# Mutual Information as a Tool for Identifying Phase Transitions in Complex Systems

Robert Wicks*, Sandra Chapman* and Richard Dendy*†

* Centre for Fusion, Space and Astrophysics, Physics Department, University of Warwick, Coventry, CV4 7AL

† UKAEA Culham Division, Culham Science Centre, Abingdon OX14 3DB

## Abstract:

Simple models for flocking in biological systems are of considerable interest from a complex systems perspective. We investigate the dynamics of the Vicsek model of interacting self propelled particles using mutual information. The mutual information of the system is found to peak at the phase transition in a similar manner to the susceptibility of the system. The mutual information can be calculated from relatively short data series and from data from just one or two particles.

## Introduction:

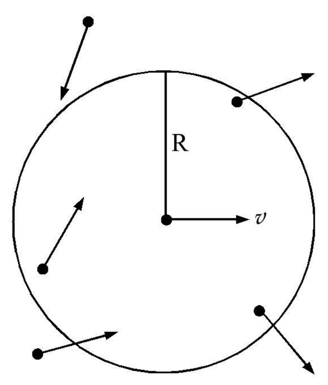
- Growing interest in the use of mutual information in complex systems, physical and life sciences.
- In natural plasmas mutual information has recently been used [1] [2] to quantify the causal linkage between strongly nonlinear timeseries.
- Mutual information has previously been shown to identify phase transitions in stationary complex systems [3].
- Here it is shown to be able to identify the phase transition of a dynamical complex system.
- We consider the Vicsek model of self-propelled particles which is a biologically motivated complex system [4].

## The Vicsek Model:

- A complex system with a phase transition.
- Simple rules govern motion of many particles but only at short ranges.
- Long range, large scale, ordered behaviour for low values of noise amplitude η.
- Disorder and small scales for high η.
- Phase transition at critical noise $\eta_c$ with clusters on all scales and interaction over all scales.

### The Rules:

$$x_{n+1} = x_n + \vec{v}\,\delta t$$
$$\theta_{n+1} = \langle\theta_n\rangle_R + \delta\theta_n$$




Where $v$ is a velocity with direction θ and $|v|$ is kept constant, average over $R$ is the average angle within a circle of radius and δθ is an **iid** random angle in the range $-\eta \le \delta\theta \le \eta$.

### Measuring Order:

- Order is defined for the Vicsek model using an order parameter φ.
- Fluctuations in the system are characterised by the susceptibility χ.
- φ → zero and χ → ∞ at the phase transition.

$$\phi = \frac{1}{Nv}\left|\sum_{i=1}^{N}\vec{v}_i\right| \qquad \chi = \sigma(\phi)^2 = \langle\phi^2\rangle - \langle\phi\rangle^2$$

The order parameter is near unity for small η, for large η the system is disordered and between there is a phase transition [5] at critical noise $\eta_c$.

## Results:

The typical behaviour of the system with noise can be seen in **figure 1**, this is a system with 3000 particles in a $50^2$ box with $R = 0.5$ and $|v| = 0.1$.

**Figure 1b)** shows φ and χ as a function of noise for this system but at much higher resolution in noise than the pictures in **figure 1**. χ peaks at the phase transition and φ decreases rapidly with noise although never reaches zero showing the effect of the finite size of the system.

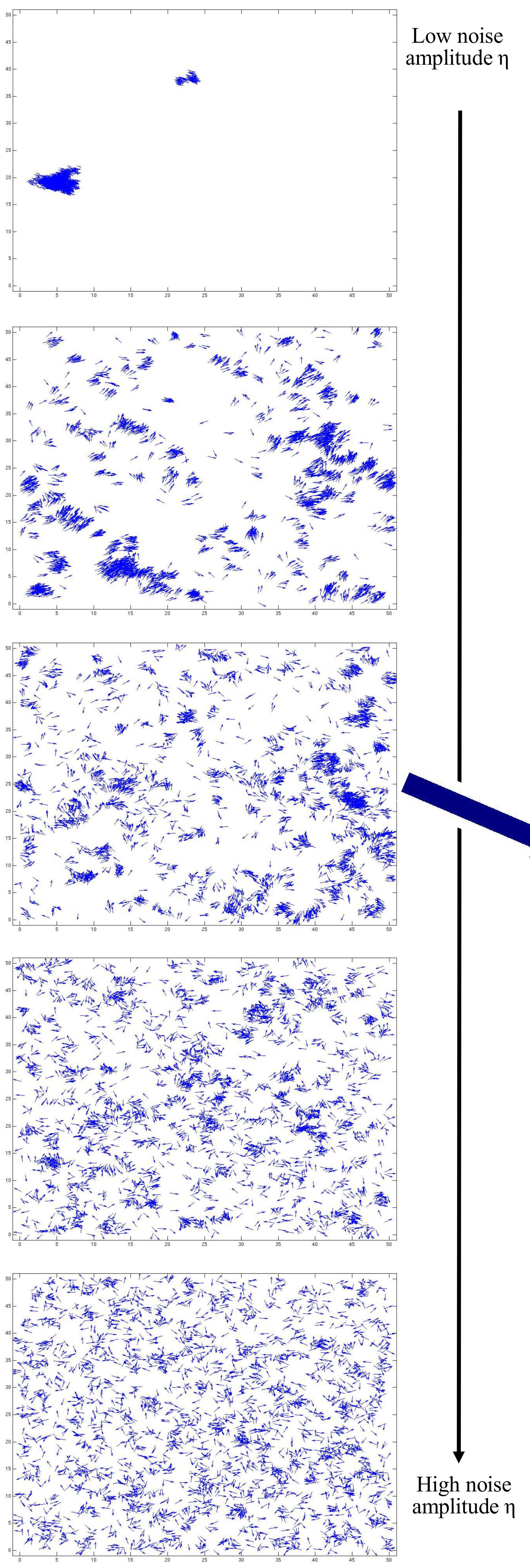
**Figure 2** shows how MI is calculated for the Vicsek system. Position, in this case x, is plotted against angle θ and the occupation probabilities $P(x_i)$, $P(\theta_j)$, and $P(x_i,\theta_j)$ are calculated. These are then used to calculate entropies H(X), H(Θ) and H(X,Θ) from which MI is calculated.

The key results are shown in **figures 3** and **4**, this is a detailed analysis of MI as a function of noise.

- MI and χ peak in approximately the same place.
- MI has smaller error bars allowing the exact position of the peak to be identified more readily than in χ.
- Peak is not in exactly the same place as MI is a measure of the clustering in space and velocity, unlike susceptibility which is the variance of the velocity.

In **figure 4** a timeseries of position and velocity has been used. Data from only 2 particles was used for a time of 10000 steps and a similar peak can be found.

**a)**




**b)**

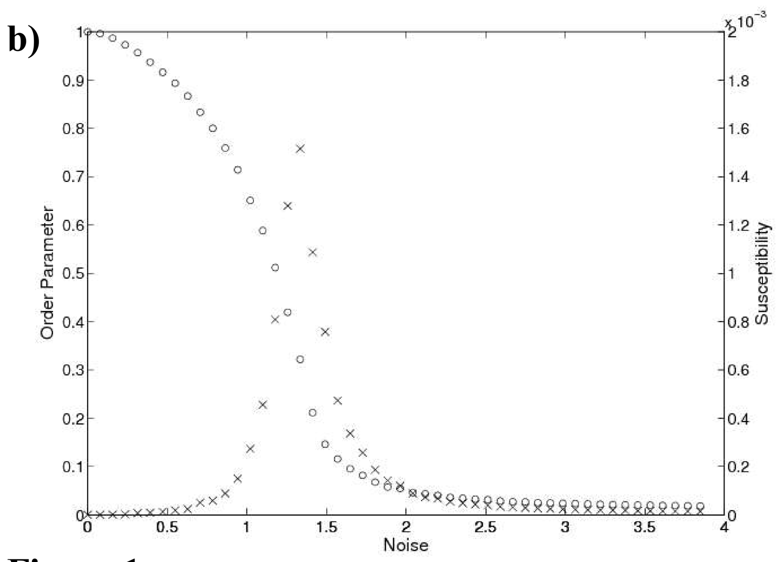



**Figure 1:**
**a)** A typical Vicsek system at five different values of η (0, 0.2π, 0.4π, 0.6π, 0.8π and π). The system has ρ = 1.2, N = 3000, $R = 0.5$ and $v = 0.1$. Arrows represent the velocity of each particle.

**b)** The final graph shows the effect of noise on the order parameter (circles) and susceptibility (crosses) for a full range of noises. The susceptibility peaks at the phase transition where without finite size effects the order parameter aught to become zero.

## References:

[1] T. K. March, S. C. Chapman, R. O. Dendy, Geophys. Res. Lett. 32, (2005).
[2] T. K. March, S. C. Chapman, R. O. Dendy, Physica D. 200, (2005).
[3] T. Vicsek, A. Czirók, E. Ben-Jacob, I. Cohen and O. Shochet, Phys. Rev. Lett. 75, 1226 (1995).
[4] H. Matsuda, K. Kudo, R. Nakamura, O. Yamakawa, T. Murata, Int. Journal of Theor. Phys. 35, No. 4, 839 (1996).
[5] A. Czirók, H. E. Stanley and T. Vicsek, J. Phys. A: Math. Gen. 30 1375-1385 (1997).

## Mutual Information:

- I(X,Y), measures quantitatively the information shared between two signals (X,Y) of almost any type.
- Signals are decomposed into alphabets of possible measurements and each 'letter' $(x_i)$ is assigned a probability $P(x_i)$.
- Signal entropy H(X) is calculated.
- Linear combinations of entropies determine the mutual information I for multiple signals.
- Unit of measurement is bits due to logarithm base two.

$$H(X) = -\sum_i P(x_i)\log_2(P(x_i))$$
$$I(X,Y) = H(X) + H(Y) - H(X,Y)$$
$$I(X,Y) = \sum_{i,j} P(x_i,y_j)\log_2\left(\frac{P(x_i,y_j)}{P(x_i)P(y_j)}\right)$$

### Using Mutual Information (MI):

- Probabilities must be calculated from the data.
- Many ways to do this, the easiest and fastest is to divide the signal up into bins and calculate occupation probabilities for the bins.
- Results using this method on the third window in **figure 1** are shown in **figure 2**.
- The number of bins used greatly affects the result, however by varying this over a range a stable solution can be found.

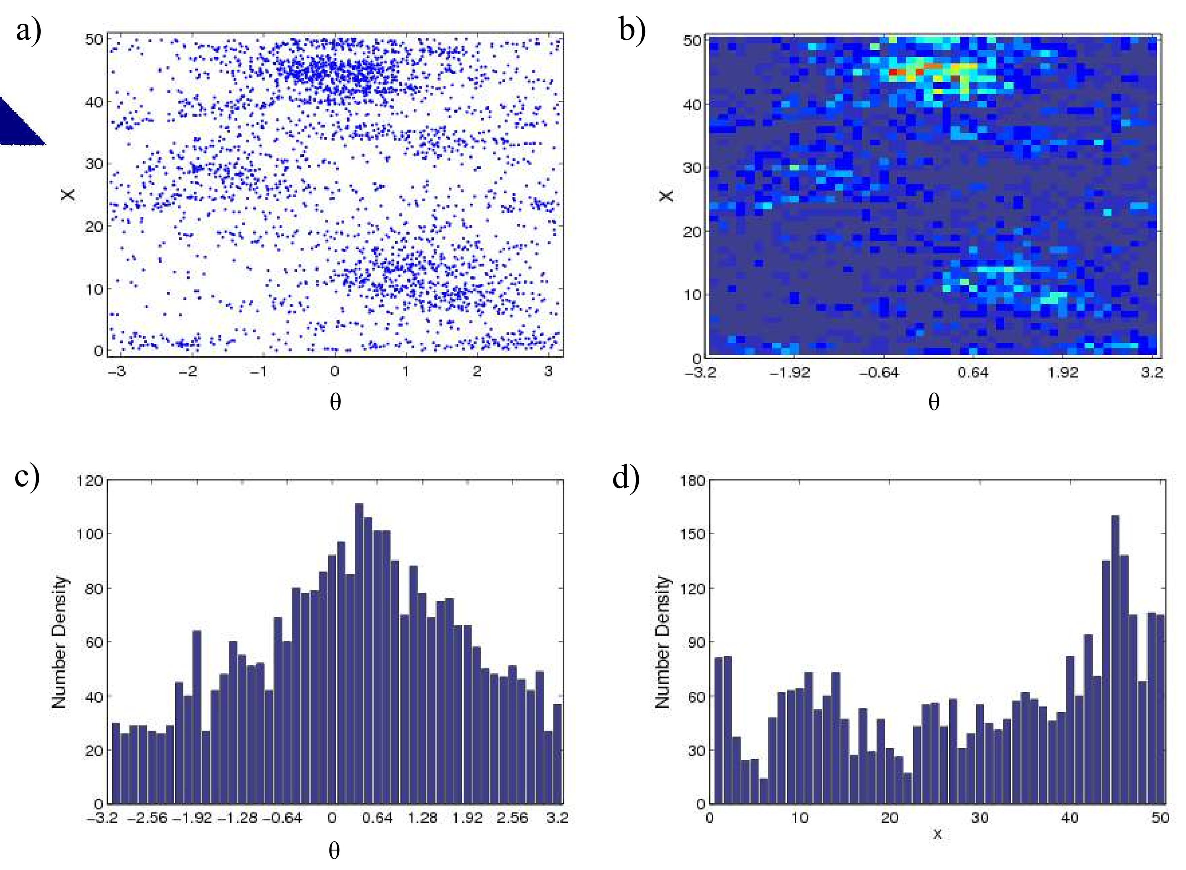



**Figure 2:** The key steps in calculating MI, a) Plot of x against θ. b) Discretisation of a) showing variation of $P(x_i,\theta_j)$ over parameter space. c) Occupation number for 50 bins in θ from which $P(\theta_j)$ is calculated. d) Occupation number for 50 bins in x form which $P(x_i)$ is calculated.

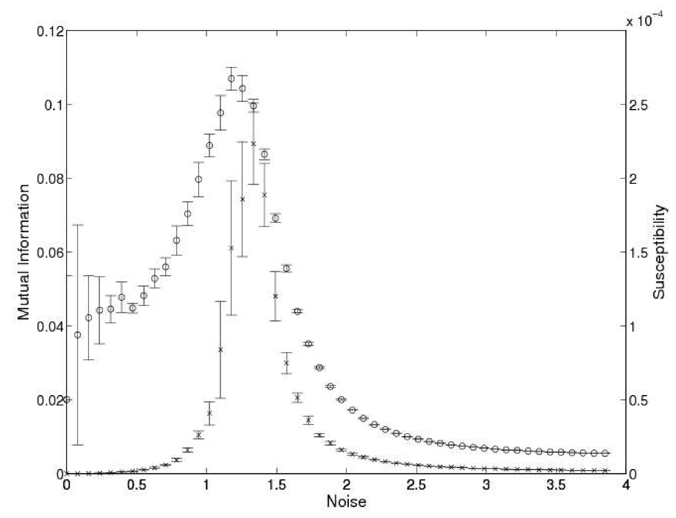



**Figure 3:** Variation of MI (circles) and susceptibility (crosses) with noise. 50,000 time steps were used and the MI and susceptibility were averaged over a further 10,000 steps.

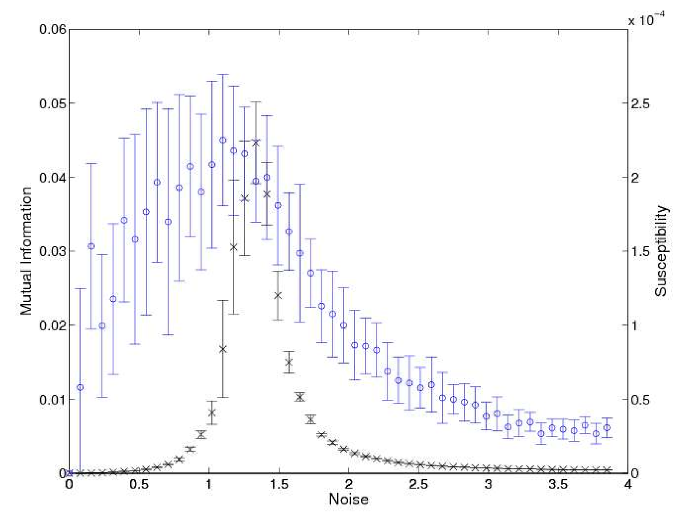



**Figure 4:** Variation of MI calculated from the timeseries of two particles' position and velocity for 10000 steps (circles) and susceptibility of the whole system (crosses) with noise.

## Conclusions:

- Mutual information is able to provide a simple method for identifying the phase transition in the Vicsek model and therefore shows potential for use in any non-equilibrium dynamical system.
- Error on MI is smaller than that on the susceptibility, it measures the correlation of the particles in velocity and position, which is only an approximate measure of where the phase transition is.
- Careful choices of number of bins must be made to give good accuracy.
- A timeseries of measurements from just one particle is enough to make an approximation of MI for the whole system.
- Phase transitions can be infered from little data. MI is of use in analysis of data that has limited numbers of observations, for example tracer particles in laminar and turbulent fluids, GPS tracking of animals and timeseries from satellites.

## Acknowledgments:

This work was supported in part by the EPSRC. RW acknowledges a PPARC CASE PhD studentship in association with UKAEA. The Centre for Scientific Computing at the University of Warwick provided the computing facilities.