

Cite this: DOI:00.0000/xxxxxxxxxx

Electronic supplementary information for: Dependence of phase behavior and surface tension on particle stiffness for active Brownian particles[†]

Nicholas Lauersdorf^{a‡}, Thomas Kolb^{b‡}, Moslem Moradi^a, Ehssan Nazockdast^a and Daphne Klotsa^{*a}

Received Date

Accepted Date

DOI:00.0000/xxxxxxxxxx

The Electronic Supplementary Information section contains additional figures, videos and data that supplement the discussion in the main text.

1 Choosing the proportionality constant, β , of the pair force

The analytical prediction of the lattice spacing (a) hinges on the balance between both the pair active force ($\mathbf{F}^{\text{pair}} = \beta \mathbf{F}^a$) and the repulsive interparticle force within the bulk dense phase ($F^{\text{WCA}}(a)$) and, therefore, our choice of the proportionality constant, β . In addition, our analytical calculations for the cluster radius (r_c), the interparticle pressure (Π^{P}), and the surface tension (γ) are strongly dependent on the lattice spacing (a) and, in turn, β . As a result, it is critical to find a β that best predicts our lattice spacing from simulation. As we show in fig. 2, one can determine β by fitting eq. 6 to our measured values for a vs F^* from simulation.

We quantify the quality of fit through measuring the coefficient of determination:

$$R^2 = 1 - \frac{\sum_i (a_i - \hat{a}_i)^2}{\sum_i (a_i - \bar{a})^2}, \quad (1)$$

where a_i is our i^{th} lattice spacing from simulation, \bar{a} is the mean lattice spacing from simulation, and \hat{a}_i is our i^{th} lattice spacing from theory (Using eq. 6 in paper). β is chosen to be where R^2 is closest to 1.0. By this standard, we obtain a strong fit ($R^2 > 0.99$) at $\beta \approx 2$ ($R^2 > 0.995$). The optimal fit is found to be when $\beta = 1.92$, resulting in a maximum $R^2 = 0.9963$.

2 Computing the dense phase composition of soft ABPs

We start with a few basic relations for the number of particles (N), simulation area (A), and area fraction (ϕ) for particles with

area, $A_p = \frac{\pi\sigma^2}{4}$ in the dense (d) and gas (g) phases.

$$N = N_d + N_g. \quad (2)$$

$$A = A_d + A_g. \quad (3)$$

$$\phi = \frac{NA_p}{A}. \quad (4)$$

We now proceed in deriving an expression for the number of particles in the dense phase (N_d). To this end, we begin with equation 3, substituting using equation 4,

$$\frac{NA_p}{\phi} = \frac{N_d A_p}{\phi_d} + \frac{N_g A_p}{\phi_g}. \quad (5)$$

We can divide through by the particle area, and rearrange to isolate, N_d ,

$$N_d = \phi_d \left(\frac{N}{\phi} - \frac{N_g}{\phi_g} \right). \quad (6)$$

To remove the dependence on N_g , we substitute equation 2 into equation 6:

$$N_d = \phi_d \left(\frac{N}{\phi} - \frac{N - N_d}{\phi_g} \right). \quad (7)$$

Through algebraic simplification, we solve for N_d :

$$\frac{N_d}{\phi_d} = \frac{N}{\phi} - \frac{N}{\phi_g} + \frac{N_d}{\phi_g}, \quad (8)$$

$$\frac{N_d}{\phi_d} - \frac{N_d}{\phi_g} = \frac{N}{\phi} - \frac{N}{\phi_g}, \quad (9)$$

$$N_d \left(\frac{1}{\phi_d} - \frac{1}{\phi_g} \right) = N \left(\frac{1}{\phi} - \frac{1}{\phi_g} \right), \quad (10)$$

$$N_d \left(\frac{\phi_g - \phi_d}{\phi_d \phi_g} \right) = N \left(\frac{\phi_g - \phi}{\phi \phi_g} \right). \quad (11)$$

^a Department of Applied Physical Sciences, University of North Carolina at Chapel Hill, USA

^b Department of Chemistry, University of North Carolina at Chapel Hill, USA

[‡] These authors contributed equally to this work.

Finally, we solve for the dense phase size as a function of system (ϕ), gas phase (ϕ_g) and dense phase (ϕ_d) area fractions in addition to the total system size (N),

$$N_d = N \left(\frac{\phi_d(\phi_g - \phi)}{\phi(\phi_g - \phi_d)} \right). \quad (12)$$

3 Instantiating Seeded Clusters

As shown by Levis *et al.*, soft particles undergo MIPS at a smaller critical cluster size than hard particles due to a lower nucleation barrier¹. However, this similarly impacts the stability of MIPS. Due to a lower nucleation barrier, softer clusters are more likely to destabilize and break apart before reaching a stable configuration. Therefore, since our softest systems rarely underwent MIPS from random initial conditions within $\tau = 300\tau_r$, we instantiated small circular clusters at a dense-phase area fraction (ϕ_d) predicted in eq. 8 of paper, of size that is at least one fourth of and no more than equal to its predicted steady state size (N_d from eq. 12), thus accelerating the process of overcoming kinetic limitations of cluster formation and, in turn, reducing the necessary computation time.

We see that instantiating initially seeded clusters does not influence whether MIPS occurs or not in our systems as systems that are not predicted to undergo MIPS quickly destabilized, forming a single, homogeneous gas phase. Similarly, the resulting phase-separated, steady-state cluster structure (fig. S1) and size (fig. S2) are independent of the initially seeded cluster size.

4 Method for Determining the Cluster Radius and Interface Width

In section §4.1, we label the dense-dilute interface as the region within the dense phase with both sharply decreasing area fraction and large inwards orientational alignment of particles' body forces. In this section, we shall detail our mathematical approach to discerning this region. The dense phase exhibits two spatial regions of varying properties: a bulk dense phase and a dense-dilute interface². The bulk dense phase has constant number density, $n(r)$, and negligible alignment towards the cluster's center of mass, $\alpha(r) = -\hat{\mathbf{p}}(r) \cdot \mathbf{r}$; whereas, the dense-dilute interface exhibits a monotonically decreasing density from the dense to the dilute phase density², resembling that of typical equilibrium liquid-gas interfaces³⁻⁵, and a high degree of alignment towards the cluster's center of mass. Exploiting these distinct trends in the interface and bulk dense phase, we have mathematically defined the interface and detail a computational procedure for determining the start and end of the dense-dilute interface layer.

Utilizing these observations, we assigned the product of alignment and number density, $\alpha(r)n(r)$ (the integrand of the true pressure, see eq. 26 in paper) as a function of distance from the cluster's center of mass, $r = \|\mathbf{r}\|$, to be our metric of interest for identifying the interface. Knowing that the maximum alignment must be within the interface of finite width, h , by our definition, we first found the radial location, r_{\max} , of the maximum of the pressure integrand, $\max(\alpha(r)n(r))$, which approximately corresponded to the location of peak alignment, α_{\max} .

With r_{\max} identified, we then defined limits for the magnitude

and slope of the pressure integrand to find the continuous region which satisfies these conditions, or the interface. To easily do this, we takes radial steps from r_{\max} either inwards towards the cluster's center of mass or outwards towards the cluster edge to identify the start of the interface ($r_c - h$) and the end of the interface (r_c) respectively. At each radial step, we measure whether that position satisfies either of two criteria. If it does, then it is considered part of the interface. However, if it is the first radial step in either direction which does not, it is labelled as an interface boundary, making all radial points between these two boundaries the interface.

Now, we will define those limits. First, we see that the magnitude of the pressure integrand within the interface region must be discernibly greater than the noise levels in our measurements, giving us the condition: $\alpha(r)n(r) \geq 0.2\max[\alpha(r)n(r)]$. However, with only this definition and despite the data being smoothed and averaged over three neighboring positions, noise peaks can similarly make our interface seem wider than it is due to our relatively sparse radial measurement, $\Delta r = 3.0$. Therefore, we also note the pressure integrand must also be sharply changing between α_{\max} and either $\alpha(r_c - h)$ or $\alpha(r_c)$ with maximum magnitude slopes in either the inward or outward directions located at $r_c - h < r_{\text{in}} < r_{\max}$ and $r_{\max} < r_{\text{out}} < r_c$ respectively. We move radially inward or outward from r_{in} and r_{out} respectively, checking whether each radial position satisfies our second criterion: $|\frac{d}{dr}[\alpha(r)\phi(r)]| \geq 0.2\max(|\frac{d}{dr}[\alpha(r)n(r)]|)$. Upon the first radial step in either direction failing to satisfy both criteria, that position is labelled as an interface boundary and we can simply take the difference to measure our interface width, h .

Though generally resembling a circular structure, the clusters' shapes evolved temporally and spatially due to the particles' dynamic nature, complicating radial measurements that assume the cluster is circular. Similar to the cluster radius fluctuating with angle from the center of mass, so too does the interface structure. Breaking of local symmetry results in varying sites of sources and sinks of particles across the interface that either adsorb or desorb particles respectively⁶, resulting in a dynamic interface width that varies for each conical surface. Due to the dissimilar cluster radii and interface structure in any direction, we performed radial measurements over twenty conical surfaces at every time step, where all particles within each specified 18° range from the center of mass were radially binned and their average alignment (fig. 5a-c) and local area fraction, $\bar{\phi}_{\text{local}}$, (fig. 5d-f) were measured as a function of distance from the cluster's center of mass. Though this measurement can be performed without discernment of the angle around the cluster's center of mass, it considerably blurs the interface and induces much greater inaccuracy for softer and less stable systems.

In order to accurately average the radial measurements of each conical surface at steady state (fig. 5), we normalized the distance from the center of mass by the measured cluster radius for each conical surface ($\|\mathbf{r}\|/r_c$). We obtained accurate local area fraction (fig. 5a, d, g) and orientational alignment (fig. 5b, e, h) at each normalized radii, demonstrating strong agreement with our analytical calculations for the area fraction (ϕ_{theory} , see eq. 8 in paper) and interparticle pressure (Π_{theory}^P , see eq. 17 in paper)

within the bulk dense phase. When the interface location is calculated using this method, the start and end of the interface approximately corresponded to where the local area fraction equals the analytical area fraction of the dense and gas phase respectively (fig. 5a-c). Calculating the beginning and end of the dense-dilute interface via this method determined the interface width, h , in our simulations when averaged over all conical surfaces and time at steady state (see SI, fig. S12).

5 Computing the surface tension of soft ABPs

Force balance dictates that the jump in force per unit area (traction) across the interface must balance against the force induced by the interface itself. Assuming that the interfacial forces are entirely due to surface tension, the equation describing this force balance reduces to

$$\Delta\hat{\mathbf{F}}_I = 2\gamma\kappa_m\hat{\mathbf{n}} + (\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla\gamma, \quad (13)$$

where $\Delta\hat{\mathbf{F}}_I = \hat{\mathbf{F}}_d - \hat{\mathbf{F}}_g$ is the jump in force per unit area across the interface, $\hat{\mathbf{n}}$ is the normal unit vector of the surface pointing outwards that is separating the dense and dilute phases, γ is the surface tension and $\kappa_m = 1/r_c$ is the mean curvature of the interface. The first and second terms on the right hand side represents the force jumps along the normal and tangential directions of the surface, respectively. Note that the tangential component becomes negligible, compared to the normal direction, in our system. Substituting $(\mathbf{I} - \hat{\mathbf{n}}\hat{\mathbf{n}}) \cdot \nabla\gamma \approx 0$ gives us

$$\hat{\mathbf{F}}_d - \hat{\mathbf{F}}_g = 2\gamma\kappa_m\hat{\mathbf{n}}, \quad (14)$$

Since all forces are acting in either direction normal to the surface, we re-write the coexistence criterion in terms of scalars:

$$\hat{F}_d - \hat{F}_g = 2\gamma\kappa_m, \quad (15)$$

As this equation represents units of force per unit area (or pressure), re-write the equation to find the surface tension, γ in terms of measurable quantities:

$$(\Pi_d - \Pi_g) \frac{r_c}{2} = \gamma, \quad (16)$$

where Π_g and Π_d are the gas and dense phase pressures respectively and $r_c = \kappa_m^{-1}$ is the cluster radius. Substituting $r_c = \left(\sqrt{\frac{\sqrt{3}}{2\pi}}\right)\sqrt{Na}$, and $\Pi_d^p = \frac{4\sqrt{3}F^a}{a}$ while approximating $\Pi_g \approx 0$, we obtain:

$$\gamma_{\text{active}} = 4\sqrt{\frac{3\sqrt{3}}{2\pi}}\sqrt{NF^a}, \quad (17)$$

which is linearly increasing with the simulation box size (\sqrt{n}), contrary to the mechanical definition of surface tension which should apply to any system. Therefore, there must be a flaw in our initial momentum equation (eq. 17), which was our treatment of the aligned interface as the surface tension and not as a body force density.

6 Determining the dependence of surface tension on softness, area fraction, and activity

So far, we have discussed how the interparticle separation, area fraction, radius, and pressure of the dense phase depend on the independent variables of our system: softness (ϵ), area fraction (ϕ), and activity (Pe). Now, we seek to extend this analysis to our calculation of the normalized surface tension, $(2\gamma)/(\Pi_d r_c)$, through the use of multivariate linear regression. Suppose we have a linear regression model such that our dependent variable, y , can be represented by a linear combination of independent variables (x_i)⁷:

$$y = y_{1,2,\dots,p} + e_{1,2,\dots,p} \quad (18)$$

where $y_{1,2,\dots,p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$ is the part of our dependent variable, y , explained by our p independent variables (x_1, x_2, \dots, x_p) with corresponding proportionality constants ($\beta_1, \beta_2, \dots, \beta_p$). $e_{1,2,\dots,p}$ is the part of y that cannot be unexplained by our linear regression model. We may judge how accurate our regression equation is at predicting y by correlating y and $x_{1,2,\dots,p}$. We estimate the proportionality constants, β_i by the method of least squares such that we obtain the best linearly fitted model for our data: $\hat{y} \equiv \hat{y}_{1,2,\dots,p}$. By definition, the multiple correlation coefficient of y on x_1, x_2, \dots, x_p is:

$$r = r_{0,1,2,\dots,p} = \text{corr}(y, \hat{y}) \quad (19)$$

where the function $\text{corr}(y, \hat{y})$ represents the correlation coefficient, r , of y on \hat{y} . The correlation coefficient of any two scalars, y on x , is defined to be:

$$\text{corr}(y, x) = r = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}} \quad (20)$$

where n is the sample size and \bar{x} and \bar{y} are the sample means of x and y respectively.

Clearly, as we increase the number of independent variables, p , the number of unique correlation coefficients quickly becomes large, scaling with $p(p+1)/2$. We account for these through the correlation matrix, $\mathbf{R}_{(p+1) \times (p+1)}$, whose elements correspond to the Pearson correlation coefficients (eq. 20 for each possible pairing of variables, $\mathbf{R} = (r_{i,j})_{0 \leq i,j \leq p}$ where:

$$r_{i,j} = \begin{cases} \text{corr}(y, x_j) & \text{if } i = 0, \\ 1 & \text{if } i = j, \\ \text{corr}(x_i, x_j) & \text{otherwise} \end{cases} \quad (21)$$

such that⁸:

$$\mathbf{R}_{(p+1) \times (p+1)} = \begin{bmatrix} 1 & r_{0,1} & r_{0,2} & \dots & r_{0,p} \\ r_{0,1} & 1 & r_{1,2} & \dots & r_{1,p} \\ r_{0,2} & r_{1,2} & 1 & \dots & r_{2,p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{0,p} & r_{1,p} & r_{2,p} & \dots & 1 \end{bmatrix} \quad (22)$$

It can be proven that⁹:

$$r_m = \sqrt{1 - \frac{|\mathbf{R}|}{r_{0,0}|r_{i,j}|_{p-1}}} \quad (23)$$

where r_m is the multiple correlation coefficient, $|\mathbf{R}|$ is the determinant of the correlation matrix, and $r_{0,0}|r_{i,j}|_{1 \leq i,j \leq p}$ corresponds to the cofactor of the $(0,0)^{\text{th}}$ element of \mathbf{R} . In this case, $r_{0,0} = 1$ and $|r_{i,j}|_{1 \leq i,j \leq p}$ is the determinant of correlations for x_1, x_2, \dots, x_p only.

Now, consider the correlation matrix of our system with variables $((2\gamma)/(\Pi_d r_c), \varepsilon, \phi, Pe)$, henceforth corresponding to (y, x_1, x_2, x_3) respectively and simplifying our linear regression to:

$$y_{1,2,3} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \quad (24)$$

Upon performing a multivariate linear regression (eq. 24) to our simulation data, our correlation matrix becomes:

$$\mathbf{R}_{4 \times 4} = \begin{bmatrix} 1.000 & -0.383 & 0.036 & -0.095 \\ -0.383 & 1.000 & -0.081 & -0.065 \\ 0.036 & -0.081 & 1.000 & 0.014 \\ -0.095 & -0.065 & 0.014 & 1.000 \end{bmatrix} \quad (25)$$

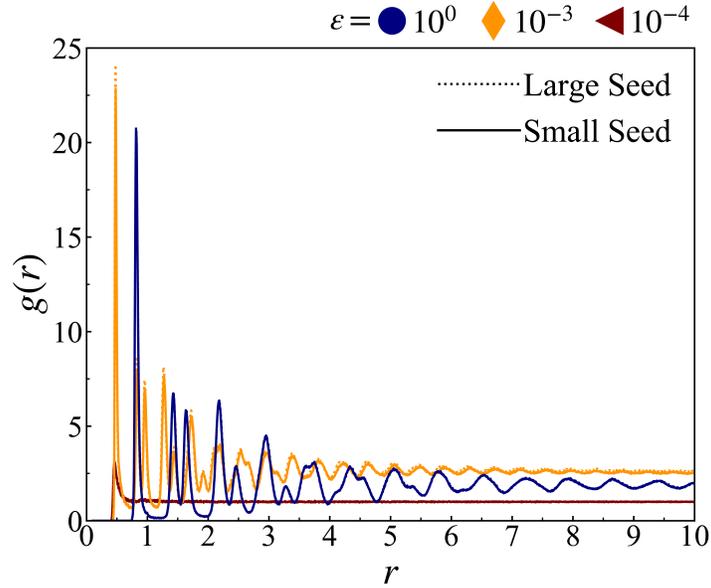
Applying eq. 23 to our simulation data (eq. 25), we find $r_m = 0.402$ and $r_m^2 = 0.161$. Though this r_m^2 already shows a relatively poor dependence of surface tension on softness, area fraction, and activity, the coefficient of determination will always increase as we incorporate more independent variables into our multivariate linear regression. Therefore, we calculate the adjusted coefficient of determination to punish a greater number of poorly chosen independent variables¹⁰:

$$r_{m,\text{adj}}^2 = 1 - \left[\frac{(1 - r_m^2)(n - 1)}{n - p - 1} \right] \quad (26)$$

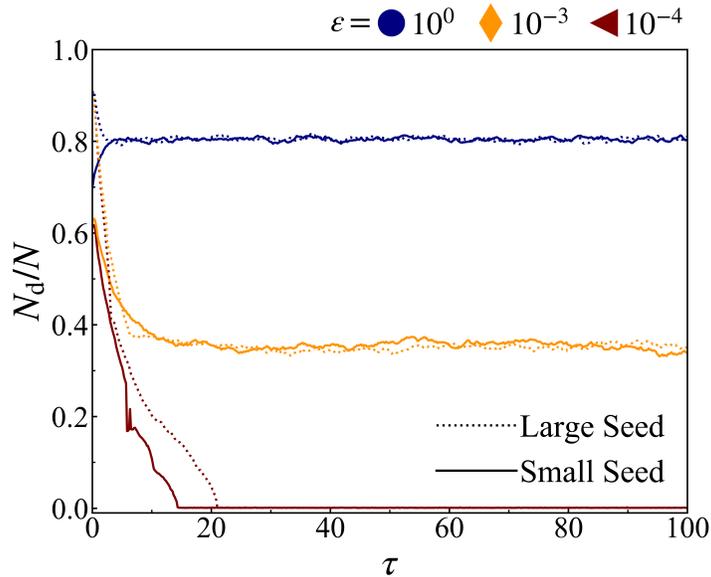
where n is the sample size and p is the number of independent variables. Upon applying eq. 26 to our results from eq. 23 ($r_m^2 = 0.161$), we find $r_{m,\text{adj}}^2 = 0.135$ where $r_{m,\text{ad}} \ll 1$, confirming that softness, area fraction, and activity are poor predictors of the normalized surface tension.

Notes and references

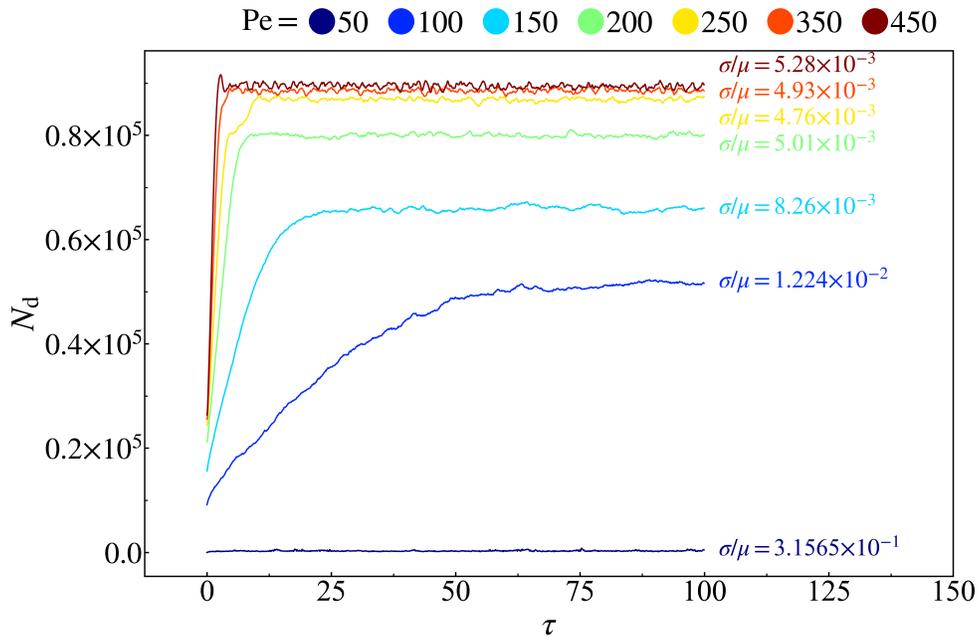
- 1 D. Levis, J. Codina and I. Pagonabarraga, *Soft Matter*, 2017, **13**, 8113–8119.
- 2 J. Bialké, J. T. Siebert, H. Löwen and T. Speck, *Physical Review Letters*, 2015, **115**, 1–5.
- 3 J. Miyazaki, J. A. Barker and G. M. Pound, *The Journal of Chemical Physics*, 1975, **64**, 3364–3369.
- 4 J. D. Weeks, *The Journal of Chemical Physics*, 1977, **67**, 3106–3121.
- 5 G. Chapela, G. Saville, S. Thompson and J. Rowlinson, 1977, **73**, 1133.
- 6 A. Wysocki, R. G. Winkler and G. Gompper, *New Journal of Physics*, 2016.
- 7 M. G. Kendall, *The Advanced Theory of Statistics*, 1948, pp. 368–385.
- 8 U. Yule and M. G. Kendall, *An Introduction to the Theory of Statistics*, 1968, pp. 281–309.
- 9 C. R. Rao, *Linear Statistical Inference and its Applications.*, 1974, pp. 266–270.
- 10 H. Theil, *Economic Forecasts and Policy*, 1961, p. 213.
- 11 T. Apostol, *Mathematical Analysis*, Addison-Wesley Publishing Company, 2nd edn, 1981, pp. 306–335.



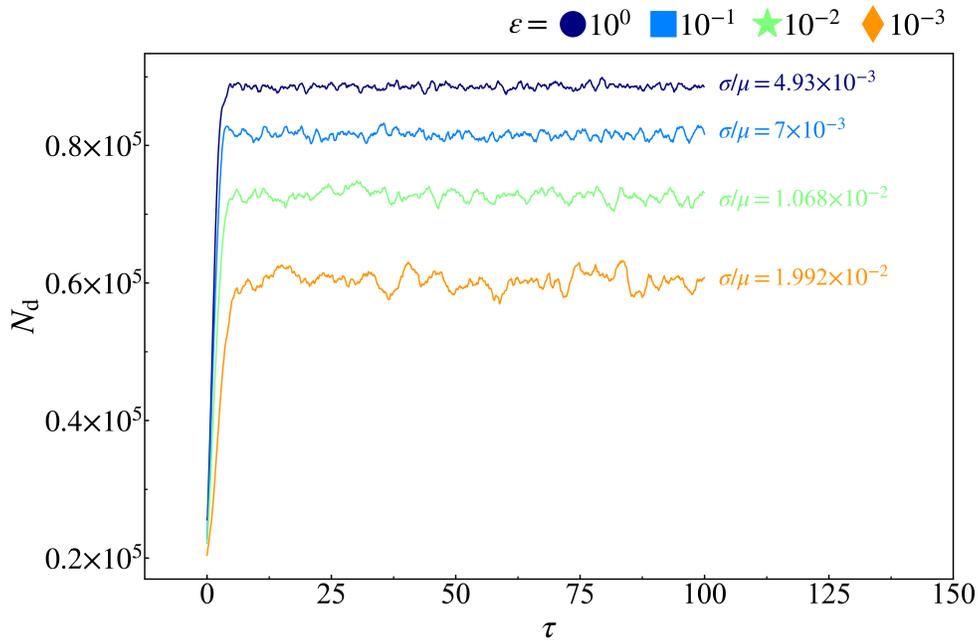
S 1 Radial density function (RDF) for three systems of constant system area fraction ($\phi = 0.55$) and activity ($Pe = 200$) with varying softness (see legend, $\epsilon = 10^0$, 10^{-3} , and $10^{-4} k_B T$). For each softness, two simulations were run with either a large (dotted line) or small (solid line) initially seeded clusters (discussed in detail in the SI, see section §3). For all simulations, RDF is measured for a single time step at $\tau_r = 300$. At $\tau = 100$, the two hardest systems ($\epsilon = 10^0$ and $10^{-3} k_B T$) had maintained steady state for $\tau \geq 50\tau_r$; whereas, the softest system ($\epsilon = 10^{-4} k_B T$) never underwent MIPS in either case. We conclude that seeding size has a negligible impact on whether MIPS occurs and, if it does occur, the resulting phase-separated lattice structure.



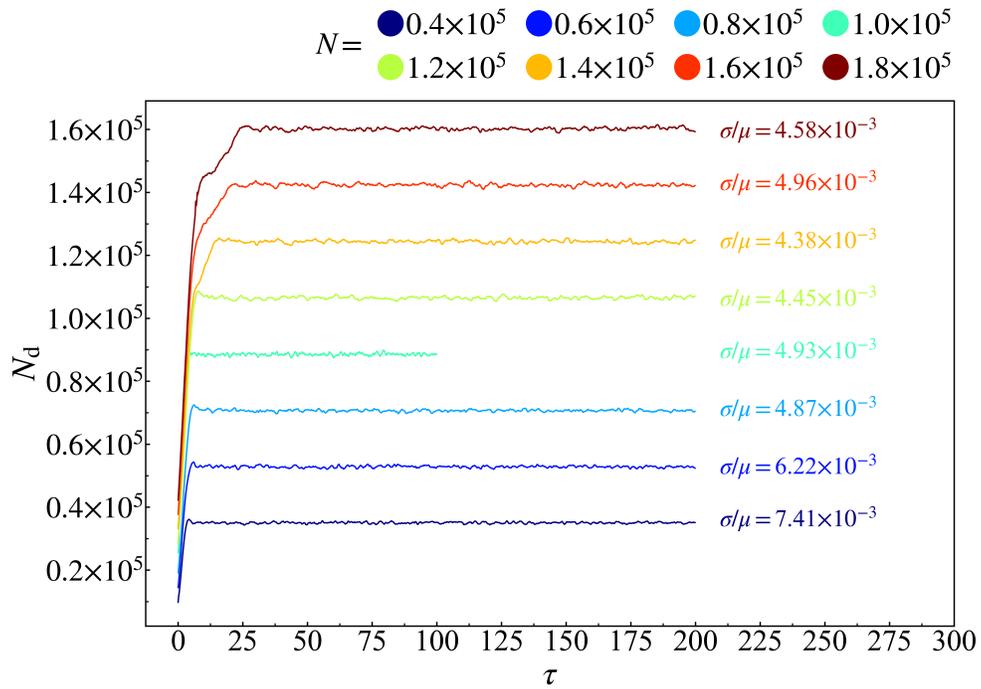
S 2 Cluster fraction (N_d/N) over time ($\tau = 100$) for three systems of constant system area fraction ($\phi = 0.55$) and activity ($Pe = 200$) with varying softness (see legend, $\epsilon = 10^0$, 10^{-3} , and $10^{-4} k_B T$). For each softness, two simulations were run with either a large (dotted line) or small (solid line) initially seeded clusters (discussed in detail in the SI, see section §3). At $\tau = 100$, the two hardest systems ($\epsilon = 10^0$ and $10^{-3} k_B T$) had maintained steady state for $\tau \geq 50\tau_r$; whereas, the softest system ($\epsilon = 10^{-4} k_B T$) never underwent MIPS in either case. We conclude that seeding size has a negligible impact on whether MIPS occurs and, if it does occur, the resulting steady state cluster size.



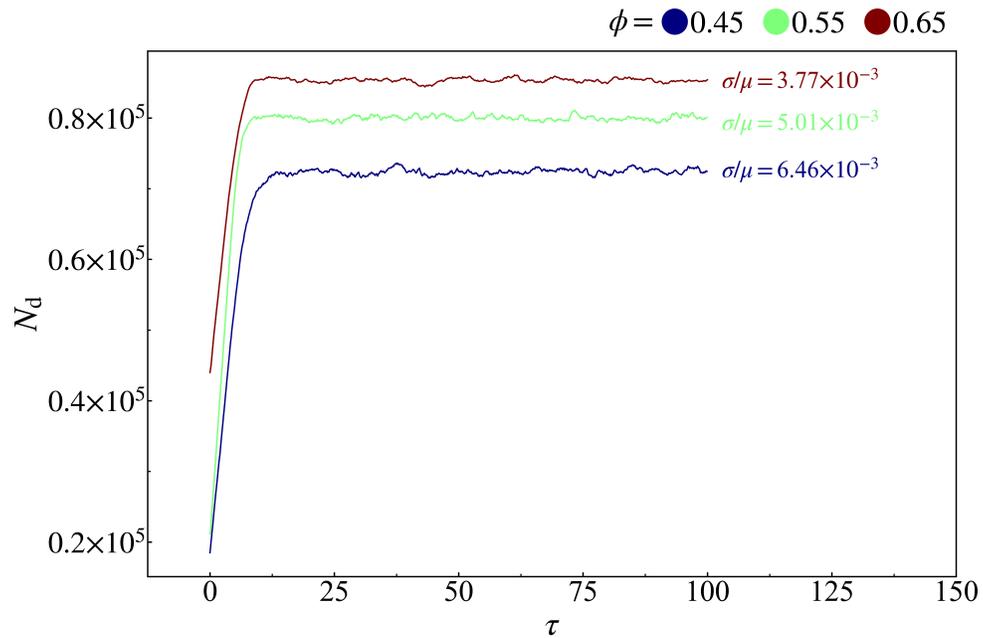
S 3 The number of particles participating in the dense phase cluster(s) over time ($\tau = 100$) with an initially seeded cluster with starting area equal to $\frac{1}{4}$ of the predicted steady-state size. Activity (Pe) is allowed to vary while system size ($N = 10^5$), area fraction ($\phi = 0.65$), and softness ($\varepsilon = 1.0$) are held constant. Activity has a large influence on dense phase size up until $Pe \geq 250$ where the increase in dense phase size as activity increases is marginal. As activity increases, the system more quickly reaches its steady state. The coefficient of variation ($C_v = \sigma/\mu$, where σ is the standard deviation and μ is the mean of the dense phase size) calculated for each system over its steady state frames. For all systems, one standard deviation corresponds to a fluctuation of less than 4% of the mean steady-state cluster size. As activity increases, the system more quickly reaches its steady state and the degree of fluctuations in system size is greatly reduced. Fluctuations in dense phase size become approximately constant at $Pe \geq 200$ in our high activity regime.



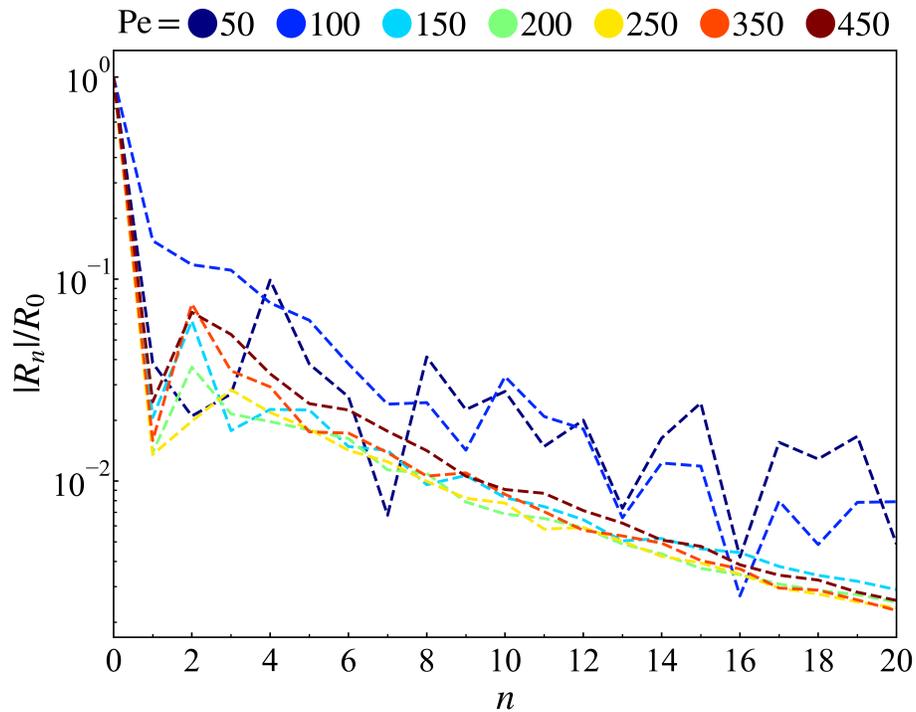
S 4 The number of particles participating in the dense phase cluster(s) over time ($\tau = 100$) with an initially seeded cluster with starting area equal to $\frac{1}{4}$ of the predicted steady-state size. Softness (ε) is allowed to vary while system size ($N = 10^5$), area fraction ($\phi = 0.65$), and activity ($Pe = 350$) are held constant. As softness increases, particles can compress more easily resulting in a smaller cluster area and, in turn, greater rate of desorption from the cluster (eq. 10 in the paper). In all systems, the system quickly reaches steady state by $\tau \approx 25$ and maintains steady state for the remaining simulation time. The coefficient of variation ($C_v = \sigma/\mu$, where σ is the standard deviation and μ is the mean of the dense phase size) calculated for each system over its steady state frames. For all systems, one standard deviation corresponds to a fluctuation of less than 2% of the mean steady-state cluster size. As particles become softer, the size of the cluster is more prone to fluctuations via avalanche events.



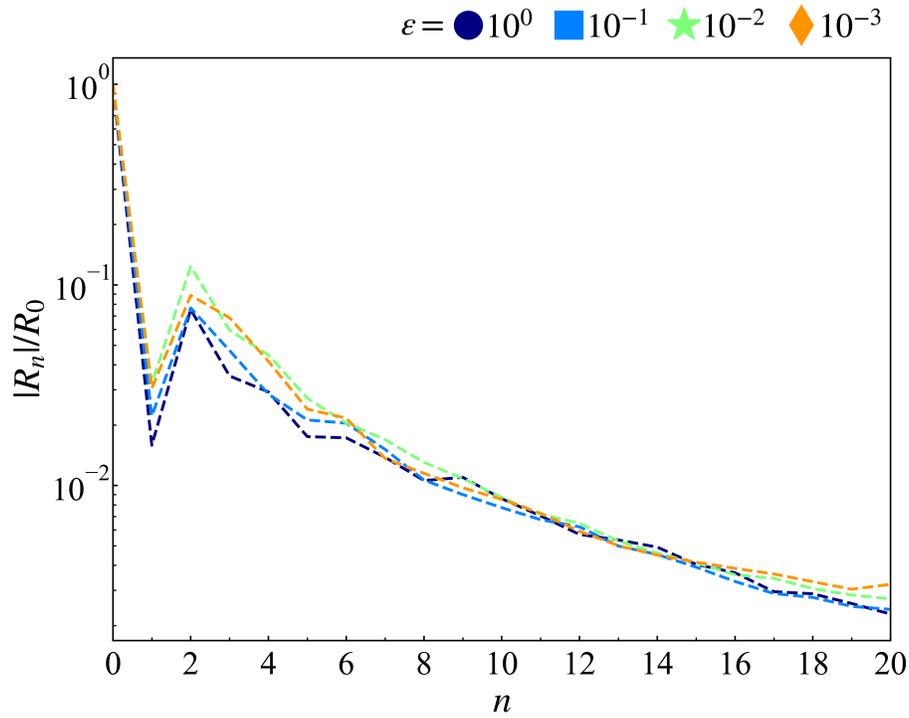
S 5 The number of particles participating in the dense phase cluster(s) over time ($\tau = 100$ for $N = 1.0^5$ and $\tau = 200$ otherwise) with an initially seeded cluster with starting area equal to $\frac{1}{4}$ of the predicted steady-state size. System size (N) is allowed to vary while softness ($\epsilon = 1.0$), area fraction ($\phi = 0.65$), and activity ($Pe = 350$) are held constant. In all systems, the system quickly reaches steady state by $\tau \approx 25$ and maintains steady state for the remaining simulation time. The steady-state dense phase size scales with the system size. The coefficient of variation ($C_v = \sigma/\mu$, where σ is the standard deviation and μ is the mean of the dense phase size) calculated for each system over its steady state frames. For all systems, one standard deviation corresponds to a fluctuation of less than 1% of the mean steady-state cluster size. At $N > 0.6 \times 10^5$, system fluctuations become independent of system size.



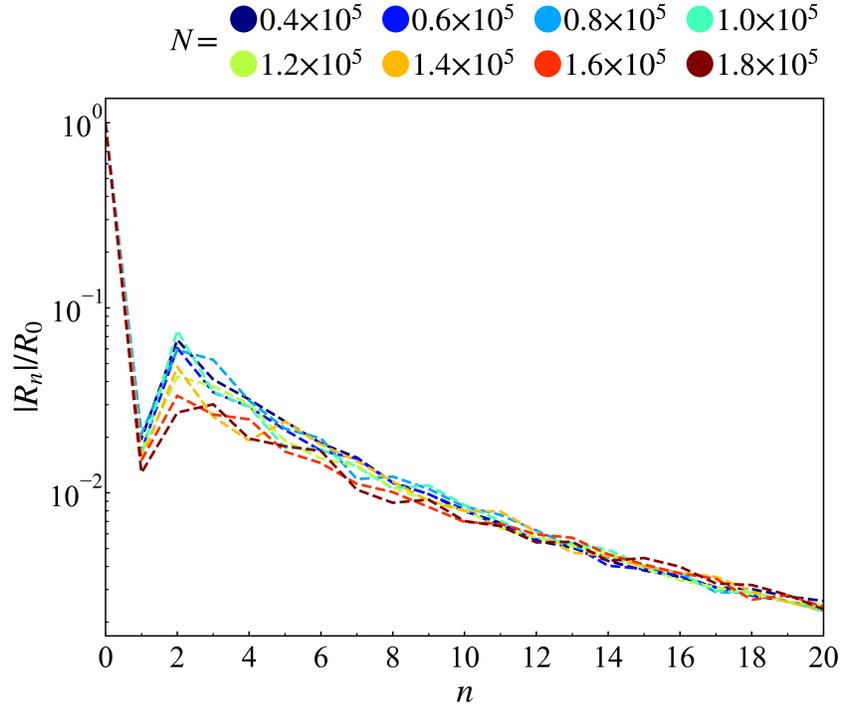
S 6 The number of particles participating in the dense phase cluster(s) over time ($\tau = 100$) with an initially seeded cluster with starting area equal to $\frac{1}{4}$ of the predicted steady-state size. Area fraction (ϕ) is allowed to vary while system size ($N = 1.0 \times 10^5$), softness ($\varepsilon = 1.0$), and activity ($Pe = 200$) are held constant. Since the area fraction is independent of phi, the additional particles participating in the dense phase simply increase the area of the cluster without influencing the density. In all systems, the system quickly reaches steady state by $\tau \approx 25$ and maintains steady state for the remaining simulation time. The coefficient of variation ($C_v = \sigma/\mu$, where σ is the standard deviation and μ is the mean of the dense phase size) calculated for each system over its steady state frames. For all systems, one standard deviation corresponds to a fluctuation of less than 1% of the mean steady-state cluster size. Fluctuations among different ϕ are comparable in magnitude.



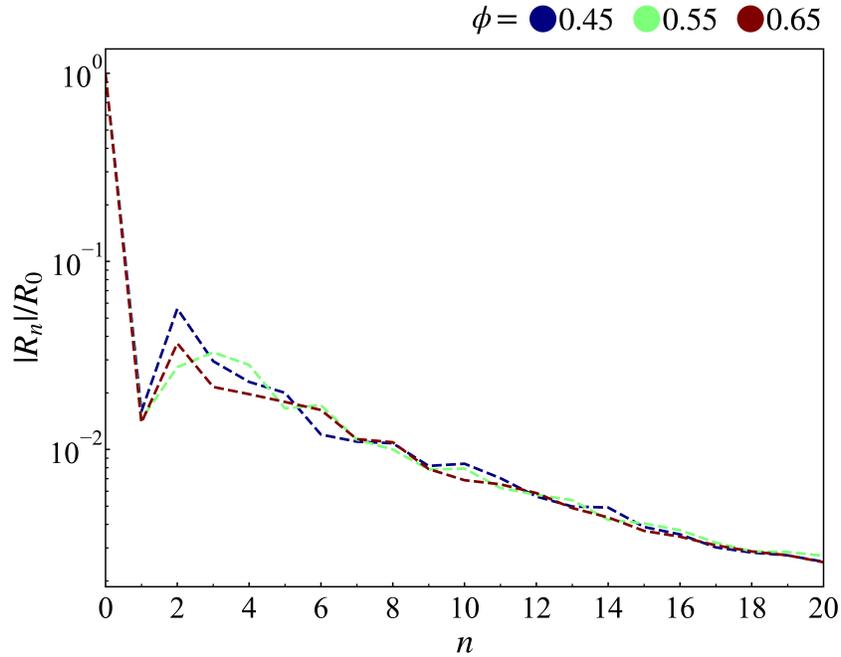
S 7 The amplitude ($R_n = \sqrt{a_n^2 + b_n^2}$ where a_n and b_n are the real and imaginary components of the n -th Fourier mode's amplitude) of the Fourier modes (n) calculated over the system's steady state and normalized by the lowest wavelength mode, $R_0 = r_c$ where r_c is the cluster radius. The average radius of the cluster's surface was measured in $d\theta = \pi/24$ increments and fit to a Fourier series to quantify the degree of surface fluctuations via the Fourier mode amplitudes¹¹. Activity (Pe) is allowed to vary while system size ($N = 1.0 \times 10^5$), softness ($\varepsilon = 1.0$), and area fraction ($\phi = 0.65$) are held constant. We see that $n = 0$ is the dominant mode for all activities. As activity increases from $Pe = 50$ to $Pe = 150$, the higher wavelength modes ($n > 0$) play a significantly decreasing role in the interface stability and, in turn, activity promotes greater stability in the cluster shape. For $Pe \geq 150$, the second highest mode is at least one order of magnitude smaller than R_0 and increasing activity further plays a negligible role in improving surface stability.



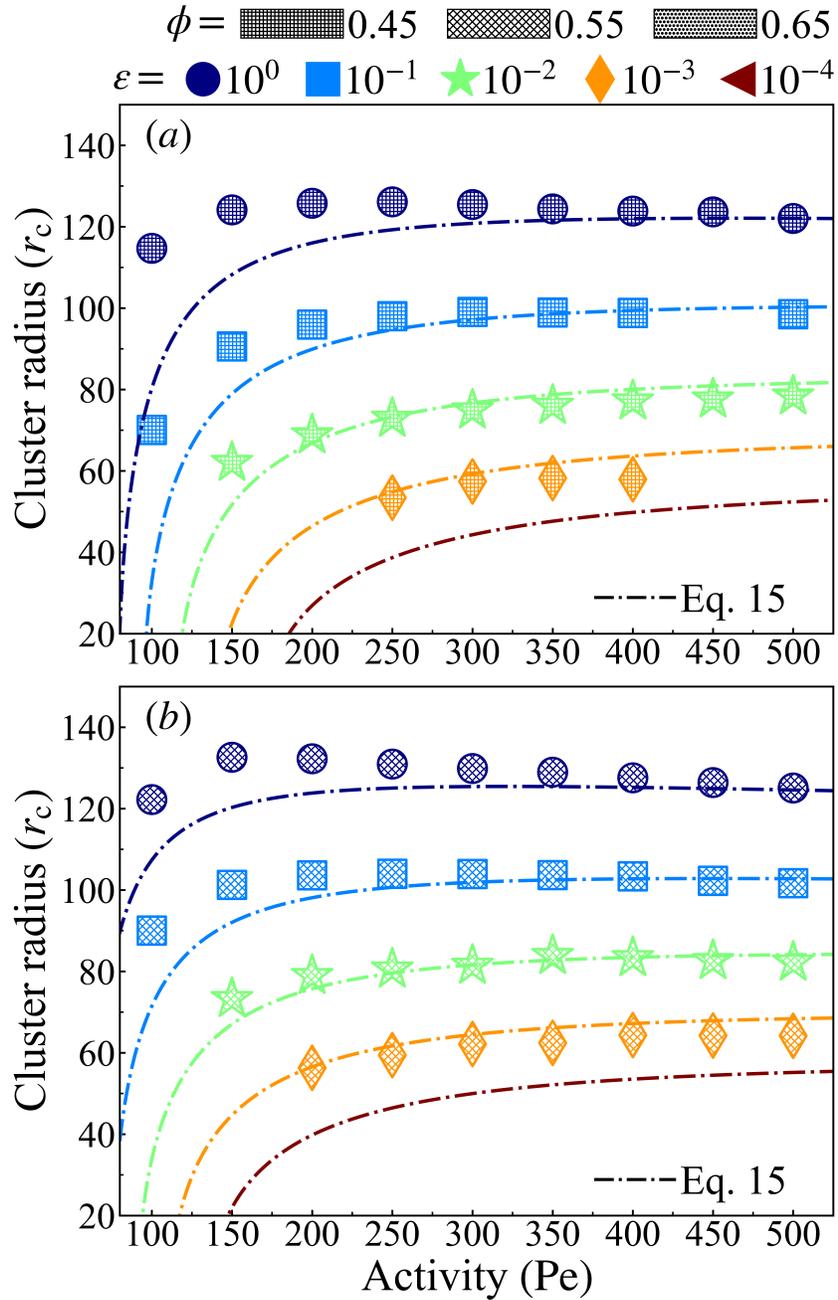
S 8 The amplitude ($R_n = \sqrt{a_n^2 + b_n^2}$ where a_n and b_n are the real and imaginary components of the n -th Fourier mode's amplitude) of the Fourier modes (n) calculated over the system's steady state and normalized by the lowest wavelength mode, $R_0 = r_c$ where r_c is the cluster radius. The average radius of the cluster's surface was measured in $d\theta = \pi/24$ increments and fit to a Fourier series to quantify the degree of surface fluctuations via the Fourier mode amplitudes¹¹. Softness (ϵ) is allowed to vary while system size ($N = 1.0 \times 10^5$), activity ($Pe = 350$), and area fraction ($\phi = 0.65$) are held constant. We see that $n = 0$ is the dominant mode for all softnesses. In addition, we see that softness plays a relatively negligible role in changing the surface stability. For all softnesses, the second highest mode is at least one order of magnitude smaller than R_0 .



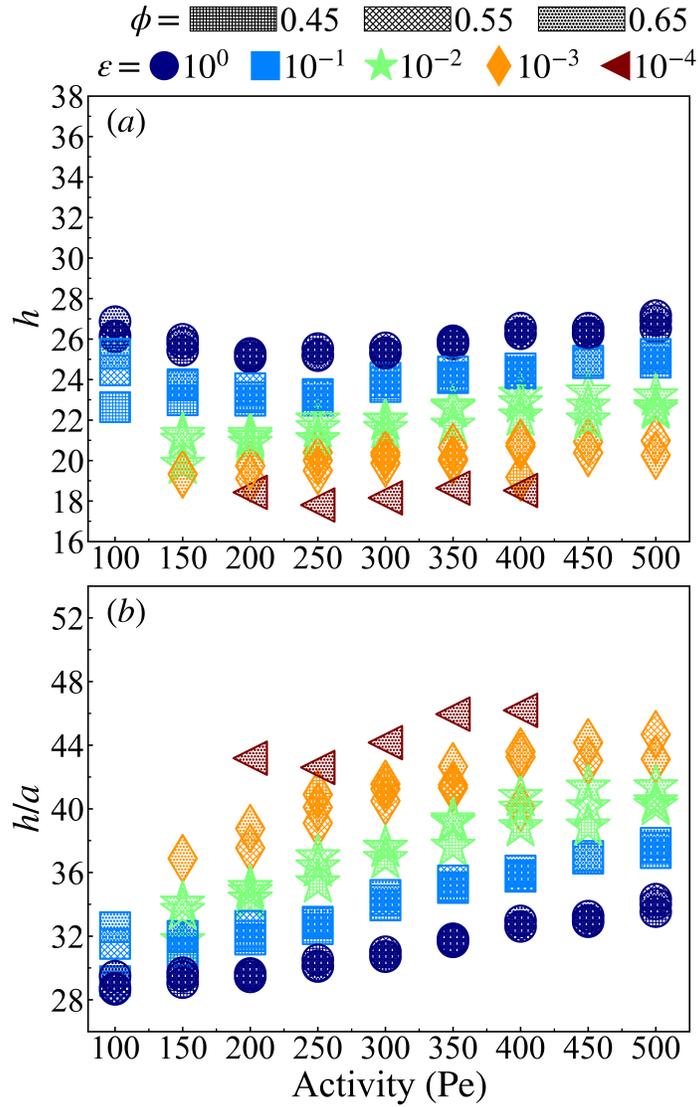
S 9 The amplitude ($R_n = \sqrt{a_n^2 + b_n^2}$ where a_n and b_n are the real and imaginary components of the n -th Fourier mode's amplitude) of the Fourier modes (n) calculated over the system's steady state and normalized by the lowest wavelength mode, $R_0 = r_c$ where r_c is the cluster radius. The average radius of the cluster's surface was measured in $d\theta = \pi/24$ increments and fit to a Fourier series to quantify the degree of surface fluctuations via the Fourier mode amplitudes¹¹. System size (N) is allowed to vary while softness ($\varepsilon = 1.0$), activity ($Pe = 350$), and area fraction ($\phi = 0.65$) are held constant. We see that $n = 0$ is the dominant mode for all system sizes. In addition, we see that system size plays a relatively negligible role in changing the surface stability. For all system sizes, the second highest mode is at least one order of magnitude smaller than R_0 .



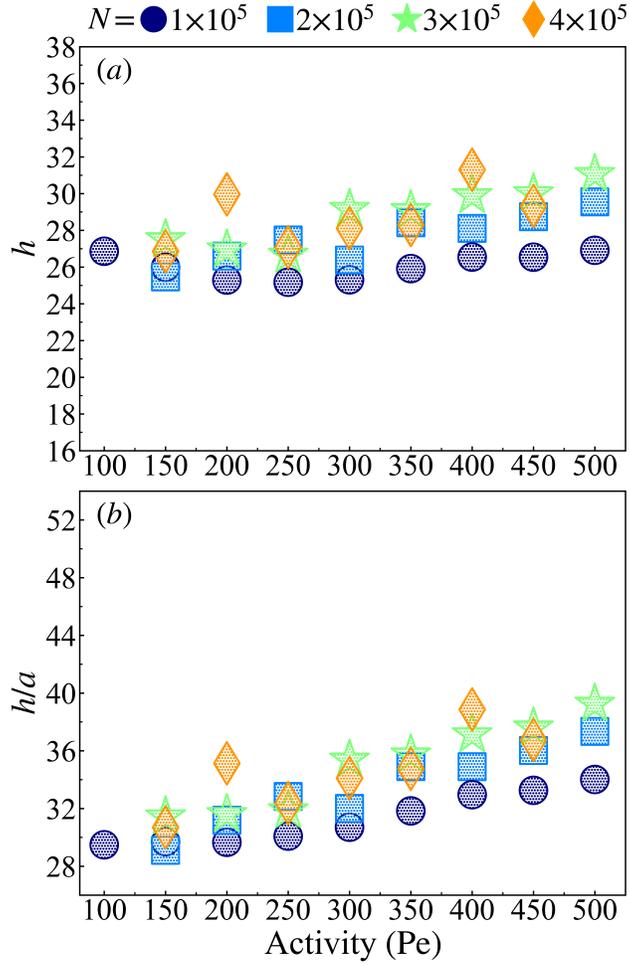
S 10 The amplitude ($R_n = \sqrt{a_n^2 + b_n^2}$ where a_n and b_n are the real and imaginary components of the n -th Fourier mode's amplitude) of the Fourier modes (n) calculated over the system's steady state and normalized by the lowest wavelength mode, $R_0 = r_c$ where r_c is the cluster radius. The average radius of the cluster's surface was measured in $d\theta = \pi/24$ increments and fit to a Fourier series to quantify the degree of surface fluctuations via the Fourier mode amplitudes¹¹. Area fraction (ϕ) is allowed to vary while softness ($\varepsilon = 1.0$), activity ($Pe = 200$), and system size ($N = 1.0 \times 10^5$) are held constant. We see that $n = 0$ is the dominant mode for all system sizes. In addition, we see that area fraction plays a relatively negligible role in changing the surface stability. For all system sizes, the second highest mode is at least one order of magnitude smaller than R_0 .



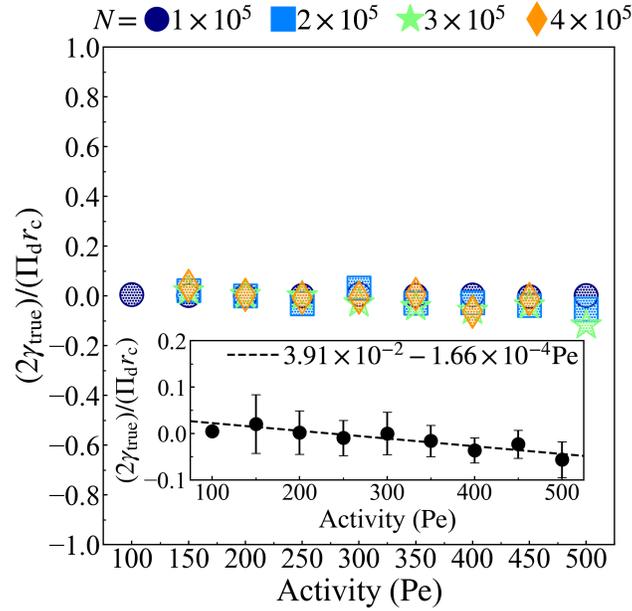
S 11 (a) Analytically computed cluster radius from eq. 15 in paper at total system area fraction of $\phi = 0.45$ with simulations at various softness (color). (b) Analytically computed cluster radius from eq. 15 at total system area fraction of $\phi = 0.55$ with simulations at various softness (color). As we reduce the area fraction, We see fewer systems that undergo MIPS in the simulations we ran, resulting in data being sparse namely for $\epsilon = 10^{-3} k_B T$ (orange diamonds) and $\epsilon = 10^{-4} k_B T$ (brown triangles).



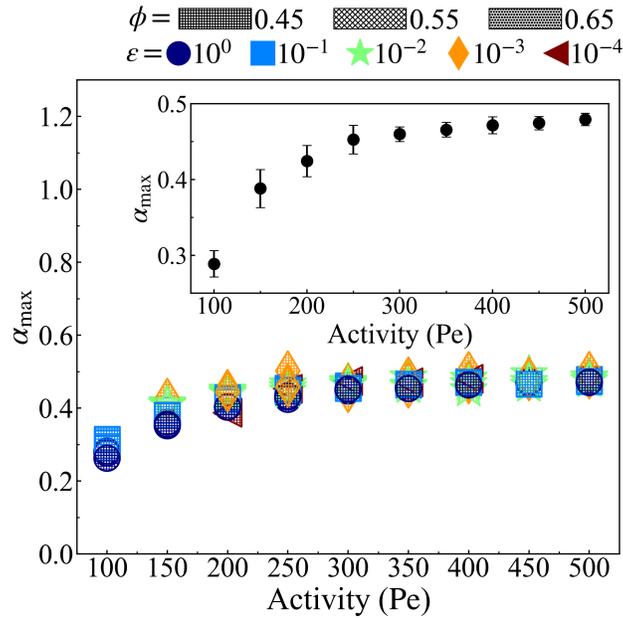
S 12 (a) Interface width calculated via the discrete approach (h) as discussed in detail in the SI, see section §3. The interface width, h , is approximately independent of activity (Pe) and area fraction (ϕ) with a weak dependence on stiffness (ϵ) despite being approximately within noise levels ($\Delta r = 3.0$). (b) When considering the dimensionless interface thickness (h/a), the interface width consists of more particles for both more active systems at constant particle stiffness and softer particle systems at constant activity.



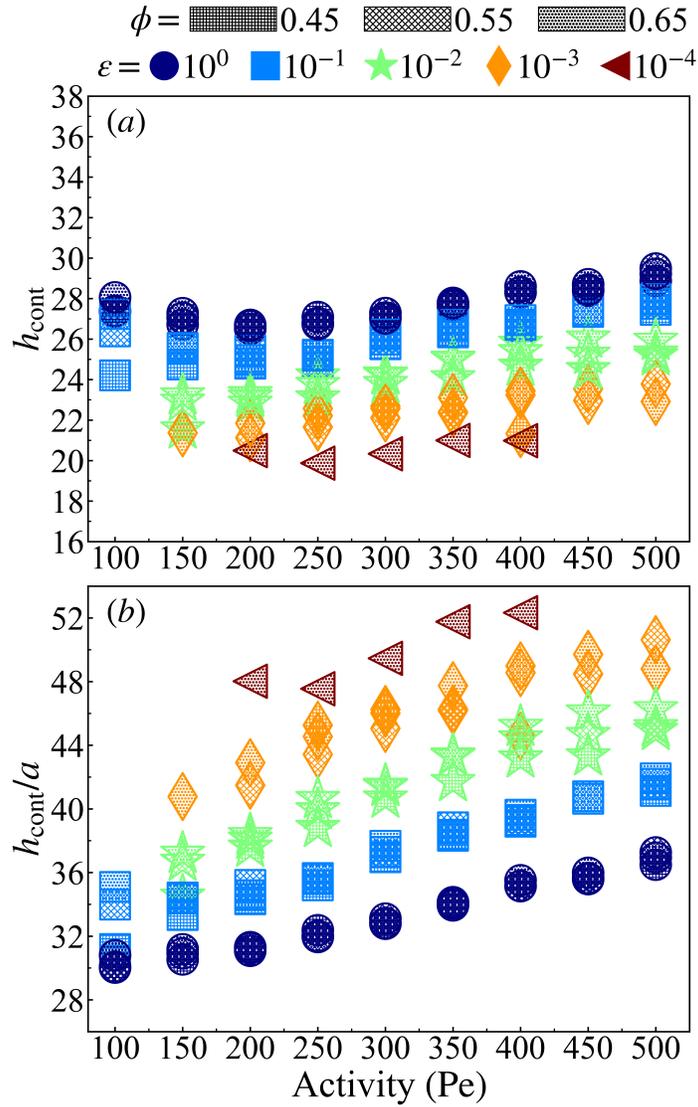
S 13 (a) Interface width calculated via the discrete approach (h), as discussed in detail in the SI, see section §3, vs the activity (Pe) for various simulation box sizes (\sqrt{N}) where $N = 10^5$ is the standard system size for all other results presented in this paper. The area fraction ($\phi = 0.65$) and softness ($\epsilon = 1.0$) are both held constant. The interface width is approximately constant at each activity for all simulation sizes with the variance in interface width being within noise levels as the step size for radial measurements is $\Delta r = 3.0$. (b) When considering the dimensionless interface thickness (h/a), the interface width consists of more particles for both more active systems at constant particle stiffness and softer particle systems at constant activity.



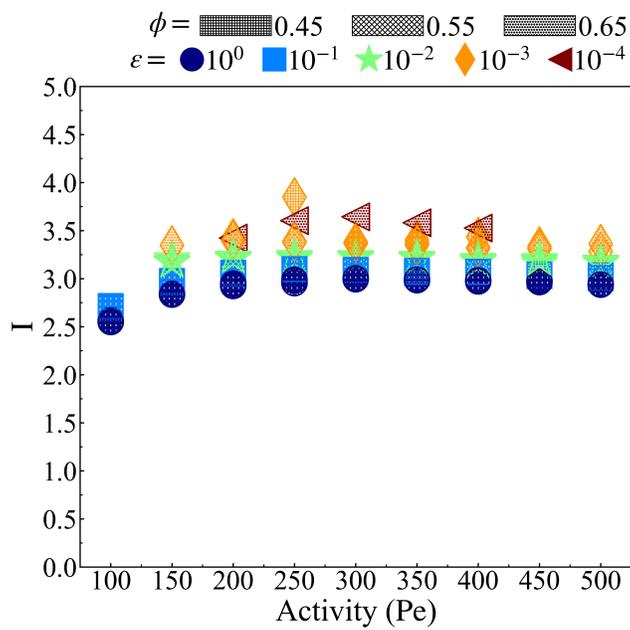
S 14 The non-dimensional surface tension, $(2\gamma_{\text{true}})/(\Pi_d r_c)$ calculated via the continuum approach using values for $n(x)$, $\alpha(x)$, r_c (Fig. 5b), and Π_d (Hollow circles from Fig. 8 calculated via the virial formulation, eq. 16 in the paper) from simulation at $\phi = 0.65$ and $\varepsilon = 1.0$ for four different systems sizes. $N = 1 \times 10^5$ is the standard system utilized for all other results presented in this paper. At all activities, γ_{true} remains approximately constant around zero with a slight bias in the negative direction. The inset shows the surface tension averaged over the system size (N) at each activity with error bars corresponding to a single standard deviation. In the inset, all surface tension measurements (colored) are fitted (dashed line) such that we do not bias low activity where fewer systems undergo MIPS. The line of best fit is found to be approximately constant, slightly negative, and near zero while being encompassed in the standard deviation at most activities.



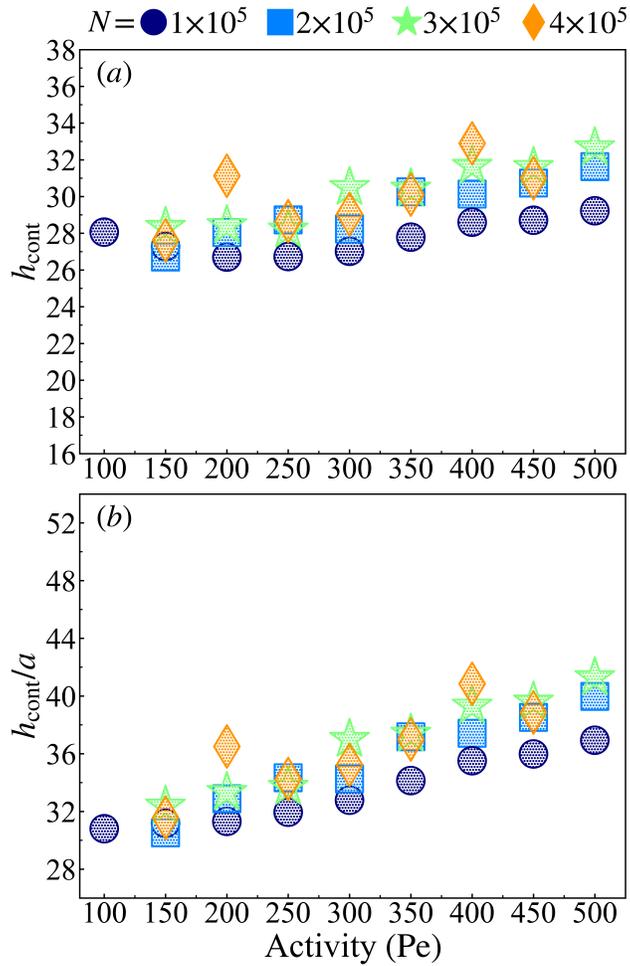
S 15 The maximum particle alignment (α_{max}) with the cluster's center of mass, occurring within the bulk-gas interface. The alignment is found to be independent of both area fraction (ϕ) and softness (ε) in addition to being roughly independent of activity (Pe) at high activities ($Pe > 150$). The inset shows the maximum alignment averaged over ϕ and ε at each activity with error bars corresponding to a single standard deviation. For our calculation of the interface width (eq. 27 in paper), the average maximum alignment is chosen to be a constant corresponding to $\alpha_{\text{max}} = 0.45$, which is the average maximum alignment at high activities ($Pe > 150$).



S 16 (a) Interface width calculated via the continuum approach (h_{cont}) with eq. 27 in paper. The interface width, h_{cont} , is approximately independent of activity (Pe) and area fraction (ϕ) with a weak dependence on stiffness (ϵ) despite being approximately within noise levels ($\Delta r = 3.0$), which is in agreement with that seen in the interface width measured via the discrete approach, h (fig. S12). (b) When considering the dimensionless interface thickness (h/a), the interface width consists of more particles for both more active systems at constant particle stiffness and softer particle systems at constant activity.



S 17 The integral, $I = \left(\int_0^1 \tilde{\phi}(\tilde{x}) \tilde{\alpha}(\tilde{x}) d\tilde{x} \right)^{-1}$, as a function of activity (Pe with varying softness (color) and area fraction (hatching)). I is found to be approximately constant at high activities ($Pe > 150$) where I is chosen to be $I = 3.0$ for calculations of the interface width via the continuum approach, see eq. 27 in paper.



S 18 (a) Interface width calculated via the continuum approach (h_{cont}) with Eq. 27 in paper vs the activity (Pe) for various simulation box sizes (\sqrt{N}) where $N = 10^5$ is the standard system size for all other results presented in this paper. The area fraction ($\phi = 0.65$) and softness ($\epsilon = 1.0$) are both held constant. The interface width is approximately constant at each activity for all system sizes with the variance in interface width being within noise levels as the step size for radial measurements is $\Delta r = 3.0$. (b) When considering the dimensionless interface thickness (h_{cont}/a), the interface width consists of more particles for both more active systems at constant particle stiffness and softer particle systems at constant activity.