Homogeneous	10.99 (0)	17.02 (0.16)	76	60
Complex	20.37 (0.09)	19.57 (0.68)	1.12	1.99
Coherent	0 (0)	15.13 (1.1)	0	2.49

TABLE 2: Parameter values used for equation (1), producing 187 different parameter configurations. Nine different initial conditions were used per configuration, producing 1,683 different simulation runs.

Parameter	Values
s	0.01, 0.08, 0.15, 0.225, 0.30, 0.375, 0.45, 0.525, 0.60, 0.675, 0.75
d	-4, -3.5, -3, -2.5, -2, -1.5, -1, -0.5, 0.01, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4
a	10

when $|s/d|$ is smaller because the ratio of the depth of the well to the range of the force is greater. Similarly, for $d > 0$, the most complex potentials are those with an intermediate sharpness to their minimum, though this balance is also biased towards the sharper curves. Note that the most significant difference between the maximum and minimum complexity curves is their range: the minimum complexity curves extend much further out. This corresponds with the low complexity regions found at low s values in Figure 7.

4. Discussion

This analysis showed that the potential energy function has a substantial effect on the visual appearance and spatial complexity of the particles. Specifically, the peak in the complexity, C (Figure 9(a)), is relatively sharp and centered around the value $|s/d| \approx 0.17$. This corresponds to a characteristic shape of the potential energy function that causes the most complex behavior, seen in Figure 10. This characteristic shape means that a complex spatial distribution is produced at a specific balance between the depth of the potential well and the range of the force. If the potential well is deep, particles are more likely to stay together, and if the force is long-ranged, particles are more likely to interact with each other. Thus, if the potential well is deep and the force is long-ranged, the particles will form one large clump. However, if the potential well is shallow and the force short-ranged, particles will not cluster much at all. When a balance between these is struck, clusters of varying sizes tend to form, and it is this kind of distribution that appears complex and which increases the value of C .

Other particle aggregation models have explored how balancing aggregation mechanisms affect the resultant patterns. For example, it was found that a balance between deposition and diffusion rates creates a “morphology phase diagram,” with unique patterns developing when the two rates are comparable [22]. Furthermore, a lattice model of particle aggregation [23] and the preferential attachment network growth algorithm [24] show that aggregation can result in a power-law distribution in the size of the clusters/connections. In the former model, it is interesting to note that one of the main parameters that control the degree of

the power-law is essentially the range of local interactions, similar to s in the present work. Furthermore, that model finds that increasing the range of interactions (β) results in a power-law mass distribution with a larger (negative) exponent (α). According to C as a measure of complexity, this would mean that longer-range interactions produce less complex patterns. While there are significant differences between this model [23] and the present work, these findings are in accordance. This may aid speculation on the effect of density in the present work, as it seems that the interaction range needs to be at some mean value relative to the system size: if the interaction becomes comparable to the system size, all particles may simply clump together. However, the most important feature of these models is that they require the input of particles to sustain their complexity; they are open systems. Preferential attachment without the input of nodes shows power-law behavior transiently but converges to a complete network [24], and similarly, all particles will clump together without continual injection in the lattice model [23]. A similar result is likely if only diffusion is allowed in [22]. It is likely that these models need to be open because they do not allow clusters to break apart once formed. Consider the present work, which was isolated (no energy change and no particle changes), but allowed clusters to break apart, and which resulted in power-law cluster size distributions (Figure 3 and Table 1), at statistical equilibrium. More generally, this result shows that complex spatial patterns can be produced in an isolated physical system at equilibrium depending on particle interactions. This is interesting because most work focuses on complexity presumed to be due to the open aspect of systems. While openness is likely necessary for high degrees of complexity, it should be considered what degree can be achieved without external exchanges. Furthermore, being isolated, the system’s entropy must have increased (or stayed the same), and yet coherent structures are spontaneously formed. This illustrates that entropy is not equivalent to disorder [25].

Additionally, this model shows that there is continuity among gravity-like, molecular-bonding-like, and billiard-ball forces, and there is continuity in the complexity of their collective behavior. The dimension the continuity between the forces runs along is essentially the degree to which the parts of the system influence each other, and the

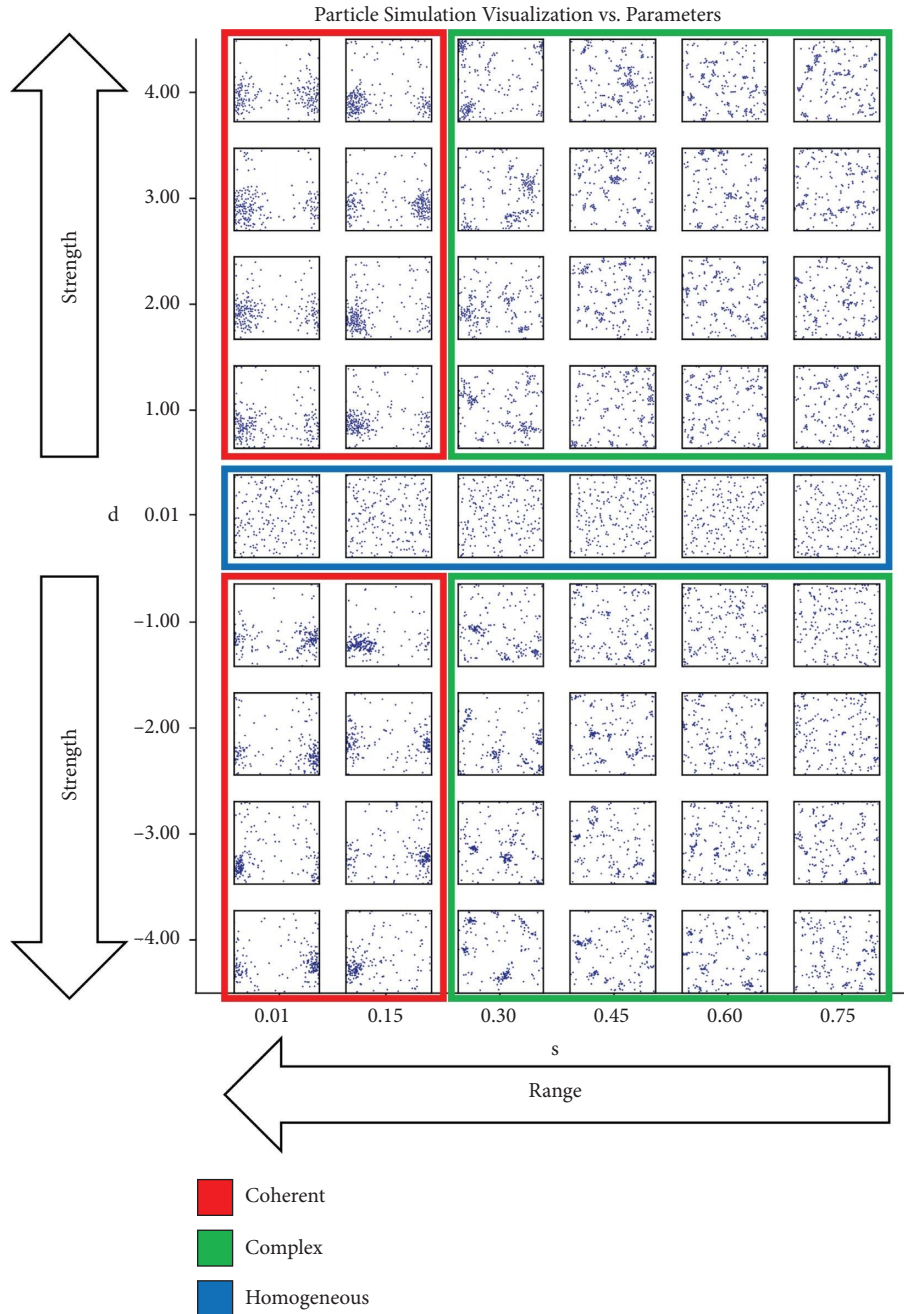


FIGURE 5: Individual frames from the videos produced from simulating the particle system under different parameters. Only a subset of parameter configurations for a single initial condition is shown. See equation (1) and Table 2. In all cases, $a = 10$. Full videos are available online (see Data Availability section). Note that periodic boundary conditions were used. The particles are in a dynamic equilibrium (see text).

most complex behavior was observed at intermediate values of this influence. The exact value of this balance likely depends on other aspects of the systems, such as the density or the boundary conditions; the balance producing complex behavior is likely a function of the other properties of the system.

While there is no consensus on how to measure complexity, C has several merits as a measure. One issue is specifying exactly where the middle between order and homogeneity is. C takes a relatively objective approach that

does not depend on the specific values of the system by transforming the curve to the log-log space. This has the effect of emphasizing information content at intermediate scales by reducing the area contributed by larger scales and the initial maximum of information present in all systems at small scales. Furthermore, C has an intelligible meaning: since it is grounded in the use of scale, it measures how “compressible” the spatial distribution of the system is, being maximal when the system resists compression and minimal when it is easily compressible. Lastly, C aligns well with

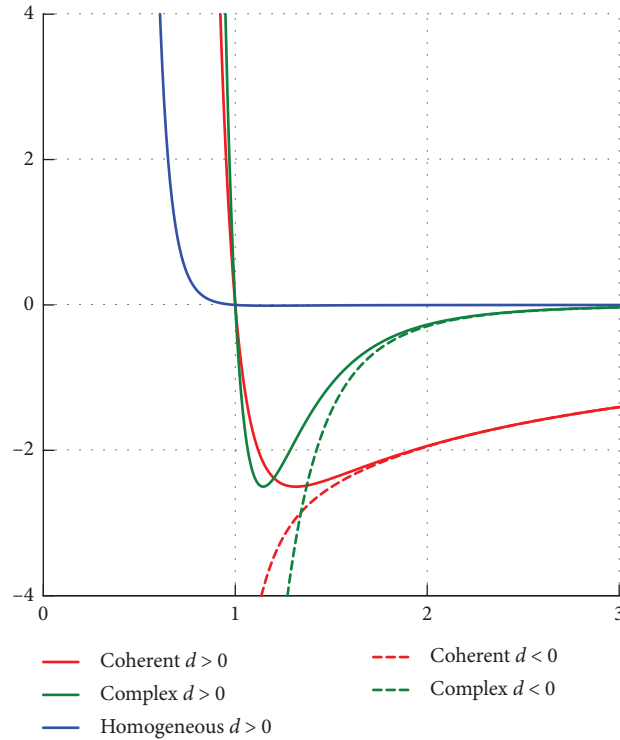


FIGURE 6: Parameterizations of equation (1) that represent each region in Figure 5. Solid lines indicate $d > 0$, and dashed lines indicate $d < 0$. s and d values in the middle of each region were chosen.

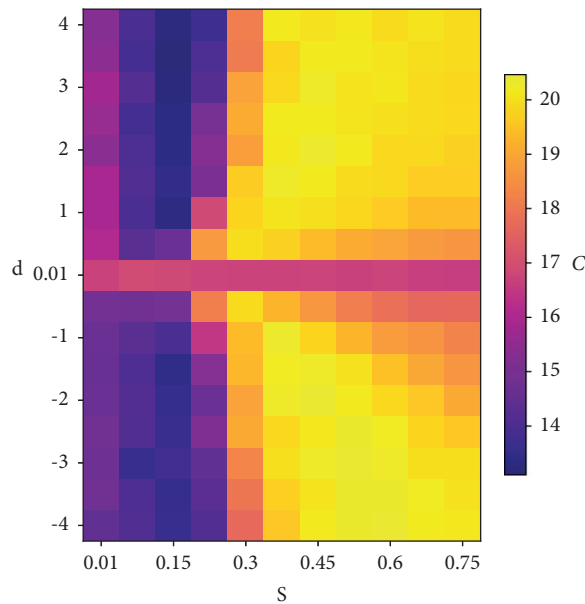


FIGURE 7: C as a function of s and d across all simulation runs. The colorscale is calibrated by the benchmark distributions.

TABLE 3: Results of Welch’s unequal t test, where the null hypothesis was that each behavior regime actually came from the same distribution. Values of approximately 0 are reported because the machine precision of floating point numbers was exceeded by the smallness of the p value.

Regime 1	Regime 2	Test-statistic	p value
Complex	Homogeneous	94.15	≈ 0
Complex	Coherent	-87.76	≈ 0
Homogeneous	Coherent	-38.85	1.50×10^{-173}

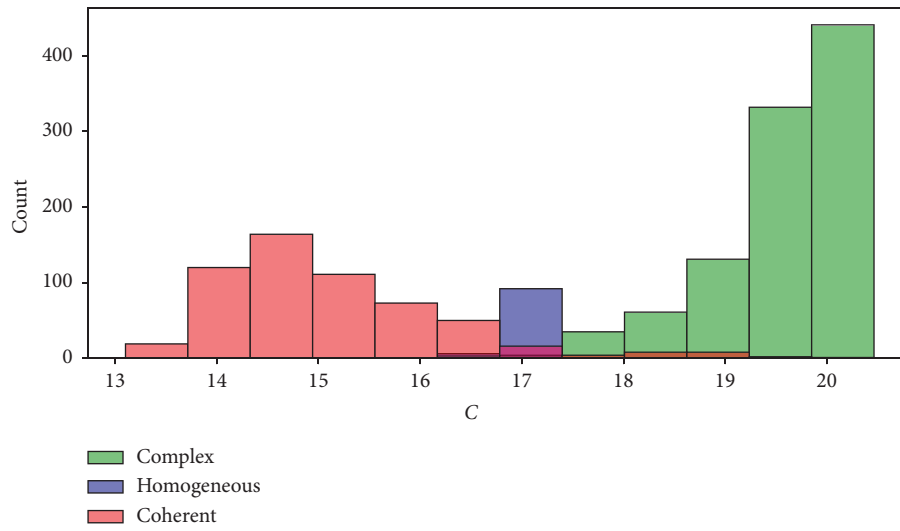
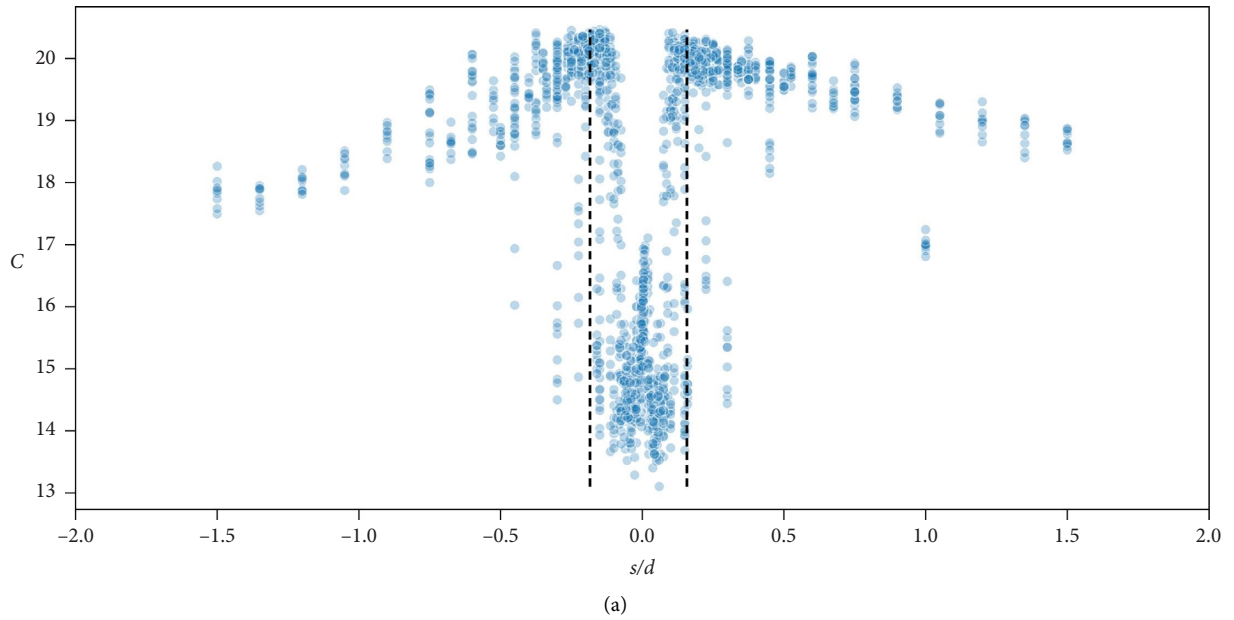


FIGURE 8: Distribution of the three manually labeled regimes of behavior defined by contiguous parameter regions over C . Each point is a simulation run that was labeled by its regime based on its parameters.



(a)
FIGURE 9: Continued.

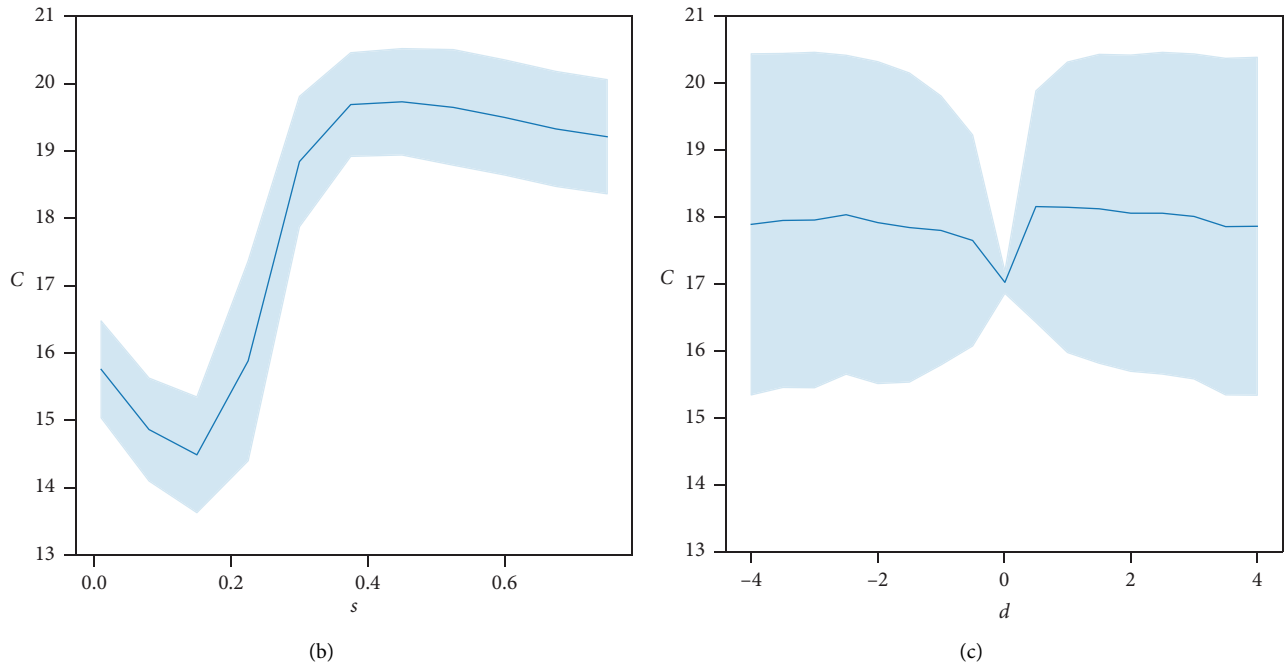


FIGURE 9: (a) C vs. the ratio of s and d . The dashed vertical lines are the average value of the x -axis of the top 10 most complex parameterizations, for positive and negative d , separately. (b) s vs. C . The error bar is standard deviation. (c) d vs. C . The error bar is standard deviation.

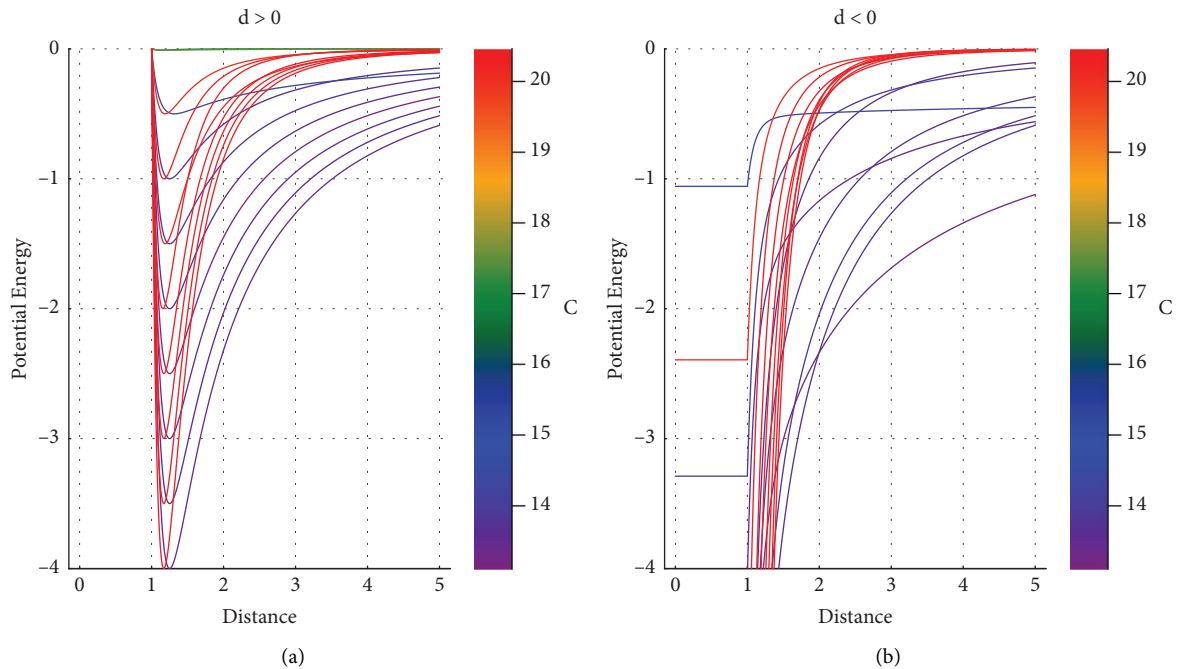


FIGURE 10: Equation (1) plotted for the s value that maximizes and minimizes C at each d value. Color indicates C . (a) positive d values (Lennard–Jones-like potential) and (b) negative d values (gravity-like potential). This shows the characteristic shape of the potential that produces the most complex and least complex behavior.

intuitions about spatial complexity matching what a visual ranking of complexity might show.

However, C only captures a subset of what could influence the perception of spatial complexity. For instance, C would not reflect the differences in the complexity of several

clusters that were colinear versus the same that were not colinear. The many aspects of a system make it difficult to capture all types of patterns with a single measure; in this case, an argument can be made either way for whether the colinear or noncolinear clusters are more complex. One

argument in accordance with the kind of information used in this paper would say that the colinear clusters are correlated because knowing two narrows down the positions of the others and C could be extended to include this information. However, other basic conceptions of what exactly constitutes a complex spatial pattern are very likely possible.

As a general remark, it is very likely that the relationships found in this work would change given different model assumptions. Specifically, these results depend on it being an isolated system, the density of the system, and the energy of the system relative to the potential energy parameters. The density is important because it affects the rate of the interaction between particles, which could change the patterns formed. Similarly, if the system's internal energy were much higher relative to the depth of the potential, stable bonds would not form as often between the particles and clustering would also not be as coherent. Being an isolated system, it evolves without interference; being nonisolated could increase the complexity, decrease it, or change the type of pattern that occurs. For instance, only varying types of clustering patterns were observed in this work. If the system were driven, other patterns may be possible. Additionally, a modeling choice was made to truncate the potential at zero distance for $d < 0$ (Section 2.1.1), which likely enabled complex patterns to form under a gravity-like potential by effectively introducing a minimum in the potential. Other choices, such as aggregating the particles on collision and preserving momentum, would likely produce different behavior types.

5. Conclusion

This work simulated a collective motion model and varied the microscopic interaction rules to explore the range of collective behavior produced. It then formalized intuitive notions about the system's complexity and ranked the collective behavior accordingly. The most complex spatial patterns were produced for parameterizations of the potential energy that corresponded to a specific balance between the depth of the potential well and the range of the force, described by the ratio $|s/d| \approx 0.17$. Additionally, the complexity measure developed, C , was argued to be a meaningful measure of the spatial complexity due to its relatively nonarbitrary method of specifying where the mean between order and homogeneity is, its use of the concept of scale, and its good alignment with an intuitive recognition of complex spatial patterns. Furthermore, a strong relationship between C and the parameters of the potential energy function was found, thus establishing a continuous relationship between gravity-like and billiard-ball like potentials, with a Lennard–Jones-like potential in the middle. Finally, these results show that power-law cluster distributions are possible in an isolated system if, it was speculated, clusters are able to break apart.

Data Availability

All simulation code is available at https://github.com/austinmarcus/collective_motion_LJ. All videos and a subset, due

to storage constraints, of time-series data are available at https://osf.io/dxywj/?view_only=6093fddb2bf743a4b89610fc724e9c96. The additional data are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to thank Christian Koertje for numerous helpful discussions and advice.

References

- [1] Y. Bar-Yam, *Dynamics of Complex Systems (Studies in Nonlinearity)*, CRC Press, Boca Raton, FL, USA, 1999.
- [2] M. E. J. Newman, "Resource letter CS-1: complex systems," *American Journal of Physics*, vol. 79, no. 8, pp. 800–810, 2011.
- [3] H. Sayama, *Introduction to the Modeling and Analysis of Complex Systems*, Open SUNY Textbooks, New York, NY, USA, 2015.
- [4] T. Vicsek and A. Zafeiris, "Collective motion," *Physics Reports*, vol. 517, no. 3-4, pp. 71–140, 2012.
- [5] H. Sayama, "Swarm chemistry," *Artificial Life*, vol. 15, no. 1, pp. 105–114, 2009.
- [6] Y. Yang, F. Turci, E. Kague, C. L. Hammond, J. Russo, and C. P. Royall, "Dominating lengthscales of zebrafish collective behaviour," *PLoS Computational Biology*, vol. 18, no. 1, Article ID e1009394, 2022.
- [7] C. W. Reynolds, "Flocks, herds and schools: a distributed behavioral model," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, pp. 25–34, New York, NY, USA, August 1987.
- [8] I. Prigogine, "Time, structure, and fluctuations," *Science*, vol. 201, no. 4358, pp. 777–785, 1978.
- [9] M. P. Allen, "Introduction to molecular dynamics simulation," *Computational soft matter: From Synthetic Polymers to Proteins*, vol. 23, no. 1, pp. 1–28, 2004.
- [10] S. A. Hollingsworth and R. O. Dror, "Molecular dynamics simulation for all," *Neuron*, vol. 99, no. 6, pp. 1129–1143, 2018.
- [11] M. E. Tuckerman and G. J. Martyna, "Understanding modern molecular Dynamics: techniques and applications," *The Journal of Physical Chemistry B*, vol. 104, no. 2, pp. 159–178, 2000.
- [12] D. Osmanović, "Spatial distributions of nonconservatively interacting particles," *Physical Review*, vol. 103, no. 2, Article ID 22610, 2021.
- [13] D. C. Rapaport, "Molecular-dynamics study of Rayleigh–Bénard convection," *Physical Review Letters*, vol. 60, no. 24, pp. 2480–2483, 1988.
- [14] D. Nerukh, V. Ryabov, and R. C. Glen, "Complex temporal patterns in molecular dynamics: a direct measure of the phase-space exploration by the trajectory at macroscopic time scales," *Physical Review*, vol. 77, no. 3, Article ID 36225, March 2008.
- [15] S. Stephan, J. Staubach, and H. Hasse, "Review and comparison of equations of state for the Lennard–Jones fluid," *Fluid Phase Equilibria*, vol. 523, Article ID 112772, 2020.
- [16] W. Kinsner, "Complexity and its measures in cognitive and other complex systems," in *Proceedings of the 2008 7th IEEE*

- International Conference on Cognitive Informatics*, pp. 13–29, IEEE, Stanford, CA, USA, August 2008.
- [17] S. Lloyd, “Measures of complexity: a nonexhaustive list,” *IEEE Control Systems*, vol. 21, no. 4, pp. 7–8, 2001.
 - [18] Y. Bar-Yam, *Complexity Rising: From Human Beings to Human Civilization, a Complexity Profile*, Encyclopedia of Life Support Systems, Abu Dhabi, UAE, 2002.
 - [19] P. Young, *The Leapfrog Method and Other Symplectic Algorithms for Integrating newton’s Laws of Motion*, University of California, Santa Cruz, CA, USA, 2013.
 - [20] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia: a fresh approach to numerical computing,” *Society of Indian Automobile Manufacturers Review*, vol. 59, no. 1, pp. 65–98, 2017.
 - [21] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pp. 226–231, AAAI Press, Portland, OR, USA, August 1996.
 - [22] P. Jensen, A.-L. Barabási, H. Larralde, S. Havlin, and H. E. Stanley, “Deposition, diffusion, and aggregation of atoms on surfaces: a model for nanostructure growth,” *Physical Review B*, vol. 50, no. 20, Article ID 15316, 15329 pages, 1994.
 - [23] H. Takayasu, I. Nishikawa, and H. Tasaki, “Power-law mass distribution of aggregation systems with injection,” *Physical Review*, vol. 37, no. 8, pp. 3110–3117, 1988.
 - [24] A.-L. Barabasi and R. Albert, “Emergence of scaling in random networks,” *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
 - [25] D. Styer, “Entropy as disorder: history of a misconception,” *The Physics Teacher*, vol. 57, no. 7, pp. 454–458, 2019.