`manuk-kepudang`: A Python framework for spatial entropy quantification in the three-dimensional Vicsek model

Sandy H. S. Herho[1,2], Gandhi Napitupulu[3,4], Iwan P. Anwar[4], Nurjanna J. Trilaksono[5], Rusmawan Suwarman[5], Dasapta E. Irawan[6*]

[1]Department of Earth and Planetary Sciences, University of California, Riverside, 92521, CA, USA.
[2]School of Systems Science and Industrial Engineering, State University of New York, Binghamton, 13902, NY, USA.
[3]Coastal Hazards and Energy System Science (CHESS) Lab, Hiroshima University, Hiroshima, 739-8529, Hiroshima Prefecture, Japan.
[4]Applied and Environmental Oceanography Research Group, Bandung Institute of Technology, Bandung, 40132, West Java, Indonesia.
[5]Atmospheric Science Research Group, Bandung Institute of Technology, Bandung, 40132, West Java, Indonesia.
[6*]Applied Geology Research Group, Bandung Institute of Technology, Bandung, 40132, West Java, Indonesia.

Contributing authors: dasaptaerwin@itb.ac.id;

Abstract

Collective motion in active matter systems arises from local interactions among self-propelled constituents, producing emergent behavior observed across scales from bacterial colonies to bird flocks. This article introduces `manuk-kepudang`, an open-source Python library that implements the three-dimensional Vicsek model alongside a suite of information-theoretic diagnostics for characterizing spatial organization. The library pairs Numba-accelerated simulation of alignment dynamics with six entropy measures (positional, orientational, local alignment, pair correlation, Voronoi, and position-velocity mutual information), which are combined into a composite Spatial Complexity Index (SCI). These metrics capture aspects of configurational and orientational disorder that scalar

order parameters alone cannot resolve. Validation across four test cases covering ordered, disordered, and critical regimes confirms that the entropy measures distinguish between collective states and reproduce the expected phenomenology of noise-driven phase transitions. We further apply nonparametric statistical methods suited to non-Gaussian velocity distributions, enabling rigorous characterization of how distributions evolve as collective alignment emerges. The library provides an accessible platform for studying collective motion, with applications ranging from biological swarm dynamics and robotic coordination to fundamental investigations of nonequilibrium phase transitions.

Keywords: Active matter, Collective motion, Spatial entropy, Vicsek model

MSC Classification: 82C22 , 82C26 , 94A17 , 65C35

1 Introduction

Collective motion constitutes one of the most visually striking and scientifically compelling manifestations of emergent behavior in nature, observable across an extraordinary range of length scales from bacterial colonies and cellular aggregates through insect swarms and fish schools to the spectacular aerial displays of starling murmurations [1–3]. These systems, despite their diverse biological origins and the vastly different physical interactions governing their constituents, exhibit remarkably universal phenomenology: the spontaneous emergence of coherent, large-scale motion from purely local interactions among individuals possessing no global knowledge of the system state [4]. Understanding the mechanisms underlying such self-organization has attracted sustained interest from physicists, biologists, and applied mathematicians alike, driven both by fundamental curiosity regarding nonequilibrium phase transitions and by practical applications in contexts ranging from tissue morphogenesis to autonomous robotic swarms [5].

The seminal model introduced by Vicsek et al. [1] has become the canonical minimal framework for theoretical and computational investigations of flocking phenomena. In this model, point particles move at constant speed while adjusting their direction of motion to align with the average orientation of neighboring particles, subject to stochastic angular perturbations representing the cumulative effect of unresolved degrees of freedom. Despite its apparent simplicity, the Vicsek model exhibits a genuine nonequilibrium phase transition between a disordered state at high noise amplitude, wherein particles move in uncorrelated directions resembling an ideal gas, and an ordered flocking state at low noise, wherein global alignment emerges spontaneously and particles collectively migrate through the simulation domain [6, 7]. The nature of this transition—whether continuous or discontinuous, and how it depends on dimensionality, density, and system size—has been the subject of extensive theoretical and numerical investigation over the past three decades [8–10].

Quantitative characterization of collective states in active matter systems has traditionally relied upon scalar order parameters measuring the degree of global alignment, supplemented by correlation functions probing the spatial structure of orientational

fluctuations [3]. While such measures capture essential aspects of the ordered and disordered phases, they provide limited insight into the heterogeneity, complexity, and information content of the instantaneous particle configuration. Information-theoretic approaches rooted in Shannon entropy offer a principled and model-independent framework for characterizing disorder and organization in spatially extended systems [11, 12]. By quantifying the statistical properties of position and velocity distributions through entropic measures, one may probe aspects of collective behavior—such as local heterogeneity, geometric packing irregularity, and position-velocity coupling—that escape detection by conventional order parameters.

The Python programming language has emerged as a dominant platform for scientific computing, owing to its accessible syntax, extensive ecosystem of numerical and visualization libraries, and strong community support for open-source development [13, 14]. The availability of high-performance array operations through NumPy, coupled with just-in-time compilation capabilities provided by Numba [15], enables Python implementations to achieve computational throughput approaching that of compiled languages while retaining the flexibility and readability that facilitate rapid prototyping, pedagogical application, and reproducible research. These attributes render Python particularly suitable for developing simulation tools intended for broad dissemination within the active matter research community, where accessibility and extensibility are valued alongside raw performance.

The present contribution introduces `manuk-kepudang`, an open-source Python library implementing the three-dimensional Vicsek model together with a comprehensive suite of spatial entropy metrics for characterizing collective motion. The library combines an efficient Numba-accelerated simulation engine with information-theoretic diagnostics including positional, orientational, local alignment, pair correlation, and Voronoi entropies, as well as position-velocity mutual information, aggregated into a composite Spatial Complexity Index (SCI) that provides a scalar measure of overall system disorder. Auxiliary analysis tools employing nonparametric statistical methods enable rigorous quantification of velocity distributions and their temporal evolution across dynamical regimes. Through systematic application to test cases spanning the ordered, disordered, and critical regions of parameter space, we demonstrate that the implemented entropy measures successfully discriminate between collective states and provide physical insight complementary to traditional order parameter characterizations [16, 17].

2 Methods

2.1 Model Description

The Vicsek model constitutes a minimal framework for describing the collective dynamics of self-propelled particles through local alignment interactions [1]. We consider an ensemble of N point particles confined within a three-dimensional cubic domain $\Omega = [0, L]^3$ subject to periodic boundary conditions. The instantaneous state of the i-th particle is specified by its position vector $\boldsymbol{r}_i(t) \in \Omega$ and velocity vector $\boldsymbol{v}_i(t) \in \mathbb{R}^3$. The fundamental constraint underlying the model is that each particle maintains a constant speed v_0, reflecting the self-propelled nature of the constituents,

such that

$$|\boldsymbol{v}_i(t)| = v_0 \tag{1}$$

holds for all particles and all times [2]. This constraint distinguishes active matter systems from passive Brownian suspensions, wherein particles continuously extract energy from their environment to sustain directed motion [4].

The interaction topology is determined by a finite sensing radius $r > 0$. At each instant t, particle i interacts with all particles lying within a sphere of radius r centered at its current position. Accounting for the periodic boundary conditions through the minimum image convention [18], the neighborhood of particle i is defined as

$$\mathcal{N}_i(t) = \left\{ j \in \{1, 2, \ldots, N\} : |\boldsymbol{r}_i(t) - \boldsymbol{r}_j(t)|_{\mathrm{PBC}} < r \right\}, \tag{2}$$

where the minimum image distance under periodic boundaries is computed via

$$|\boldsymbol{r}_i - \boldsymbol{r}_j|_{\mathrm{PBC}} = |\boldsymbol{\Delta}_{ij}|, \quad \Delta_{ij}^{\alpha} = r_i^{\alpha} - r_j^{\alpha} - L\,\mathrm{round}\left(\frac{r_i^{\alpha} - r_j^{\alpha}}{L}\right), \tag{3}$$

with $\alpha \in \{x, y, z\}$ indexing the Cartesian components and round$(\cdot)$ denoting rounding to the nearest integer. The neighborhood $\mathcal{N}_i(t)$ includes particle i itself since the self-distance vanishes.

The dynamical evolution proceeds through a discrete-time update scheme. At each time step, particles adjust their velocity direction based on the local alignment field and subsequently translate according to the updated velocity. The alignment interaction embodies the tendency of each particle to orient its motion along the mean direction of its neighbors, a mechanism observed across diverse biological systems from bird flocks to bacterial swarms [3, 5]. To construct the alignment signal, we first compute the vector sum of velocities within the neighborhood,

$$\boldsymbol{S}_i(t) = \sum_{j \in \mathcal{N}_i(t)} \boldsymbol{v}_j(t), \tag{4}$$

which aggregates the directional information from all interacting particles. The magnitude of this sum depends on both the number of neighbors and the degree of local alignment among them.

A crucial step in the Vicsek dynamics is the extraction of the pure directional content from the aggregate velocity. Rather than using the raw sum $\boldsymbol{S}_i(t)$, we normalize it to obtain a unit vector representing the mean orientation,

$$\hat{\boldsymbol{n}}_i(t) = \frac{\boldsymbol{S}_i(t)}{|\boldsymbol{S}_i(t)|} = \frac{\sum_{j \in \mathcal{N}_i(t)} \boldsymbol{v}_j(t)}{\left|\sum_{j \in \mathcal{N}_i(t)} \boldsymbol{v}_j(t)\right|}. \tag{5}$$

This normalization ensures that the alignment signal carries unit weight irrespective of the local particle density, rendering the subsequent noise amplitude η a proper measure of angular fluctuation strength relative to the alignment tendency [6, 8].

Stochastic fluctuations enter the dynamics through an angular noise term that perturbs the deterministic alignment. At each update, particle i draws an independent random unit vector $\boldsymbol{\xi}_i(t)$ uniformly distributed on the unit sphere $\mathbb{S}^2$. This isotropic noise is generated by sampling three independent standard Gaussian variates $\zeta_i^x, \zeta_i^y, \zeta_i^z \sim \mathcal{N}(0,1)$ and normalizing the resulting vector,

$$\boldsymbol{\xi}_i(t) = \frac{\boldsymbol{\zeta}_i(t)}{|\boldsymbol{\zeta}_i(t)|}, \quad \boldsymbol{\zeta}_i(t) = (\zeta_i^x(t), \zeta_i^y(t), \zeta_i^z(t))^{\mathsf{T}}. \tag{6}$$

The noise vectors are statistically independent across particles and uncorrelated in time, satisfying $\langle \boldsymbol{\xi}_i(t) \rangle = \mathbf{0}$ and $\langle \boldsymbol{\xi}_i(t) \cdot \boldsymbol{\xi}_j(t') \rangle = \delta_{ij}\delta_{tt'}$ [16].

The competition between alignment and noise is governed by the dimensionless noise amplitude $\eta \geq 0$. The new direction of motion emerges from the vector sum of the normalized alignment direction and the scaled noise contribution,

$$\boldsymbol{w}_i(t) = \hat{\boldsymbol{n}}_i(t) + \eta\, \boldsymbol{\xi}_i(t). \tag{7}$$

When $\eta = 0$, the dynamics reduces to deterministic alignment wherein each particle exactly adopts the mean neighbor orientation. As η increases, the stochastic perturbation progressively disrupts the alignment tendency, eventually overwhelming it for $\eta \gg 1$ and yielding effectively random reorientations [7].

To satisfy the constant-speed constraint (1), the combined direction vector $\boldsymbol{w}_i(t)$ must be projected onto the sphere of radius v_0. This projection is accomplished through normalization followed by rescaling,

$$\boldsymbol{v}_i(t + \Delta t) = v_0\, \frac{\boldsymbol{w}_i(t)}{|\boldsymbol{w}_i(t)|}. \tag{8}$$

Substituting the expressions (5) and (7) into (8) yields the complete velocity update equation,

$$\boldsymbol{v}_i(t + \Delta t) = v_0\, \hat{\boldsymbol{u}} \left(\frac{\sum_{j \in \mathcal{N}_i(t)} \boldsymbol{v}_j(t)}{\left| \sum_{j \in \mathcal{N}_i(t)} \boldsymbol{v}_j(t) \right|} + \eta\, \boldsymbol{\xi}_i(t) \right), \tag{9}$$

where the operator $\hat{\boldsymbol{u}}(\cdot)$ denotes normalization to unit magnitude, $\hat{\boldsymbol{u}}(\boldsymbol{a}) = \boldsymbol{a}/|\boldsymbol{a}|$. Equation (9) constitutes the first governing equation of the three-dimensional Vicsek model.

Following the velocity update, each particle undergoes ballistic displacement along its new direction of motion. The position evolves according to

$$\boldsymbol{r}_i(t + \Delta t) = \boldsymbol{r}_i(t) + \boldsymbol{v}_i(t + \Delta t)\, \Delta t, \tag{10}$$

where $\Delta t > 0$ is the discrete time step. The use of the updated velocity $\boldsymbol{v}_i(t + \Delta t)$ rather than the pre-update value ensures immediate response to the reorientation [9]. To maintain confinement within the simulation domain while preserving translational

invariance, periodic boundary conditions are imposed through the modular operation

$$\boldsymbol{r}_i(t+\Delta t) = [\boldsymbol{r}_i(t) + \boldsymbol{v}_i(t+\Delta t)\,\Delta t] \mod L, \tag{11}$$

applied component-wise such that $r_i^\alpha \in [0, L)$ for each Cartesian coordinate. Equation (11) represents the second governing equation completing the dynamical specification.

The macroscopic state of the system is characterized by the polar order parameter $\varphi(t)$, which quantifies the degree of global velocity alignment [1]. This observable is defined as the magnitude of the mean velocity normalized by the maximum possible value,

$$\varphi(t) = \frac{1}{Nv_0}\left|\sum_{i=1}^{N} \boldsymbol{v}_i(t)\right|. \tag{12}$$

The normalization ensures $\varphi \in [0, 1]$ by virtue of the triangle inequality and the constraint (1). In the disordered regime, particle velocities point in uncorrelated directions and the vector sum nearly vanishes due to statistical cancellation, yielding $\varphi \approx 0$. Conversely, when particles move collectively as a coherent flock, their velocities align and the sum approaches its maximum magnitude Nv_0, giving $\varphi \approx 1$. The order parameter thus serves as an indicator of the phase transition between disordered and ordered states that occurs as the noise amplitude η crosses a critical threshold η_c dependent on the particle density $\rho = N/L^3$ [6, 10].

The complete dynamics is summarized by the coupled map

$$\boldsymbol{v}_i(t+\Delta t) = v_0\,\hat{\boldsymbol{u}}\left(\langle\hat{\boldsymbol{v}}\rangle_{\mathcal{N}_i} + \eta\,\boldsymbol{\xi}_i(t)\right), \tag{13a}$$

$$\boldsymbol{r}_i(t+\Delta t) = [\boldsymbol{r}_i(t) + \boldsymbol{v}_i(t+\Delta t)\,\Delta t] \mod L, \tag{13b}$$

where $\langle\hat{\boldsymbol{v}}\rangle_{\mathcal{N}_i} = \boldsymbol{S}_i/|\boldsymbol{S}_i|$ denotes the normalized local mean velocity defined in (5). Equations (13a)–(13b) are applied synchronously to all N particles at each discrete time step, with the neighborhoods $\mathcal{N}_i(t)$ evaluated using the positions at time t prior to the update. The model is thus specified by six parameters: the particle number N, the domain size L, the intrinsic speed v_0, the interaction radius r, the noise amplitude η, and the time step Δt. These combine into the relevant dimensionless groups $\tilde{\rho} = Nr^3/L^3$ measuring the mean number of neighbors and $\tilde{v} = v_0\Delta t/r$ comparing displacement to interaction range, which together with η govern the phase behavior of the system [8].

2.2 Numerical Implementation

The governing equations derived in the preceding section constitute a discrete-time dynamical system that is directly amenable to numerical integration. The inherently discrete nature of the Vicsek model—wherein updates occur at fixed intervals Δt rather than continuously—obviates the need for sophisticated temporal discretization schemes typically required for ordinary differential equations (ODEs). Nevertheless, careful attention must be paid to the order of operations, the treatment of periodic boundaries, and the efficient computation of local neighborhoods to ensure both correctness and computational tractability.

The `manuk-kepudang` library implements the three-dimensional Vicsek model following a synchronous update protocol, wherein the states of all N particles are updated simultaneously based on the configuration at the current time step [19]. This stands in contrast to asynchronous schemes where particles are updated sequentially, which can introduce spurious correlations and break the symmetry of the interaction rules [20]. Denoting the discrete time index by n such that $t = n\Delta t$, the synchronous update proceeds by first computing all new velocities $\{\boldsymbol{v}_i^{n+1}\}_{i=1}^N$ from the current state $\{(\boldsymbol{r}_i^n, \boldsymbol{v}_i^n)\}_{i=1}^N$, and subsequently advancing all positions using the newly computed velocities.

The velocity update requires the identification of neighbors within the interaction radius r for each particle. For particle i at position $\boldsymbol{r}_i^n$, the neighborhood set $\mathcal{N}_i^n$ defined in (2) is constructed by evaluating the minimum image distance to all other particles under periodic boundary conditions according to (3). The squared distance $d_{ij}^2 = |\boldsymbol{\Delta}_{ij}|^2$ is compared against r^2 to determine membership in $\mathcal{N}_i^n$. This brute-force approach entails $\mathcal{O}(N^2)$ distance evaluations per time step, which remains computationally feasible for the system sizes considered herein ($N \lesssim 10^3$) but would necessitate cell-list or tree-based acceleration for larger ensembles [21].

Upon determination of the neighborhoods, the local velocity sum $\boldsymbol{S}_i$ is accumulated according to (4), where the summation includes particle i itself. A critical numerical consideration arises when $|\boldsymbol{S}_i|$ approaches zero, which can occur in highly disordered configurations where neighbor velocities cancel. To prevent division by zero during the normalization step (5), a threshold $\epsilon = 10^{-10}$ is employed: if $|\boldsymbol{S}_i| < \epsilon$, a random unit vector is substituted for the alignment direction $\hat{\boldsymbol{n}}_i$. This safeguard ensures numerical stability without appreciably affecting the dynamics, as such cancellations are exceedingly rare except in pathological configurations.

The stochastic component of the dynamics requires the generation of random unit vectors uniformly distributed on the unit sphere $\mathbb{S}^2$ as prescribed by (6). Following the method of Marsaglia [22], three independent standard normal variates are drawn and subsequently normalized. The random number generation is performed using the Mersenne Twister algorithm as implemented in the NumPy library [23], which provides a period of $2^{19937} - 1$ and satisfies stringent statistical tests for randomness [24]. The independence of noise samples across particles and time steps is essential for faithfully representing the uncorrelated fluctuations inherent to the model.

The construction of the perturbed direction vector $\boldsymbol{w}_i$ follows (7), superposing the alignment signal with the scaled noise contribution. The subsequent normalization and rescaling to enforce the constant-speed constraint proceeds according to (8). In the rare event that $|\boldsymbol{w}_i| < \epsilon$ due to exact cancellation between alignment and noise, a fresh random unit vector scaled by v_0 is assigned to maintain well-defined dynamics. This numerical treatment preserves the statistical properties of the original model while ensuring robust behavior across all parameter regimes.

The position update follows directly from the kinematic relation (13b). Each Cartesian component is advanced by the displacement $v_i^{\alpha,n+1}\Delta t$, and the periodic boundary conditions are enforced through the modular operation prescribed in (11). The floor function implementation ensures that particles exiting through one face of the cubic domain re-enter through the opposite face, consistent with the toroidal topology

implied by periodic boundaries. The use of the updated velocity $\boldsymbol{v}_i^{n+1}$ rather than the pre-update value ensures immediate response to the reorientation, as specified in the governing equations.

The order parameter φ^n quantifying collective alignment is evaluated at each saved time step according to (12). The vector summation over all particle velocities is performed in double-precision floating-point arithmetic to minimize accumulation of round-off errors, particularly relevant for large N where many terms of alternating sign may partially cancel. The normalization by Nv_0 ensures that the order parameter remains bounded within the interval $[0, 1]$ as guaranteed by the triangle inequality and the constant-speed constraint.

The initialization of the particle ensemble proceeds by drawing positions uniformly within the simulation domain and velocities from an isotropic distribution on the sphere of radius v_0. Specifically, for each particle i, the initial position components are sampled as $r_i^{\alpha,0} \sim \mathcal{U}(0, L)$, where $\mathcal{U}(a, b)$ denotes the uniform distribution on the interval $[a, b]$. The initial velocities are generated by the same spherical sampling procedure used for the noise vectors, drawing Gaussian variates and normalizing before scaling to the prescribed speed. This initialization ensures an initially disordered state with $\langle\varphi^0\rangle \sim N^{-1/2}$ in expectation, from which the system may spontaneously develop order depending on the noise amplitude and density.

Computational efficiency is achieved through just-in-time (JIT) compilation of the velocity update kernel using the Numba library [15]. The core loop iterating over particle pairs is decorated with compilation directives that translate the Python code to optimized machine instructions and distribute the outer loop across available CPU cores via OpenMP-style parallelization. This approach yields speedups of one to two orders of magnitude compared to pure Python implementations while maintaining the readability and flexibility of high-level code [25, 26]. The computational complexity per time step scales as $\mathcal{O}(N^2)$ due to the all-pairs distance computation, with a prefactor reduced by vectorization and parallel execution.

The algorithm implemented in `manuk-kepudang` proceeds as follows for each time step. Given the current configuration, the minimum image distances from each particle to all others are computed and compared against the interaction radius to identify neighborhoods. The velocity sums over these neighborhoods are accumulated and normalized to extract the local alignment directions, with numerical safeguards applied when the magnitude falls below the threshold. Random unit vectors are generated for all particles and combined with the alignment directions according to the noise amplitude, followed by normalization and rescaling to produce the updated velocities. After all velocities have been computed synchronously, the positions are advanced and wrapped into the primary simulation cell. The order parameter is evaluated and stored at the specified save interval before proceeding to the next iteration.

To validate the implementation and demonstrate the range of collective behaviors accessible to the model, four test cases spanning distinct dynamical regimes are provided with the `manuk-kepudang` distribution. These cases differ in the noise amplitude η and system size while sharing common values $v_0 = 0.5$ and $\Delta t = 1.0$ for the speed and time step, respectively.

Test Case 1 probes the ordered phase with parameters $N = 200$, $L = 10.0$, $r = 2.0$, and $\eta = 0.3$. The low noise amplitude permits the alignment interaction to dominate, driving the system toward a state of collective motion. The dimensionless density $\tilde{\rho} = Nr^3/L^3 = 1.6$ ensures a sufficient number of neighbors per particle to sustain coherent alignment. This case serves to verify that the implementation correctly reproduces the ordered, flocking state expected at subcritical noise.

Test Case 2 examines the disordered phase by increasing the noise amplitude to $\eta = 2.0$ while retaining identical geometric parameters ($N = 200$, $L = 10.0$, $r = 2.0$). At this noise level, the stochastic perturbation overwhelms the alignment signal, preventing the emergence of global order. This case confirms that the noise implementation correctly disrupts collective motion when the noise-to-signal ratio exceeds the critical threshold.

Test Case 3 investigates finite-size effects by enlarging the system to $N = 500$ particles in a domain of side $L = 15.0$ with interaction radius $r = 2.5$ and intermediate noise $\eta = 0.5$. The resulting density $\tilde{\rho} \approx 2.3$ places the system within the ordered regime, while the larger particle number permits observation of more pronounced collective structures. The extended simulation duration of 600 time steps accommodates the longer relaxation times associated with larger systems.

Test Case 4 targets the critical regime near the order-disorder transition by setting $\eta = 1.0$ with $N = 300$, $L = 12.0$, and $r = 2.0$. Near criticality, the system exhibits large fluctuations as it intermittently samples ordered and disordered configurations [7]. The extended duration of 800 time steps allows observation of these fluctuations and provides sufficient statistics for characterizing the metastable dynamics.

The parameter values for all test cases are summarized in Table 1. These configurations collectively span the phase diagram of the three-dimensional Vicsek model, enabling validation of the implementation across the full range of accessible collective behaviors.

Table 1 Parameter values for the four test cases implemented in `manuk-kepudang`. All cases use $v_0 = 0.5$ and $\Delta t = 1.0$.

Case	Regime	N	L	r	η	Steps
1	Ordered	200	10.0	2.0	0.3	500
2	Disordered	200	10.0	2.0	2.0	500
3	Large system	500	15.0	2.5	0.5	600
4	Critical	300	12.0	2.0	1.0	800

The correctness of the numerical implementation is verified by monitoring the constant-speed constraint, which must hold exactly at all times. The implementation enforces this constraint explicitly through the normalization step, and numerical tests confirm that deviations remain below machine precision in double-precision arithmetic. Additionally, the statistical isotropy of the initial conditions and the disordered phase is verified by confirming that the time-averaged velocity components vanish within statistical uncertainty, indicating no spurious directional bias in the random number generation or update procedure.

2.3 Spatial Entropy and Complexity Metrics

Beyond the scalar order parameter φ defined in (12), a comprehensive characterization of collective motion requires quantification of the spatial organization and heterogeneity inherent to the particle ensemble. Information-theoretic measures rooted in Shannon entropy provide a principled framework for such characterization, offering model-independent metrics that capture distinct aspects of order and disorder [11]. The `manuk-kepudang` library implements a suite of six complementary entropy measures together with a composite SCI, each probing a different facet of the system's configurational and dynamical state.

The positional entropy H_{pos} quantifies the uniformity of the spatial distribution of particles within the simulation domain. To evaluate this measure, the cubic box of side L is discretized into M^3 cells of equal volume, where M denotes the number of bins per dimension. The occupation number n_k of cell k is tallied for all $k \in \{1, 2, \ldots, M^3\}$, and the empirical probability of finding a particle in cell k is estimated as $p_k = n_k/N$. The Shannon entropy of this discrete distribution is then computed as

$$H_{\mathrm{pos}} = -\sum_{k=1}^{M^3} p_k \ln p_k, \tag{14}$$

where the sum extends only over occupied cells to avoid the indeterminate form $0 \ln 0$, which is conventionally taken as zero in the limit. A uniform distribution of particles across all cells yields the maximum entropy $H_{\mathrm{pos}}^{\max} = \ln(\min(N, M^3))$, while clustering into a single cell produces zero entropy. The normalized positional entropy $\tilde{H}_{\mathrm{pos}} = H_{\mathrm{pos}}/H_{\mathrm{pos}}^{\max} \in [0, 1]$ thus serves as a dimensionless measure of spatial uniformity, with values approaching unity indicating homogeneous dispersion and values near zero signifying aggregation or clustering [27]. The implementation employs $M = 10$ bins per dimension by default, yielding $10^3 = 1000$ cells, which provides adequate resolution for typical system sizes while avoiding excessive noise from sparsely populated bins. The histogram construction is performed using the multidimensional binning capabilities of the NumPy library [23].

The orientational entropy H_{orient} measures the isotropy of velocity directions on the unit sphere $\mathbb{S}^2$. Each velocity vector $\boldsymbol{v}_i$ is first normalized to unit magnitude $\hat{\boldsymbol{v}}_i = \boldsymbol{v}_i/|\boldsymbol{v}_i|$ and subsequently converted to spherical coordinates (θ_i, ϕ_i), where the polar angle $\theta_i = \arccos(\hat{v}_i^z) \in [0, \pi]$ is measured from the positive z-axis and the azimuthal angle $\phi_i = \arctan 2(\hat{v}_i^y, \hat{v}_i^x) + \pi \in [0, 2\pi]$ is measured in the xy-plane. The unit sphere is partitioned into $M_\theta \times M_\phi$ bins with uniform spacing in θ and ϕ, and the occupation probabilities are computed accounting for the non-uniform solid angle element $\mathrm{d}\Omega = \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\phi$. Specifically, the raw histogram counts n_{jk} in bin (j, k) are weighted by the inverse of the solid angle factor $\sin\theta_j$, where θ_j denotes the central polar angle of bin j, to obtain a density that would be uniform for an isotropic distribution. The Shannon entropy of the resulting probability distribution is

$$H_{\mathrm{orient}} = -\sum_{j=1}^{M_\theta} \sum_{k=1}^{M_\phi} \tilde{p}_{jk} \ln \tilde{p}_{jk}, \tag{15}$$

where $\tilde{p}_{jk}$ represents the solid-angle-corrected probability. An isotropic distribution of velocities yields maximum entropy, characteristic of the disordered phase, while perfect alignment of all velocities along a single direction produces zero entropy, corresponding to the fully ordered state. The default binning employs $M_\theta = 18$ and $M_\phi = 36$, providing approximately 10° resolution in both angular coordinates. Normalization to the interval $[0, 1]$ is performed analogously to the positional entropy.

The local alignment entropy H_{local} characterizes the heterogeneity of orientational order across the system by examining the distribution of local order parameters. For each particle i, a local order parameter φ_i^{loc} is defined as the magnitude of the normalized velocity sum within its neighborhood,

$$\varphi_i^{\mathrm{loc}} = \frac{1}{|\mathcal{N}_i| v_0} \left| \sum_{j \in \mathcal{N}_i} \boldsymbol{v}_j \right|, \tag{16}$$

where $|\mathcal{N}_i|$ denotes the number of neighbors of particle i including itself. This quantity ranges from zero, when neighbor velocities cancel exactly, to unity, when all neighbors move in the same direction. The collection $\{\varphi_i^{\mathrm{loc}}\}_{i=1}^N$ is histogrammed into M_{loc} bins spanning the interval $[0, 1]$, and the Shannon entropy of the resulting distribution is computed as

$$H_{\mathrm{local}} = -\sum_{m=1}^{M_{\mathrm{loc}}} p_m \ln p_m, \tag{17}$$

where p_m is the fraction of particles with local order parameter falling in bin m. A homogeneous system in which all particles experience the same local environment yields a sharply peaked distribution with low entropy, while a heterogeneous system with coexisting ordered and disordered regions produces a broad distribution with high entropy. This measure is particularly informative near the phase transition, where spatial coexistence of phases leads to enhanced heterogeneity [6]. The neighborhood identification employs the same interaction radius r used in the dynamical evolution, and the efficient neighbor search is performed using the k-d tree algorithm as implemented in the SciPy library [28], which reduces the computational complexity from $\mathcal{O}(N^2)$ to $\mathcal{O}(N \log N)$ for each query.

The pair correlation entropy H_{pair} is derived from the radial distribution function $g(r')$, which measures the probability of finding a particle at distance r' from a reference particle relative to an ideal gas at the same density [29]. The radial distribution function is computed by histogramming pairwise distances into M_r radial bins spanning the interval $[0, L/2]$, with the upper limit set by the minimum image convention. For each bin centered at r'_m with width Δr, the histogram count is normalized by the expected count for a uniform distribution,

$$g(r'_m) = \frac{n_m}{N \cdot 4\pi r'^2_m \Delta r \cdot \rho}, \tag{18}$$

where n_m is the number of particle pairs with separation in bin m and $\rho = N/L^3$ is the number density. For an ideal gas without spatial correlations, $g(r') = 1$ for all r',

while deviations from unity indicate short-range order (peaks at preferred separations) or exclusion (depletion at small distances). The pair correlation entropy quantifies the deviation from ideal gas behavior through the integral

$$S_{\text{pair}} = -\int_0^{L/2} g(r') \ln g(r')\, 4\pi r'^2 \, \mathrm{d}r', \tag{19}$$

which is evaluated numerically using the trapezoidal rule. Rather than using this integral directly, the implementation employs a normalized measure based on the integrated deviation from unity,

$$H_{\text{pair}} = 1 - \frac{\int_0^{L/2} |g(r') - 1|\, r'^2 \, \mathrm{d}r'}{\int_0^{L/2} r'^2 \, \mathrm{d}r'}, \tag{20}$$

which yields values near unity for an ideal gas (disordered) and smaller values for systems with pronounced structural correlations (ordered). This formulation ensures robust behavior when $g(r')$ passes through zero or exhibits large fluctuations at small r' due to limited sampling.

The Voronoi entropy H_{Vor} quantifies geometric disorder through the distribution of Voronoi cell volumes. The Voronoi tessellation partitions space into convex polyhedra, each containing all points closer to a given particle than to any other [30]. In a perfectly ordered crystalline arrangement, all Voronoi cells have identical volumes, while in a disordered configuration, cell volumes exhibit substantial variation. The Voronoi tessellation is computed using the Qhull algorithm as interfaced through the SciPy library [28, 31]. To approximate periodic boundary conditions, the particle configuration is replicated in the six cardinal directions, and the tessellation is performed on this extended set; only cells corresponding to particles in the primary box are retained for analysis. The cell volumes $\{V_i\}_{i=1}^{N}$ are extracted from the convex hull of each cell's vertices, and the coefficient of variation $c_v = \sigma_V / \langle V \rangle$ is computed, where σ_V and $\langle V \rangle$ denote the standard deviation and mean of the volume distribution, respectively. The Voronoi entropy is then defined as

$$H_{\text{Vor}} = 1 - \exp(-c_v), \tag{21}$$

which maps the coefficient of variation to the interval $[0, 1]$ with the limiting behaviors $H_{\text{Vor}} \to 0$ as $c_v \to 0$ (uniform cells, ordered) and $H_{\text{Vor}} \to 1$ as $c_v \to \infty$ (highly variable cells, disordered). This exponential mapping ensures smooth behavior and bounded range while preserving the monotonic relationship between geometric variability and the entropy measure.

The mutual information $I(\boldsymbol{r}; \boldsymbol{v})$ between position and velocity quantifies the statistical dependence of a particle's direction of motion on its location within the simulation domain. In a spatially homogeneous system, the velocity distribution is independent of position, yielding zero mutual information, while the emergence of spatial structure—such as coherent domains with locally aligned velocities—introduces correlations that increase the mutual information [12]. The mutual information is computed from the

joint and marginal entropies as

$$I(\boldsymbol{r};\boldsymbol{v}) = H(\boldsymbol{r}) + H(\boldsymbol{v}) - H(\boldsymbol{r},\boldsymbol{v}), \tag{22}$$

where $H(\boldsymbol{r})$ is the entropy of the discretized position distribution, $H(\boldsymbol{v})$ is the entropy of the discretized velocity direction distribution, and $H(\boldsymbol{r},\boldsymbol{v})$ is the joint entropy. The position space is discretized into M_{MI}^3 cubic cells, while velocity directions are classified into the eight octants defined by the signs of the velocity components, yielding a joint distribution over $8M_{\mathrm{MI}}^3$ states. The normalized mutual information $\tilde{I} = I/\min(H(\boldsymbol{r}), H(\boldsymbol{v})) \in [0,1]$ is reported, with values near zero indicating spatial homogeneity and values approaching unity indicating strong position-velocity coupling. The default implementation uses $M_{\mathrm{MI}} = 8$ bins per spatial dimension.

The SCI provides a composite measure of disorder by combining the six individual entropy measures into a single scalar. Each component metric is designed to yield values near unity for disordered configurations and near zero for ordered configurations, enabling straightforward averaging. The SCI is defined as the arithmetic mean

$$\mathrm{SCI} = \frac{1}{6}\left(\tilde{H}_{\mathrm{pos}} + \tilde{H}_{\mathrm{orient}} + \tilde{H}_{\mathrm{local}} + H_{\mathrm{pair}} + H_{\mathrm{Vor}} + (1 - \tilde{I})\right), \tag{23}$$

where the mutual information enters with opposite sign since high mutual information corresponds to spatial structure rather than disorder. The equal weighting of components reflects the absence of prior knowledge favoring any particular measure, though alternative weightings could be employed for specific applications. The SCI provides a single-number summary of the overall degree of spatial disorder, complementing the traditional order parameter φ which captures only the global alignment of velocities. Importantly, the SCI and φ probe distinct aspects of the system state: a configuration with all particles aligned but uniformly distributed yields high φ and intermediate SCI, while a configuration with clustered particles moving in random directions yields low φ and intermediate SCI with different component contributions.

The temporal evolution of these metrics is computed by evaluating each measure at every saved time step of the simulation trajectory. For a trajectory comprising T snapshots, this yields time series $\{H_{\mathrm{pos}}^n\}_{n=1}^T$, $\{H_{\mathrm{orient}}^n\}_{n=1}^T$, and so forth, which may be analyzed to characterize the relaxation dynamics and steady-state fluctuations. Summary statistics including the mean, standard deviation, minimum, maximum, and final values are computed over the second half of each time series to exclude the initial transient and focus on the statistically stationary regime. These summary statistics, together with the full time series, are exported in both CSV and NetCDF formats for subsequent analysis and visualization.

The computational cost of evaluating the full metric suite scales as $\mathcal{O}(N^2)$ per time step, dominated by the pairwise distance calculations required for the radial distribution function and the local alignment field. The Voronoi tessellation contributes an additional $\mathcal{O}(N \log N)$ cost, while the remaining histogram-based measures require only $\mathcal{O}(N)$ operations. For the system sizes and trajectory lengths considered in the test cases, the total metric computation time is comparable to or exceeds the simulation time itself, reflecting the comprehensive nature of the diagnostic suite. Progress

feedback is provided through the tqdm library to monitor the metric computation during extended analyses.

2.4 Data Analysis

Beyond the entropy metrics implemented within the `manuk-kepudang` library, a comprehensive understanding of collective motion dynamics necessitates detailed examination of velocity component distributions. The velocity field $\{\boldsymbol{v}_i\}_{i=1}^{N}$ encodes the instantaneous dynamical state of the particle ensemble, and the statistical properties of its Cartesian components (v_x, v_y, v_z) reveal fundamental characteristics of the ordered and disordered phases [32]. The analysis described in this section was performed using auxiliary Python scripts employing the SciPy statistical library [28], operating on trajectory data exported from `manuk-kepudang` in NetCDF format.

The probability density function of each velocity component is estimated nonparametrically using kernel density estimation (KDE), which constructs a smooth approximation to the underlying distribution without imposing parametric assumptions [33]. For a sample $\{v_i^\alpha\}_{i=1}^{N}$ of component $\alpha \in \{x, y, z\}$, the KDE takes the form

$$\hat{f}^\alpha(v) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{v - v_i^\alpha}{h}\right), \tag{24}$$

where $K(\cdot)$ denotes the kernel function and $h > 0$ is the bandwidth parameter controlling the degree of smoothing. We employ the Gaussian kernel $K(u) = (2\pi)^{-1/2} \exp(-u^2/2)$ with bandwidth selected according to Scott's rule $h = N^{-1/5}\hat{\sigma}$, where $\hat{\sigma}$ is the sample standard deviation [34]. This choice balances bias and variance optimally for distributions close to Gaussian while remaining robust for moderate departures from normality. The resulting density estimates are normalized to unit maximum for comparative visualization across regimes with differing concentration.

The distributional properties of velocity components are assessed through a battery of nonparametric statistical tests selected for their robustness to non-normality and their complementary diagnostic power. The Shapiro–Wilk test [35] evaluates the null hypothesis that a sample originates from a normally distributed population, computing a test statistic W that measures the correlation between the ordered sample values and the expected order statistics of a standard normal distribution. Rejection of normality at conventional significance levels ($p < 0.05$) indicates that the velocity component distribution deviates from Gaussian form, which bears implications for the applicability of parametric statistical methods and reveals aspects of the underlying dynamics such as the emergence of preferred directions in the ordered phase.

To assess whether the three velocity components share a common distribution—as expected from isotropy in the disordered phase—we employ the Kruskal–Wallis H-test [36], a nonparametric analogue of one-way analysis of variance. The test statistic

$$H = \frac{12}{n(n+1)} \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1), \tag{25}$$

where R_j denotes the sum of ranks for group j, n_j is the group size, and $n = \sum_j n_j$ is the total sample size, follows approximately a chi-squared distribution with $k-1$ degrees of freedom under the null hypothesis of identical population distributions. Significant departure from this null indicates anisotropy in the velocity field, characteristic of the ordered flocking state wherein particles preferentially move along a common direction.

Pairwise comparisons between velocity components are conducted using the Mann–Whitney U-test [37], which evaluates whether one distribution is stochastically greater than another without assuming specific parametric forms. The test statistic U counts the number of times an observation from the first sample precedes an observation from the second in the combined ranking, with significance assessed against the null hypothesis of equal location. This test complements the omnibus Kruskal–Wallis test by identifying which specific component pairs exhibit distributional differences when anisotropy is detected.

The homogeneity of variance across velocity components is tested using Levene's test, which assesses the null hypothesis that population variances are equal. Unlike Bartlett's test, Levene's test remains valid under substantial departures from normality, making it appropriate for the potentially non-Gaussian velocity distributions encountered in collective motion systems. The test transforms the data to absolute deviations from group medians and performs a one-way analysis of variance on these transformed values, yielding an F-statistic whose significance indicates heteroscedasticity when rejected.

Dependence structure among velocity components is quantified through Spearman's rank correlation coefficient [38], which measures the strength and direction of monotonic association without requiring linear relationships. For components α and β, the coefficient

$$\rho_{\alpha\beta} = 1 - \frac{6\sum_{i=1}^{N} d_i^2}{N(N^2-1)}, \tag{26}$$

where d_i denotes the difference between ranks of v_i^α and v_i^β, ranges from -1 (perfect negative monotonic relationship) to $+1$ (perfect positive monotonic relationship), with zero indicating absence of monotonic association. In the disordered phase, isotropy implies near-zero correlations between components, whereas the ordered phase may exhibit correlations reflecting the geometry of the collective motion direction.

Temporal evolution of the velocity distributions is assessed by comparing samples from early and late simulation times using the two-sample Kolmogorov–Smirnov test [39]. The test statistic

$$D_{n,m} = \sup_v \left| F_n^{(1)}(v) - F_m^{(2)}(v) \right|, \tag{27}$$

where $F_n^{(1)}$ and $F_m^{(2)}$ are the empirical cumulative distribution functions of the two samples, quantifies the maximum vertical distance between the distributions. Large values of $D_{n,m}$ indicate that the velocity distribution has evolved substantially over the simulation duration, as expected during relaxation from random initial conditions toward the stationary state characteristic of each dynamical regime.

To complement significance testing with a measure of practical importance, we compute Cliff's delta [40] as an effect size quantifying the degree of temporal change. This nonparametric measure calculates the probability that a randomly selected observation from the final time exceeds one from the initial time, minus the reverse probability:

$$\delta = \frac{\#(v_i^{\text{end}} > v_j^{\text{begin}}) - \#(v_i^{\text{end}} < v_j^{\text{begin}})}{N_{\text{end}} \cdot N_{\text{begin}}}, \tag{28}$$

where the counts enumerate all pairwise comparisons between samples. The magnitude of δ is interpreted according to established thresholds: $|\delta| < 0.147$ indicates negligible effect, $0.147 \leq |\delta| < 0.33$ small effect, $0.33 \leq |\delta| < 0.474$ medium effect, and $|\delta| \geq 0.474$ large effect [41]. Unlike standardized mean differences, Cliff's delta remains meaningful for non-normal distributions and is bounded within $[-1, 1]$, facilitating interpretation across diverse dynamical regimes.

The statistical analysis is applied at three temporal snapshots for each simulation case: the initial configuration ($t = 0$), the midpoint ($t = T/2$), and the final state ($t = T$), where T denotes the total simulation duration. This sampling strategy captures both the transient relaxation dynamics and the approach to statistical stationarity, enabling assessment of how distributional properties evolve as the system organizes from random initial conditions into its characteristic collective behavior pattern. For each snapshot, the full complement of tests is computed, with significance assessed at the conventional $\alpha = 0.05$ level and multiple comparison corrections applied where appropriate.

3 Results and Discussion

The four test cases summarized in Table 1 were executed on a workstation equipped with an Intel Core i7-8550U processor running Fedora Linux 39, with computational times ranging from approximately 355 seconds for Case 1 to 931 seconds for Case 3, the majority of which was consumed by the entropy metric computations rather than the dynamical simulation itself. The resulting trajectories, comprising between 501 and 801 temporal snapshots depending on the case, provide a comprehensive dataset for examining the collective behavior across distinct dynamical regimes.

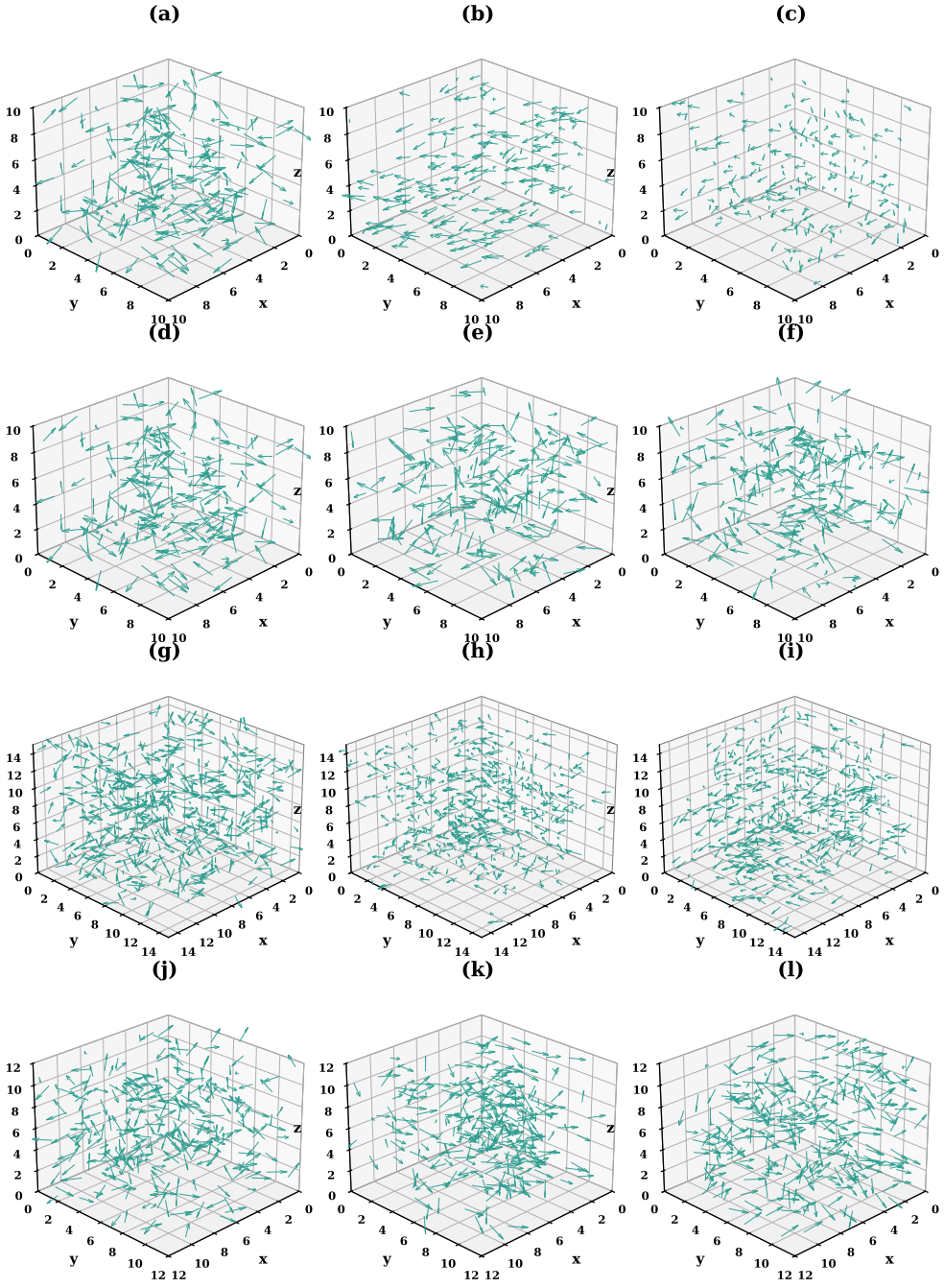

Fig. 1 Three-dimensional visualization of particle configurations and velocity fields at three temporal snapshots for each simulation case. Rows correspond to Case 1 (a–c), Case 2 (d–f), Case 3 (g–i), and Case 4 (j–l), while columns represent the beginning ($t = 0$), middle ($t = T/2$), and end ($t = T$) of each simulation. Arrows indicate instantaneous velocity vectors with the velocity magnitude $|\boldsymbol{v}|$, which remains constant at $v_0 = 0.5$ by construction. The emergence of collective alignment is visually apparent in Case 1 and Case 3, where velocity vectors progressively orient along a common direction, whereas Case 2 maintains isotropic orientations throughout the simulation duration.

The evolution of the polar order parameter φ reveals the expected bifurcation between ordered and disordered states as governed by the noise amplitude η. In Case 1 with $\eta = 0.3$, the order parameter undergoes a rapid transition from its initial value of $\varphi_0 \approx 0.054$—consistent with the $N^{-1/2}$ scaling anticipated for uncorrelated initial conditions—to a steady-state value exceeding 0.95 within approximately 50 time units, indicating the spontaneous emergence of coherent flocking motion. This behavior stands in stark contrast to Case 2, where the elevated noise amplitude $\eta = 2.0$ suppresses any tendency toward alignment, maintaining φ in the range 0.01–0.26 throughout the 500 time-unit duration with a mean value of approximately 0.106. The visual manifestation of this dichotomy is readily apparent in Figure 1, where panels (a–c) and (d–f) display the temporal progression of velocity orientations for the ordered and disordered cases, respectively. The ordered phase exhibits a striking coalescence of velocity vectors into a narrow directional cone by mid-simulation, whereas the disordered phase retains an essentially isotropic distribution of orientations at all sampled times.

Case 3, designed to probe finite-size effects with $N = 500$ particles in an enlarged domain, likewise achieves an ordered state with final order parameter $\varphi \approx 0.886$, though the relaxation timescale appears somewhat extended relative to Case 1, consistent with the expectation that larger systems require more time to establish long-range correlations [42]. Perhaps most instructive is Case 4, wherein the noise amplitude $\eta = 1.0$ places the system in the vicinity of the order-disorder transition. The order parameter in this case fluctuates around $\varphi \approx 0.5$ with a standard deviation of approximately 0.06, neither achieving the strong alignment of the subcritical regime nor collapsing to the near-zero values characteristic of the supercritical disordered phase. This intermediate behavior suggests that the chosen parameters situate the system near the critical noise amplitude η_c, where the competing tendencies toward alignment and randomization are approximately balanced [17].

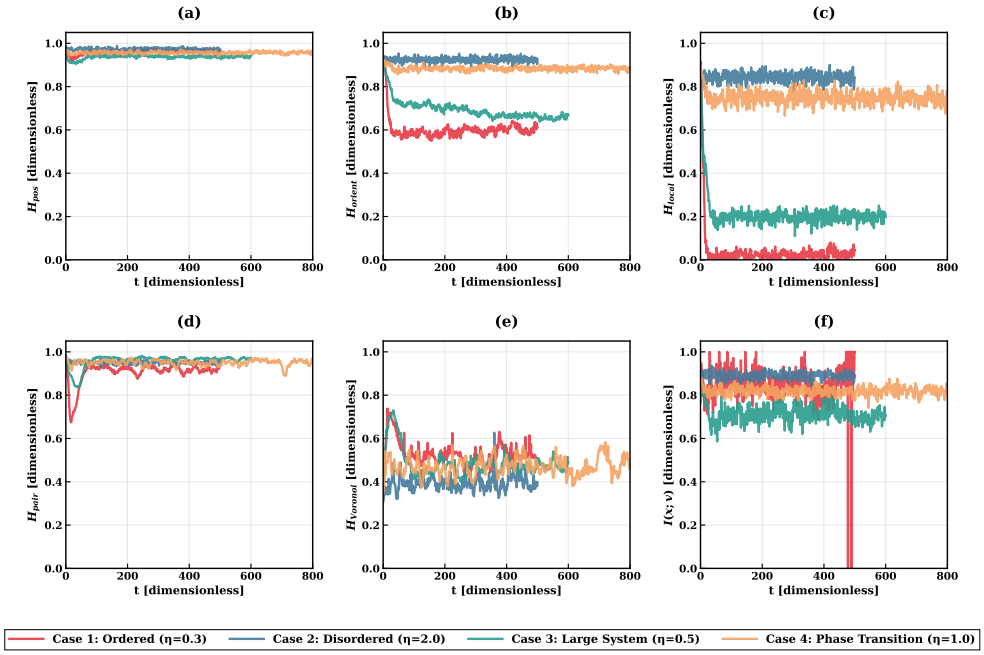

Fig. 2 Temporal evolution of the six entropy measures implemented in `manuk-kepudang` across the four simulation cases. Panels show (a) positional entropy $\tilde{H}_{\text{pos}}$, (b) orientational entropy $\tilde{H}_{\text{orient}}$, (c) local alignment entropy $\tilde{H}_{\text{local}}$, (d) pair correlation entropy H_{pair}, (e) Voronoi entropy H_{Vor}, and (f) position-velocity mutual information $\tilde{I}(\boldsymbol{r};\boldsymbol{v})$. All quantities are normalized to the interval $[0, 1]$, with higher values indicating greater disorder or heterogeneity. The clear separation between cases in panels (b) and (c) demonstrates the discriminatory power of orientational and local alignment entropies for characterizing the order-disorder transition.

The temporal evolution of the six entropy measures, displayed in Figure 2, reveals substantial variation in their sensitivity to the collective state of the system. The positional entropy $\tilde{H}_{\text{pos}}$, shown in panel (a), remains uniformly high across all cases with values in the range 0.91–0.99, indicating that the spatial distribution of particles maintains near-uniform coverage of the simulation domain irrespective of the dynamical regime. This observation is consistent with the absence of cohesive interactions in the standard Vicsek model, which lacks any mechanism for density segregation or clustering [6]. The modest fluctuations observed, particularly the transient dip in Case 1 near $t \approx 20$, likely reflect finite-size effects in the histogram estimation rather than genuine spatial inhomogeneity.

In contrast, the orientational entropy $\tilde{H}_{\text{orient}}$ exhibits pronounced differentiation between cases, as evident in panel (b) of Figure 2. Case 1 displays a rapid decrease from approximately 0.94 to a steady-state value near 0.61 within the first 100 time units, reflecting the concentration of velocity directions into a narrow angular region as collective alignment develops. Case 2 maintains high orientational entropy throughout ($\tilde{H}_{\text{orient}} \approx 0.92$), consistent with the persistence of isotropic velocity distributions

in the disordered phase. The intermediate cases exhibit behavior commensurate with their respective noise amplitudes: Case 3 achieves a final orientational entropy of approximately 0.67, while Case 4 fluctuates around 0.88, indicating partial but incomplete alignment. These results suggest that the orientational entropy serves as an effective proxy for the order parameter, capturing the degree of velocity alignment through an information-theoretic lens.

The local alignment entropy $\tilde{H}_{\mathrm{local}}$, presented in panel (c), emerges as perhaps the most discriminating measure among those implemented. In Case 1, this quantity plummets from approximately 0.79 to values approaching zero within approximately 30 time units, indicating that the distribution of local order parameters $\{\varphi_i^{\mathrm{loc}}\}$ collapses to a delta-function-like peak near unity—that is, nearly all particles experience a highly aligned local environment. The disordered Case 2 maintains elevated local alignment entropy ($\tilde{H}_{\mathrm{local}} \approx 0.84$) throughout, reflecting the broad distribution of local order parameters when neighbor velocities are uncorrelated. Case 3 and Case 4 occupy intermediate positions, with Case 3 eventually achieving low values similar to Case 1, and Case 4 remaining elevated near 0.75. The steep initial transient observed in all cases reflects the reorganization from random initial conditions, during which the local environment of each particle undergoes rapid evolution before the system settles into its characteristic stationary state.

The pair correlation entropy H_{pair}, derived from the radial distribution function $g(r')$, exhibits more subtle variation across cases, as shown in panel (d). All cases maintain values in the range 0.67–0.98, with Case 1 displaying a pronounced transient dip to approximately 0.67 near $t \approx 16$ before recovering to steady-state values around 0.94. This transient may reflect the temporary emergence of short-range positional correlations as the velocity field organizes, though the effect appears to be short-lived. The generally high values across all cases indicate that the spatial structure remains close to that of an ideal gas, consistent with the absence of excluded-volume or attractive interactions in the Vicsek model.

The Voronoi entropy H_{Vor}, displayed in panel (e), quantifies the geometric regularity of the particle packing through the distribution of Voronoi cell volumes. All cases exhibit an initial increase from relatively low values ($H_{\mathrm{Vor}} \approx 0.31$–$0.37$) toward higher steady-state values in the range 0.38–0.53, with considerable fluctuations throughout the simulation. The initial low values reflect the relatively uniform random initialization, which by chance produces a more regular packing than might be expected in equilibrium. The subsequent increase and fluctuation suggest that the dynamical evolution introduces geometric heterogeneity through the ballistic motion of particles, though no clear separation between ordered and disordered cases is apparent in this measure.

The position-velocity mutual information $\tilde{I}(\boldsymbol{r};\boldsymbol{v})$, shown in panel (f), probes the coupling between spatial location and velocity direction. High values indicate that knowledge of a particle's position provides information about its velocity direction, as would occur if spatially coherent velocity domains existed within the system. The results reveal generally high mutual information across all cases ($\tilde{I} \approx 0.70$–0.92), with Case 1 displaying occasional spikes to unity and a pronounced transient drop near $t \approx 478$. The physical interpretation of these fluctuations remains somewhat

unclear, though they may reflect finite-size effects in the discretization scheme used to estimate the mutual information, particularly when the velocity distribution is highly concentrated in a few angular bins.

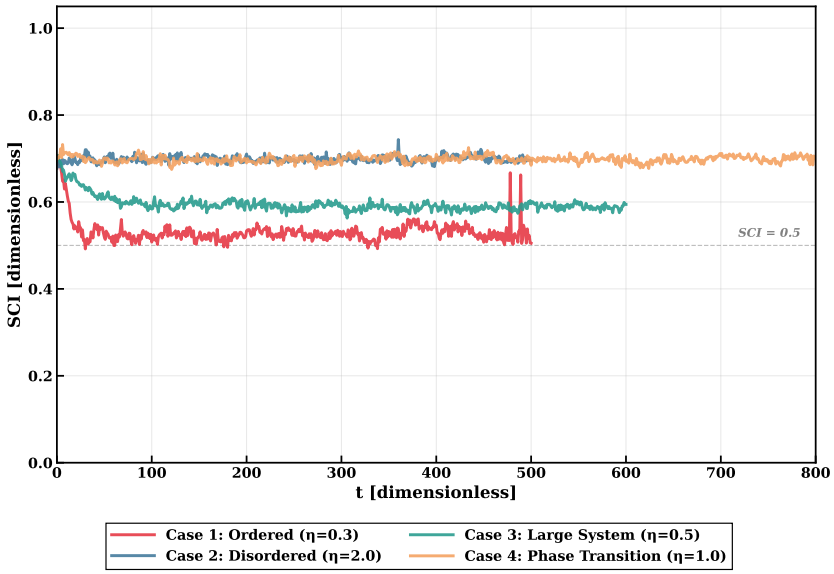

Fig. 3 Temporal evolution of the SCI for the four simulation cases. The SCI is computed as the arithmetic mean of the six component entropies according to equation (23), with values approaching unity indicating high disorder and values approaching zero indicating strong spatial and orientational order. The horizontal dashed line marks SCI = 0.5 for reference. Case 1 (ordered) achieves the lowest steady-state SCI (≈ 0.53), while Cases 2 and 4 maintain elevated values near 0.70, and Case 3 occupies an intermediate position (≈ 0.59).

The SCI, constructed as the arithmetic mean of the six component measures according to equation (23), provides a composite characterization of system disorder as displayed in Figure 3. The SCI successfully discriminates between the ordered and disordered regimes, with Case 1 achieving a final value of approximately 0.505 and Case 2 maintaining values near 0.695 throughout. Notably, the Spearman correlation between SCI and the noise amplitude η across the four cases yields $\rho = 1.0$, indicating perfect rank-order correspondence between the imposed noise level and the resulting disorder as quantified by the SCI. This monotonic relationship suggests that the SCI provides a physically meaningful measure of the system's proximity to the ordered state, complementing the traditional order parameter φ while incorporating information about spatial structure and heterogeneity that φ alone cannot capture.

The Kruskal–Wallis test applied to the SCI time series across all four cases yields a test statistic $H \approx 1918$ with $p < 0.001$, indicating significant differences between cases. Pairwise Mann–Whitney U-tests reveal that all case pairs exhibit statistically significant differences ($p < 0.001$) except for Cases 2 and 4, which cannot be distinguished at conventional significance levels ($p \approx 0.73$). This latter finding is instructive:

despite their different noise amplitudes ($\eta = 2.0$ versus $\eta = 1.0$), both cases reside in or near the disordered regime and thus exhibit similar steady-state SCI values, though their transient behaviors and order parameter distributions differ substantially.

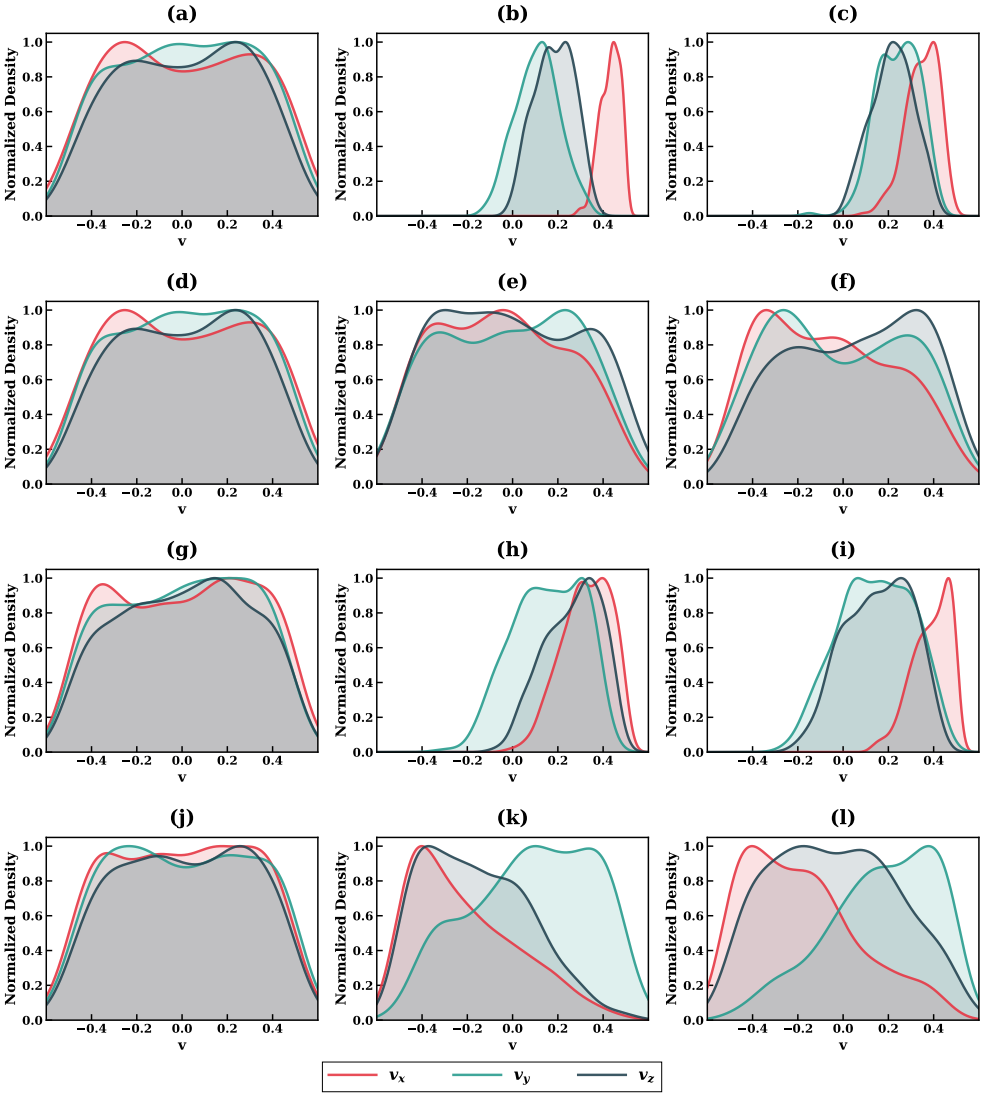

Fig. 4 Kernel density estimates of velocity component distributions at three temporal snapshots for each simulation case. Rows correspond to Case 1 (a–c), Case 2 (d–f), Case 3 (g–i), and Case 4 (j–l), with columns representing beginning, middle, and end times as in Figure 1. Each panel displays normalized KDE curves for v_x (red), v_y (teal), and v_z (dark blue). The transition from broad, overlapping distributions at $t = 0$ to narrow, separated peaks in the ordered cases (rows 1 and 3) reflects the emergence of a preferred direction of collective motion, whereas the disordered case (row 2) maintains broad, symmetric distributions throughout.

The velocity component distributions, estimated via kernel density estimation and displayed in Figure 4, provide a complementary perspective on the collective dynamics that illuminates the microscopic basis of the order-disorder transition. At the initial time $t = 0$, all cases exhibit broad, roughly symmetric distributions spanning the kinematically accessible range $[-v_0, v_0]$ for each component, consistent with the isotropic initialization procedure. The distributions are not precisely uniform due to the constant-speed constraint, which induces correlations among the Cartesian components; nevertheless, the substantial overlap among v_x, v_y, and v_z distributions reflects the absence of any preferred direction in the initial state.

The temporal evolution of these distributions diverges dramatically between the ordered and disordered regimes. In Case 1, panels (b) and (c) of Figure 4 reveal a striking transformation: by mid-simulation, the v_x distribution has contracted to a narrow peak centered near $+0.4$, while v_y and v_z have shifted toward smaller positive values with considerably reduced variance. This pattern indicates the emergence of a collective velocity oriented predominantly along the positive x-direction with modest positive components in y and z. The Shapiro–Wilk test applied to these distributions yields $W < 0.7$ with $p < 0.001$, decisively rejecting the null hypothesis of normality and confirming that the aligned state produces highly non-Gaussian velocity statistics.

The Kruskal–Wallis test comparing the three velocity components within each snapshot provides a rigorous assessment of isotropy. At $t = 0$, the test yields $H \approx 0.3$ with $p \approx 0.86$ for Case 1, indicating no statistically significant difference among components—as expected for an isotropic initialization. By the final time, however, the same test yields $H \approx 280$ with $p < 0.001$, reflecting the profound anisotropy induced by collective alignment. Case 2 exhibits qualitatively different behavior: the Kruskal–Wallis statistic remains modest throughout ($H < 5$, $p > 0.05$), consistent with the maintenance of statistical isotropy in the disordered phase. These findings corroborate the visual impression from Figure 4 and demonstrate that standard nonparametric tests can effectively diagnose the symmetry-breaking transition characteristic of flocking systems.

The pairwise Spearman correlations between velocity components reveal additional structure in the aligned state. In Case 1 at the final time, the correlations $\rho_{xy} \approx 0.45$, $\rho_{xz} \approx 0.38$, and $\rho_{yz} \approx 0.52$ are all statistically significant ($p < 0.001$), indicating that the constant-speed constraint induces dependencies among components when the distribution is concentrated near a particular direction on the velocity sphere. In the disordered Case 2, by contrast, correlations remain near zero ($|\rho| < 0.1$) and statistically insignificant, as the broad angular distribution samples the velocity sphere sufficiently uniformly to decorrelate the Cartesian projections.

The temporal evolution of velocity distributions is quantified through the two-sample Kolmogorov–Smirnov test comparing initial and final distributions for each component. In Case 1, all three components yield $D > 0.5$ with $p < 0.001$, indicating that the distributions have changed substantially over the simulation duration. The associated Cliff's delta effect sizes range from 0.45 to 0.78, corresponding to medium-to-large effects in the standard classification scheme [41]. Case 2 exhibits much smaller changes, with $D < 0.15$ and $|\delta| < 0.15$ (negligible effect) for all components, confirming

that the disordered state represents a stationary distribution that is established early and maintained throughout the simulation.

The results presented here demonstrate that the `manuk-kepudang` library successfully reproduces the phenomenology of the three-dimensional Vicsek model across the ordered, disordered, and critical regimes. The entropy metrics implemented within the library provide quantitative characterization of distinct aspects of the collective state, with the orientational entropy and local alignment entropy emerging as particularly effective discriminators of the phase transition. The composite SCI offers a convenient scalar summary that correlates monotonically with the imposed noise amplitude and complements the traditional order parameter in characterizing system behavior.

Several limitations of the present study warrant acknowledgment. The system sizes employed ($N \leq 500$) are modest by contemporary standards, and finite-size effects may influence the quantitative values of the entropy measures and their fluctuations [43]. The precise location of the critical noise amplitude η_c depends on the particle density and interaction radius in a manner that has not been systematically explored here, and the parameters of Case 4 were chosen heuristically rather than through careful finite-size scaling analysis. Furthermore, the computational cost of the entropy metrics—which dominated the total execution time in all cases—may limit the applicability of the full diagnostic suite to very large systems or long trajectories without algorithmic improvements such as spatial hashing or parallel evaluation.

Despite these limitations, the agreement between the observed behavior and theoretical expectations for the Vicsek model provides confidence that the implementation is correct and that the entropy metrics offer meaningful physical insight. The sharp transition in orientational entropy observed in Cases 1 and 3, the maintenance of high entropy in Case 2, and the intermediate fluctuating behavior in Case 4 collectively paint a coherent picture of the order-disorder transition that is consistent with extensive prior literature on the Vicsek model and its variants [3, 9]. The `manuk-kepudang` library thus provides a validated platform for investigating collective motion phenomena and their information-theoretic characterization in three-dimensional active matter systems.

4 Conclusion

This work has presented `manuk-kepudang`, an open-source Python library for simulating the three-dimensional Vicsek model and characterizing collective motion through a suite of information-theoretic entropy measures. The library implements the canonical flocking dynamics with local alignment interactions and angular noise, providing a validated computational framework for investigating the order-disorder transition that emerges in self-propelled particle systems. The six entropy metrics introduced—positional, orientational, local alignment, pair correlation, Voronoi, and position-velocity mutual information—together with the composite SCI, offer complementary perspectives on the spatial and orientational organization of the particle ensemble that extend beyond traditional scalar order parameters. The orientational and local alignment entropies proved particularly effective at discriminating between ordered and disordered phases, while the SCI demonstrated monotonic correspondence

with the imposed noise amplitude across all tested regimes. The statistical analysis framework, employing nonparametric tests appropriate for the non-Gaussian velocity distributions characteristic of flocking systems, provides rigorous methodology for quantifying distributional changes and temporal evolution. Although computational costs associated with the entropy metrics currently exceed those of the dynamical simulation itself, the library nonetheless offers a practical and accessible tool for researchers investigating collective behavior in active matter systems, with potential applications spanning biological swarms, robotic coordination, and the fundamental physics of nonequilibrium phase transitions.

Acknowledgements

This work was supported by the Dean's Distinguished Fellowship Award from the University of California, Riverside and the PPMI ITB 2025.

AI Disclosure

This manuscript utilized Claude Sonnet 4.5 (Anthropic) for English language editing and grammar refinement. All scientific content, including conceptual development, theoretical derivations, data analysis, figure generation, and interpretation of results, was solely conducted by the authors. The authors take full responsibility for the accuracy and integrity of the research presented herein.

Open Research

The `manuk-kepudang` library is freely available as open-source software under the MIT License. The complete source code, documentation, and usage examples are maintained at https://github.com/sandyherho/manuk-kepudang. Users may install the library directly from the Python Package Index (PyPI) via `pip install manuk-kepudang`, with the package hosted at https://pypi.org/project/manuk-kepudang/.

All Python scripts for data analysis, statistical computations, and figure generation presented in this article are archived in a supplementary repository at https://github.com/sandyherho/suppl-manuk-kepudang. This repository enables full reproducibility of the results and figures reported herein.

The complete simulation outputs, including trajectory data in NetCDF format, figures in PNG, PDF, and EPS formats, statistical reports in plain text, computation log files, and animated visualizations in GIF format, are permanently archived and publicly accessible at the Open Science Framework (OSF): https://doi.org/10.17605/OSF.IO/JNYU2.

References

[1] Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., Shochet, O.: Novel Type of Phase Transition in a System of Self-Driven Particles. Physical Review Letters **75**(6), 1226 (1995). https://doi.org/10.1103/PhysRevLett.75.1226

[2] Ramaswamy, S.: The Mechanics and Statistics of Active Matter. Annual Review of Condensed Matter Physics **1**, 323–345 (2010). https://doi.org/10.1146/annurev-conmatphys-070909-104101

[3] Vicsek, T., Zafeiris, A.: Collective Motion. Physics Reports **517**(3–4), 71–140 (2012). https://doi.org/10.1016/j.physrep.2012.03.004

[4] Marchetti, M.C., Joanny, J.F., Ramaswamy, S., Liverpool, T.B., Prost, J., Rao, M., Simha, R.A.: Hydrodynamics of soft active matter. Reviews of Modern Physics **85**(3), 1143–1189 (2013). https://doi.org/10.1103/RevModPhys.85.1143

[5] Couzin, I.D., Krause, J., James, R., Ruxton, G.D., Franks, N.R.: Collective Memory and Spatial Sorting in Animal Groups. Journal of Theoretical Biology **218**(1), 1–11 (2002). https://doi.org/10.1006/jtbi.2002.3065

[6] Chaté, H., Ginelli, F., Grégoire, G., Raynaud, F.: Collective motion of self-propelled particles interacting without cohesion. Physical Review E **77**(4), 046113 (2008). https://doi.org/10.1103/PhysRevE.77.046113

[7] Grégoire, G., Chaté, H.: Onset of Collective and Cohesive Motion. Physical Review Letters **92**(2), 025702 (2004). https://doi.org/10.1103/PhysRevLett.92.025702

[8] Ginelli, F.: The Physics of the Vicsek Model. The European Physical Journal Special Topics **225**, 2099–2117 (2016). https://doi.org/10.1140/epjst/e2016-60066-8

[9] Chaté, H.: Dry Aligning Dilute Active Matter. Annual Review of Condensed Matter Physics **11**, 189–212 (2020). https://doi.org/10.1146/annurev-conmatphys-031119-050752

[10] Toner, J., Tu, Y., Ramaswamy, S.: Hydrodynamics and phases of flocks. Annals of Physics **318**(1), 170–244 (2005). https://doi.org/10.1016/j.aop.2005.04.011

[11] Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal **27**(3), 379–423 (1948). https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[12] Cover, T.M., Thomas, J.A.: Elements of Information Theory, 2nd edn. John Wiley & Sons, Hoboken (2005). https://doi.org/10.1002/047174882X

[13] Herho, S., Anwar, I., Herho, K., Dharma, C., Irawan, D.: Comparing Scientific Computing Environments for Simulating 2D Non-Buoyant Fluid Parcel Trajectory under Inertial Oscillation: A Preliminary Educational Study. Indonesian Physical Review **7**(3), 451–468 (2024). https://doi.org/10.29303/ipr.v7i3.335

[14] Herho, S., Fajary, F., Herho, K., Anwar, I., Suwarman, R., Irawan, D.: Reappraising double pendulum dynamics across multiple computational platforms. CLEI

Electronic Journal **28**(1) (2025). https://doi.org/10.19153/cleiej.28.1.10

[15] Lam, S.K., Pitrou, A., Seibert, S.: Numba: a LLVM-Based Python JIT Compiler. In: Proceedings of the Second Workshop on the LLVM Compiler Infrastructure in HPC, pp. 1–6. ACM, New York (2015). https://doi.org/10.1145/2833157.2833162

[16] Romanczuk, P., Bär, M., Ebeling, W., Lindner, B., Schimansky-Geier, L.: Active Brownian particles: From individual to collective stochastic dynamics. The European Physical Journal Special Topics **202**, 1–162 (2012). https://doi.org/10.1140/epjst/e2012-01529-y

[17] Aldana, M., Dossetti, V., Huepe, C., Kenkre, V.M., Larralde, H.: Phase Transitions in Systems of Self-Propelled Agents and Related Network Models. Physical Review Letters **98**(9), 095702 (2007). https://doi.org/10.1103/PhysRevLett.98.095702

[18] Allen, M.P., Tildesley, D.J.: Computer Simulation of Liquids, 2nd edn. Oxford University Press, Oxford (2017). https://doi.org/10.1093/oso/9780198803195.001.0001

[19] Huberman, B.A., Glance, N.S.: Evolutionary games and computer simulations. Proceedings of the National Academy of Sciences **90**(16), 7716–7718 (1993). https://doi.org/10.1073/pnas.90.16.7716

[20] Schönfisch, B., de Roos, A.: Synchronous and Asynchronous Updating in Cellular Automata. Biosystems **51**(3), 123–143 (1999). https://doi.org/10.1016/S0303-2647(99)00025-8

[21] Hockney, R.W., Eastwood, J.W.: Computer Simulation Using Particles. CRC Press, Boca Raton (1988). https://doi.org/10.1201/9780367806934

[22] Marsaglia, G.: Choosing a Point from the Surface of a Sphere. The Annals of Mathematical Statistics **43**(2), 645–646 (1972). https://doi.org/10.1214/aoms/1177692644

[23] Harris, C.R., Millman, K.J., Walt, S.J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N.J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M.H., Brett, M., Haldane, A., Río, J.F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., Oliphant, T.E.: Array programming with NumPy. Nature **585**, 357–362 (2020). https://doi.org/10.1038/s41586-020-2649-2

[24] Matsumoto, M., Nishimura, T.: Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Transactions on Modeling and Computer Simulation **8**(1), 3–30 (1998). https://doi.org/10.1145/272991.272995

[25] Herho, S., Kaban, S.N., Irawan, D.E., Kapid, R.: Efficient 1D Heat Equation Solver: Leveraging Numba in Python. Eksakta: Berkala Ilmiah Bidang MIPA **25**(2), 126–137 (2024). https://doi.org/10.24036/eksakta/vol25-iss02/487

[26] Herho, S.H.S., Trilaksono, N.J., Fajary, F.R., Napitupulu, G., Anwar, I.P., Khadami, F., Irawan, D.E.: kh2d-solver: A Python library for idealized two-dimensional incompressible Kelvin-Helmholtz instability. Applied and Computational Mechanics **19**(2) (2025). https://doi.org/10.24132/acm.2025.1040

[27] Batty, M.: Spatial Entropy. Geographical Analysis **6**(1), 1–31 (1974). https://doi.org/10.1111/j.1538-4632.1974.tb01014.x

[28] Virtanen, P., Gommers, R., Oliphant, T.E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., Walt, S.J., Brett, M., Wilson, J., Millman, K.J., Mayorov, N., Nelson, A.R.J., Jones, E., Kern, R., Larson, E., Carey, C.J., Polat, .I., Feng, Y., Moore, E.W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E.A., Harris, C.R., Archibald, A.M., Ribeiro, A.H., Pedregosa, F., Mulbregt, P.: SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature Methods **17**, 261–272 (2020). https://doi.org/10.1038/s41592-019-0686-2

[29] Hansen, J.-P., McDonald, I.R.: Theory of Simple Liquids, 4th edn. Academic Press, Oxford (2013). https://doi.org/10.1016/C2010-0-66723-X

[30] Okabe, A., Boots, B., Sugihara, K., Chiu, S.N., Kendall, D.G.: Spatial Tessellations: Concepts and Applications of Voronoi Diagrams. John Wiley & Sons, Chichester (2000). https://doi.org/10.1002/9780470317013

[31] Barber, C.B., Dobkin, D.P., Huhdanpaa, H.: The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software **22**(4), 469–483 (1996). https://doi.org/10.1145/235815.235821

[32] Cavagna, A., Cimarelli, A., Giardina, I., Parisi, G., Santagati, R., Stefanini, F., Viale, M.: Scale-free correlations in starling flocks. Proceedings of the National Academy of Sciences **107**(26), 11865–11870 (2010). https://doi.org/10.1073/pnas.1005766107

[33] Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Routledge, New York (1998). https://doi.org/10.1007/978-1-4899-3324-9

[34] Scott, D.W.: Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York (1992). https://doi.org/10.1002/9780470316849

[35] Shapiro, S.S., Wilk, M.B.: An analysis of variance test for normality (complete samples). Biometrika **52**(3–4), 591–611 (1965). https://doi.org/10.1093/biomet/52.3-4.591

[36] Kruskal, W.H., Wallis, W.A.: Use of Ranks in One-Criterion Variance Analysis. Journal of the American Statistical Association **47**(260), 583–621 (1952). https://doi.org/10.1080/01621459.1952.10483441

[37] Mann, H.B., Whitney, D.R.: On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. The Annals of Mathematical Statistics **18**(1), 50–60 (1947). https://doi.org/10.1214/aoms/1177730491

[38] Spearman, C.: The Proof and Measurement of Association Between Two Things. The American Journal of Psychology **15**(1), 72–101 (1904). https://doi.org/10.2307/1412159

[39] Smirnov, N.: Table for Estimating the Goodness of Fit of Empirical Distributions. The Annals of Mathematical Statistics **19**(2), 279–281 (1948). https://doi.org/10.1214/aoms/1177730256

[40] Cliff, N.: Dominance statistics: Ordinal analyses to answer ordinal questions. Psychological Bulletin **114**(3), 494–509 (1993). https://doi.org/10.1037/0033-2909.114.3.494

[41] Romano, J., Kromrey, J.D., Coraggio, J., Skowronek, J.: Appropriate Statistics for Ordinal Level Data: Should We Really Be Using t-Test and Cohen's d for Evaluating Group Differences on the NSSE and Other Surveys? Annual Meeting of the Florida Association of Institutional Research, 1–33 (2006)

[42] Baglietto, G., Albano, E.V.: Nature of the order-disorder transition in the Vicsek model for the collective motion of self-propelled particles. Physical Review E **80**(5), 050103 (2009). https://doi.org/10.1103/PhysRevE.80.050103

[43] Baglietto, G., Albano, E.V.: Finite-size scaling analysis and dynamic study of the critical behavior of a model for the collective displacement of self-driven individuals. Physical Review E **78**(2), 021125 (2008). https://doi.org/10.1103/PhysRevE.80.050103