



# Designing self-assembling kinetics with differentiable statistical physics models

Carl P. Goodrich<sup>a,b,1,2</sup> , Ella M. King<sup>c,1</sup> , Samuel S. Schoenholz<sup>d</sup> , Ekin D. Cubuk<sup>d</sup>, and Michael P. Brenner<sup>a,c,d</sup> 

<sup>a</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138; <sup>b</sup>Institute of Science and Technology Austria, A-3400 Klosterneuburg, Austria; <sup>c</sup>Physics Department, Harvard University, Cambridge, MA 02138; and <sup>d</sup>Brain Team, Google Research, Mountain View, CA 94043

Edited by Steve Granick, Institute for Basic Science, Ulju-gun, Ulsan, South Korea, and approved January 25, 2021 (received for review November 20, 2020)

**The inverse problem of designing component interactions to target emergent structure is fundamental to numerous applications in biotechnology, materials science, and statistical physics. Equally important is the inverse problem of designing emergent kinetics, but this has received considerably less attention. Using recent advances in automatic differentiation, we show how kinetic pathways can be precisely designed by directly differentiating through statistical physics models, namely free energy calculations and molecular dynamics simulations. We consider two systems that are crucial to our understanding of structural self-assembly: bulk crystallization and small nanoclusters. In each case, we are able to assemble precise dynamical features. Using gradient information, we manipulate interactions among constituent particles to tune the rate at which these systems yield specific structures of interest. Moreover, we use this approach to learn nontrivial features about the high-dimensional design space, allowing us to accurately predict when multiple kinetic features can be simultaneously and independently controlled. These results provide a concrete and generalizable foundation for studying nonstructural self-assembly, including kinetic properties as well as other complex emergent properties, in a vast array of systems.**

self-assembly | colloids | inverse design

Kinetic features are a critical component of a wide range of biological and material functions. The complexity seen in biology could not be achieved without control over intricate dynamic features. Viral capsids are rarely infectious if they assemble too quickly (1), protein folding requires deft navigation and control of kinetic traps (2), and crystal growth is largely controlled by the relative rates of different nucleation events (3). However, the design space of kinetic features remains largely unexplored. Self-assembly traditionally focuses on tuning structural properties, and while there have been significant successes in developing complex structural features (4–9), there has been little exploration of dynamical features in the same realm. We demonstrate quantitative control over assembly rates and transition rates in canonical soft matter systems. In doing so, we begin to explore the vast design space of dynamical materials properties.

To intelligently explore the design space of dynamical features, we rely on recent advances in automatic differentiation (AD) (10–14), a technique for efficiently computing exact derivatives of complicated functions automatically, along with advances in efficient and powerful implementations of AD algorithms (15–17) on sophisticated hardware such as GPUs (graphical processing units) and TPUs (tensor processing units). While the development of these software packages is driven by the machine learning community, a distinct, nondata-driven approach is emerging that combines AD with traditional scientific simulations. Rather than training a model using preexisting data, AD connects observables to physical parameters by directly accessing physical models. This enables intelligent navigation of high-dimensional phase spaces. Initial examples utilizing this approach include optimizing the dispersion of photonic crystal waveguides (18), the invention of new coarse-grained algorithms

for solving nonlinear partial differential equations (19), the discovery of molecules for drug development (20), and greatly improved predictions of protein structure (21).

Here, we begin to explore a class of complex inverse design problems that traditionally has been hard to access. Using AD to train well-established statistical physics-based models, we design materials for dynamic, rather than structural, features. We also use AD to gain theoretical insights into this design space, allowing us to predict the extent of designability of different properties. AD is an essential component of this approach (22) because we rely on gradients to connect physical parameters to complex emergent behavior. While there are other approaches for obtaining gradient information (e.g., finite difference approximations), AD calculates exact derivatives and more importantly, can efficiently handle large numbers of parameters. Furthermore, the theoretical insights we develop rely on accurate calculations of the Hessian matrix of second derivatives, for which finite difference approaches are insufficient.

We start in *Tuning Assembly Rates of Honeycomb Crystals* by considering the bulk crystallization of identical particles into both honeycomb and triangular lattices. By differentiating over entire molecular dynamics (MD) trajectories with respect to interaction parameters, we are able to tune their relative crystallization rates. We proceed with a more experimentally relevant example in *Tuning Transition Rates in Colloidal Clusters*, where

PHYSICS

## Significance

Engineering at the nanoscale is rich and complex: researchers have designed small-scale structures ranging from smiley faces to intricate sensors. However, designing specific dynamical features within these structures has proven to be significantly harder than designing the structures themselves. Biology, on the other hand, demonstrates fine-tuned kinetic control at nearly all scales: viruses that form too quickly are rarely infectious, and proper embryonic development depends on the relative rate of tissue growth. Clearly, kinetic features are designable and critical for biological function. We demonstrate a method to control kinetic features of complex systems and apply it to two classic self-assembly systems. Studying and optimizing for kinetic features, rather than static structures, opens the door to a different approach to materials design.

Author contributions: C.P.G., E.M.K., S.S.S., E.D.C., and M.P.B. designed research; C.P.G. and E.M.K. performed research; C.P.G., E.M.K., S.S.S., and E.D.C. contributed new reagents/analytic tools; C.P.G. and E.M.K. analyzed data; and C.P.G., E.M.K., S.S.S., E.D.C., and M.P.B. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nd/4.0/).

<sup>1</sup>C.P.G. and E.M.K. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. Email: carl.goodrich@ist.ac.at.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2024083118/-/DCSupplemental>.

Published March 2, 2021.

we consider small colloidal nanoclusters that can transition between structural states. By optimizing over the transition barriers, we obtain precise control over the transition dynamics. Together, *Tuning Assembly Rates of Honeycomb Crystals* and *Tuning Transition Rates in Colloidal Clusters* demonstrate that kinetic features can be designed in classic self-assembly systems while maintaining the constraint of specific structural motifs. The use of AD provides a way to directly access the design space of kinetic features by connecting emergent dynamics with model parameters. This connection reveals insight into the high-dimensional design landscape, enabling us to make accurate predictions in *Simultaneous Training of Multiple Transitions* regarding when large numbers of kinetic features of the colloidal clusters can be controlled simultaneously. Finally, we conclude in *Discussion* with an analysis of both the specific and broad impact of these results.

### Tuning Assembly Rates of Honeycomb Crystals

We start with the classic self-assembly problem introduced by Rechtsman et al. (23, 24) of assembling a honeycomb lattice out of identical isotropic particles. The honeycomb lattice is a two-dimensional analog of the diamond lattice, which exhibits desirable mechanical and photonic properties, but its low density makes assembling such a crystal a particularly difficult problem. Nevertheless, Rechtsman et al. (23) successfully designed a honeycomb lattice using a potential of the form

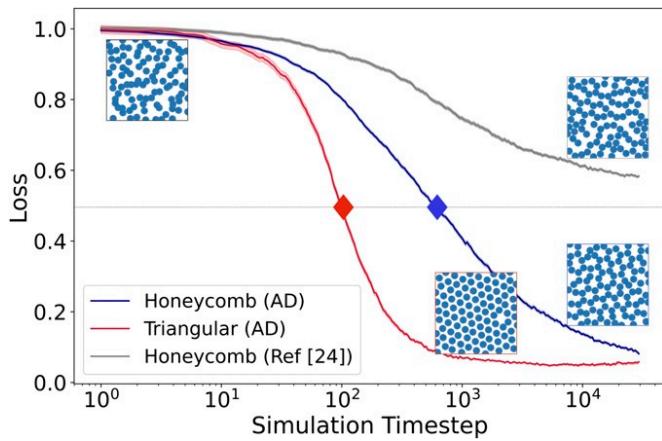
$$U(r) = \frac{b_0}{r^{12}} - \frac{a_0}{r^{10}} + a_1 e^{-a_2 r} - b_1 e^{-b_2(r-a_3)^2}. \quad [1]$$

They fixed the parameters  $b_0 = 5$ ,  $b_1 = 0.4$ , and  $b_2 = 40$  and optimized  $a_0$ ,  $a_1$ ,  $a_2$ , and  $a_3$  using simulated annealing (23, 24) at zero temperature. We optimize the same set of four free parameters. However, we optimize via AD rather than simulated annealing and work toward a different goal: tuning the rate of assembly at finite temperature while still enforcing the structural constraint that a honeycomb lattice is assembled.

We consider a system of  $N = 100$  particles that interact via Eq. 1 and evolve under overdamped Langevin dynamics at constant temperature (SI Appendix). The blue curve in Fig. 1 shows the assembly dynamics for a specific set of parameters. Specifically, it shows a “honeycomb loss function,”  $L_H$ , that measures the onset of crystallization by identifying the most “honeycomb-like” region, which is a leading indicator of bulk crystallization (SI Appendix).  $L_H \approx 1$  for a randomly arranged system and decreases to  $L_H = 0$  for a perfect honeycomb lattice. Going forward, we will define the crystallization rate  $k_H \equiv 1/t_H$ , where  $t_H$  is the time at which  $L_H$  drops below 0.5 (indicated by the blue diamond in Fig. 1).

The parameters in Fig. 1 were not chosen randomly but are the results of an initial optimization procedure. The starting parameter values input into the optimization procedure match the starting parameters of Rechtsman et al. (23, 24) and do not yield a honeycomb lattice structure on the timescale we consider. To perform the optimization, as explained in SI Appendix, we use AD to calculate the gradient of  $L_H$  after 3,300 simulation time steps. With the gradient in hand, we iteratively minimize  $L_H$  using a standard optimization procedure. Although not targeting a dynamical rate, these results are our first example of AD-based optimization over an MD simulation. Note that this is slightly different than the aim of Rechtsman et al. (23, 24), who were focused on the zero-temperature ground state, and not surprisingly, their results (gray curve in Fig. 1) are not optimal for finite time, finite temperature assembly.

While the system forms a honeycomb lattice at low-volume fractions, the same system with the same interaction parameters will assemble into a triangular lattice at higher-volume fractions. The “triangular loss function,”  $L_T$ , shown in red in Fig. 1



**Fig. 1.** Assembly process for honeycomb and triangular lattices. Two different loss functions are pictured: a honeycomb loss and a triangular loss. The honeycomb loss is shown in blue for the parameters found using AD and in gray for the parameters found by Rechtsman et al. (24), whereas the triangular loss is shown in red. The two corresponding potentials are shown in SI Appendix. Each curve is averaged over 200 independent simulations, while the triangular lattice simulations are performed under a higher density than the honeycomb lattice simulations. The three images on the right-hand side are representative examples of the assembled lattice structures. The bottom two images use the same set of parameters as found using AD but have different volume fractions. The image in the upper left shows a sample random initial configuration. In order to extract assembly rates, we note the time at which the lattices are half-assembled, demarcated with a diamond.

can similarly be used to define a second crystallization rate  $k_T$ . These rates are measured relative to intrinsic timescales that are nontrivially coupled to the potential. To regularize the intrinsic timescale, we seek to control  $k_H$  relative to  $k_T$ .

We use AD over simulations of the assembly process to optimize for our desired assembly dynamics. To understand how we perform this optimization, consider the function

$$\mathcal{S}_{\text{MD}}(a, R, \rho, t), \quad [2]$$

which runs an MD simulation for  $t$  simulation steps at density  $\rho$ , starting from an initial set of positions  $R$ , and with interactions given by Eq. 1 with parameters  $a = \{a_0, a_1, a_2, a_3\}$ . The function returns the final set of positions  $R_t$  after  $t$  simulation time steps. Using this notation, Fig. 1 shows  $L_H(t) \equiv \mathcal{L}_H(\mathcal{S}_{\text{MD}}(a, R_0, \frac{2}{3}\rho_T, t))$  in blue and  $L_T(t) \equiv \mathcal{L}_T(\mathcal{S}_{\text{MD}}(a, R_0, \rho_T, t))$  in red, where  $\mathcal{L}_{H,T}(R)$  are functions that return the respective loss function for a set of positions over time.

The calculation of  $L_H(t)$  [or  $L_T(t)$ ] is analogous to a simple feed-forward neural network with  $t$  hidden layers corresponding to each time step and where  $\mathcal{L}_H$  is applied to the last hidden layer to obtain the output. However, rather than using a set of variable weights and biases to move from one layer to the next, we use the discretized equations of motion to integrate the dynamics. Furthermore, the variables that we train over,  $a$ , are physical parameters (i.e., that define the interaction potential) rather than the millions of weights and biases that typically comprise a neural network. By using AD to propagate the gradient of, for example,  $L_H(t)$  through each simulation step to obtain  $\frac{d}{da} L_H(t)$ , we are able to train a stochastic MD simulation in a way similar to standard neural networks. Importantly, however, this is a data-free approach with “baked-in physics” where results are not only interpretable but correspond to physical parameters.

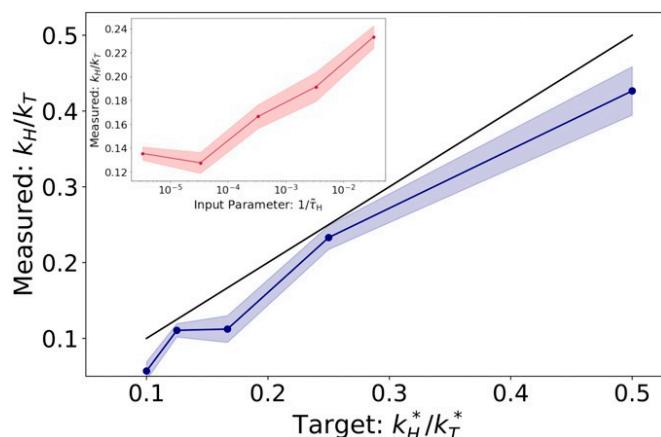
We now seek to find parameters,  $a$ , such that the rate of assembly of the honeycomb lattice at density  $\rho_H$  is  $k_H^*$  and the rate

of assembly of the hexagonal lattice at density  $\rho_T$  is  $k_T^*$ . This is equivalent to minimizing the squared loss,  $L = (L_T(t_T^*) - \frac{1}{2})^2 + (L_H(t_H^*) - \frac{1}{2})^2$ . The gradient  $\frac{dL}{da}$ , which involves differentiating over two ensembles of MD simulations, is used in conjunction with the RMSProp (25) stochastic optimization algorithm to minimize  $L$ . Fig. 2 shows the results of this procedure, where we have fixed  $k_H^* = 0.01$  and varied  $k_T^*$ . While the ratio  $k_H/k_T$  is not identical to the target ratio, we have obtained nontrivial control over the rates using only four parameters.

These results were generated using “forward-mode” AD, in which memory usage is independent of the length of the simulation but computation time scales with the number of parameters. Forward-mode AD quickly becomes extremely time intensive for more complex systems. A more scalable approach is to use “reverse-mode” AD, where the computation time scales favorably with the number of parameters. In the reverse-mode formulation, however, the entire simulation trajectory must be stored in memory, which severely limits the size and length of the simulation. While techniques such as gradient rematerialization (26) can be used to mitigate this, we have developed an alternative strategy (*SI Appendix*) where only the last 300 time steps are differentiated over (Fig. 2, *Inset*). This method provides a more robust solution for large-scale optimization and is similar in spirit to truncated backpropagation used in language modeling (27) and metalearning (28).

### Tuning Transition Rates in Colloidal Clusters

We now turn to another ubiquitous dynamical feature in physical and biological systems. Spontaneous transitions between distinct structural configurations are crucially important in many natural processes from protein folding to allosteric regulation and transmembrane transport, and the rates of these transitions are critical for their respective function. While the prediction of transition rates from energy landscapes is well studied (29, 30), there have been few attempts to control transition rates. This is largely because a general understanding of how changes to interactions affect emergent rates is lacking. Such an understanding, as well as the inverse problem of computing energy landscapes for a specific rate, is essential both for understanding biophysical functionality and for exploring feature space in physical systems.



**Fig. 2.** Ratio of assembly rates as a function of target ratio. Using forward-mode AD, we obtain nontrivial control over assembly rates (black line indicates perfect agreement). *Inset* shows the results using backward-mode AD (*SI Appendix*). The backward-mode results are achieved by adjusting a tuning parameter  $1/\tau_H$ . By changing  $1/\tau_H$ , we obtain indirect control over the assembly rates by differentiating over only the final 300 simulation steps. Although less precise, this approach is more scalable to systems with many parameters.

In this section, we focus on a simple system that exhibits such transitions and can be realized experimentally. Clusters of micrometer-scale colloidal particles are ideal model materials both theoretically and in the laboratory. Furthermore, advances in DNA nanotechnology (31–33) have enabled precise control over binding energies between specified particles by coating the colloids with specific DNA strands. As demonstrated by Hormoz and Brenner (34) and Zeravcic et al. (35) and confirmed experimentally by Collins (36), such control over the binding energies allows for high-yield assembly of specific clusters under thermal noise. The assembly of one stable cluster over the others is obtained by choosing the binding energies  $B_{\alpha\beta}$  between particle pairs  $(\alpha, \beta)$  so that the target structure is the ground state. While there is significant interest in avoiding kinetic traps in order to maximize yield (37), there have been few attempts to control persistent kinetic features of these systems.

We aim to control the rate of transitioning between distinct states. We do this in two ways: first, by simultaneously tuning the energy of the connecting saddle point and the state energies themselves and second, by directly tuning the Kramers approximation for the transition rates, which also takes the curvature of the energy landscape into account. We then consider the question of how many different kinetic features can be tuned simultaneously. To address this, we develop a constraint-based theory analogous to rigidity percolation that predicts when simultaneous control is possible and reveals insight into the nature of the design landscape.

For concreteness going forward, we consider clusters of  $N = 7$  spheres that interact via a short-ranged Morse potential with binding energy  $B_{\alpha\beta}$  and with dynamics given by the overdamped Langevin equation with friction coefficient  $\gamma$  (*SI Appendix*). For positive  $B_{\alpha\beta}$ , Arkus et al. (38) identified all possible stable states up to permutations, each with 15 stabilizing contacts and no internal floppy modes. From these, we pick states that are separated by only a single energy barrier and use the  $\frac{1}{2}N(N-1) = 21$  different  $B_{\alpha\beta}$  as adjustable parameters to control the transition kinetics.

The transition rate from state  $i$  to state  $j$  is

$$k_{ij} = \nu_{ij} e^{-\beta \Delta E_{it}}, \quad [3]$$

where  $\Delta E_{it}$  is the energy barrier, and  $\beta \equiv (k_B T)^{-1}$  with  $k_B$  as the Boltzmann constant and  $T$  as the temperature.  $\nu_{ij}$  is a nontrivial prefactor that we can approximate using Kramers theory (29, 30):

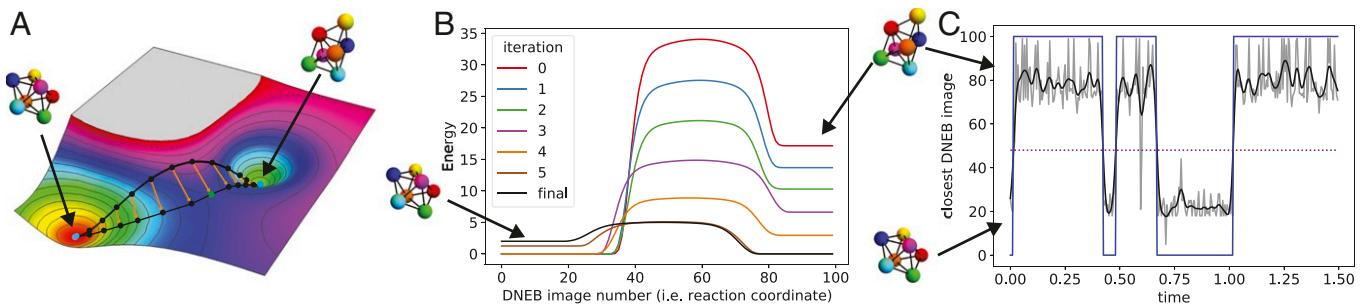
$$\nu_{ij} = \frac{\omega_b}{2\pi\gamma} \frac{\prod_{\ell} \omega_{\ell}^A}{\prod_{\ell'} \omega_{\ell'}^S}, \quad [4]$$

where  $\omega_b$  and  $\{\omega_{\ell}^S\}$  are the frequencies of the unstable and stable vibrational modes, respectively, at the saddle point and  $\{\omega_{\ell}^A\}$  are the frequencies at state  $i$ . We will demonstrate control over the transition kinetics in two ways: 1) by controlling the forward and backward energy barriers  $\Delta E_{it}$  and  $\Delta E_{jt}$  and 2) by controlling directly the forward and backward transition rates  $k_{ij}$  and  $k_{ji}$  using the Kramers approximation.

In order to proceed, we first have to find the transition state separating states  $i$  and  $j$ . This is done using the doubly nudged elastic band (DNEB) method (39), which is illustrated in Fig. 3A and described in *SI Appendix*. The transition state, along with the initial and final states, allows us to calculate the energy barriers as well as the vibrational frequencies necessary for Eq. 4.

With this in hand, we next compute a loss function,  $L$ . When targeting the energy barriers, we use

$$L = (\Delta E_{it} - \Delta E_{it}^*)^2 + (\Delta E_{jt} - \Delta E_{jt}^*)^2, \quad [5]$$



**Fig. 3.** (A) Illustration of the DNEB method (39) for finding the transition state between two local energy minima, which in our case, corresponds to metastable seven-particle clusters. Given such an energy landscape, we construct a series of  $n_t = 100$  images that span the two minima and are connected by high-dimensional springs. Keeping the two end points fixed to the minima, the energy of the ensemble is minimized collectively. The resulting  $n_t$  images give the steepest descent path from one minimum to the other, traversing the saddle point. Thus, the image with the largest energy approximates the saddle point, or transition state, and we can use the image number as a one-dimensional representation of the steepest descent path, or reaction coordinate. However, note that the spacing between images is not imposed, and so, the image number is not a direct measure of distance. *SI Appendix* has more details. (B) Energy (in  $k_BT$ ) along the steepest descent path after successive iterations of the optimization algorithm. For clarity, the minimum energy along the path is subtracted. Note that the regions of constant energy at the beginning and end of the path correspond to global rotations of the cluster. (C) Verification of the transition rates via MD simulation. Periodic snapshots from MD simulations are mapped to the steepest descent path found through the DNEB calculation (gray). Noise is reduced using a Butterworth low-pass filter (black) and then binarized (blue). The dashed line represents the threshold for binarizing the signal and corresponds to the image with the highest energy (i.e.,  $R_t$ ). Transition rates are calculated from the dwell times of the binarized signal.

where  $\Delta E_{it}^*$  and  $\Delta E_{jt}^*$  are the desired energy barriers. When targeting the transition rates directly, we use

$$L = (k_{ij} - k_{ij}^*)^2 + (k_{ji} - k_{ji}^*)^2, \quad [6]$$

where  $k_{ij}$  and  $k_{ji}$  are the Kramers rates and  $k_{ij}^*$  and  $k_{ji}^*$  are the targets. Finally, we optimize  $L$  using reverse-mode AD to calculate the 21 derivatives  $\frac{dL}{dB_{\alpha\beta}} \Big|_{R_i, R_j, R_t}$ , where  $R_{i,j,t}$  are the positions of the particles in the respective states. This gradient is fed into a standard optimization algorithm [RMSProp (25)] to minimize the loss, and  $R_t$  is recalculated only as necessary (in practice, every 50 optimization steps [*SI Appendix*]), which we collectively refer to as a single “iteration”). The entire process (18 iterations) takes roughly 3 minutes on a Tesla P100 GPU.

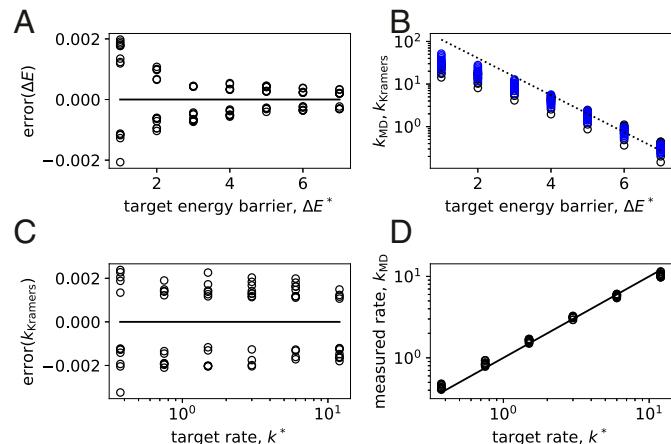
As a first example, we use Eq. 5 to target energy barriers of  $3k_B T$  and  $5k_B T$  between two of the stable  $N = 7$  particle clusters. Fig. 3B shows the energy along the reaction coordinate (i.e., the steepest descent path that connects the two metastable states via the transition state), after various iterations of the above procedure. The final energy differences of  $\Delta E_{it} = 2.998k_B T$  and  $\Delta E_{jt} = 5.001k_B T$  are both within 0.1% of their target and can be further refined by continuing the optimization procedure. Fig. 4A shows the results of 49 distinct target combinations and demonstrates that similar accuracy can be obtained over a wide range of energies.

The connection between the energy barriers and the resulting transition rates is predicted using the Kramers approximation, which is validated via MD simulations where we extract the rates directly (Fig. 3C and *SI Appendix*). Fig. 4B shows both rates ( $k_{\text{Kramers}}$  and  $k_{\text{MD}}$ ) as a function of the target energy barrier. While the validation agrees well with the Kramers prediction, the prefactor  $\nu$  is not constant, underscoring the difficulty of targeting energy barriers as a proxy for transition rates without a quantitative model for  $\nu$ .

An alternative is to target  $k_{\text{Kramers}}$  directly, which we do using the rate-based loss function (Eq. 6). As shown in Fig. 4C, we obtain a typical error of 0.2% after the optimization. As before,  $k_{\text{MD}} \approx k_{\text{Kramers}}$ , and Fig. 4D shows that the full dynamics of the MD simulations agree extremely well with the target dynamics. Thus, we have succeeded in quantitatively designing the transition kinetics of colloidal-like clusters and can do so accurately over a wide range of rates. By optimizing over experimentally

relevant parameters, these results give a direct prediction that can be tested in clusters of DNA-coated colloids.

**Simultaneous Training of Multiple Transitions.** In the above results, transition kinetics are designed while simultaneously imposing specific structural constraints. While most self-assembly approaches focus on obtaining a specific structure, we design for both structure and kinetics simultaneously. Indeed, we have designed for two kinetic features, namely the forward energy barrier/rate and the backward energy barrier/rate. Is it possible to take this further and simultaneously specify many kinetic features? Since particles are distinguishable, permutations



**Fig. 4.** Training energy barriers and transition rates. (A) After 18 iterations of our procedure, the energy barrier  $\Delta E$  is trained to within 0.2% of the target barrier  $\Delta E^*$ . Shown is the error  $[\text{error}(\Delta E) \equiv (\Delta E - \Delta E^*)/\Delta E^*]$  as a function of the target. (B) The transition rates measured via MD simulation ( $k_{\text{MD}}$ ; black data) are not exactly proportional to  $e^{-\beta \Delta E^*}$  (the dashed line has a slope of  $-\beta$ ), indicating that the prefactor  $\nu$  in Eq. 3 is not constant. However, this prefactor is captured reasonably well by the Kramers approximation ( $k_{\text{Kramers}}$ ; blue data). (C) After 18 iterations of training on the Kramers rate  $k_{\text{Kramers}}$ , we obtain an accuracy within 0.2% of the target rate  $k^*$ . (D) The rates measured via MD simulation  $k_{\text{MD}}$  agree very well with the target rate  $k^*$  (the solid line corresponds to  $k_{\text{MD}} = k^*$ ). Thus, we are able to accurately and quantitatively design the transition kinetics of the seven-particle clusters we consider.

represent distinct states, and thus, there are many thousands of possible transitions in our seven-particle clusters. What determines which subsets of transitions can be simultaneously and independently controlled?

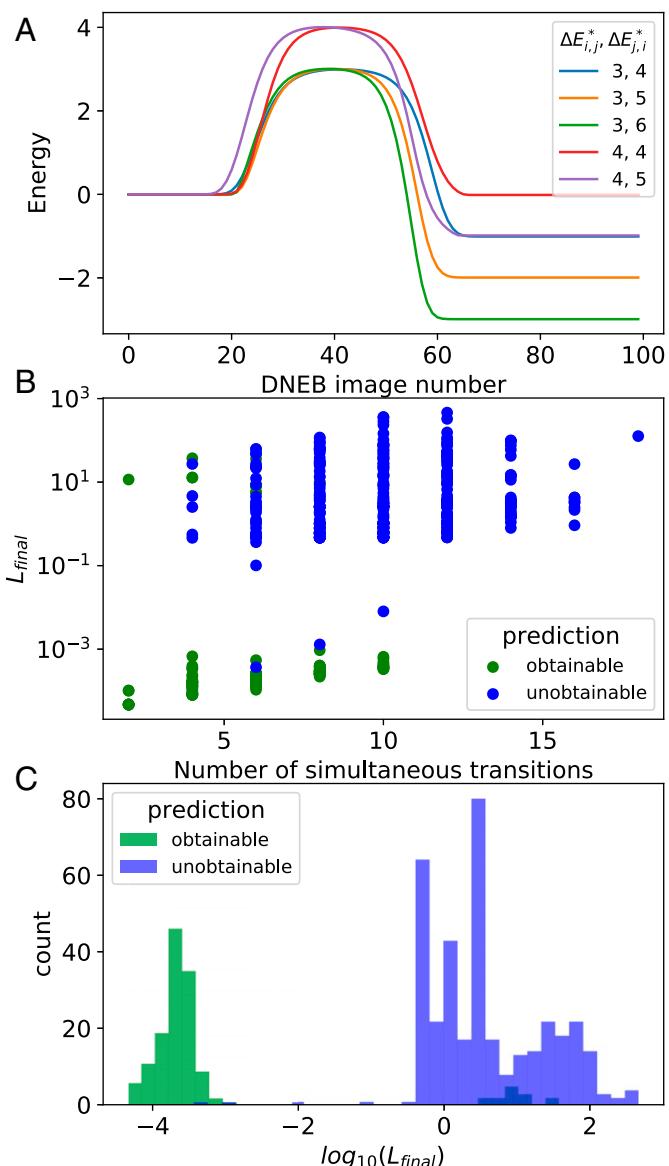
To address this question, we develop an analogy to network rigidity. Each kinetic feature we would like to impose appears as a constraint in the loss function. If we want to impose  $x$  kinetic features, then we attempt to minimize a loss function analogous to Eq. 5 with  $x$  terms of the form  $(\Delta E - \Delta E^*)^2$ .<sup>4</sup> This constraint satisfaction problem has a strong analogy to network rigidity (40), where physical springs constrain the relative positions of nodes through an interaction proportional to  $(r - r_0)^2$ , with  $r$  as the distance between particles and  $r_0$  as the rest length of the spring. When attempting to place a spring, we can look to the eigenvalues of the Hessian matrix of the total energy to determine whether the new constraint can be satisfied. Each “zero mode” (eigenvector of the Hessian with zero eigenvalue) represents a degree of freedom that can be adjusted slightly while maintaining all of the constraints. The relevant question is whether the new constraint can be satisfied with only these free degrees of freedom. If this is the case, then imposing the constraint causes the system to lose a zero mode. However, if the additional constraint cannot be satisfied simultaneously with the other constraints, then the number of zero modes remains the same, and the constraint is called “redundant.” Thus, if the number of nonzero modes equals the number of constraints, than all of the constraints can be satisfied simultaneously.

We employ this same approach in our system to determine if a set of constraints (desired kinetic features) can be simultaneously satisfied by a given set of variables (the 21 binding energies  $B_{\alpha\beta}$ ). To proceed with a given set of desired kinetic features, we start with a random initial set of binding energies and write down a temporary loss function where the targets are equal to the current energy barriers. This is analogous to placing a spring between two nodes such that the spring’s rest length is equal to the separation between the nodes. It also guarantees that the Hessian matrix  $\frac{d^2 L}{dB_{\alpha\beta} dB_{\alpha'\beta'}}$  is positive semidefinite and that the zero modes correspond to unconstrained degrees of freedom. We employ AD to calculate the Hessian matrix, which we then diagonalize and count the number of zero and nonzero modes. If the number of nonzero modes equals the number of constraints, then we predict that the desired set of kinetic features can be simultaneously obtained.

To test this prediction, we have picked an initial structure and nine adjacent structures, with the intention to simultaneously specify the energy barriers of various subsets of the nine forward and nine backward transitions. As a first example, we pick a set of 10 transitions (5 forward and 5 backward) where the Hessian has 10 nonzero modes, meaning our theory predicts that our optimization should be successful. This is confirmed in Fig. 5A, which shows the energy along the five transition pathways after simultaneously optimizing for all 10 barriers. The legend shows the target energy barriers,  $\Delta E_{ij}^*$  and  $\Delta E_{ji}^*$ , for both the forward and backward transitions, respectively. Each observed barrier is within 0.3% of their respective target and becomes more accurate with additional training using smaller training rates.

This result is generalized in Fig. 5B and C, where we show the loss function  $L_{\text{final}}$  after optimization for all 511 subsets of the nine transition pathways. Cases where we predict that the kinetic features can be simultaneously obtained (i.e., that  $L_{\text{final}} \rightarrow 0$ ) are shown in green, while cases where we predict that the kinetic features cannot be simultaneously obtained are shown in blue.

<sup>4</sup>Or analogous to Eq. 6 with  $x$  terms of the form  $(k - k^*)^2$ . For simplicity going forward, we will focus on imposing kinetic features via energy barriers.



**Fig. 5.** (A) Energy (in  $k_B T$ ) along the steepest descent paths connecting a single starting structure with five adjacent structures, obtained through the DNEB method (Fig. 3 and *SI Appendix*). These results are after the simultaneous optimization of all 10 energy barriers (target energy barriers [in  $k_B T$ ] are shown in the legend). The observed energy barriers are within 0.3% of their targets. (B) The loss function after attempting to simultaneously control the dynamics of multiple transitions, as a function of the number of transitions. Due to the form of the loss function (Eq. 5), simultaneous control is successful when  $L_{\text{final}} \rightarrow 0$ . In practice, when we obtain  $L_{\text{final}} < 10^{-2}$ , this can be further reduced by additional optimization with smaller learning rates, indicating that the transition dynamics can indeed be controlled to a desired precision. The data are colored according to the Hessian-based prediction discussed in the text: green (blue) data correspond to cases where we predicted that simultaneous control could be (could not be) obtained. (C) Histogram of  $L_{\text{final}}$  for all data shown in B.

While our predictions are not perfect, we find that they are correct roughly 97% of the time.

A key difference between our system and network rigidity is that we are trying to predict a highly nonlocal feature of the function space, namely whether or not a loss function can be minimized to zero. Conversely, rigidity in elastic networks is a local property of the energy landscape. We extrapolate from the local curvature to predict highly nonlocal behavior, without proper

justification. The fact that this nevertheless works quite well indicates that there is indeed some persistent internal structure in the function space we are considering, despite its considerable complexity. We leave a more in-depth examination of this structure to future studies.

## Discussion

Rather than being limited to structural motifs, our results expand the design space of colloids with specific interactions to include complex kinetics. Understanding the connection between interactions and emergent dynamical properties is hugely important in biological physics, materials science, and nonequilibrium statistical mechanics. Moreover, achieving in synthetic materials the same level of complexity and functionality as found in biology requires detailed control over kinetic features.

We demonstrate quantitative control over kinetic features in two self-assembly systems. In the classic problem of assembling a honeycomb lattice using monotonic, spherically symmetric particles, we achieve nontrivial control over lattice assembly rates. Understanding and controlling crystallization kinetics is critical to the design of many real materials and is intimately related to the statistics of defects. Furthermore, since a honeycomb crystal is a diatomic lattice where neighboring particles have different orientational order, assembling a honeycomb lattice out of identical particles is a notoriously difficult problem. Our ability not only to assemble such a lattice but to control its assembly rate relative to a triangular lattice reveals an essential connection between assembly dynamics and particle interactions that we are able to exploit for kinetic design.

Second, by considering small clusters of colloidal-like particles with specific interactions, we discover that transition kinetics are far more designable than was previously thought. Since spherically symmetric particles provide a foundational understanding not only in self-assembly but in fields throughout physics, our results suggest the same potential for designing transition dynamics in numerous other systems.

Moreover, the methodology we present is applicable beyond the study of colloidal kinetics. At its core, this method directly extracts the effect of experimental or model parameters on an emergent property, such as the kinetic rates that we focus on. It does so by measuring the derivative of the emergent property with respect to the underlying parameters. Unlike finite difference approaches to taking derivatives, these calculations

are both efficient and exact, enabling, for example, the Hessian calculation that is essential to our analysis in *Simultaneous Training of Multiple Transitions*. We have shown that despite the stochasticity inherent to the thermal systems we consider, these derivatives are predictive. In other words, the validity of the linear regime accessed by the gradient calculation is not overwhelmed by stochastic noise. Thus, we can cut through the highly complex dependence of emergent dynamical features on model parameters (such as interaction energies, size distributions, temperature and pressure schedules, or even particle shapes) and use this dependence to control behavior.

Thus, we have shown that gradient computations over statistical physics calculations using AD are possible, efficient, and sufficiently well behaved for optimization of kinetic features. This approach is complimentary to other inverse approaches, such as “on-the-fly” optimization methods (41) where iterative simulation results are used to update parameter values at each step in order to suppress competitor structures and promote the desired structure. However, unlike for structural features, it is difficult to intuit the connection between interactions and kinetics. By revealing this connection, the differentiable statistical physics models presented here provide a framework for understanding emergent behavior. Indeed, our work merely scratches the surface of what can be achieved using the principles of kinetic design. This approach scales favorably with the number of input parameters, enabling the design of increasingly complex phenomena. We envision engineering materials with emergent properties, exploring the vast search space of biomimetic function and manipulating entire phase diagrams.

## Materials and Methods

MD simulations and AD calculations are performed using a fully differentiable MD package, called JAX MD (22), that has recently been developed by two of us and has hardware acceleration and ensemble vectorization built in. *SI Appendix* has more information on our methods and procedures.

**Data Availability.** All study data are included in the article and/or *SI Appendix*.

**ACKNOWLEDGMENTS.** We thank Agnese Curatolo, Megan Engel, Ofer Kimchi, Seong Ho Pahng, and Roy Frostig for helpful discussions. This material is based on work supported by NSF Graduate Research Fellowship Grant DGE1745303. This research was funded by NSF Grant DMS-1715477, Materials Research Science and Engineering Centers Grant DMR-1420570, and Office of Naval Research Grant N00014-17-1-3029. M.P.B. is an investigator of the Simons Foundation.

1. J. D. Perlmutter, M. F. Hagan, Mechanisms of virus assembly. *Annu. Rev. Phys. Chem.* **66**, 217–239 (2015).
2. L. M. Alexander, D. H. Goldman, L. M. Wee, C. Bustamante, Non-equilibrium dynamics of a nascent polypeptide during translation suppress its misfolding. *Nat. Commun.* **10**, 2709 (2019).
3. P. Tan, N. Xu, L. Xu, Visualizing kinetic pathways of homogeneous nucleation in colloidal crystallization. *Nat. Phys.* **10**, 73–79 (2014).
4. K. E. Dunn *et al.*, Guiding the folding pathway of DNA origami. *Nature* **525**, 82–86 (2015).
5. G. M. Whitesides, J. P. Mathias, C. T. Seto, Molecular self-assembly and nanochemistry: A chemical strategy for the synthesis of nanostructures. *Science* **254**, 1312–1319 (1991).
6. T. Douglas, M. Young, Host-guest encapsulation of materials by assembled virus protein cages. *Nature* **393**, 152–155 (1998).
7. S. I. Stupp *et al.*, (1997) Supramolecular materials: Self-organized nanostructures. *Science* **276**, 384–389.
8. B. G. Lohmeijer, U. S. Schubert, Supramolecular engineering with macromolecules: An alternative concept for block copolymers. *Angew. Chem. Int. Ed.* **41**, 3825–3829 (2002).
9. J. D. Hartgerink, J. R. Granda, R. A. Milligan, M. R. Ghadiri, Self-assembling peptide nanotubes. *J. Am. Chem. Soc.* **118**, 43–50 (1996).
10. A. E. Bryson, W. F. Denham, A steepest-ascent method for solving optimum programming problems. *J. Appl. Mech.* **29**, 247–257 (1962).
11. A. Bryson, *Applied Optimal Control: Optimization, Estimation and Control* (Taylor & Francis, 1975).
12. S. Linnainmaa, Taylor expansion of the accumulated rounding error. *BIT Numer. Math.* **16**, 146–160 (1976).
13. D. E. Rumelhart, G. E. Hinton, R. J. Williams, Learning representations by back-propagating errors. *Nature* **323**, 533–536 (1986).
14. A. G. Baydin, B. A. Pearlmutter, A. A. Radul, J. M. Siskind, Automatic differentiation in machine learning: A survey. *J. Mach. Learn. Res.* **18**, 5595–5637 (2017).
15. A. Paszke *et al.*, “Pytorch: An imperative style, high-performance deep learning library” in *Advances in Neural Information Processing Systems 32*, H. Wallach *et al.*, Eds. (Curran Associates, Inc., 2019), pp. 8024–8035.
16. M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, Version 2. <https://www.tensorflow.org/>. Accessed 1 September 2020.
17. J. Bradbury *et al.*, JAX: Composable transformations of Python+NumPy programs, Version 0.2.5 (2018). <https://github.com/google/jax#citing-jax>. Accessed 22 February 2021.
18. M. Minkov *et al.*, Inverse design of photonic crystals through automatic differentiation. *ACS Photonics* **7**, 1729–1741 (2020).
19. Y. Bar-Sinai, S. Hoyer, J. Hickey, M. P. Brenner, Learning data-driven discretizations for partial differential equations. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15344–15349 (2019).
20. K. McCloskey *et al.*, Machine learning on DNA-encoded libraries: A new paradigm for hit-finding. [arXiv:2002.02530](https://arxiv.org/abs/2002.02530) (31 January 2020).
21. A. W. Senior *et al.*, Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).
22. S. S. Schoenholz, E. D. Cubuk, JAX, M.D.: End-to-end differentiable, hardware accelerated, molecular dynamics in pure python. [arXiv:1912.04232v1](https://arxiv.org/abs/1912.04232v1) (9 December 2019).
23. M. Rechtsman, F. Stillinger, S. Torquato, Designed interaction potentials via inverse methods for self-assembly. *Phys. Rev.* **73**, 011406 (2006).

24. M. C. Rechtsman, F. H. Stillinger, S. Torquato, Optimized interactions for targeted self-assembly: Application to a honeycomb lattice. *Phys. Rev. Lett.* **95**, 228301 (2005).

25. T. Tieleman, G. Hinton, Lecture 6.5-rmsprop: Divide the gradient by a running average of its recent magnitude. *COURSERA: Neural Networks Mach. Learn.* **4**, 26–31 (2012).

26. A. Griewank, A. Walther, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation. *ACM Trans. Math Software* **26**, 19–45 (2000).

27. T. Mikolov, M. Karafiat, L. Burget, J. Černock, S. Khudanpur, “Recurrent neural network based language model” in *Eleventh Annual Conference of the International Speech Communication Association (ISCA) [International Speech Communication Association]*, 2010, pp. 1045–1048.

28. M. Al-Shedivat *et al.*, Continuous adaptation via meta-learning in nonstationary and competitive environments. arXiv:1710.03641v1 (10 October 2017).

29. H. A. Kramers, Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7**, 284–304 (1940).

30. P. Hanggi, P. Talkner, M. Borkovec, Reaction-rate theory: Fifty years after Kramers. *Rev. Mod. Phys.* **62**, 251–341 (1990).

31. C. A. Mirkin, R. L. Letsinger, R. C. Mucic, J. J. Storhoff, A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607–609 (1996).

32. M. P. Valignat, O. Theodoly, J. C. Crocker, W. B. Russel, P. M. Chaikin, Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 4225–4229 (2005).

33. P. L. Biancaniello, A. J. Kim, J. C. Crocker, Colloidal interactions and self-assembly using DNA hybridization. *Phys. Rev. Lett.* **94**, 058302 (2005).

34. S. Hormoz, M. P. Brenner, Design principles for self-assembly with short-range interactions. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 5193–5198 (2011).

35. Z. Zeravcic, V. N. Manoharan, M. P. Brenner, Size limits of self-assembled colloidal structures made using specific interactions. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 15918–15923 (2014).

36. J. Collins, “Self-assembly of colloidal spheres with specific interactions,” PhD thesis, Harvard University, Cambridge, MA (2014).

37. K. E. Dunn *et al.*, Guiding the folding pathway of DNA origami. *Nature* **525**, 82–86 (2015).

38. N. Arkus, V. N. Manoharan, M. P. Brenner, Deriving finite sphere packings. *SIAM J. Discrete Math.* **25**, 1860–1901 (2011).

39. S. A. Trygubenko, D. J. Wales, A doubly nudged elastic band method for finding transition states. *J. Chem. Phys.* **120**, 2082–2094 (2004).

40. D. J. Jacobs, M. F. Thorpe, Generic rigidity percolation: The pebble game. *Phys. Rev. Lett.* **75**, 4051–4054 (1995).

41. Z. M. Sherman, M. P. Howard, B. A. Lindquist, R. B. Jadrach, T. M. Truskett, Inverse methods for design of soft materials. *J. Chem. Phys.* **152**, 140902 (2020).