

Proofreading mechanism for colloidal self-assembly

Qian-Ze Zhu¹, Chrisy Xiyu Du^{1,2}, Ella M. King³, and Michael P. Brenner^{1,3}¹*School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02139, USA*²*Mechanical Engineering, University of Hawai'i at Mānoa, Honolulu, Hawaii 96822, USA*³*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 16 December 2023; accepted 6 November 2024; published 9 December 2024)

Designing components that can robustly self-assemble into structures with biological complexity is a grand challenge for material science. Proofreading and error correction is required to improve assembly yield beyond equilibrium limits, using energy to avoid kinetic traps in the energy landscape. Here, we introduce an explicit two-staged proofreading scheme for patchy particle colloidal assemblies that substantially improves assembly yield and robustness. The first stage implements local rules whereby particles increase their binding strengths when they detect a local environment corresponding to a desired target. The second stage corrects remaining errors, adding a reverse pathway inspired by kinetic proofreading. The scheme shows significant yield improvements, eliminating kinetic traps, giving a much broader temperature range with high yield. Additionally, the scheme is robust against quenched disorder in the components. Our findings illuminate a pathway for advancing the programmable design of synthetic living materials, potentially fostering the synthesis of novel biological materials and functional behaviors.

DOI: [10.1103/PhysRevResearch.6.L042057](https://doi.org/10.1103/PhysRevResearch.6.L042057)

Living materials exhibit more complex structures and dynamics than those we create synthetically. For example, enzymes exquisitely regulate metabolic pathways [1]; microtubules and virus shells self-assemble and disassemble on cue [2–5]; in DNA proofreading, polymerase reduces binding errors well below the equilibrium yield [6]. A longstanding goal is to create synthetic materials with this level of complexity. An essential component of a strategy uses programmable building blocks, with examples ranging from DNA nanotechnology [7–9], to biological assemblies such as clathrin, where triskelia are programed to interact for emergent behaviors [10]. In all of these cases, building blocks are programed to achieved properties that far exceed functionality of a random set. Yet even with programmable components, there are limits to the complexity of the emergent structure [11–13], with random errors corrupting yield [9].

Traditional colloidal self-assembly processes passively follow equilibrium dynamics, where the lack of active error correction leads to low yields and kinetic traps [14]. To create more robust assembly pathways, we hypothesized that it is necessary to spend energy to proofread errors during assembly, and aimed to construct a simple system to test and develop proofreading strategies. Since enzymatic activity requires bond directionality, isotropic interactions [15,16] are insufficient, so we focused on assembly with patchy particles [see Fig. 1(a)], powerful enough to capture the complex properties yet simple enough to design [17]. In our scheme,

the interaction strength and range of each patch is programed to interact with other patches. Our model for proofreading is inspired by microtubule growth [2,3] where bonds strengthen or weaken depending on their local environment through energy-dependent processes (such as ATP/GTP hydrolysis). To capture this, we give each particle a *state*, dictating how the patches interact with other particles. Each particle has an identical initial state; yet as assembly proceeds, particles measure their local neighborhood and correspondingly change state, analogous to allostery [18]. State change costs energy, since changing state gives patches different interaction strengths, ranges, and (potentially) locations. Our key question is whether it is possible to program state changes to improve assembly yield.

For a given structure, we aim to find a series of transitions of patches that optimize assembly yield. The design space is enormous. To find the optimal parameters, we leverage automatic differentiation methods [19], the algorithmic underpinning of modern machine learning (ML). Parameter optimization during training is effectively analogous to natural evolution. Previous research has optimized interaction strengths and patch locations for equilibrium assembly *without state change*, demonstrating the ability to efficiently find patch locations and strengths for designing nontrivial structures [20]. The optimization scheme with state change is much more difficult, given the need to find the logical rules for decision making.

Inspired by microtubule growth, we aim to develop rules to robustly grow a two-dimensional ladderlike structure. Each individual patchy particle consists of a central particle interacting via a soft sphere potential and three patches interacting via a Morse potential [Fig. 1(a)]. We carry out a molecular dynamics simulation under constant temperature and

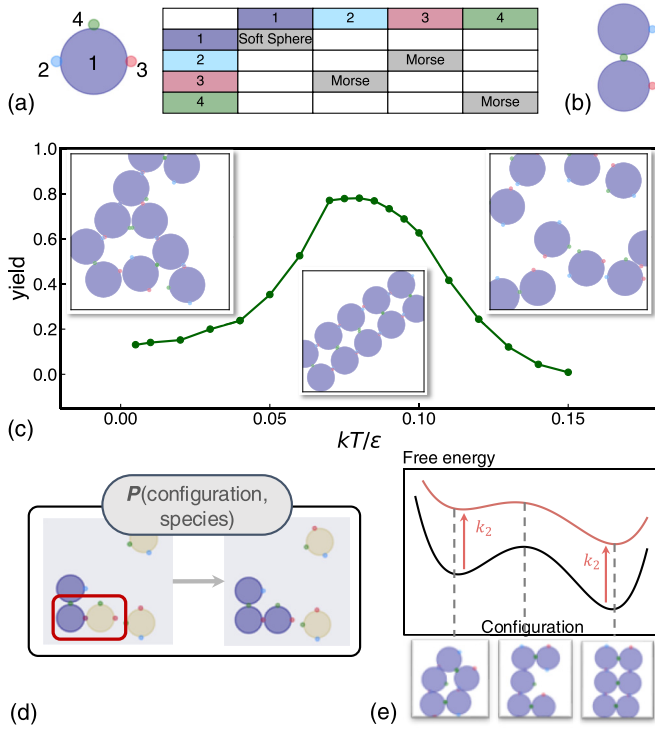


FIG. 1. System and two-stage proofreading mechanism. (a) Left: Individual patchy particle consisting of a central sphere (1: purple) and three distinct patches (2: blue; 3: red; 4: green). Right: Interaction matrix illustrating intercomponent interactions. Gray squares represent a soft sphere or Morse potential as specified. Blank squares denote no interactions. (b) Schematic illustration of the seed particle. (c) Yield curve for a conventional monospecies system. The x axis is the temperature $k_B T$ rescaled by the interaction potential energy ϵ . Inset: Exemplary final configurations of 40 000-step forward simulations in a (left) kinetic regime, (middle) optimal regime, and (right) entropically favored monomer regime. (d) Illustration of the state change mechanism. Yellow represents weaker interactions, and purple represents stronger interactions. Only correct bonds (highlighted by the red box) are reinforced. (e) Illustration of the full proofreading mechanism via the engineer of free-energy landscape. Insets display (left) a local defect, (middle) the configuration at the energy barrier, and (right) part of a correctly assembled structure.

volume condition with periodic boundary condition [see Supplemental Material (SM) [21] and an accompanying PYTHON notebook [22] for details]. We begin the assembly with a seed consisting of two rigidly connected particles, with valid contacts only in one direction [Fig. 1(b)]. Figure 1(c) shows the yield curve for assembly of this structure in a system with ten particles. Yield is defined as the ratio of the number of error-free assemblies with size 10 to the total number of assemblies (see SM and the PYTHON notebook [22] for a detailed implementation of the yield function). In the low-temperature regime (left), the structures are kinetically trapped, while in the high-temperature regime (right) dissociated monomers are entropically favored. The maximum yield is around 80%.

We design the proofreading scheme to inhibit errors from occurring because of local minima, whereby particles form a ring with a smaller number of contacts than the dense ladder [see the left inset of Fig. 1(e)]. To inhibit this, particles must

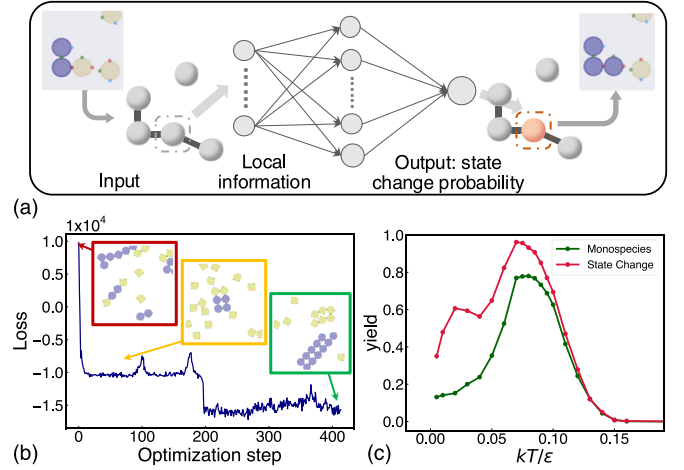


FIG. 2. State change. (a) Schematic illustration of GNN as the update rule for state change. (b) The loss trajectory during GNN training. Inset: 100 000-step forward simulation using the indicated parameters at optimization step 0 (red), 50 (yellow), and 400 (green). (c) Yield curves for (green) conventional monospecies and (red) state change systems. The x axis is temperature rescaled by the dominant interaction energy scale. Note for the state change system, the interaction energy scale refers to the strong bond strength E_2 .

strengthen their bonds when the *local environment* meets desired criteria. This process is depicted in Fig. 1(d): The initial state of the particle (yellow, species 1) has weak interactions E_1 . When the correct bonds occur in its local environment [red box in Fig. 1(d)], it changes state (purple, species 2) strengthening interactions to E_2 . Errors still occur after this stage because of mismatches between different local environments. We therefore introduce a second stage, inspired by kinetic proofreading [23], whereby we add a reverse pathway to decrease the bond strength of randomly chosen particles. This modulates the free-energy landscape to favor the transition from local minima to the target structure [Fig. 1(e)].

We demonstrate in what follows with properly chosen transition rules, this two-stage proofreading mechanism substantially increases the yield of the desired structures over a much broader temperature range.

State change. In this first stage of our proofreading scheme, we need a local update rule for each particle that determines whether the local environment matches the desired target. The space of local rules is enormous and choosing the correct rule is critical: With rules that are either too permissive or too stringent, assembly yield degrades (see SM). We search local rules by parametrizing the state change probability as a general function of the geometry of the local environment. It is convenient to use a graph neural network (GNN) [24], encoding the local environment into graph nodes and edge embeddings, with the node encoding neighboring particle states, whereas the edge includes pairwise distances between particle centers and patches [Fig. 2(a)]. GNNs are particularly amenable to training. We choose a cutoff range of three particle radii, covering nearest neighbors and next neighbors in the target structure. We train the GNN parameters to improve the target yield, using automatic differentiation to back-propagate the gradient of a loss from a batch of forward molecular

dynamics simulations. Details regarding the specific architecture of GNN can be found in SM. We also included a detailed implementation of the GNN we used in the accompanying PYTHON notebook [22]. We emphasize that the specific GNN architecture is not essential to our proofreading scheme. The GNN serves merely as a general function representation, allowing us to encode an arbitrary function of a particle and its neighbors. The GNNs implemented here have 79 parameters in total, which we denoted by $\hat{\theta}$. To optimize these 79 parameters, we use JAX-MD, an end-to-end differentiable and GPU-accelerated molecular dynamics (MD) engine [25]. We measure the performance of the update rule during training with a loss function indicating how far the final structure maps to the target, $\mathcal{L} = \mathcal{L}(s_1, s_2, \dots, s_N; \hat{\theta})$, where s_i represents the state of particle i , including position, orientation, species, and other information. The loss function has two terms (see SM), one penalizing deviations from the compact ladder structure, with the other encouraging elongation of the structure.

Since state change is an intrinsically discrete process, to ensure differentiability, we adapt the REINFORCE algorithm (also referred to as the score function) [26]. Given the probability $\rho[s_i(t); \hat{\theta}]$ that particle i goes through state change, the objective of the optimization is an ensemble average of the loss function at the final simulation step t , namely

$$\langle \mathcal{L} \rangle = \int_{\text{configuration space}} \mathcal{D}\mathbf{s}(t) \mathcal{L}(s_1(t), \dots, s_N(t); \hat{\theta}) \times P[s_1(t), \dots, s_N(t); \hat{\theta}], \quad (1)$$

where $P[s_1(t), \dots, s_N(t); \hat{\theta}]$ is the joint probability of the final configuration of the whole system $[s_1(t), \dots, s_N(t)]$ with the current parameters $\hat{\theta}$,

$$P[s_1(t), \dots, s_N(t); \hat{\theta}] = \prod_{\tau=0}^t \prod_{i=1}^N \rho[s_i(\tau); \hat{\theta}]. \quad (2)$$

Using the REINFORCE algorithm, we can take gradients of the loss function with respect to the parameter $\hat{\theta}$,

$$\begin{aligned} \nabla_{\hat{\theta}} \langle \mathcal{L} \rangle &= \int \mathcal{D}\mathbf{s}(t) \nabla_{\hat{\theta}} [\mathcal{L}(\mathbf{s}(t); \hat{\theta}) P(\mathbf{s}(t), \hat{\theta})] \\ &= \int \mathcal{D}\mathbf{s}(t) P(\nabla_{\hat{\theta}} \log P) \mathcal{L} + \int \mathcal{D}\mathbf{s}(t) P \nabla_{\hat{\theta}} \mathcal{L} \\ &= \langle (\nabla_{\hat{\theta}} \log P) \mathcal{L} \rangle + \langle \nabla_{\hat{\theta}} \mathcal{L} \rangle, \end{aligned} \quad (3)$$

where the first term is the REINFORCE gradient, and the second term is the conventional gradient.

We optimize state change parameters with a system consisting of 21 particles at temperature $k_B T = 1.0$, with weak interaction $E_1 = 4.0$, and strong interaction $E_2 = 50.0$. Following Ref. [20], we use Adam optimizer to run 400 optimization steps with a concatenated learning rate scheme of (0.005, 0.001), where each optimization step runs 512 replicates of 41 000-step forward simulation, and back-propagates the loss over the last 1000 simulation steps. With these simplifications, we can store the trajectories in memory on an NVIDIA A100 Tensor Core GPU with 80 GB of RAM. The basic construction of the mechanism and optimization of the GNN update rule are included in the accompanying PYTHON notebook [22].

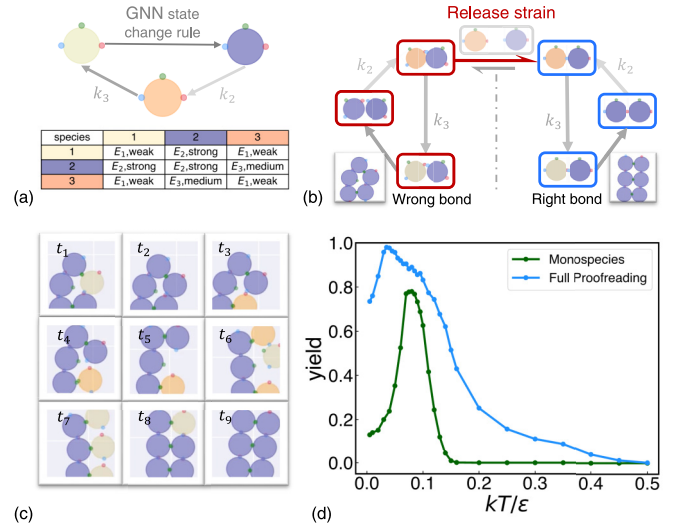


FIG. 3. Full proofreading mechanism. (a) Schematic illustration of the full proofreading mechanism, including both the state change and the reverse pathway. Lower panel: Interaction matrix between the three particle species. (b) Step-by-step illustration of the proofreading dynamics. (c) Snapshots of the proofreading dynamics in an increasing-time order. (d) Yield curve for the proofreading mechanism compared with the monospecies assembly. The x axis is temperature rescaled by the dominant interaction energy scale. Note for proofreading mechanism, the interaction energy scale refers to the medium bond strength E_3 .

We find that the learning rule does not converge when initialized with random initial parameters. Instead we pretrain the model by initializing it with logical rules for the desired local structure (see SM), which allows optimization to rapidly converge [Fig. 2(b)]. The insets show the final configurations of 100 000-step forward simulations at different optimization stages. Curiously, the training dynamics exhibits an initial plateau in the loss trajectory before converging to its final state. This indicates that the system is initially trapped in a local minimum, which inhibits growth [see the yellow inset of Fig. 2(b)]. It then makes a few attempts to jump out of the local minimum (intermittent spikes), and finally converges to an optimal growth rule.

Figure 2(c) shows the yield curve for the state change scheme compared with the monospecies assembly. To fairly compare with Fig. 1(c), we rescale the temperature with respect to the strong interaction energy E_2 after state change. The state change scheme greatly improves the yield, especially in the regime where structures are kinetically trapped.

Reverse pathway. Yet, the state change scheme still leaves errors. Since state change causes bonds to become irreversible, these errors are frozen in (see SM). We find two types of errors: Either free-floating particles accidentally undergo a state change, or dislocations form due to incommensurate local rules [Fig. 3(b) left panel].

The second stage of our proofreading scheme aims to fix these issues by adding a reverse pathway to weaken bonds. Inspired by kinetic proofreading [23], we randomly choose particles to transition to an intermediate species 3 with moderate bond strength $E_3 = 15.0$. The transition dynamics and

interaction matrix between these three species is shown in Fig. 3(a). We choose the probability of a species 2 (purple) particle to transition into species 3 (orange) particle to be $k_2 = 2 \times 10^{-5}$, and the probability of a species 3 (orange) particle to transition into species 1 (yellow) particle is $k_3 = 2 \times 10^{-4}$. This reverse process is identical for all particles in the structure, and does not depend on whether the bonding environment is correct.

We reason that if a particle's bonding environment is correct, it will quickly reform its original structure through the state change mechanism [the right loop of Fig. 3(b)]. If the bonding environment is in a local minimum, it has a chance to correct itself, breaking an incorrect bond, and ultimately stabilizes in the correct structure through the state change mechanism, as shown in Fig. 3(b). Snapshots in Fig. 3(c) and the accompanying Supplemental Material videos [21] clearly show the error correcting process. The reverse pathway effectively modulates the free-energy landscape, elevating the activation barrier for escaping local minima to a level that is comparable to the thermal energy ($\Delta E_{LM} \lesssim k_B T$), while the energy depth of the ground state remains significantly greater than the thermal energy ($\Delta E_{GS} \gg k_B T$), as illustrated in Fig. 1(c). This results in a net transition from the wrong bonds to the right bonds, thereby substantially improving the yield.

Figure 3(d) shows the yield as a function of rescaled temperature for the full two-stage proofreading scheme compared with the conventional monospecies assembly. For a fair comparison, we again rescale the temperature with respect to the dominant energy scale, which is the intermediate interaction E_3 for proofreading. The proofreading scheme not only greatly improves the yield, but also greatly broadens the optimal temperature range where the yield is relatively high.

Quenched disorder. In addition to thermal noise, there may also be intrinsic disorder in the patch geometry. Is our proofreading scheme robust to such errors? To model this quenched disorder, we perturb the location of the patches in individual components away from their optimal (90° angle) geometry. For every patch ($\alpha = 1, 2, 3$) on every particle ($i = 1, \dots, N$) we move the angle by $\varepsilon_{i\alpha}$ from its regular location, where $\{\varepsilon_{i\alpha}\}$ are drawn from a uniform distribution bounded by the magnitude of the quenched disorder. Note that in this scheme every particle in the simulation is manufactured slightly differently. Here, the magnitude of disorder is set to be 0.2 (radian). Figure 4(a) illustrates two examples of such disordered particles. Figure 4(b) shows a correctly assembled structure using these disordered particles. Disorder makes it substantially less likely to form error-free structures.

We use the identical proofreading scheme described above, without retraining the parameters and again calculate the yield curves for each of the three cases (conventional monospecies, state change, full two-stage proofreading) with the disordered components. To fairly compare the yield with intrinsic disorder in the system, we loosen the threshold in yield definition (see SM). Figure 4(c) shows the yield for the conventional monospecies assembly (green curve) decreases substantially with quenched disorder. Both the state change (red curve) and the full proofreading (blue curve) recover high yields, demonstrating the robustness of the proofreading mechanism.

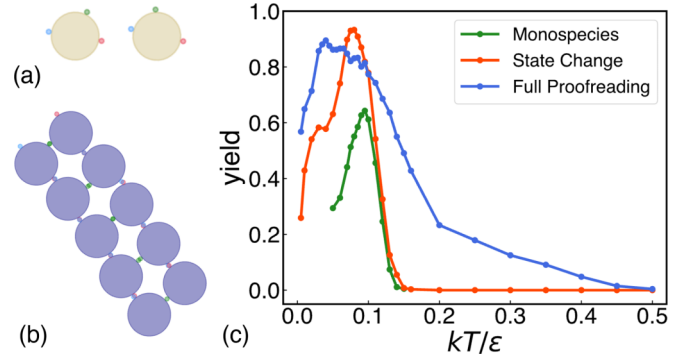


FIG. 4. Quenched disorder. (a) Examples of individual particles with quenched disorder. (b) A correctly assembled structure with quenched disorder. (c) Yield curve for (green) conventional monospecies, (red) state change mechanism, and (blue) full proofreading mechanism. The x axis is temperature rescaled by the dominant interaction energy scales.

Discussion. Here, we have demonstrated that it is possible to program proofreading mechanisms to substantially improve yield during self-assembly. Although we have studied a particular example of colloidal assembly, the principles uncovered here are general with much broad applicability. Our basic idea is conceptually simple: Every component can validate its local environment, and when the environment is correct, it undergoes a state change to increase bond strengths. Such conformational changes usually require energy expenditure and are common in biology. One example is allosteric interactions in proteins [27,28], where the binding of a molecule at a specific site dramatically changes the configuration of the protein. Our model also closely parallels the functioning of DNA replication machineries in eukaryotic cells. Stable nucleotide addition to new DNA strands by polymerase action is mirrored in our state change mechanism, while our proofreading mechanism implements the ability to sense structural inconsistencies and remove incorrect bonds, resembling the MutSα-EXO1 pathway in eukaryotes [29]. Our model requires substantial tuning to work effectively: We accomplished this by parametrizing the decision rule with a neural network and then directly optimizing it through molecular dynamics simulation of the assembly process. The optimization converges only by employing techniques developed in ML, such as loss function design, model pre-training, and hyperparameter tuning. Given that evolution has had billions of years to explore parameter space, these ML techniques are crucial for emulating these mechanisms. The resulting state change scheme is highly efficient but still suffers errors. To fix these, we introduce a proofreading scheme inspired by Hopfield's kinetic proofreading [23]. Particles that have undergone a state change randomly transition to states with weaker bond strengths which allow these errors to be corrected. From this configuration, the probability of correcting an error is much higher than the probability of forming a new error, and hence this results in nearly perfect yield.

While the conceptual foundation of these ideas have long existed [18,23], this time the proofreading mechanism has been explicitly programed in a specific model of

self-assembly. Goals for future work should apply these schemes to other structures such as the assembly of three-dimensional (3D) shells that can encapsulate or release their contents, or to other functional behaviors, such as self-replication and disassembly.

Our study highlights how ML techniques can be seamlessly applied to physics systems, aiding the discovery of interesting properties. While the details of ML infrastructure are not critical, their role in identifying the right parameter regimes is essential. Pretraining the GNN using logical decision-making rules parallels pretraining large ML models and is crucial for ensuring convergence within a practical time frame. The use of differentiable molecular dynamics simulations with patchy particles to carry out optimization opens up enormous possibilities for the design of novel components [20]. Furthermore, the ability to program interactions and state change opens the door to even more complicated structures and behaviors. As examples, we could design components to engineer the yield shape itself, by optimizing the temperatures where particular structures occur. Alternatively, the system could be programmed to exhibit multifarious assembly [30], whereby different seeds nucleate different structures. To facilitate the

development of these possibilities, we have open sourced the code coupling molecular dynamics simulations to the training of local rules [22].

This work is a step to a longstanding goal of building synthetic materials with the properties of living ones [16]. Computational design is a necessary but not sufficient condition for implementing them in experiments. We now must develop these computational approaches in systems that can be experimentally realized, and tune the computational models to reflect the experimental system accurately enough that we can use computation to guide experiments. The development of such a system is an important and unsolved problem. Possible experimental systems range from magnetic handshake materials [31], to DNA origami [8,9].

Acknowledgments. We thank Francesco Mottes for suggesting that we consider quenched disorder, and all members of the Brenner group for a stimulating environment to use automatic differentiation technologies for optimizing dynamical problems. This material is based upon work supported by the Office of Naval Research (ONR N00014-17-1-3029, ONR N00014-23-1-2654), the NSF Grant DMR-1921619, and the NSF AI Institute of Dynamic Systems (No. 2112085).

- [1] E. A. Newsholme and C. Start, *Regulation in Metabolism* (Wiley, New York, 1973).
- [2] M. Knossow, V. Campanacci, L. A. Khodja, and B. Gigant, The mechanism of tubulin assembly into microtubules: Insights from structural studies, *iScience* **23**, 101511 (2020).
- [3] V. VanBuren, L. Cassimeris, and D. J. Odde, Mechanochemical model of microtubule structure and self-assembly kinetics, *Biophys. J.* **89**, 2911 (2005).
- [4] R. F. Bruinsma, G. J. Wuite, and W. H. Roos, Physics of viral dynamics, *Nat. Rev. Phys.* **3**, 76 (2021).
- [5] P. Buzón, S. Maity, P. Christodoulis, M. J. Wiertsema, S. Dunkelbarger, C. Kim, G. J. Wuite, A. Zlotnick, and W. H. Roos, Virus self-assembly proceeds through contact-rich energy minima, *Sci. Adv.* **7**, eabg0811 (2021).
- [6] L. J. Reha-Krantz, Dna polymerase proofreading: Multiple roles maintain genome stability, *Biochim. Biophys. Acta, Proteins Proteomics* **1804**, 1049 (2010).
- [7] R. Iinuma, Y. Ke, R. Jungmann, T. Schlichthaerle, J. B. Woehrstein, and P. Yin, Polyhedra self-assembled from dna tripods and characterized with 3D DNA-paint, *Science* **344**, 65 (2014).
- [8] Y. Ke, S. M. Douglas, M. Liu, J. Sharma, A. Cheng, A. Leung, Y. Liu, W. M. Shih, and H. Yan, Multilayer DNA origami packed on a square lattice, *J. Am. Chem. Soc.* **131**, 15903 (2009).
- [9] P. W. Rothmund, Folding DNA to create nanoscale shapes and patterns, *Nature (London)* **440**, 297 (2006).
- [10] K. L. Morris, J. R. Jones, M. Halebian, S. Wu, M. Baker, J.-P. Armache, A. Avila Ibarra, R. B. Sessions, A. D. Cameron, Y. Cheng *et al.*, Cryo-EM of multiple cage architectures reveals a universal mode of clathrin self-assembly, *Nat. Struct. Mol. Biol.* **26**, 890 (2019).
- [11] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, Size limits of self-assembled colloidal structures made using specific interactions, *Proc. Natl. Acad. Sci. USA* **111**, 15918 (2014).
- [12] A. Murugan, J. Zou, and M. P. Brenner, Undesired usage and the robust self-assembly of heterogeneous structures, *Nat. Commun.* **6**, 6203 (2015).
- [13] M. H. Huntley, A. Murugan, and M. P. Brenner, Information capacity of specific interactions, *Proc. Natl. Acad. Sci. USA* **113**, 5841 (2016).
- [14] M. F. Hagan and G. M. Grason, Equilibrium mechanisms of self-limiting assembly, *Rev. Mod. Phys.* **93**, 025008 (2021).
- [15] G. Meng, N. Arkus, M. P. Brenner, and V. N. Manoharan, The free-energy landscape of clusters of attractive hard spheres, *Science* **327**, 560 (2010).
- [16] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, Colloquium: Toward living matter with colloidal particles, *Rev. Mod. Phys.* **89**, 031001 (2017).
- [17] A. B. Pawar and I. Kretzschmar, Fabrication, assembly, and application of patchy particles, *Macromol. Rapid Commun.* **31**, 150 (2010).
- [18] J.-P. Changeux, Allostery and the Monod-Wyman-Changeux model after 50 years, *Annu. Rev. Biophys.* **41**, 103 (2012).
- [19] A. G. Baydin, B. A. Pearlmutter, A. A. Radul, and J. M. Siskind, Automatic differentiation in machine learning: A survey, *J. Mach. Learn. Res.* **18**, 1 (2018).
- [20] E. M. King, C. X. Du, Q.-Z. Zhu, S. S. Schoenholz, and M. P. Brenner, Programming patchy particles for materials assembly design, *Proc. Natl. Acad. Sci. USA* **121**, e2311891121 (2024).
- [21] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.6.L042057> for details regarding the simulation, yield definition, and construction and optimization of the GNN.
- [22] https://github.com/brennergrouppaper_notebooks/blob/main/Proofreading_self_assembly.ipynb (2023).
- [23] J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity, *Proc. Natl. Acad. Sci. USA* **71**, 4135 (1974).

- [24] P. W. Battaglia, J. B. Hamrick, V. Bapst, A. Sanchez-Gonzalez, V. Zambaldi, M. Malinowski, A. Tacchetti, D. Raposo, A. Santoro, R. Faulkner *et al.*, Relational inductive biases, deep learning, and graph networks, [arXiv:1806.01261](https://arxiv.org/abs/1806.01261).
- [25] <https://github.com/jax-md/jax-md> (2023).
- [26] R. J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Mach. Learn.* **8**, 229 (1992).
- [27] R. Nussinov, C.-J. Tsai, and J. Liu, Principles of allosteric interactions in cell signaling, *J. Am. Chem. Soc.* **136**, 17692 (2014).
- [28] D. Ni, N. Liu, and C. Sheng, Allosteric modulators of protein–protein interactions (PPIs), in *Protein Allostery in Drug Discovery*, edited by J. Zhang and R. Nussinov, Advances in Experimental Medicine and Biology Vol. 1163 (Springer, Singapore, 2019), pp. 313–334.
- [29] G.-M. Li, Mechanisms and functions of DNA mismatch repair, *Cell Res.* **18**, 85 (2008).
- [30] A. Murugan, Z. Zeravcic, M. P. Brenner, and S. Leibler, Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures, *Proc. Natl. Acad. Sci. USA* **112**, 54 (2015).
- [31] R. Niu, C. X. Du, E. Esposito, J. Ng, M. P. Brenner, P. L. McEuen, and I. Cohen, Magnetic handshake materials as a scale-invariant platform for programmed self-assembly, *Proc. Natl. Acad. Sci. USA* **116**, 24402 (2019).