## PAPER

# Gradient-based active learning for intelligent discovery of colloidal phase diagrams

Sumeet Vadhavkar, [iD] *[a] Pradeep Bajracharya, [iD] [a] Md Shakil Zaman,[a] Poornima Padmanabhan [iD] *[b] and Linwei Wang [iD] [a]

Rapid and accurate prediction of the colloidal phase diagrams is necessary for controlling their structure and properties. Mapping these multi-dimensional phase diagrams with molecular dynamics (MD) is costly, especially when sharp, localized transition boundaries demand dense sampling to resolve the boundaries. We target this challenge by actively learning the boundary where the order parameter – the number of nearest neighbours, $N_{nn}$ – undergoes an abrupt change. We describe colloids interacting *via* isotropic pair potentials with tunable range of interaction and aim to identify the boundary between the dilute and condensed phases. We model $N_{nn} = f(\mathbf{x})$ with a Gaussian process (GP) and derive a novel gradient-aware acquisition strategy that prioritizes locations with large and/or uncertain spatial derivatives of the GP posterior. The gradient-aware strategy is based on the thermodynamic principle of large susceptibility, *i.e.*, large derivative of the order parameter, at phase boundaries. A simple ellipse-based exclusion heuristic helps spread acquisitions along the boundary manifold. Across 1D and 2D planes of the phase diagram, the proposed strategy localizes phase boundaries accurately with significantly fewer MD runs than dense grid search and outperforms random acquisition or standard acquisition strategies unaware of the gradient of the GP. The framework can be readily extended to higher-dimensional space and support multi-fidelity MD simulations, providing a practical route to sample-efficient and physics-aligned discovery of colloidal phase diagrams.

### Design, System, Application

Colloidal particles self-assemble into a wide variety of equilibrium and non-equilibrium structures. Small changes in their interparticle interactions can affect the phase diagram significantly. In this work, we have developed a gradient-based active learning acquisition framework to efficiently identify the phase boundary between dilute and condensed phases. The system under investigation is comprised of colloidal particles at several volume fractions interacting with an isotropic potential of varying strength and interaction range. Even such a simple potential leads to a computationally expensive, three-dimensional phase diagram. We have applied the gradient-based active learning framework to quickly target the boundary in one-dimensional and two-dimensional slices of the phase diagram within a few dozen simulations. Short molecular dynamics simulations yield a number of nearest neighbors metric that changes abruptly at the phase boundary, making it suitable for gradient-based acquisition strategies. This rapid exploration will enable the rapid prediction of multi-dimensional phase diagrams of colloidal particles.

## 1. Introduction

Colloidal systems, characterized by microscopic particles dispersed within a solvent, permeate our world.[1,2] Their unique properties stem from their diverse phase behaviors, encompassing states resembling solids, liquids, gels, and glasses. Accurately predicting the phases of colloidal systems offers a powerful tool for controlling their properties. This capability has transformative potential for various fields, including materials science, drug delivery, and environmental engineering.[3–5]

Computational models, such as Molecular Dynamics (MD)[6] simulations, are commonly employed to simulate colloidal particles under varying conditions (*e.g.*, temperature, concentration) in order to construct phase diagrams. In this context, the parameter space comprises all relevant tunable inputs (such as temperature, concentration, and interaction strength) that dictate the system's phase behavior. By systematically probing this parameter space, researchers can identify phase boundaries, predict material properties, and significantly reduce the need for extensive physical experimentation.[7] Traditionally, grid search is used, which exhaustively samples the chosen range of each

*[a] College of Computing and Information Sciences, Rochester, NY 14623, USA. E-mail: sv6234@rit.edu*

*[b] Department of Chemical Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA. E-mail: poornima.padmanabhan@rit.edu; Tel: +1 585 475 4877*
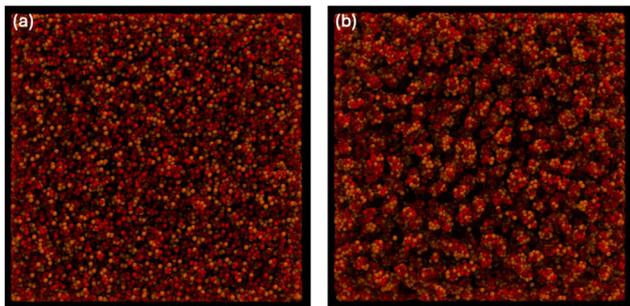
**Fig. 1** Snapshots of particle positions at the end of the simulation. Particles are colored according to their radii. (a) An instance of low $N_{nn}$ = 2.88 at $V^*$ = 2.08, $\phi$ = 0.05, $\kappa a$ = 29.5, showing a homogeneous, well-dispersed phase. (b) An instance of high $N_{nn}$ = 7.28 at $V^*$ = 4.19, $\phi$ = 0.05, $\kappa a$ = 30.0, showing clusters which is a hallmark of the condensed phase.

parameter to provide high-resolution information to construct the phase diagram.[8]

Although often faster than physical experimentation, MD simulations can still be computationally expensive.[9,10] This high cost becomes a significant limitation when employing grid search methods to explore phase diagrams. This is particularly true when dealing with high-dimensional parameter spaces, where the number of grid points required increases rapidly with the number of dimensions —a phenomenon known as the 'curse of dimensionality.'[11,12] Specifically, if a parameter space is divided into $n$ segments per dimension, the total number of grid points for a $d$-dimensional space becomes $n^d$. This rapid growth in the number of required simulations leads to a significant increase in computational cost, making grid search inefficient and even infeasible for large-scale problems. Moreover, phase diagrams are characterized by regions that are almost piecewise smooth with sharp transition boundaries between different phases.[13,14] To capture these sharp boundaries of phase transition necessitate a high grid resolution,[15] which exacerbates the curse of dimensionality while wasting significant redundant computations outside the region of interest (*i.e.*, boundaries of phase transition). Finally, simple MD simulations are prone to incorrect predictions due to the lack of adequate sampling or convergence of results, which often necessitates additional, more computationally expensive methods such as the Gibbs ensemble and Gibbs–Duhem integration methods to accurately determine phase boundaries.[16,17]

While the aforementioned methods are well-suited for systems at equilibrium (stable state), they become invalid or insufficient in rapidly evolving, non-equilibrium environments. Hence, they may not be suitable for the increasingly important fields of gels, glasses, or active matter systems that are not at equilibrium.[18,19] The dynamic nature of these non-equilibrium systems poses additional challenges that traditional grid-search methods may not adequately address.

Therefore, there is a pressing need for more efficient techniques that can minimize the computational burden of grid-search approaches while identifying approximate phase boundaries in compute-intensive, high-dimensional, and non-equilibrium systems. The approximate phase boundaries can subsequently be used for more detailed investigation through appropriate advanced techniques and/or physical experimentation. Active learning presents a promising solution to this challenge.[20,21] When learning to approximate a function over an input parameter space, an active learner will strategically select new points in the input space to query the input–output values of the function of interest, such that the number of points needed to learn the function can be minimized. More specifically, consider a function $y = f(\boldsymbol{\theta})$, $f$: $\mathbb{R}^n \rightarrow \mathbb{R}$, where $y$ is the function value of interest (*e.g.*, a measure of the system's phase) and $\boldsymbol{\theta}$ is the set of parameters under which the system is simulated. Instead of approximating $f$ passively from random or uniform samples, active learning uses an acquisition function (AF) to guide the intelligent selection of $\boldsymbol{\theta}_i$ to query the true value of $y_i = f(\boldsymbol{\theta}_i)$, considering a balance between two competing goals: exploration and exploitation. Exploration involves sampling points in unexplored regions of the parameter space to gain new information (often driven by the uncertainty in the current approximation of $f$), while exploitation focuses on refining the model in regions where the function values are of the most interest, typically the optimum (*e.g.*, maximum or minimum of $y$ over the space of $\boldsymbol{\theta}$). In other words, standard AF design in active learning is focused on the magnitude and uncertainty of the approximated output of function $f$.[22,23] This is typically done using a Gaussian Process (GP) to model $f$ which allows the predictive mean and variance of the GP to be analytically calculated and used in the AF. They have been successfully applied to search the parameter space of a complex system to discover the region of interest with a minimum number of simulations or experiments.[24,25]

In learning to approximate the phase diagram, however, the region of interest is not the phase but the boundary of phase transitions that are known to be sharp and localized. In learning to approximate the phase diagram, one can alternately describe it as identifying the boundary (manifold) where the order-parameter of interest undergoes a sharp transition. In thermodynamics, susceptibility is the derivative of the order parameter with respect to a thermodynamic property of the system that is often one of the axes of the phase diagram. Characteristically, the susceptibility exhibits a sharp peak at a phase transition. For multi-dimensional phase diagrams, one would analogously consider the gradient of the order parameter as the susceptibility vector. Standard AF design, not concerned with the rate of change (*i.e.*, gradient)[22,23] of the underlying function of interest, is thus not suited for capturing the narrow boundaries of phase transitions if directly applied. Moreover, standard AFs often focus on selecting individual query points in isolation over the parameter space, without considering a broader region of interest. This can lead to sampling localized clusters of

points near a phase-transition boundary without accurately mapping its complete manifold.

A recent effort attempts to address these challenges by modeling the phase-transition boundaries as a classification boundary to be learned by active learning.[26] It involves casting the problem as a classification task where, upon querying the MD simulation at a particular point in the parameter space, a label (*e.g.*, −1 or 1) is assigned based on the simulation output to indicate the current phase of the colloid at that point. A GP is used to approximate these binary labels across the parameter space, with the AF designed to guide the selection of new query points towards regions with high uncertainty (*i.e.*, high variance of the GP) or near the classification boundary (*i.e.* predicted label value close to 0). While demonstrating promising results in directing computational resources toward refining the transition region, this approach relies on a heuristic strategy that approximates the sharp boundary of phase transitions as a classification boundary between two discrete values. This approximation, partly dependent on a threshold to binarize an otherwise continuous phase diagram, introduces errors in the delineation of phase transition boundary. Furthermore, the assumption that the phase transition will occur at points where the value of $y$ approaches 0 is not necessarily physically realistic. Rather, one must account for some physics-informed criterion such as the behavior of the susceptibility, rather than a unique value of the order parameter, at a phase transition.

In this work, we propose a novel active learning framework for rapidly localizing and approximating the sharp boundaries of phase transition in colloidal systems with a small number of queries for MD simulations. Our approach features two main contributions. First, we derive an analytical gradient-based AF that focuses on the first-order derivatives of the GP approximation of the phase diagram. In other words, we focus on the underlying rate of change of the approximated order parameter over the parameter space, considering both its magnitude and uncertainty, to guide the acquisition of query points towards the phase transition boundary. Second, to prevent excessive and localized sampling in high-gradient regions, we introduce a heuristic strategy that prevents new queries within an elliptical neighborhood of existing points in the parameter space. This strategy promotes the identification of the complete phase-transition manifold while further enhancing computational efficiency.

We validated our framework on one-dimensional (1D) and two-dimensional (2D) colloidal systems, in comparison to uniform grid sampling and active learning with commonly used upper-confidence-bound (UCB) AF without considering the gradient of the underlying function. Our results demonstrated rapid and reliable convergence to accurate phase diagrams—even with randomly chosen initial samples. We further investigated the effect of the length scale of the GP and the radii of the ellipse. Looking ahead, our method can be naturally extended to three-dimensional (3D) systems and further integrated into multi-fidelity active learning

where lower-fidelity simulations provide fast lower-resolution approximations while higher-fidelity simulations can be dedicated to refine critical transition zones as guided by active learning.

## 2. Problem formulation

In this work, we target the phase diagrams of attractive colloids with highly tunable interactions. For many colloidal systems, the interactions can be tuned in several ways including patchiness, range of attraction, shape anisotropy, and many others.[27] To showcase our method, the interaction potentials chosen in this study vary in interaction strength and range, while remaining isotropic. Phase diagrams produced by such potentials are known to exhibit a multitude of equilibrium (dilute gas-like, dense liquid-like, and crystalline solid) as well as non-equilibrium (gel, attractive glass, repulsive glass) phases.[15,28] We will limit the scope of this paper to delineating the phase boundary between the dilute phase and condensed phases, which can be obtained unambiguously for most experimental systems by detecting changes in turbidity and/or rheology. In simulations, condensed phases can be characterized by a dramatic increase in local volume fraction or local structural correlations. To incorporate additional phases in the phase diagram, further physics-based classification criteria can be developed.

We utilize molecular dynamics (MD) simulations to study the self-assembly of colloidal particles interacting with a tunable Morse potential given by Eqn. S1 in the SI. Our input parameters to the MD simulations represent temperature changes, volume fraction, and the range of interaction. Mathematically, the parameters are scaled attractive energy $V^* = V_0/k_\mathrm{B}T$, volume fraction of the colloids $\phi$, and $\kappa a$ that controls the range of interaction relative to the radius of the particle $a$. In summary, each simulation tracks the evolution of 62 500 polydisperse colloidal particles of average radius $a$ in implicit solvent undergoing Brownian motion, at the specified points in the phase diagram (attractive energy, volume fraction, and range) for 4000 Brownian times. The MD simulations are executed using LAMMPS with computing resources from research computing at RIT.[10,29,30] Additional simulation details are available in the SI. At the end of each simulation, we track the local structure around each particle and compute the number of nearest neighbors $N_\mathrm{nn}$ from the radial distribution function (Eqn. S2 in the SI). In the literature, similar order parameters have been used to distinguish dilute and condensed phases. For instance, the discrete number of particles in the neighborhood of a single particle, also called contact number, have been used to identify the water to vapor transition[31,32] and to classify non-equilibrium colloidal phases (gels and glasses)[15,33] The exact value of this contact number, $N_\mathrm{nn}$, at the phase boundary is often taken to be 6, the mid-way point between the maximum allowable range,[31,32] but more generally, a rapid change of the order parameter will occur at the phase boundary which

would naturally lend itself to a gradient-based acquisition method. Two sample snapshots of the particles are shown in Fig. 1, where low and high $N_{nn}$ are able to distinguish dilute and condensed phases. The phase diagram we are approximating is thus represented by the following function:

$$N_{nn} = f(V^*, \phi, \kappa a). \tag{1}$$

## 3. Methodology

Rather than exhaustively sampling the entire parameter space $\mathbf{x} = (V_0, \kappa, \phi)$ to approximate $N_{nn} = f(\mathbf{x})$, we employ an active learning strategy to adaptively select points in the parameter space to query MD simulations as described in the SI, in order to learn a GP approximation of $f(\mathbf{x})$ and localize the phase transition boundary with a minimal number of MD simulation queries.

Our active learning approach leverages a gradient-based AF, which uses both the magnitude and uncertainty of the first-order derivatives of the approximated $f(\mathbf{x})$ to prioritize regions where sharp transitions are likely. In doing so, we efficiently map the boundaries of phase transitions within a phase diagram with fewer simulations, focusing computational resources on parameter combinations most indicative of phase boundaries. As summarized in Algorithm 1, the active learning process involves the following main components. First, the GP is initialized with an initial set of labeled points. It then enters an iterative process of optimizing an AF to choose new parameter values, at which we query the MD simulations and update the GP with newly acquired data. The iteration continues until a convergence criteria is met.

---

**Algorithm 1** Active Learning Procedure for Identifying Phase Boundaries

---

**Require:** Predefined labeled initial points representing various states of the colloidal material

1: **Initialization:** Initialize with a set of predefined labeled initial points
2: **repeat**
3:     **Selection of New Query Points:**
4:     Use a gradient-based Upper Confidence Bound (UCB) acquisition function to select new query points
5:     **MD Simulation Queries:**
6:     Perform additional simulations or experiments at the selected query points to obtain new data
7:     **GP Update:**
8:     Update the GP model with the newly acquired data
9:     **Stopping Criteria Check:**
10:     Monitor the rolling mean of the GP's predictions over consecutive windows
11: **until** the change in the rolling mean is consistently below a predefined threshold

---

### 3.1. Iterative Gaussian process updates

To model $N_{nn} = f(\mathbf{x})$ as defined earlier, we approximate it as a GP as:

$$f \sim GP(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')), \tag{2}$$

where $\mu(\mathbf{x})$ is the mean function representing the expected value at $\mathbf{x}$, and $k(\mathbf{x}, \mathbf{x}')$ is the covariance function (or kernel) describing the correlation between function values at $\mathbf{x}$ and $\mathbf{x}'$. In our approach, we use the Radial Basis Function (RBF) kernel, defined as:

$$k(\mathbf{x}, \mathbf{x}') = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|^2}{2\ell^2}\right), \tag{3}$$

where $\ell$ is the length scale parameter of the kernel. The RBF kernel is chosen for its smoothness and ability to model complex, non-linear functions.

Given training data $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$, the joint distribution of the observed outputs $\mathbf{y} \in \mathbb{R}^n$ and the (scalar) function value $y_* = f(\mathbf{x}_*)$ at a single test point $\mathbf{x}_*$ is

$$\begin{pmatrix} \mathbf{y} \\ y_* \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ 0 \end{pmatrix}, \begin{pmatrix} K(X, X) & K(X, \mathbf{x}_*) \\ K(\mathbf{x}_*, X) & K(\mathbf{x}_*, \mathbf{x}_*) \end{pmatrix} \right), \tag{4}$$

where $X = [x_1, \ldots, \mathbf{x}_n]^\top$, $K(X, X)$ is the $n \times n$ kernel matrix, $K(X, \mathbf{x}_*) \in \mathbb{R}^{n \times 1}$, and $K(\mathbf{x}_*, \mathbf{x}_*) \in \mathbb{R}$. Conditioning yields the posterior predictive mean and variance at $\mathbf{x}_*$:

$$\begin{aligned} \mu_*(\mathbf{x}_*) &= K(\mathbf{x}_*, X)[K(X, X)]^{-1}\mathbf{y}, \\ \sigma_*(\mathbf{x}_*) &= K(\mathbf{x}_*, \mathbf{x}_*) - K(\mathbf{x}_*, X)[K(X, X)]^{-1}K(X, \mathbf{x}_*). \end{aligned} \tag{5}$$

The training is initialized with $p$ labeled pairs $(\mathbf{x}, y)$ (see section 3.4), and increases by two at each acquisition (see section 3.2). The GP is iteratively updated with each newly acquired simulation result.

### 3.2. Gradient-based acquisition function design

To target the acquisition of simulations around the *unknown* boundary of phase transition, we design the AF to focus on the spatial derivatives of the GP. Fortunately, the derivative of a GP is also a GP: for the RBF kernel used in this paper, the derivative with respect to $\mathbf{x}$ is:

$$\frac{\partial k(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x}} = -\frac{(\mathbf{x} - \mathbf{x}')}{\ell^2} k(\mathbf{x}, \mathbf{x}'), \tag{6}$$

and the covariance between the derivatives at points $\mathbf{x}$ and $\mathbf{x}'$ is obtained by differentiating the kernel twice:

$$\frac{\partial^2 k(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x} \partial \mathbf{x}'} = \left(\frac{1}{\ell^2} - \frac{(\mathbf{x} - \mathbf{x}')^2}{\ell^4}\right) k(\mathbf{x}, \mathbf{x}'). \tag{7}$$

Let $\nabla y_*$ represent the derivative of $y$ with respect to $\mathbf{x}$ at the test point $\mathbf{x}_*$. The joint distribution of $\mathbf{y}$ and $\nabla y_*$ at the test points is:

$$\begin{pmatrix} \mathbf{y} \\ \nabla y_*|_{\mathbf{x}=\mathbf{x}_*} \end{pmatrix} \sim \mathcal{N}\left( \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} K(X, X) & \left.\frac{\partial K(X, \mathbf{x})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_*} \\ \left.\frac{\partial K(\mathbf{x}, X)}{\partial \mathbf{x}}\right|_{\mathbf{x}=\mathbf{x}_*} & \left.\frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x} \partial \mathbf{x}'}\right|_{\mathbf{x}=\mathbf{x}'=\mathbf{x}_*} \end{pmatrix} \right). \tag{8}$$

Therefore, the predictive mean and standard deviation of the derivative at a test point $\mathbf{x}$ are:

$$\nabla\mu(\mathbf{x}_*) = \frac{\partial K(\mathbf{x}, X)}{\partial \mathbf{x}}\bigg|_{\mathbf{x}=\mathbf{x}_*}[K(X,X)]^{-1}\mathbf{y},$$

$$\sum\nolimits_{\nabla}(\mathbf{x}_*) = \frac{\partial^2 K(\mathbf{x}, \mathbf{x}')}{\partial \mathbf{x}\partial \mathbf{x}'}\bigg|_{\substack{\mathbf{x}=\mathbf{x}_*\\ \mathbf{x}'=\mathbf{x}_*}}$$
$$- \frac{\partial K(\mathbf{x}, X))}{\partial \mathbf{x}}\bigg|_{\mathbf{x}=\mathbf{x}_*}[K(X,X)]^{-1}\frac{\partial K(X, \mathbf{x}')}{\partial \mathbf{x}'}\bigg|_{\mathbf{x}'=\mathbf{x}_*}.$$

$$(9)$$

Note that the GP derivatives are calculated separately for each dimension in the parameter space, where regions of high derivative magnitude are expected to be narrow and localized, providing essential information for phase-change detection. We thus acquire points based on the spatial derivative per dimension, exploiting the high gradient magnitude along that dimension along with its variance. For dimension $j$ of a $n$-dimensional $\mathbf{x}$, this is achieved *via* a derivative-based UCB acquisition function expressed as:

$$\alpha_j(\mathbf{x}_*) = \left|\nabla_{x_j}\mu(\mathbf{x}_*)\right| + \beta\sqrt{\left[\Sigma_\nabla(\mathbf{x}_*)\right]_{jj}}, \quad j = 1, \ldots, n.$$

$$(10)$$

For each coordinate $j$, we choose $\mathbf{x}_*^{(j)}$ that maximizes $\alpha_j$ over a candidate set, producing up to $n$ points per iteration (after ellipse filtering) as described in the following section. The hyperparameter $\beta$ controls the balance between exploitation of high-gradient regions and exploration of regions the algorithm has yet to visit: in this paper, the value of $\beta$ is set to 1.

### 3.3. Heuristics to capture boundary manifold

By exploiting high gradient magnitudes, $\alpha(\mathbf{x})$ runs the risk of selecting localized dense clusters of points within the phase transition boundary, without being able to capture its complete manifold efficiently. To prevent this, we create ellipses around the points that have already been selected to represent regions that are no longer considered for future acquisitions. Our ellipse-based strategy is inspired by the work in Dai & Glotzer,[26] where circles are constructed around past acquisition points to prevent placement of new acquisition points. In our gradient-based acquisition, the region of interest is directional. For instance, if the goal is to capture a phase transition boundary with a steep gradient along the $x$-direction, the region of interest will be narrow along $x$ but extended along $y$. In this case, designing an ellipse with its long axis aligned to the perpendicular $y$-direction allows query points to be more widely distributed along $y$ while maintaining tight coverage along $x$. Namely, to capture the high $x$-gradient, points are given vertical ellipses to force exploration in $y$. Conversely, to capture the $y$-gradient, points are given horizontal ellipses to force exploration in $x$. This dual strategy enables the efficient tracing of the boundary's complex curvature. The exact

**Table 1** Optimal ellipse radii for each 2D plane. "N/A" indicates that dimension is not part of that plane

|  | $\phi$ *vs.* $V^*$ | $\kappa a$ *vs.* $V^*$ | $\phi$ *vs.* $\kappa a$ |
|---|---|---|---|
| $\kappa a$ | N/A | 1 | 20 |
| $\phi$ | 0.05 | N/A | 1 |
| $V^*$ | 1 | 1 | N/A |

radii of each ellipse for each dimension of acquisition are listed in Table 1 and they are tuned using domain expertise. Broadly speaking, larger magnitudes of the variable will result in larger exclusion radii.

### 3.4. Initial samples and stopping criteria

**Initial samples.** For each two-dimensional (2D) parameter, we begin by selecting five points at random from within the specified parameter bounds. This random initialization ensures that we explore a broad portion of the phase space without bias. However, if prior domain knowledge is available (*e.g.*, approximate ranges or known transition regions), it can be used to guide a more informed selection of the initial points. Such an approach may lead to faster convergence, as the GP model starts with samples in physically relevant areas rather than spending simulations locating them. Other studies have also chosen points at the edges or boundaries as a convenient initial sample choice.[34]

Prior to initialization, we define the GP prior as a zero-mean function with the RBF kernel as described earlier. A white noise kernel is added to this to account for observation noise from the MD simulations. To provide a data-driven starting point, the RBF kernel's initial length scale for each dimension is set heuristically to the median of the pairwise distances of the initial five samples.

**Stopping criteria.** When learning to approximate the phase diagram and identify phase boundaries, the GP's predictions should stabilize as the active learning process converges. We monitor this with the rolling mean $\mu_{\text{rolling}}(t)$ of the predicted function calculated as a moving average over a specified window of recent acquisitions:

$$\mu_{\text{rolling}}(t) = \frac{1}{w}\sum_{i=t-w+1}^{t}\mu_i,$$

$$(11)$$

where $\mu_i$ represents the mean prediction of the GP model at iteration $i$ and $w$ is the window size.

To determine if the model has stabilized, the change in the rolling mean is compared to a predefined threshold $\varepsilon$. If the change in the rolling mean over consecutive windows is consistently below $\varepsilon$, it indicates that the model's predictions are stabilizing. Formally, the process can be considered converged if

$$|\mu_{\text{rolling}}(t) - \mu_{\text{rolling}}(t - w)| < \varepsilon$$

$$(12)$$

for several consecutive windows.

In the current work, we performed all experiments with a predefined number of iterations, while evaluating the validity of the stopping criteria.

# 4. Experiments and results

In this section, we demonstrate the effectiveness of our gradient-based active learning framework for identifying phase boundaries in colloidal systems. We present results for identifying boundaries of phase transition over both one-dimensional (1D) and two-dimensional (2D) parameter space using MD simulations.

## 4.1. Experimental setup

**Data generation.** We considered parameter ranges in $V^*$, $\phi$, and $\kappa a$ that defined the Morse potential and volume fraction relevant to colloidal interactions. In the 1D setup, we fixed two parameters (*e.g.*, $\phi = 0.20$ and $\kappa a = 30$) and varied $V^*$ over.[1,5] For the 2D setup, we fixed one parameter and varied the other two over practical ranges, for example $\phi \in [0.01, 0.50]$, $\kappa a \in [10, 60]$, and $V^* \in [0.1, 5]$. Our Molecular Dynamics (MD) simulations produced an observable $N_{nn}$ (number of nearest neighbors), from which we identified a phase transition boundary.

**Implementation details.** There are several important hyperparameters in our active learning method. First, the hyperparameter to optimize is the lengthscale of the GP. Compared to regular active learning, the focus on the sharp gradient region increases the risk of overfitting the lengthscale hyperparameters to such local regions and result in highly fluctuating GP functions. Therefore, instead of naive likelihood-based optimization of this hyperparameter,

we adopt a dynamic scheme to adjust the lengthscale values throughout the active learning process, utilizing the spatial distribution of the training points over time. More specifically, we initially calculate a lengthscale value in each dimension by examining pairwise distances among the early sample points, discarding any zero distances, and then taking the median of the remainder. At specified intervals (*e.g.*, every few active learning iterations), we recompute this median distance to reflect newly explored points and multiply it by preset factors to define the lower and upper bounds. This dynamic approach prevents the length scale from shrinking too much or growing excessively, striking a balance between localized resolution and broader search.

Another set of important hyperparameters is the ellipse radii to avoid excessive clustering of acquisition points and promote exploration of the phase-transition boundary. Their values are determined based on domain expertise. Table 1 lists the radii values we use for each 2D parameter space ($\phi$ *vs.* $V^*$, $\kappa a$ *vs.* $V^*$, and $\phi$ *vs.* $\kappa a$). Each row represents individual parameters, while the columns correspond to the planes over which the phase diagrams are being computed. For instance, in the first row, $\kappa a$ is left unused in constructing the phase diagram on the $\phi$ *vs.* $V^*$ plane, and the value of the ellipse radius is set to 1 and 20 along the $\kappa a$ direction for identifying the phase-transition boundary in the $\kappa a$ *vs.* $V^*$ and $\phi$ *vs.* $\kappa a$ planes, respectively.

**Grid search based ground truth.** A dense uniform sampling in the parameter space is performed, which we also treat as the ground truth for the phase diagrams. In the 1D experiment, 20 equidistant points are used for estimating the phase diagram. In 2D, an equispaced grid of 100 points are used.
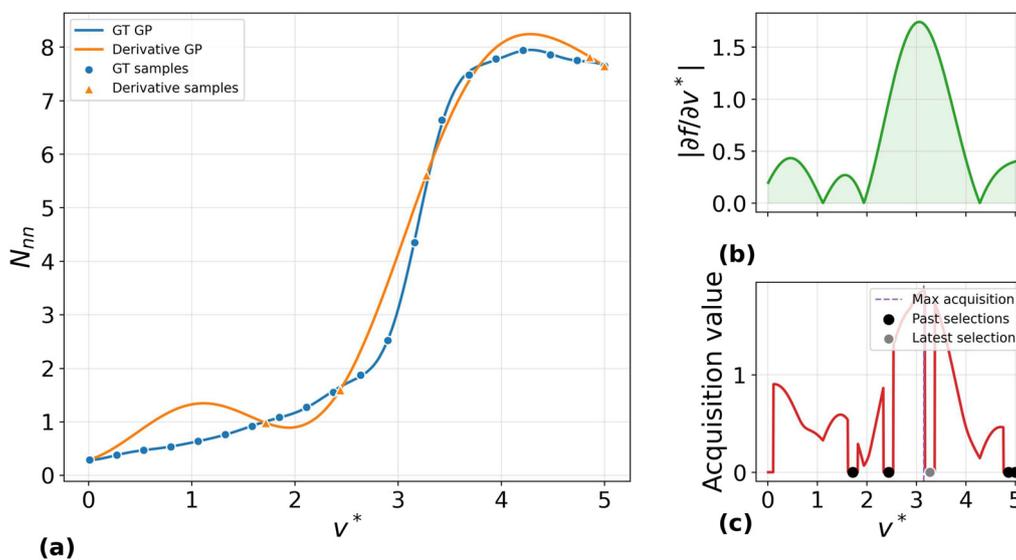


**Fig. 2** One-dimensional phase diagram with $\phi = 0.20$ and $\kappa a = 30$, comparing the derivative based acquisition method to grid search. (a): The derivative based method was able to get a decent solution in just 8 simulations achieving an HD 0.087 of and mean squared error of 0.19, compared to the reference established from 20 MD simulations. (b): The derivative of the GP mean highlights region of high gradient magnitude at the phase transition boundary. (c): The acquisition function peaks sharply in this same region, guiding the next sample to the most informative location.

**Baselines for comparison of methods.** We compare the presented gradient-based acquisition method with alternative acquisition methods including:

• *Acquisition based on the mean of the GP.* We consider a standard UCB acquisition function that does not use the gradient information and simply targets the mean of the GP. Given by the following equation:

$$\alpha(\mathbf{x}) = |\mu(\mathbf{x})| + \beta \times \sigma(\mathbf{x}) \qquad (13)$$

where $\beta$ is a constant that can be tuned to tune exploration. For our experiments we use $\beta = 1$.

• *Random sampling.* In this approach, the model is constructed by incrementally fitting a GP to a growing set of sampling points. In each iteration, the training set is augmented with two new points randomly sampled from a pre-defined set of 100 uniformly-sampled points in the parameter space that was used in constructing the reference phase diagram. Because the candidate points for sampling define our reference, the performance of this approach is expected to be stronger than true random sampling in the grid, constituting a stronger baseline for comparison.

**Metrics.** We evaluated the performance of all methods (gradient-based active learning and baselines) against the ground truth by:

• Accuracy is measured by the discrepancy between the predicted and ground-truth phase boundaries using the Hausdorff distance (HD). Hausdorff distance measures how far two shapes are from each other. Given two sets of points $A$ and $B$ (*e.g.*, two contours), it looks at the point in $A$ that is farthest from any point in $B$, and the point in $B$ that is farthest from any point in $A$; the larger of those two numbers is the Hausdorff distance. We use the usual Euclidean distance between points.

• Efficiency measured by the number of MD simulations needed to achieve the desired accuracy in delineating the phase-transition boundaries. Obtained from a dense grid of 100 MD simulations.

### 4.2. 1D results

We first illustrate our method in a simplified 1D scenario by fixing $\phi = 0.20$ and $\kappa a = 30$, and varying $V^* \in [0.1, 5]$. We select two query points at the ends of the range and evaluate them *via* MD simulations to initialize the GP. Fig. 2(a) shows that within just four active learning simulations (left panel), our gradient-based approach reconstructs the phase transition remarkably well. Compared to the reference established by the grid search method with 20 MD simulations (blue line), the presented active learning is able to achieve an HD score of 0.087 and a mean-squared error of 0.19 using a fraction of simulations (4 *vs.* 20). Fig. 2(b), showing the derivative of the GP mean, highlights that the algorithm strategically adds new query points precisely in the region of the highest gradient, which corresponds to the phase boundary. This targeted sampling

is driven by the acquisition function Fig. 2(c), which peaks sharply in the area deemed most informative for the next evaluation.

### 4.3. 2D results

We next extend our approach to 2D phase diagrams by fixing one parameter and varying the other two, including the ($\phi$, $\kappa a$) plane at $V^* = 4$ (Fig. 2), the ($V^*$, $\kappa a$) plane at $\phi = 0.05$ (Fig. 2), and the ($V^*$, $\phi$) plane at $\kappa a = 50$ (Fig. 2). In each figure, the leftmost column shows the reference phase diagram learned using 100 MD simulations, where the colorbar codes the values of $N_{nn}$. The sequences of images demonstrate the updated approximation of the phase diagram (top row) and its contour maps (bottom row) as active learning continues.

**Ground truth for 2D phase diagrams.** Fig. 2 shows the phase diagram for fixed $V^* = 4$, where the majority of the phase diagram shows the stable condensed phase with high $N_{nn}$ in red. A clear dilute phase with low $N_{nn}$ appears for high $\kappa a$ and low $\phi$ in the top-left, showing that the phase boundary is near the edge of the phase diagram. Delineating boundaries near the bounds of the phase diagram is particularly challenging as (i) randomly selected points on the plane will exhibit high $N_{nn}$ and predict low gradients, and (ii) gradients tend to be poorly defined near the edges. Therefore, this plane is expected to be the most challenging for the gradient-based method. Fig. 2 shows the phase diagram for fixed $V^* = 0.05$. A dilute phase with low $N_{nn}$ appears for low $V^*$ and a condensed phase with high $N_{nn}$ appears for high $V^*$. The gradient appears to vary primarily in the $V^*$ direction with a smaller gradient in the $\kappa a$ direction appearing near the edge. Finally, Fig. 2 shows the well-known phase diagram of colloids with $V^*$ related to inverse temperature. Regions in high $V^*$ exhibit condensed phases, while the phase boundary occurs diagonally from the top left to the middle right. The diagonal nature suggests that gradients in both directions will need to be accurately sampled by the actively learning methodology. The ground truth is qualitatively similar to well-known phase diagrams published in the literature.[15,28] The three planes exhibit different kinds of gradients and are chosen to demonstrate the applicability of the methodology.

**Accuracy and efficiency of the gradient-based acquisition.** The quantitative HD metrics for the visual results in Fig. 3 are summarized in Fig. 4. As shown, for the phase diagram at $V^* = 4$ (Fig. 3(a)), the presented gradient-based acquisition is able to predict the high-gradient phase boundary in the top-left by simulation 13, directing subsequent acquisition points in that region (along with the boundary in the bottom of the phase diagram). By simulation 19, the phase boundary in the top left has been reasonably captured (with an HD close to 0.1 compared to the reference). The subsequent acquisitions focus on refining both this boundary and that at the bottom, achieving an HD $\leq$ 0.1 by simulation 40. A detailed examination of the contour maps in Fig. 3(a) shows that most
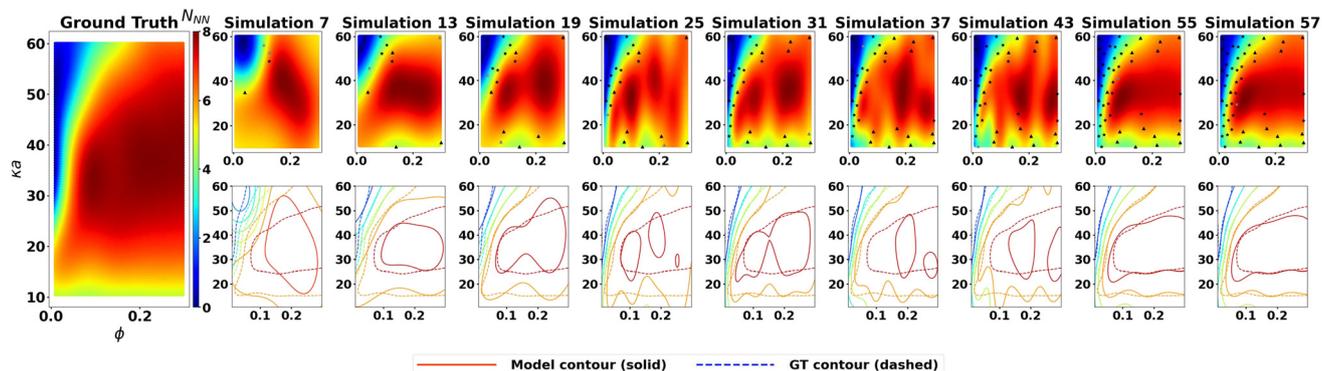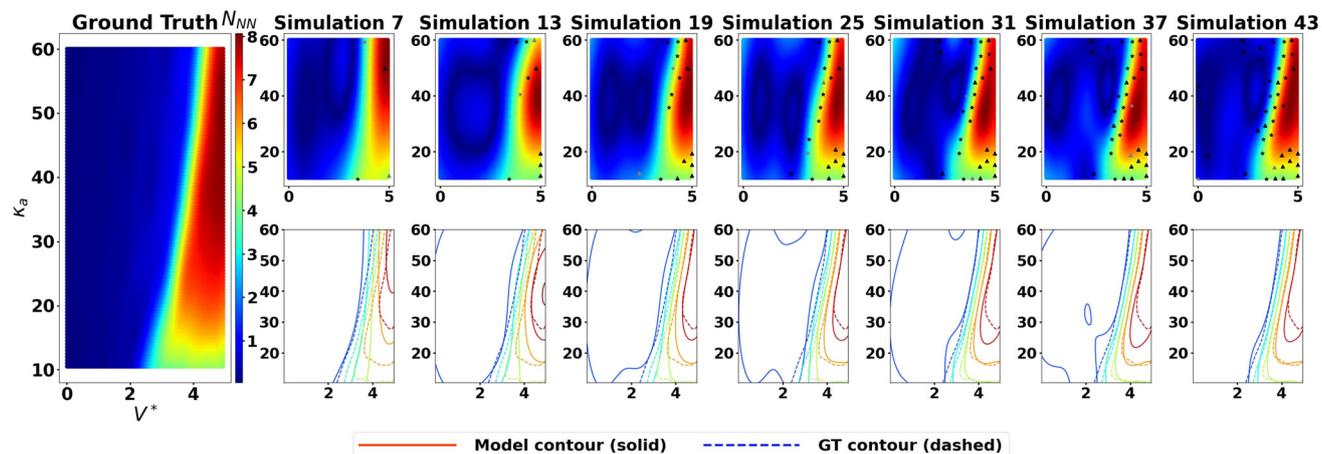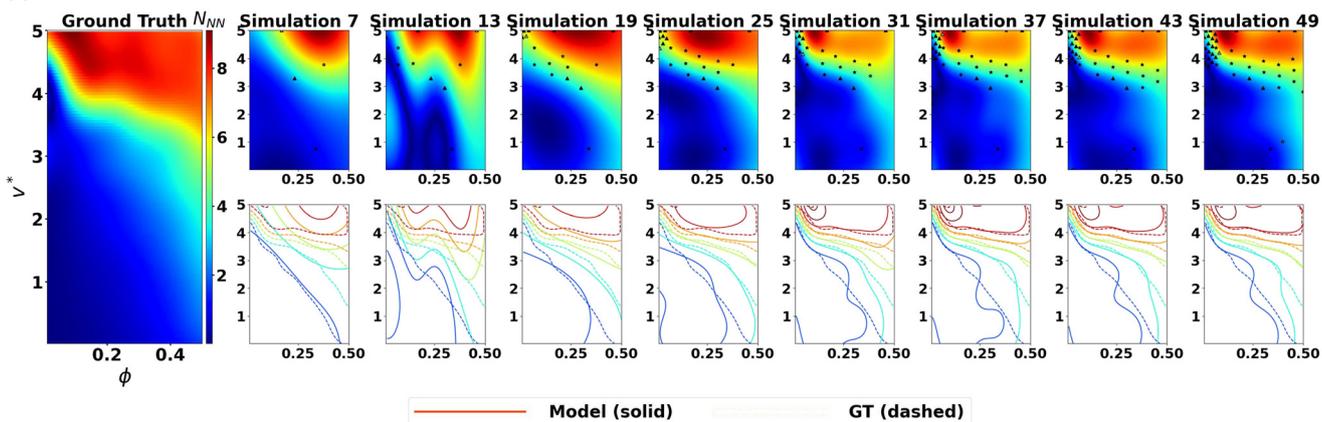
(a) Two dimensional plane in $\phi$ and $\kappa a$ with fixed $V^* = 4$



(b) Two dimensional plane in $V^*$ and $\kappa a$ with fixed $\phi = 0.05$



(c) Two dimensional plane in $\phi$ and $V^*$ with fixed $\kappa a = 50$

**Fig. 3** Gradient-based active-learning progression for three two-dimensional planes. In each panel, the leftmost column is the ground-truth phase diagram (100-simulation grid); subsequent columns show the GP mean (top) and contour comparison (bottom) as active learning proceeds.

of the challenges come from an accurate approximation of the boundary at the bottom, while the top-left phase boundary is well localized and remains stable since simulation 25.

For the phase diagram at $\phi = 0.05$, the primary boundary with the largest gradient in the $V^*$ direction is rapidly captured by simulation 13, while the secondary boundary in the orthogonal direction in the $\kappa a$ direction near the edge

takes a bit longer until simulation 19 to be localized, achieving a HD $\leq 0.1$ beyond 17 simulations.

For the phase diagram at $\kappa a = 50$, the diagonal phase boundary is again quickly detected at simulation 13, with subsequent acquisitions directed towards this region. By simulation 19, the diagonal structure of the phase boundary is well captured with a HD $\leq 0.1$ and, by simulation 24, the detailed shape of the boundary near the top left of the
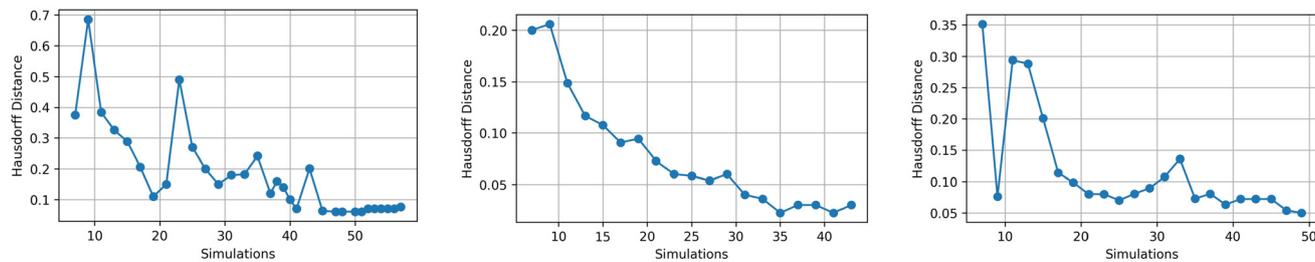
(a) Two dimensional plane in $(\phi, \kappa a)$ at $V^* = 4$

(b) Two dimensional plane in $(V^*, \kappa a)$ at $\phi = 0.05$

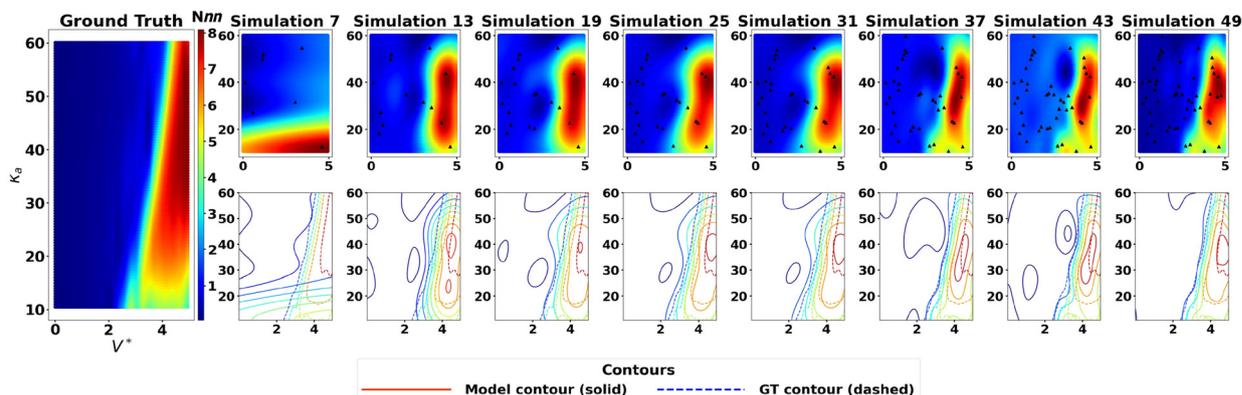(c) Two dimensional plane in in $(V^*, \phi)$ at $\kappa a = 50$

**Fig. 4** Hausdorff distance (HD) between the predicted and ground-truth phase boundary against the number of sequential MD simulations.

diagram is well captured, with a HD well below 0.1 throughout the rest of the simulations.

Overall, as summarized in Fig. 4, the presented gradient-based acquisition is able to detect the primary boundary within 9–11 simulations and, by simulation 17–21, accurately capture the detailed shape of the phase boundary with a HD $\leq 0.1$ if the primary boundary is not near the edge of the phase diagram. This means a reduction of over 80% computation compared to the reference (17 *vs.* 100 simulations). In the more challenging case where the primary phase boundary in near the edge of the phase diagram, the

presented method is still able to accurately capture the primary boundary by simulation 24 (70% reduction of computation) and delineate even the secondary boundary in about 40 simulations (50% computation). This ability to localize phase boundary near edges can be important when applying to physical experimentation, where the feasible windows of parameters could be constrained by other practical factors.

**Comparison to alternative acquisition methods.** Fig. 5 provide examples of results from the two alternative methods where new parameter points are acquired randomly in



(a) Two dimentional plane in $V^*$ and $\kappa a$ with fixed $\phi = 0.05$ using randomly sampled points
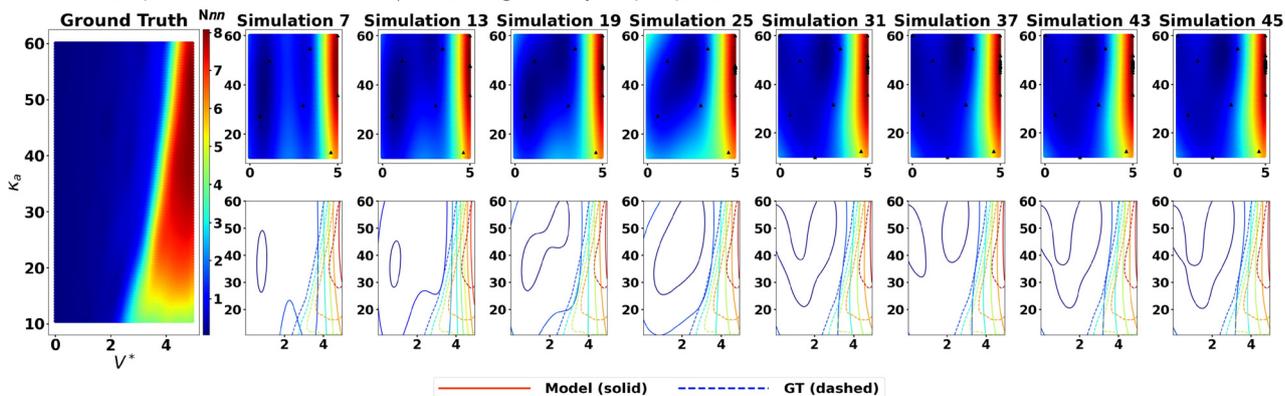


(b) Two dimentional plane in $V^*$ and $\kappa a$ with fixed $\phi = 0.05$ using acquisition based on mean targeted UCB.

**Fig. 5** (a): Progression of active learning with random acquisition for $\phi = 0.05$. (b) Progression of active learning with mean targeted UCB for $\phi = 0.05$.
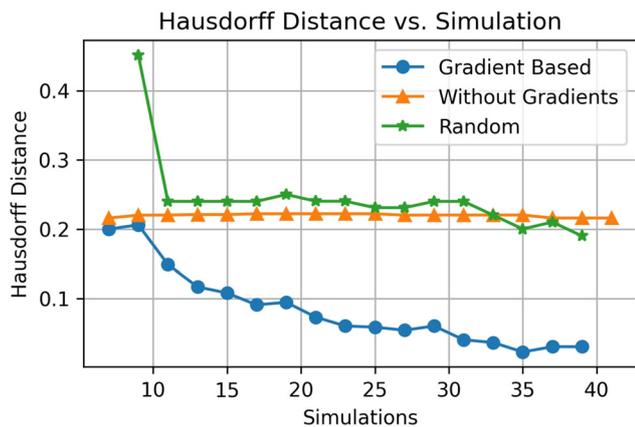
**Fig. 6** Comparison of HD metrics across three strategies for the plane in and $V^*$ and $\kappa a$ with fixed $\phi = 0.05$ as the number of simulations increases. Each curve reports HD between the predicted and ground-truth phase boundary.

Fig. 5(a) or by targeting the max mean of the GP as in standard UCB in in Fig. 5(b). Additional results for two other planes are included in Fig. S3 of the SI.

As shown in Fig. 5(a), for the phase diagram at $\phi = 0.05$, the random sampling approach struggles to correctly capture the top left of the transition boundary. This is in contrast with the gradient-based acquisition in Fig. 2 that starts to refine this boundary at simulation 13 and correctly captures its detail by simulation 19.

As shown in Fig. 5(b), for the phase diagram at $\phi = 0.05$, acquisition explicitly targeting a high mean quickly captures the top left boundary as early as simulation 7, but fails to learn the bottom right boundary which was along the edge and high in gradients in the $y$ direction. As a result, it was not able to correctly capture the shape of the phase-transition boundary even with more than 45 simulations.

These visual observations are confirmed by the quantitative HD metrics in Fig. 6, which shows that the HD metrics by the gradient-based acquisition drops significantly faster and stays lower compared to the two alternative methods. An important advantage of the gradient-based acquisition is that it does not require prior knowledge or assumption about the $N_{nn}$ value on which to locate the phase boundary; it remains effective even when such knowledge is absent. We compared this method to a baseline where we explicitly used the knowledge that the boundary lies at $N_{nn} = 6$ and the gradient method was comparable to this method. Additional details are in Fig. S4 and S5 of the SI.

### 4.4. Additional method analysis

**Stopping criterion.** Although we present fixed-iteration results in this paper, our method naturally supports stopping once the GP's predictions stabilize. Section 3.4 describes a rolling-mean approach to detect when changes remain below a threshold $\varepsilon$ for multiple iterations. For each two-dimensional phase diagram, we monitor six rolling statistics:

1. GP rolling mean (leftmost panel) – the moving average of the GP mean over all parameter pairs.

2. Mean-error to ground truth – the absolute difference between the GP mean and the reference mean.

3. GP rolling mean of $\partial f/\partial x_1$ or $\partial f/\partial x_2$ – moving average of the derivative along each direction.

4. Derivative-error in $x_1$ or $x_2$ – absolute difference between the derivative and the ground truth in each direction.

An example is shown in Fig. 7 here the rolling mean drops below $\varepsilon = 0.2$ which is the threshold for stopping at this point the RSME has achieved a small value close to 0.5 indicating good results. Beyond this point the GP is further refined but at the cost of additional simulations. In each case, a rolling-mean threshold $\varepsilon = 0.2$ would have stopped the run before the maximum number of simulations is reached. Moreover, when the rolling mean is under the threshold $\varepsilon = 0.2$, not only is the mean difference with the ground truth low but also the difference in the means of derivatives in both directions is low. See supplementary information for more comprehensive convergence metrics. For $\phi = 0.05$, this happens at simulation 27. This suggests that under the threshold value, we get accurate phase boundary across both dimensions. Fine-tuning $\varepsilon$ therefore offers a principled way to trade off efficiency and accuracy depending on the application need.

**Ellipse-based heuristic.** To further evaluate the effect of the ellipse-based heuristic, we compare our presented method with and without the ellipse-based heuristic. Fig. 8 visualises the progression of gradient-based AL across iterations without the ellipse-based heuristics. Here, symbols marked as * are acquisition points in the $x$-direction and symbols marked as $\Delta$ are in the $y$-direction; the ellipses are not used in the acquisition method, but displayed to
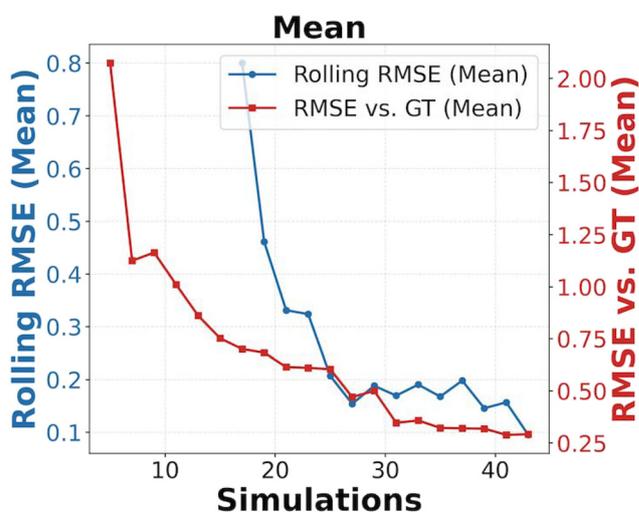


(a) The plane in $V^*$ and $\kappa a$ with fixed $\phi = 0.05$

**Fig. 7** Convergence diagnostics for the phase diagram at $\phi = 0.05$, including the rolling mean of the GP prediction (blue) and its deviation from the ground-truth (red).
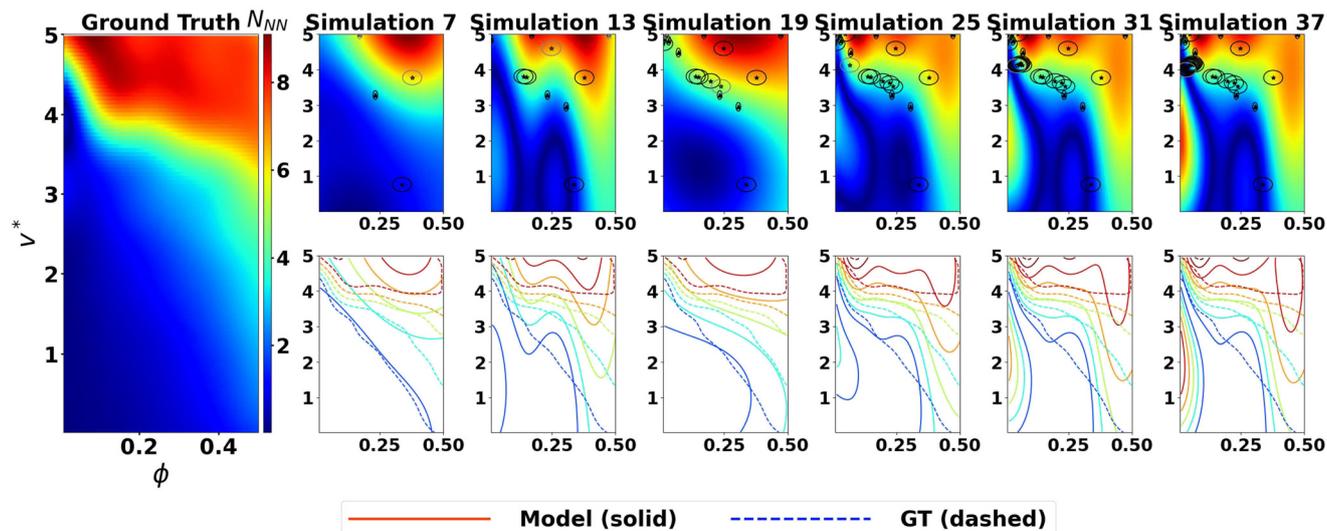
View Article Online

MSDE · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · · Paper

**Fig. 8** Active-learning progression on the phase diagram for $\kappa a = 50$ without the ellipse-exclusion rule. Black markers denote sampled points. Successive queries cluster along a single segment of the phase boundary, leaving other regions under-sampled and delaying global refinement. The ellipses shown are not used in the acquisition method, but displayed to illustrate the clustering of successive acquisition points which does not happen when the ellipse-based heuristics are used.

illustrate the clustering of successive acquisition points which does not happen when the ellipse-based heuristics are used. As shown, successive acquisitions tend to concentrate along one portion of the boundary, delaying explorations of the phase boundary elsewhere. In contrast, ellipse heuristics helps to distribute acquisitions along the entire boundary, yielding higher accuracy with fewer MD runs.

Fig. 9 summarizes the quantitative results for the fixed $\kappa a = 50$ plane with and without ellipse-based heuristics. The HD metrics show that the ellipse-heuristics strategy reaches an error <0.1 within 9 simulations and stabilizes within 17, whereas the one without drops quickly to HD <0.1 within 13 simulations but is unable to stabilize and reaches an HD as
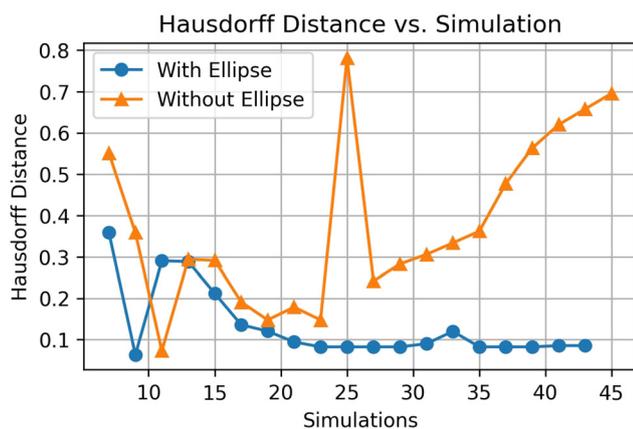


**Fig. 9** Hausdorff distance between the predicted and reference $N_{nn}$ = 6 contour on a two-dimensional plane in $\phi$ and $V^*$ with fixed $\kappa a = 50$. The ellipse rule (blue) accelerates convergence and maintains a low error, whereas omitting the rule (orange) slows convergence and leaves a higher residual error.

high as 0.7 within subsequent simulations. If necessary, one can adjust the rate of convergence by adjusting the exclusion ellipse radii.

## 5. Discussion

### Convergence on random initialization

A noteworthy outcome of our two-dimensional experiments was the method's robustness to its initial conditions. Each experiment was initialized using only five randomly sampled points, providing the model with a sparse and potentially uninformative starting set without any prior knowledge of the phase boundary. Despite this uninformed initialization, the results (section 4.3) demonstrated that our approach not only consistently converged to a high-fidelity solution but did so with remarkable sample efficiency. This suggested that the algorithm was not highly sensitive to the initial sampling and could reliably perform its exploratory-exploitative search, a significant practical advantage for real-world applications where well-chosen initial samples were not available.

### Importance of domain knowledge

Although this strategy was largely data-driven, domain knowledge was critical for setting the scaling factors and bounding the length scale appropriately. Colloidal phase boundaries could be sudden and steep, and incorrect parameter choices might have caused the model to either miss the transition or oversample irrelevant regions. By combining approximate knowledge of parameter ranges ($\phi$, $\kappa a$, $V^*$) with our ellipse-based heuristic and adaptive length-scale updates, the algorithm more reliably focused on steep gradients where phase changes occurred, reducing the computational effort needed to map narrow transition zones,

particularly when the phase boundaries were close to the upper and lower bounds of the parameters.

In practice, the phase diagram of a given colloidal system may also be drawn using different representations of the variables in each dimension. In the present study, the temperature dependence was reported using $V^* = V_0/k_B T$ corresponding to inverse temperature, but it is common practice to use dimensionless temperature ($T^* = k_B T/V_0$), or the reduced second virial coefficient, which is a non-linear function of $V_0$ and $T$.[15] In such representations, the new range of variables labeling that dimension will inform the heuristic choice of parameters, namely, the length-scale parameter in the kernel (eqn (3)) and the radius of the exclusion ellipse.

### Improvements to the acquisition protocol

Two of the acquisition methods investigated presented distinct advantages which might suggest a superior hybrid acquisition approach, especially when the target value is unknown *apriori* and the simulations are expensive. In the hybrid approach, the gradient-based acquisition could be used to converge on the phase boundary in a few iterations, from which the target value is also learned. Following this, the target-based acquisition can be utilized to accelerate convergence. This might be particularly advantageous in case of order parameters that are (i) experimentally relevant and (ii) for which the value at the phase boundary is not easily known, such as the magnitude of structure factor for equilibrium phase boundaries or storage/loss moduli for non-equilibrium boundaries.

### Applications to other phase diagrams & future work

The presented active-learning methodology was suitable for rapid discovery of phase-transition boundaries in applications where the underlying phase diagrams were expected to be piecewise smooth with localized, sharp boundaries between phases. Such sharp boundaries are observed in numerous soft material systems such as mixtures containing polymers, colloids, solvents, biomaterials, synthetic polypeptides, and proteins. This methodology can be applied to those systems as well as novel material systems, provided that the phase boundary is sharp. In contrast, target-value-based acquisition function requires a well-defined metric to identify the transition region, which may be unavailable when novel material systems have not undergone complete experimental characterization.

In this work, the phase diagram was represented by a one-dimensional scalar value, $N_{nn}$, and approximated by a GP, to distinguish two phases (a dense and dilute phase). This could be extended in future work to allow phase diagrams to be represented by multi-dimensional variables, either for a more complete characterization of a single phase boundary of interest, or for characterizing multiple phase boundaries using different metrics. This can be achieved by replacing a single-output GP with multi-output GPs[35,36] or a neural network.[37,38] If the intent is to use the multi-dimensional output to identify a single phase boundary, the proposed gradient-based acquisition can be extended to consider the vector output with respect to each input dimension. For example, the characterization of glassy phases might require identification of both local structure and dynamics; local structure ($N_{nn}$) alone would be insufficient. If the intent is to use each dimension of the output vector to identify different phase boundaries, the proposed gradient-based acquisition can be applied to each output dimension, thereby running several acquisition simulations per iteration while sharing information from the acquired simulations. For example, to incorporate the glassy phase in the phase diagram in addition to the dilute and condensed phases, one would use this approach to simultaneously identify at least two boundaries – the dilute-condensed line and the fluid-glass line.

In the current method, to avoid excessive placement of local clusters of acquisition points near high-gradient regions, we adopted heuristic ellipses to encourage acquisitions to be evenly placed along the phase-transition boundary. Future work could improve this heuristic by designing an acquisition function that searches for the high gradient manifold of the phase diagram, instead of searching for individual points of high gradient.

Finally, the presented method could be used in a multi-fidelity simulation setting, where coarse-scale simulations would provide rapid initial localization of the phase-transition boundary, followed by more complex simulations actively selected near the initial boundary to iteratively refine its delineation. The high-fidelity simulations could take one of two approaches: refinement of the order parameter that describe the molecular behavior, or refinement of the estimation of the thermodynamic properties of the phase boundary itself. In the first approach, physics-informed techniques using principal component analysis (PCA) as described in ref. 38 can be used to construct high-fidelity order parameters near the vicinity of the phase boundary that can elucidate underlying complex mechanisms of the phase transition. In the second approach, expensive high-fidelity simulations such as coexistence simulations or free energy methods to compute the underlying energy landscapes can be performed. More recently, the expense of high-fidelity simulations was shown to be dramatically reduced by the use of generative thermodynamic maps (TMs) to predict free energy distributions in the context of phase transitions,[39,40] but more work is needed in this area to improve the accuracy of the distributions.

## 6. Conclusions

Mapping colloidal phase diagrams with molecular dynamics (MD) remains a central bottleneck for materials design because the regions of primary scientific value the phase transition boundaries are typically sharp, localized, and embedded within otherwise smooth portions of parameter

space. Dense grid-based sampling can resolve such narrow transition zones, but at a computational cost that grows rapidly with dimensionality and often wastes simulations far from the boundary. In this work, we introduced a gradient-based active learning framework designed specifically to accelerate the discovery of phase boundaries in colloidal phase diagrams while maintaining physical interpretability that phase transitions are more universally characterized by high susceptibility of the order parameter rather than a unique value of the order parameter. Compared to existing active learning approaches used in phase boundary detection,[26] our proposed method automatically discovers the unknown region of sharp phase transition (*i.e.*, high gradient magnitude) from the learned phase diagram, without requiring a known value to target the phase boundary nor relying on heuristic assumptions to approximate the phase boundary as a boundary between binary classes as in ref. 26.

Across one-dimensional and two-dimensional planes of the phase diagram, the proposed gradient-aware strategy consistently localized phase boundaries with substantially fewer MD simulations than dense grid search. In addition, it outperformed random acquisition and conventional GP acquisition strategies that operate only on the predicted mean and variance of $N_{nn}$ and are not explicitly sensitive to sharp spatial transitions. Importantly, because the approach uses gradient information rather than assuming a fixed target level of the order parameter, it remains effective even when the precise boundary value is uncertain or varies across conditions – an advantage for exploratory studies where a universal threshold may not be known *a priori*.

Beyond the immediate results, the framework provides a practical route toward scalable, physics-aligned phase-diagram discovery. The gradient-based acquisition principle naturally extends to higher-dimensional parameter spaces, where grid search becomes prohibitive, and it is compatible with multi-fidelity workflows in which inexpensive, lower-accuracy simulations provide coarse localization while high-fidelity MD is reserved for refining critical boundary segments. More broadly, the methodology is applicable to other soft-matter and self-assembly systems where phase behavior is piecewise smooth with localized transition zones and where an order parameter exhibiting sharp changes can be defined. Taken together, these findings demonstrate that incorporating gradient information into active learning enables sample-efficient, boundary-focused exploration of colloidal phase diagrams and offers a computationally efficient alternative to exhaustive simulation campaigns.

## Author contributions

All authors drafted and edited the manuscript. Poornima Padmanabhan and Linwei Wang contributed to the design of the study. Poornima Padmanabhan contributed with domain knowledge expertise (MD simulations and colloidal phase diagrams). Sumeet Vadhavkar, Pradeep Bajracharya, Md Shakil Zaman, and Linwei Wang contributed to the development, implementation, and evaluation of active learning methods.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

The data that supports the findings of this article are available upon reasonable request to the corresponding authors. Supplementary information (SI) is available and contains detailed molecular dynamics simulation methodology and additional results not included in the manuscript. See DOI: https://doi.org/10.1039/d5me00233h.

## Acknowledgements

## References

1  W. B. Russel, D. A. Saville and W. R. Schowalter, *Colloidal Dispersions*, Cambridge University Press, 1989.

2  R. J. Hunter, *Foundations of Colloid Science*, Oxford University Press, 1993.

3  G. M. Whitesides and J. C. Love, *Scientific American*, 2002, vol. 285, pp. 40–47.

4  D. Peer, J. M. Karp, S. Hong, O. C. Farokhzad, R. Margalit and R. Langer, *Nat. Nanotechnol.*, 2007, **2**, 751–760.

5  W.-X. Zhang, *J. Nanopart. Res.*, 2012, **5**, 323–333.

6  D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, 3rd edn, 2023, pp. 97–124.

7  D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Academic Press, San Diego, 2nd edn, 2001.

8  W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes 3rd Edition: The Art of Scientific Computing*, Cambridge University Press, Cambridge, 3rd edn, 2007.

9  M. Karplus and J. A. McCammon, *Nat. Struct. Biol.*, 2002, **9**, 646–652.

10  S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.

11  R. Bellman, *Dynamic Programming*, Princeton University Press, 2010.

12  M. Keijzer, *Genet. Program. Evolvable Mach.*, 2005, **6**, 259–269.

13  K. Binder, *Rep. Prog. Phys.*, 1987, **50**, 783–859.

14  B. Smit and D. Frenkel, *J. Chem. Phys.*, 1992, **94**, 5663–5669.

15  S. M. Fenton, P. Padmanabhan, B. K. Ryu, T. T. D. Nguyen, R. N. Zia and M. E. Helgeson, *Proc. Natl. Acad. Sci. U. S. A.*, 2023, **120**, e2215922120.

16 A. Z. Panagiotopoulos, *Mol. Phys.*, 1987, **61**, 813–826.

17 D. A. Kofke, *J. Chem. Phys.*, 1993, **98**, 4149–4162.

18 S. Ramaswamy, *Annu. Rev. Condens. Matter Phys.*, 2010, **1**, 323–345.

19 M. C. Marchetti, J.-F. Joanny, S. Ramaswamy, T. B. Liverpool, J. Prost, M. Rao and R. A. Simha, *Rev. Mod. Phys.*, 2013, **85**, 1143–1189.

20 B. Settles, *Computer Sciences Technical Report 1648*, University of Wisconsin-Madison, 2009.

21 T. Lookman, P. V. Balachandran, D. Xue and R. Yuan, *npj Comput. Mater.*, 2019, **5**, 21–37.

22 C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, MA, 2006.

23 J. Snoek, H. Larochelle and R. P. Adams, *Adv. Neural Inf. Process. Syst.*, 2012, **25**, 2951–2959.

24 J. Dhamala, P. Bajracharya, H. J. Arevalo, J. L. Sapp, B. M. Horacek, K. C. Wu, N. A. Trayanova and L. Wang, *Med. Image Anal.*, 2020, **62**, 101670.

25 C. Meisenzahl, K. Gillette, A. J. Prassl, G. Plank, J. L. Sapp and L. Wang, *Comput. Biol. Med.*, 2024, **182**, 109201.

26 C. Dai and S. C. Glotzer, *J. Phys. Chem. B*, 2020, **124**, 1275–1284.

27 S. C. Glotzer and M. J. Solomon, *Nat. Mater.*, 2007, **6**, 557–562.

28 E. Zaccarelli, *J. Phys.: Condens. Matter*, 2007, **19**, 323101.

29 A. P. Thompson, H. M. Aktulga, R. Berger, D. S. Bolintineanu, W. M. Brown, P. S. Crozier, P. J. in 't Veld, A. Kohlmeyer, S. G. Moore, T. D. Nguyen, R. Shan, M. J. Stevens, J. Tranchida, C. Trott and S. J. Plimpton, *Comput. Phys. Commun.*, 2022, **271**, 108171.

30 Rochester Institute of Technology, Research Computing Services, 2025, https://www.rit.edu/researchcomputing.

31 V. K. Shen and P. G. Debenedetti, *J. Chem. Phys.*, 2001, **114**, 4149–4159.

32 S. L. Meadley and F. A. Escobedo, *J. Chem. Phys.*, 2012, **137**, 074109.

33 P. Padmanabhan and R. Zia, *Soft Matter*, 2018, **14**, 3265–3287.

34 D. O. Abranches, E. J. Maginn and Y. J. Colón, *AIChE J.*, 2023, **69**, e18141.

35 S. Jahangiri, Y. Kontar and D. S. Oliver, *Comput. Methods Appl. Mech. Eng.*, 2023, **405**, 115846.

36 H. H. Bruun and C. E. Rasmussen, *Proc. R. Soc. A*, 2021, **477**, 20210052.

37 P. Bajracharya, D. O'Hara, C. Meisenzahl, K. Gillette, A. J. Prassl, G. Plank, J. L. Sapp and L. Wang, *Computing in Cardiology*, 2025, vol. 52, p. 067.

38 R. B. Jadrich, B. A. Lindquist, W. D. Piñeros, D. Banerjee and T. M. Truskett, *J. Chem. Phys.*, 2018, **149**, 194110.

39 L. Herron, K. Mondal, J. S. Schneekloth and P. Tiwary, *Proc. Natl. Acad. Sci. U. S. A.*, 2024, **121**, e2321971121.

40 E. R. Beyerle and P. Tiwary, *Phys. Rev. Lett.*, 2025, **135**, 068102.