THE UNIVERSITY OF CHICAGO

PATTERN RECOGNITION THROUGH MOLECULAR COMPUTATION

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF PHYSICS

BY
JACKSON DENSBORN O'BRIEN

CHICAGO, ILLINOIS
DECEMBER 2021

To my family: past, present, and future

TABLE OF CONTENTS

LIST OF FIGURES v

LIST OF TABLES xvi

ACKNOWLEDGMENTS xvii

ABSTRACT xviii

1 INTRODUCTION 1

2 TEMPORAL PATTERN RECOGNITION THROUGH ANALOG MOLECULAR COMPUTATION 4

3 PATTERN RECOGNITION IN THE NUCLEATION KINETICS OF NONEQUILIBRIUM SELF-ASSEMBLY 18

4 CONTROL OF NUCLEATION THROUGH CONCENTRATION PATTERNS IN HETEROGENEOUS SELF-ASSEMBLY 32

5 DISCUSSION AND OUTLOOK 42

A SUPPLEMENTARY INFORMATION: TEMPORAL PATTERN RECOGNITION THROUGH ANALOG MOLECULAR COMPUTATION 43

B SUPPLEMENTARY INFORMATION: PATTERN RECOGNITION IN THE NUCLEATION KINETICS OF NON-EQUILIBRIUM SELF-ASSEMBLY 54

REFERENCES 71

LIST OF FIGURES

2.1 **Schematic of temporal feature extraction**. The three time-varying patterns of a molecular concentration $c(t)$ differ in the number of pulses n, the time period T (i.e., time between onset of pulses), and duty fraction δ (i.e., pulse width relative to time period) but have the same total integrated exposure $nT\delta$ (there are n pulses of width $T\delta$). (a) Circuits that respond to the total integrated exposure are unable to distinguish the three signals (i.e. the three patterns shown result in the same readout). (b) We seek designed molecular systems that can extract independent temporal features n, T, δ and report each of them independently. For example, such feature decoders must report pulse count without regard to pulse width and report duty fraction without regard to the time period. As suggested in the panel above, such a system could distinguish the three signals despite their equal total exposure. 5

2.2 **A pulse count decoder implemented by an Incoherent Feed-Forward Loop that responds only to step ups in input** (but not step downs). (a) Reaction network to decode pulse number. (b) DNA Strand Displacement implementation of mechanism in (a) with waste products suppressed. Species $\bar{I}$ is the negation of input I as in dual-rail logic, i.e., $\bar{I}$ is high when I is low and vice-versa. (c) The flux G (defined in SI Figures 1 and 3)leading to production of P shows a stereotyped response to pulses of I that is independent of the width and separation of such pulses in I. The stereotyped response is due to the 'incoherent' regulation of G by I; I exerts a direct fast positive effect on G, causing a rapid rise, but also exerts an indirect delayed negative effect on G by suppressing A (green). (d) Output P integrates and reports the number of stereotyped responses of G. Consequently, changing the (e) duty fraction or (g) time period of pulses in I has no impact on the output P which does change with (f) pulse count. 8

2.3 **A duty fraction decoder implemented by a timed decay mechanism that computes a moving average.** (a) Reaction network to decode duty fraction. (b) DNA Strand Displacement implementation of mechanism in (a). See SI for kinetic parameters and circuit that includes waste products. (c) Sample time traces for a typical input signal I from simulations of the DNA network. (d) Changing the duty fraction of input I changes steady state value of output P. (e,f) Changing time period T or pulse count n do not affect the readout P. . . . 9

2.4 **Time period decoder implemented by a combining an Incoherent Feed-Forward Loop with a timed decay.** (a) Reaction network to decode time period. (b) DNA Strand Displacement implementation of mechanism in (a) with waste products suppressed. Species $\bar{I}$ is the negation of input I as in dual-rail logic. (c) As with the pulse count decoder, G shows a stereotyped response of fixed width τ_a to pulses of I. However, the output P, due to its decay, now computes the duty fraction τ_a/T of the train of stereotyped pulses in G. Consequently, changing the (d) duty fraction or (e) number of pulses in I has no impact on the output P which does change with (f) time period T. 10

2.5 **Output of DNA-based decoders with realistic kinetic parameters on a library of temporal patterns that systematically vary in pulse number, duty fraction and time period.** Each decoder circuit's output changes significantly in response to changes in only one of the three features. For example, the duty fraction decoder changes by a factor of 4× as duty fraction changes from 0.2 to 0.8 but barely changes as time period changes from 5000 seconds to 15000 seconds. Limits of operation are consistent with formulas derived in the respective Results section. 14

3.1 **Molecules with multiple stable assemblies can naturally recognize high-dimensional concentration patterns through nucleation kinetics.** Promiscuous interactions allow the same molecules to spatially co-localize in distinct combinations to form distinct structures. Then, nucleation (traversal of the barrier from monomers to the wells corresponding to finished structures) can select between structures by discriminating patterns in the concentrations of different molecules; e.g., high concentrations for species 1, 2, and 3 (pattern 1) selectively nucleates the red assembly wherein they are co-localized, but high concentrations for 1, 2, and 7 (pattern 2) kinetically selects for the orange assembly (thicker arrows indicate stronger nucleation). As in an associative neural network, this system can be naturally extended to recognize more classes of concentration patterns by adding promiscuous molecular interactions—e.g., through evolution or engineering—that co-localize the same species in new spatial combinations. . . 19

3.2 **A multifarious mixture of 917 molecular species can assemble three distinct structures from one set of molecules.** (a) 42-nucleotide DNA strands self-assemble into 2-d structures by forming bonds with four complementary strands in solution. The strands can be abstracted as square tiles with four different 10 or 11 nucleotide glues determined by strand sequence. (b) One pool of 917 tile types assembles into three distinct structures, H, A and M. While each tile occurs only once in each structure, the shared purple species re-occur in multiple shapes, in distinct spatial arrangements. (c) Annealing an equal mix of all tiles results in a mixture of fully and partially assembled H, A and M, imaged by atomic force microscope (AFM). 20

3.3 **Simulations and theory show that enhanced concentration of shared tiles selectively enhances nucleation rate of structures in which high concentration tiles are co-localized.** (a) A pattern enhancing the concentration of shared tiles co-localized in H but relatively dispersed in A and M. Nucleation pathways for H climb a lower nucleation barrier, with smaller critical nucleation seeds, as seen from a stochastic nucleation simulation. (b) Regions predicted to participate in nucleation by the simulation for three concentration patterns (lighter colors correspond to higher participation) (c) Varying temperature trades off selectivity and the scale of pattern recognition for speed of recognition, even with constant tile monomer concentrations. (d) With tile monomer depletion, if assembly occurs at temperatures $T < T^H_{nuc}$ and $T < T^H_{grow}$ allowing nucleation and growth of on-target structures (here, H) but too high $T > T^A_{nuc}$ for off-target nucleation (A, shown, or M), a winner-take-all (WTA) effect enhances selectivity compared to systems with no shared components (right) illustrated using simulations of a toy model, see Extended Figure B.7. 23

3.4 **Enhanced concentrations of shared tiles can selectively nucleate structures in which those tiles are co-localized.** (a) We incorporated four different fluorophore/quencher pairs on adjacent tiles in four locations on each structure, allowing the choice of any four locations as labels: quenching of a label indicates growth of that local region on that structure. (b) Samples were annealed with the temperature protocol shown here, spending a majority of the experiment between 48 and 46 °C. (c) We prepared 37 different patterns of concentration (three shown here) that had 12 or 13 shared tiles of increased concentration (16.6x) in checkerboard pattern in a particular 5×5 location in H, A, or M. Fluorophore quenching in multiple samples, and AFM imaging, showed that many patterns resulted in selective nucleation of the shape with the checkerboard pattern; results for each location are summarized in (d) and (e). (f) Patterns enhance nucleation on the shape with localized high concentrations while simultaneously decreasing nucleation of the other two shapes for most experiments. Here, values are given relative to the fractional quenching observed for an equimolar SHAM Mix, described in Extended Figure B.6. 25

3.5 **The distinct co-localization patterns in the three structures are expressive enough to classify up to 18 images, even after distortion, where each pixel value is mapped to a tile concentration.** (a)-(e) Training: Selecting 18 images (30×30 pixels), we computationally searched for one pixel-to-tile assignment that would map each image to a concentration pattern with high concentrations localized in the shape associated with the image and distributed in other shapes; (f) For testing, we also considered 18 distorted variations of these images, using the same assignment. (g) In all trained images, and most distorted images, both fluorescence and AFM results showed that the associated concentration pattern for each image resulted in selective nucleation of the correct corresponding shape. The results across all 36 pattern recognition experiments for both AFM imaging and fluorescence monitoring are summarized, averaged over different samples with different fluorophore configurations. Arrows indicate relative fraction of quenching time or number of shapes counted in AFM images. Line color indicates desired shape for the pattern (purple: uniform SHAM Mix at 60 nM) in each sample. Dashed lines - no significant quenching or shapes seen under AFM. (h) Sample fluorescence traces and AFM images for the 'Horse' concentration pattern. (i) Same data as (g); here, proximity to triangle corners indicate relative fractions for different shapes. 27

4.1 **Nucleation pathways and rate-limiting steps in multi-component self-assembly can change dramatically with small changes in the pattern of unequal concentrations of components.**. The rate-limiting nucleation step of self-assembly is set by a balance of bulk and boundary effects, and is the lowest of the highest free energy state along each possible assembly pathway (a min-max problem). With uneven concentrations, the energy landscape is rugged and critical seeds can take on unusual shapes due to the favorable kinetic attachments of high concentration components that are co-localized on the structure. While the min-max problem on a rugged landscape is a computationally difficult problem, we can determine dominant nucleation pathways and thus rate by rare-event trajectory sampling methods. 35

4.2 **Subtle changes in concentration patterns lead to significantly altered kinetic pathways and overall nucleation rates** (a) Here we compare four patterns with the same number of high (20x) concentration components (shown in orange) on a 16×32 grid at a fixed $G_{se} = 5.2$: the cut outs below show the same 9×9 red-outlined region. Each pattern differs from adjacent patterns only in the location of 4 high concentration species which have only been shifted by one tile on the lattice. Yet each small change causes nucleation rates to differ by a factor of ~ 10, as shown in the bar graph on the right. (b) Energy along sampled trajectories for checkerboard pattern (orange) and cross pattern (blue). (inset: configurations along representative trajectories (bold).) (c) The forward probability p_f describes the probability assemblies of a particular size will evolve to a structure of the next size (usually one tile larger) without dissociating back to monomers. Heat maps shown along the top (checkerboard) and bottom (cross) indicate the probability a particular tile is included in an assembly of a given size. Up until size 5, the cross pattern has a significantly enhanced p_f because of the 5 contiguous high concentration tiles located at the center of the cross. This initial advantage more than compensates for lower p_f for the cross at later assembly stages. (d) Critical seeds along sampled trajectories for the cross pattern (top) and the checkerboard pattern (bottom). (left) Composite critical seed heat maps show the probability a particular tile is included in the sampled critical seeds. (middle) 6 example critical seed assemblies are shown for each pattern. (right) Histogram of energy and size of critical seeds. Critical seeds for cross are larger,on average, than those for the checkerboard but have lower energy because seeds for the cross can readily accommodate a contiguous region of high concentration tiles. 36

4.3 **Temperature sets a trade-off between speed and length-scale resolution.** (a) Random patterns with a fixed number of high (20x) concentration components (shown in orange) were generated such that the high concentration species fit inside a $k \times k$ square. The cut outs below show 2 such examples for $k = 8$. (b) (top) Average nucleation rates increase for higher G_{se} and for smaller k localization. (bottom) Violin plots for $k = 5$ show the distributions of these rates reaches a maximal functional range at intermediate G_{se} ($\log_{10}$ of ratio to mean rate for each G_{se}). (c) Discrimination (the ratio of nucleation rates from Pattern 1 to Pattern 2 from panel (a)) decreases as nucleation rates increases. (d) Energy trajectories of assembly at the G_{se} values shown with dashed lines in (c) ($G_{se} = 5.1$ on top, $G_{se} = 6.1$ on bottom)(Pattern 1 in green and Pattern 2 in purple). Example growth states for a particular trajectory (bold) are shown along the top for Pattern 2 and the bottom for Pattern 1 (e) (left) For 4000 sampled critical seeds, a scatter plot of size against energy for both patterns at $G_{se} = 5.1$ is shown (Pattern 1 in green, Pattern 2 in purple). Example critical seeds are shown for the different patterns at this G_{se} (Pattern 1 on the right column). The purple and green heat maps at the bottom of each column indicate the probability each tile is included in a critical seed. (right) Same as the left but for $G_{se} = 6.1$. 38

4.4 **Interplay between temperature and nucleation scale can lead to significant inversions of relative nucleation rates.** (a) Two patterns with an equal number of high concentration (20×) species are shown. (b) (top) Nucleation rates for both patterns are shown as a function of G_{se}. Two G_{se} values of interest are denoted with vertical lines, and the corresponding critical seeds are shown in panel (c). (bottom) Forward probabilities as a function of λ show that at low G_{se} the Solid w/ Scatter pattern faces a significant barrier after growing out the high concentration region in comparison with the Outline pattern. However, at larger G_{se}, the high concentration region in the Solid w/ Scatter pattern is sufficient to allow favorable growth to larger structures, and it reaches this favorable growth more quickly than for the Outline pattern. 40

A.1 **Full version of DNA strand displacement reaction network for pulse counting decoder including waste products.** The two species in red had their dynamics directly modulated to the parameters of the input series, with sp_1 (referred to as $\bar{I}$ in the main text) pulsing exactly out of phase with sp_0 (I in the main text). Graphs and labels are generated automatically within the Visual DSD software [9]. See Table A.1 for a list of initial concentrations. Note that the flux G plotted in Figure 2 is defined $G \equiv \sqrt{\mathrm{sp}_2 * \mathrm{sp}_5}$. 45

A.2 **Full version of DNA strand displacement reaction network for duty cycle decoding, including waste products.** This circuit effectively takes the moving average of the dynamics of sp_0 and reports it in sp10. See main text Figure 3 for analysis and Figure 5 for performance. Initial concentrations of bolded species are listed in Table A.2. 48

A.3 **Full version of DNA strand displacement reaction network for time period decoding, including waste products.** By taking the moving average of the incoming flux $G \equiv \sqrt{\mathrm{sp}_2 * \mathrm{sp}_5}$, this circuit decodes the period of *sp_0*. Initial concentrations of bolded species are reported in Table A.3. 50

B.1 **Proofreading tile set design and tile assignment map. (a)** Our systems are designed to grow in a regime where a tile attaching by at least two bonds is favorable, but a tile attaching by one bond is not ("threshold 2"). Motivated by self-healing tile systems [101], we seek a tile set where no correct partial assembly should ever allow an undesired tile to attach by two or more bonds, though undesired attachments by one bond are allowed, such that any favorable attachment to a partial assembly will be correct. **(b)** In addition to tiles attaching favorably by 2 bonds to growing facets, new facets in the system will only be created by tiles attaching unfavorably by one bond, and then being stabilized by further, favorable growth. At a site where tile T would correctly attach by one bond, a tile U might be able to attach incorrectly by the same bond. T would correctly be stabilized by the subsequent attachment of V by two bonds, but U might be able to be stabilized as well if there is a tile W that can attach to it and shares the same glue as V. Thus, if for every pair of tiles that can bind to each other (eg, $T + V$), there is no other pair of binding tiles (eg, $U + W$) that share two glues on the same edges of the tiles, then any tile that attaches by one bond to an assembly will either be the correct tile, or will not allow a subsequent attachment, and will likely detach. This is equivalent to "second-order sensitivity" with all directions treated as inputs, functioning as a as a form of "proofreading" [77, 76]. We created a multifarious tile system by first starting with three shapes constructed entirely of unique tiles, then repeatedly attempting **(c)** to "merge" tiles in different shapes by constraining the sequences of their domains to be identical, and checking whether each merge of two tiles results in a tile system that does not have any tile pairs violating criteria in (a) and (b). From multiple trials of the merging process, we selected the smallest result, **(d)**, containing 917 tiles. Tiles in the system were designed with the single-stranded tile (SST) motif in [69], with two alternating tiles motifs of 11 nt and 12 nt domains. 62

B.2 **Suppression of chimeric growth through tile set design.** We contrast assembly errors in three distinct tile sets: **(a, top)** the proofreading tile set with an inert boundary used in experiments (described in Figure 3.2); **(a, middle)** a simple checkerboard tile set with a strictly alternating shared and unique tile pattern for each shape, where unique tiles can be seen as mediating different interactions between shared tiles; and **(a, bottom)** an edge-guarded checkerboard in which we additionally enforce inert bonds around each shape's perimeter. **(b)** Schematic shows the two distinct kinds of chimeric structures (e.g., part-H, part-A) seen in simulation due to promiscuous interactions; chimeric structures can grow either before full assembly of the target structure or emerge spontaneously from the edge of a properly formed structure. Chimeras like those illustrated along the lower path are held together by just a few bonds and will quickly break apart (species with unintended bonds shown in red): this corresponds to the observed sharp drops in the mass trajectories of (c)-(e). **(c)-(e)** For each tile set, we performed kinetic growth simulations, starting from a pre-formed 5×5 seed, for a single seed localized on H, were performed using XGrow (with chunk fission) [94] with uniform tile concentrations corresponding to $G_{mc} = 9.5$, and varying bond energies relative to k_BT, G_{se}. The size of the assembly (in units of the size of the fully formed H) is shown as a function of time. **(c)** For small $G_{se} = 5.3$ (i.e., high temperature, slow growth), no chimeras are observed on the simulated timescales for any tile set. **(d)** For intermediate $G_{se} = 5.9$ (all 6 checkerboard trajectories still result in chimeras, while no errors are observed on the timescale probed for the guarded checkerboard or proofreading-satisfying tile set. **(e)**: Large $G_{se} = 6.6$, (i.e., low temperatures, fast growth) leads to chimeras with all 3 tile sets; chimeras are seen in all runs for checkerboard structures (red traces), 4 of the 6 runs for guarded checkerboard structures (green traces) and 1 of the 6 runs for proofreading-satisfying structures form chimeras). Simulations were carried out with model detailed in Extended Figure B.4 [102] with $\alpha = 0$. 63

B.3 **Fluorophore quenching as a measure of nucleation and growth. (a)** shows positions and types of all fluorophore/quencher pairs available for use; only one of each type of fluorophore can be used in a single sample; four selections of fluorophores were used in different samples. **(b)** shows expected behavior of fluorophore labels on shapes as one shape nucleates and grows. **(c)** shows fluorescence data for non-quenching (fluorophore tile only, orange) and quenching (5×5 lattice around fluorophore and quencher tiles, blue) controls for the ATTO647N fluorophore/quencher pair on A. Here, the temperature ramps linearly from 49°C to 35°C at a rate of $0.1 \frac{^\circ\text{C}}{\text{min}}$, with all tiles at 50 nM, and each sample has its fluorescence normalized to its maximum value independently. **(d)** shows an example of fluorescence growth time measurements. Each fluorophore signal, in each sample, is independently normalized to its maximum value during the experiment, and the time between the point where the signal goes below 0.9 ("10% quenching") and the end of the experiment is measured ("growth time"). These times are then summed for all fluorophores, in all four samples, on each shape, resulting in a growth time for each shape, and, when normalized to the sum of all growth times, a relative growth time for each shape. See Methods above for fluorophore details. 64

B.4 **Stochastic Greedy Nucleation Model, based on repeated stochastic simulations.** The **(a)** frequently-used kinetic Tile Assembly Model (kTAM) [68, 102] has rates for tile attachment and detachment events based on tile and assembly diffusion and total binding strength of correct attachments a tile can make at a lattice site. These rates can be used **(b)** to derive a free energy for any tile assembly in a system, and, assuming fixed monomer concentrations, an equilibrium concentration for any assembly. [104] showed that the equilibrium concentration of the highest-energy assembly along a nucleation trajectory under this assumption provides an upper bound for nucleation rate through that trajectory, with or without fixed monomer concentrations. However, in a large system, considering all possible intermediate assemblies and all pathways, including many that are extremely unlikely, would be infeasible. Thus, we developed the Stochastic Greedy Nucleation Model to generate stochastically-chosen paths of tile attachments. Starting from a single tile (chosen with probability proportional to relative concentration), **(c)** whenever the assembly is in a state A_{stable} where there is no tile attachment that would be favorable (have $\Delta G < 0$), one of the possible unfavorable (with $\Delta G \geq 0$) attachments is stochastically chosen, resulting in a higher-G state A_{unstable}. Then, all subsequent possible $\Delta G < 0$ attachments are made, resulting in the next A'_{stable} state; for our system of unique tiles for each site in the lattice, this sequence of favorable steps has a unique resulting assembly. The process repeats until all tiles in a shape are attached, which results **(d)** in a trajectory with a maximum-G assembly that can be used to bound the rate of nucleation through that particular trajectory. By using this process to collect many trajectories, and then repeating the entire process for each of the three shapes in the system, we can estimate nucleation rates dependent upon temperature **(e)**, with the assumption that tile monomer concentrations do not deplete, and that the trajectories found are a reasonable representation of likely trajectories. For temperature ramps, we determined a reference temperature for each shape at which the model predicted a nucleation rate fast enough to exceed some threshold (orange line), and used this to compare to fluorescence results in Extended Figure B.5(d) and B.8(b). 65

B.5 **Nucleation and growth with 'flag' patterns of enhanced concentration.** 36 different concentration patterns with enhanced concentrations of shared tiles in 5×5 regions were prepared, each with four different standard sets of fluorophores in four samples, and grown using two temperature protocols **(a)**: a ramp focusing on 48°C to 46°C over 100 hours, and a hold at 47°C. Using growth times as described in Extended Figure B.3, fluorescence data for many samples in both experiments showed preference for the desired shapes **(b, c)**, but with considerable variation in selectivity and total amount of growth. No statistically significant correlation was found between the nucleation model prediction for temperature of on-target nucleation and the time of on-target shape quenching in the temperature ramp experiment **(d)**. **(e-g)** show details of three patterns, with concentration patterns (1), weighted critical nucleus free energy starting from particular tiles (2), nucleation-model-estimated nucleation rates (3), temperature hold (4) and temperature ramp (5) experiment fluorescence results, and (6) AFM images from the temperature hold experiments. 66

B.6 **Evidence of winner-take-all in flag experiments (a)** Normalized fluorescence for ROX fluorophore in the H sample and the FAM fluorophore in the A sample for a uniform concentration pattern (i.e. no flag). We define the quenching Q for each fluorophore as the difference between the maximum value (always 1 due to normalization) and the value of the fluorescence when the ramp protocol reaches 46 °C. **(b)** Same as in (a) but for the Afs12 concentration pattern . **(c)** Same data as 3.4 (f) with the data points labeled by their corresponding concentration patterns. The horizontal axis values are calculated by averaging the fold changes of the 5 on-target fluorophores (4 in the sample with the fluorophores on the flagged structure and 1 fluorophore from the '3' sample) while the vertical axis is calculated by averaging over the 10 off-target fluorophores (8 from the samples with fluorophores on the non-flagged shape and the 2 non-corresponding fluorophores from the '3' sample). 67

B.7 **Simple model illustrates winner-take-all dynamics. (a)** A 2N + 1 state ($N = 50$) master equation models the evolution from monomers along two one-dimensional chains of states corresponding to the growth of two distinct structures (kTAM rates are used for attachment/detachment, see Extended Figure B.4) after a non-reversible nucleation step (Arrhenius' barrier crossing formula assumed). **(b)** There are high c_h and low c_l concentrations initially present in the system, with H assumed to have a region of localized high concentration species. Nucleation of H is assumed in a single location, and thus it's states correspond to growth of specific locations in H. In A, states correspond to fragments of a particular size with random composition, leading to distinct depletion dynamics. Species can be totally shared between the two structures (S) or totally unique in each structure (NS). **(c)-(e)** Extended schematics from main text Fig. 3d showing nucleation rates as a function of temperature (assuming fixed concentrations). Here, nucleation and growth temperatures are defined when rate curves become equal to inverse experimental timescale, $1/\tau$ (where the nucleation rates have been scaled by the low concentration in the system to match units). When H structures are able to nucleate and grow at temperatures before A nucleation becomes significant, winner-take-all dynamics are possible with many H's forming and depleting shared tiles, effectively decreasing $T^{\mathrm{A}}_{\mathrm{nuc}}$. **(f)** Selectivity, quantified as the fraction of structure mass in the H branch relative to the total mass in the H and A branches at the final time point, is plotted for the three temperature protocols shown in **(g)** for systems with shared species (S, dots) and systems with no shared species (NS, x's). **(h), top** Nucleation rates for the slow ramp temperature protocol are shown over the duration of the simulation. Note that when the components are shared between H and A, depletion keeps the nucleation rate of the A structure small even at low temperatures. **(h), bottom** The concentrations in state H_N and A_N (solid lines) and the average values of c_l in A and H (dashed lines) are shown for the slow ramp protocol. 68

B.8 **Classification of images viewed as concentration patterns** 36 different concentration patterns, derived from a mapping of 36 grayscale images, were run using a ramp between 48°C to 45°C over approximately 150 hours. **(a)** Three pattern examples, with source image, concentration pattern, nucleation model nucleation rate starting from particular tiles, nucleation model nucleation rates, fluorescence results, and AFM images. **(b)** Across all patterns there was some correlation between the on-target nucleation temperature predicted by the nucleation model and on-target shape quenching time. **(c)**Total AFM shape counts for each sample. . . 69

B.9 **Pattern recognition capacity** To analyze the pattern-recognition capabilities of the designed tile set, a map-training algorithm was run for increasingly larger sets of random images. **(a)** Random images are generated by permuting a list of 900 pixel values (with matching histograms). As in experiments, patterns correspond to 30x30 images with the same number of tiles at each of 10 possible grayscale values. **(b)** Before training, bright pixels (i.e., high concentrations) are randomly assigned to tiles across the three shapes. **(c)** The map training algorithm attempts to create regions of localized high concentrations on a length scale k on the targeted shape (here, H) while dispersing any high concentration tiles on the off-target shapes to prevent spurious nucleation. Note this nucleation center could occur at many different spatial locations on the H **(d)** When the number of images becomes large (18 images per shape shown here), the training algorithm is still able to create a scale k nucleation site on the target shape (H in this example), but the concentration of high concentration tiles on the desired shape is less apparent visually. **(e)** As the number of images in the set increases, the performance of the training algorithm decreases. For larger k, the pixel-tile map can exploit higher-order correlations and can thus accommodate more images. Accuracy is calculated by taking the ratio of the nucleation rate of the target shape over the nucleation rates of all three shapes (calculated using the nucleation model shown in Extended Figure B.4 at $G_{se} = 5.4$ which roughly corresponds to a temperature of 48.6 °C). Here, the shaded areas around each line correspond to the standard deviation of the accuracy contribution over the images in each set. 70

LIST OF TABLES

A.1 Pulse Counter Initial Conditions 46

A.2 Duty Fraction Decoder Initial Conditions 49

A.3 Period Decoder Initial Conditions 49

ACKNOWLEDGMENTS

It is hard to believe this journey is finally reaching its conclusion.

First, I should thank my fiancée, Celine Spinka. Her love and unwavering support carried me through the most difficult periods of these last few years. This achievement truly belongs to both of us, and I look forward to achieving many more wonderful things together in the years to come.

The amazing dedication and selflessness of my parents, Kevin and Linda, enabled me to pursue my academic passions to this exhaustive end. Throughout my life, they have often put my own happiness above their own, and I will always be grateful for the opportunities they have provided me.

I have been lucky to work with several amazingly talented scientists in my time here at UChicago. I want to thank the entire Murugan lab. I always had a reliable network to discuss with if I was stuck on a problem or considering a new direction for a project. I am very grateful to my collaborators from outside the university as well: Rebecca Schulman and Sam Schaffter (now at NIST) of Johns Hopkins University, Erik Winfree and Constantine Evans (now at Maynooth University) of the California Institute of Technology, and Orit Peleg and Chantal Nguyen of the University of Colorado at Boulder.

Finally, I am thankful to my committee for their direction and advice. In particular, I want to thank my advisor Arvind Murugan. Throughout my time in the group, Arvind was very dedicated to my personal development, ensuring I attended conferences and workshops at every possible opportunity. He always demanded simple, intuitive explanations regardless of the complexity of a problem. While I often found this incredibly challenging, I will always carry this approach with me. This distillation forces a careful examination and abstraction of the problem solving process, and carries the added benefit of easing the communication of new ideas to others.

ABSTRACT

Understanding the computational capacity of molecular systems is important both for basic insights into information processing in biological systems and the design of synthetic nano/micro-scale technologies. Here, we design systems that, instead of mimicking conventional engineered computational systems, exploit intrinsic physical dynamics to carry out pattern recognition autonomously. Using chemical reaction networks and self-assembly of many heterogeneous species, we demonstrate pattern recognition capabilities on pulsatile temporal inputs and high-dimensional concentration patterns. Throughout, we emphasize the ways in which our physical mechanisms are naturally suited to the particular computational challenges. This work not only provides molecular computational solutions for concrete problems, but also helps broaden the paradigm about how, when, and where computation can occur in molecular systems.

CHAPTER 1

INTRODUCTION

Pattern recognition is an important form of computation where a system is able to parse some lower dimensional interpretation from a complex input. In modern machine learning, pattern recognition algorithm uses include both classification and regression, and these algorithms have been employed for tasks ranging from classifying images of handwritten digits [1] to earthquake prediction from time-series data [2]. The human brain is constantly recognizing complex patterns. From recognizing the faces of our family and friends to parsing speech, pattern recognition underlies many of the capabilities of human intelligence [3].

Interestingly, many non-neural biological systems can recognize patterns autonomously. The bacteria *E. Coli* are famously capable of chemotaxis [4] where these organisms are able sense and accordingly move up gradients of nutrient concentrations. In some cases, cells use temporal patterns to encode identity and intensity of external stimuli. For example, p53 pulse duration and number of pulses changes whether the cell has been damaged by γ or UV radiation [5]. While the full mechanistic details of how cells precisely decode these patterns into meaningful, differentiated gene expression responses are still not known, it has been demonstrated that modulating the temporal dynamics of the Msn2 transcription factor in yeast can selectively express one of four different downstream genes [6].

This thesis explores basic molecular systems that are explicitly capable of pattern recognition and discrimination intrinsically due to their designed physical interactions and chemical reactions. It builds upon the wealth of molecular computational applications, dating back to Adleman's DNA-based solution to the famous Hamiltonian path problem [7]. In the systems described here, DNA also plays critical role; the designability and availability of arbitrary DNA sequences enables experimental realizations for each abstract computational mechanism. Such implementations were either explicitly achieved or discussed at length to show that the mechanisms described here function even under real (bio)physical constraints on,

for example, reaction rates.

Neural networks and machine learning algorithms show a remarkable ability to learn pattern recognition from training examples. The hardware-based systems here can be similarly trained using mechanisms ranging from training on a computer (e.g., gradient descent) to systems that can physically learn *in situ* through adaptive weights. New molecules can be synthesized dynamically to effect new behavior in response to these example inputs, typically through help of enzymes (see, e.g.[8]). Learning interactions autonomously holds exciting future directions for the kinds of systems and problems presented here.

Chapter 2 provides examples of chemical reaction networks that are capable of extracting specific features from a temporal concentration pattern. The mechanisms provided are inspired by known reaction network motifs from systems biology but are analyzed in new light for their ability to effectively process a chosen family of temporal patterns. A minimal reaction network and corresponding set of ordinary differential equations are presented for each mechanism and analyzed in a manner quite similar to the transient analysis that is typically seen in standard textbooks on electrical circuits. For each reaction network, a corresponding implementation is provided using DNA strand displacement reactions, and the functionality of these implementations is confirmed in simulation using Visual DSD software [9].

Chapter 3 demonstrates the inherent computational power in self-assembly systems for classification of high-dimensional concentration patterns. We study a kinetic model of multifarious self-assembly where a single set of monomer components is able to stably form three distinct spatially arranged lattices. Then, concentration patterns imposed on these unique components are shown to be able to control which of those three lattices will preferentially form during a temperature anneal. Analogous to an artificial neural network classifying images with continuous pixel intensities as input, this system takes in a list of concentrations and produces a weighted output of three discrete classifications, i.e. the relative abundance of the three distinct stable lattice structures. We experimentally explored different regimes

of this theory using a system of nearly 1000 distinct species of single-stranded DNA oligonucleotides. Experimental methods for monitoring of nucleation and growth of these structures in real time are introduced, and simple, heuristic computational models are introduced and discussed to provide intuition about the computation taking place during the competitive nucleation of these different assemblies.

Chapter 4 takes this idea of computation in nucleation further and demonstrates the ability for heterogeneous self-assembly to distinguish subtle spatial differences in concentration patterns. Instead of a multifarious system, a system with a single designed stable lattice is considered. By utilizing rare-event sampling techniques along with a kinetic model of self-assembly, nucleation rates for different concentration patterns are estimated computationally. It is shown that even seemingly small changes to a concentration pattern can cause nucleation to take place orders of magnitude faster. The relationship between temperature, assembly speed, nucleation length scale, and pattern discrimination is explored for this system through a few illustrative examples.

CHAPTER 2

TEMPORAL PATTERN RECOGNITION THROUGH ANALOG MOLECULAR COMPUTATION

This chapter is reproduced with permission from ACS Synthetic Biology [1]

Abstract: Living cells communicate information about physiological conditions by producing signaling molecules in a specific timed manner. Different conditions can result in the same total amount of a signaling molecule, differing only in the pattern of the molecular concentration over time. Such temporally coded information can be completely invisible to even state-of-the-art molecular sensors with high chemical specificity that respond only to the total amount of the signaling molecule. Here, we demonstrate design principles for circuits with temporal specificity, that is, molecular circuits that respond to specific temporal patterns in a molecular concentration. We consider pulsatile patterns in a molecular concentration characterized by three fundamental temporal features - time period, duty fraction and number of pulses. We develop circuits that respond to each one of these features while being insensitive to the others. We demonstrate our design principles using general chemical reaction networks and with explicit simulations of DNA strand displacement reactions. In this way, our work develops building blocks for temporal pattern recognition through molecular computation. Recent breakthroughs in synthetic biology have led to molecular sensors that report on the local environment in cells by detecting signaling molecules with high *chemical* specificity [10, 11, 12, 13].

However, by themselves, such sensors can be completely blind to temporally coded information in cells and tissues. In fact, living cells often communicate information about physiological conditions by producing a signaling molecule in a specific timed manner [5, 14, 15]. For example, many rapid pulses of nuclear p53 in mammalian cells indicates γ radia-

1. **O'Brien, J.** and Murugan, A. ACS Synth. Biol. 2019, 8, 4, 826–832, Published 3/5/2019

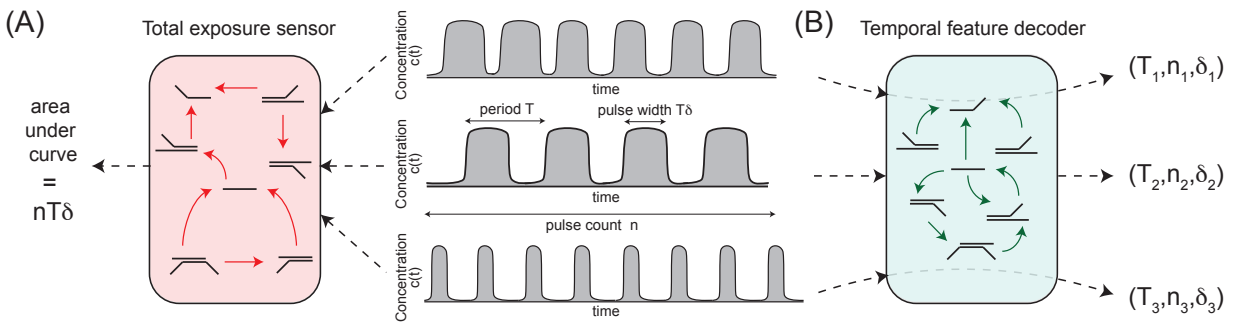

Figure 2.1: **Schematic of temporal feature extraction**. The three time-varying patterns of a molecular concentration $c(t)$ differ in the number of pulses n, the time period T (i.e., time between onset of pulses), and duty fraction δ (i.e., pulse width relative to time period) but have the same total integrated exposure $nT\delta$ (there are n pulses of width $T\delta$). (a) Circuits that respond to the total integrated exposure are unable to distinguish the three signals (i.e. the three patterns shown result in the same readout). (b) We seek designed molecular systems that can extract independent temporal features n, T, δ and report each of them independently. For example, such feature decoders must report pulse count without regard to pulse width and report duty fraction without regard to the time period. As suggested in the panel above, such a system could distinguish the three signals despite their equal total exposure.

tion damage and leads to cell cycle arrest, while a longer sustained single pulse of nuclear p53 indicates UV damage and leads to programmed cell death [5]. Thus different biological conditions can result in the same total amount of a signaling molecule, differing only in the pattern of the molecule's concentration over time [6]. Such biological conditions cannot be distinguished by a sensor that responds to the total amount (or exposure) to a target molecule, even if the sensor has high chemical specificity.

In this article, we demonstrate design principles for molecular circuits with *temporal* specificity, i.e., molecular circuits that respond to specific temporal features in the concentration of an input molecule, instead of the total exposure to that input molecule. We show chemical reaction networks satisfying these principles and derive constraints on their rate constants. We then find explicit implementations of these abstract reaction networks using DNA strand displacement reactions and verify the performance of such DNA circuits using

simulation software explicitly designed for this purpose [9]. We anticipate that this work will also be of use for other synthetic biology constructs based on enzymes and transcriptional gates [16, 17] and for analyzing naturally occurring temporal decoding mechanisms in cells [5].Several recent works have studied how temporally varying signals can be processed by biochemical systems [18, 19, 20, 21, 22]. Here we provide a single framework to independently decode all the features of a three-parameter family of pulsatile inputs.

Recognizing temporal patterns can be interpreted as an analog operation that is naturally suited for molecular computation. Molecular circuits of digital gates are robust and have solved remarkable problems[23, 24, 25] and even mimicked neural networks[26]. However, analog operations represent information directly in continuous physical variables like concentration or time intervals[27, 28], making analog devices simpler and naturally suited to temporal pattern recognition problems[29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 8, 35, 41].

Further, analog circuits can naturally exploit molecular transients for computation as shown in several works [42, 43, 44]. Transients are particularly useful for temporal pattern recognition since they can serve as natural 'rulers' to measure and process information encoded as time intervals. Even on a digital computer, temporal pattern recognition problems are best solved using recurrent networks with internal transients[45]. In this way, our approach here exploits a feature inherent to molecular devices – namely, finite timescale transients – to perform temporal pattern recognition.

Results

Time-varying signals are high dimensional and can vary in endless ways. Here, we restrict our study here to pulsatile patterns $c(t)$ in the concentration of an input molecule I (see Figure 2.1) characterized by three independent temporal features - time period T (i.e., time between onset of pulses), duty fraction δ (i.e., pulse width relative to duration between pulses)

and the number of pulses n (i.e. length of the pulse train). It is possible to parameterize this family of inputs using pulse width instead of duty fraction. However, doing so would restrict the values of pulse width possible for a given value of period to prevent pulses from overlapping. As shown in Fig.2.1, different temporal patterns can have distinct n, T, δ but have the same total area = $nT\delta$. Hence a naive sensor that is only sensitive to the total amount of I would respond in the same way to all patterns shown.

Here, we demonstrate decoders that can report on each of these temporal features n, T, δ independently - e.g., the duty fraction readout should be independent of the time period and the number of pulses while the number of pulses readout should not depend on the width or separation of the pulses.

We will assume that the amplitude of the input signal is fixed and will not consider amplitude fluctuations. A novel mechanism to buffer amplitude fluctuations was proposed recently[35]; such a stereotyping mechanism can be used upstream of the temporal decoders proposed here. The three independent numbers n, T, δ form a complete independent 'basis' for the family of patterns considered here. Other combinations of these features might be relevant for specific applications (e.g., pulse width $T\delta$ instead of duty fraction δ) but the principles behind the basis set of decoders should allow development of sensors for such features as well.

Pulse count decoder

We begin with a molecular circuit that can count the number of pulses n seen, without regard to the width of each pulse or the separation between pulses.

To count pulses n in this manner, we first seek a circuit that produces a stereotyped response to each pulse that is independent of pulse width and separation. To do so, we take inspiration from 'biochemical adaptation' mechanisms used e.g., in bacterial chemotaxis

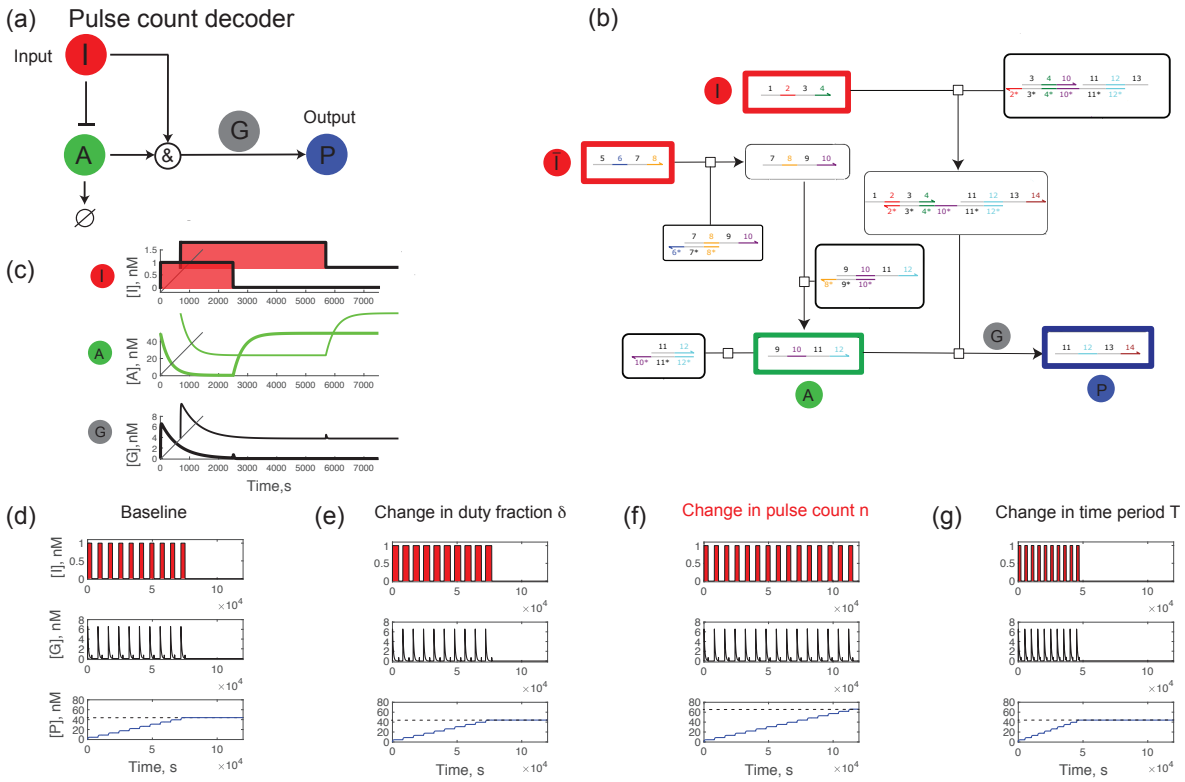

Figure 2.2: **A pulse count decoder implemented by an Incoherent Feed-Forward Loop that responds only to step ups in input** (but not step downs). (a) Reaction network to decode pulse number. (b) DNA Strand Displacement implementation of mechanism in (a) with waste products suppressed. Species $\bar{I}$ is the negation of input I as in dual-rail logic, i.e., $\bar{I}$ is high when I is low and vice-versa. (c) The flux G (defined in SI Figures 1 and 3)leading to production of P shows a stereotyped response to pulses of I that is independent of the width and separation of such pulses in I. The stereotyped response is due to the 'incoherent' regulation of G by I; I exerts a direct fast positive effect on G, causing a rapid rise, but also exerts an indirect delayed negative effect on G by suppressing A (green). (d) Output P integrates and reports the number of stereotyped responses of G. Consequently, changing the (e) duty fraction or (g) time period of pulses in I has no impact on the output P which does change with (f) pulse count.

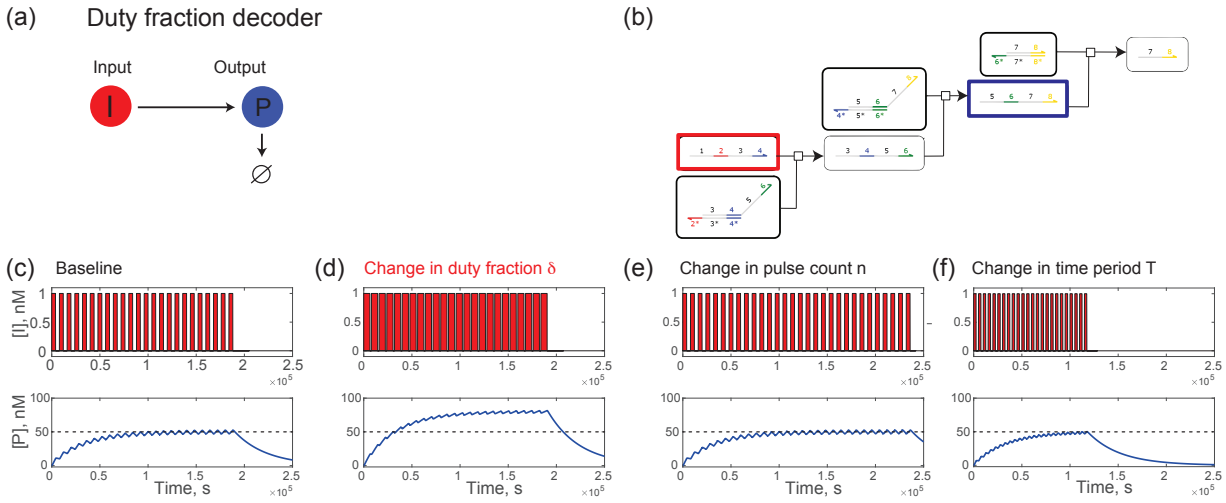

Figure 2.3: **A duty fraction decoder implemented by a timed decay mechanism that computes a moving average.** (a) Reaction network to decode duty fraction. (b) DNA Strand Displacement implementation of mechanism in (a). See SI for kinetic parameters and circuit that includes waste products. (c) Sample time traces for a typical input signal I from simulations of the DNA network. (d) Changing the duty fraction of input I changes steady state value of output P. (e,f) Changing time period T or pulse count n do not affect the readout P.

[46]. Such molecular circuits show a transient response to step *changes* in an input signal but return to their prior state and are insensitive to the steady state value of the input. Incoherent Feed-Forward Loops (IFFL) and Negative Feedback loops are two common molecular circuits that carry out adaptation in biology [47]. We focus on IFFLs which have been recently implemented with DNA strand displacement reactions [48].

However, adaptive circuits in the biological literature often respond in an equal and opposite manner to both the rising and falling edges of each input pulse[46, 49]. To count pulses, we desire an *asymmetric* response to step ups and step downs in the input. For example, a circuit that responds only to step ups but ignores step downs and steady values of the input would naturally count pulses. Such asymmetric adaptation[50, 51, 52, 53] can be naturally achieved in a reaction network if the adaptive variable has a resting concentration at zero;

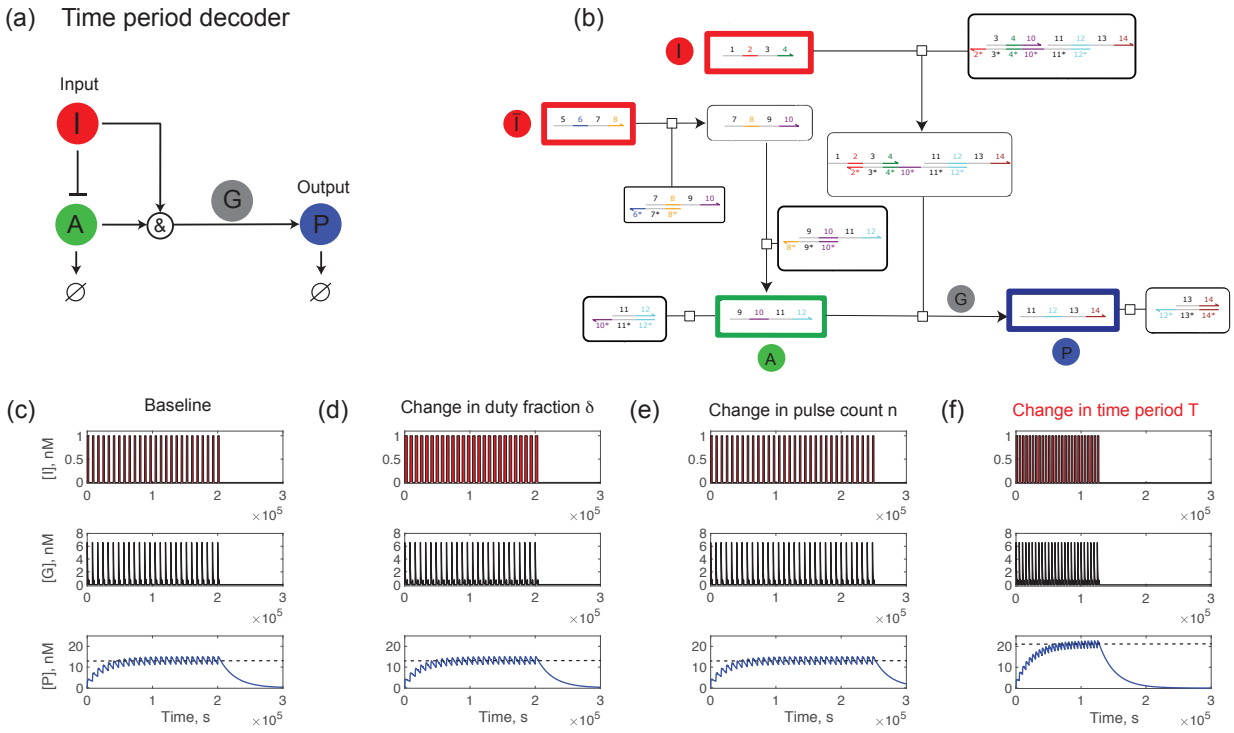

Figure 2.4: **Time period decoder implemented by a combining an Incoherent Feed-Forward Loop with a timed decay.** (a) Reaction network to decode time period. (b) DNA Strand Displacement implementation of mechanism in (a) with waste products suppressed. Species $\bar{I}$ is the negation of input I as in dual-rail logic. (c) As with the pulse count decoder, G shows a stereotyped response of fixed width τ_a to pulses of I. However, the output P, due to its decay, now computes the duty fraction τ_a/T of the train of stereotyped pulses in G. Consequently, changing the (d) duty fraction or (e) number of pulses in I has no impact on the output P which does change with (f) time period T.

then step downs can have little effect since concentrations cannot fall below zero.

Here, we present an Incoherent Feed-Forward Loop circuit with an *asymmetric* adaptive response to step ups and downs:

$$\mathrm{I} + \mathrm{A} \xrightarrow{k_1} \mathrm{P} \tag{2.1}$$

$$\mathrm{A} \xrightarrow{\lambda_1} \phi \tag{2.2}$$

$$\bar{\mathrm{I}} \xrightarrow{k_2} \mathrm{A} \tag{2.3}$$

Here $\bar{I}$ refers to a species that is the negation of the input I as in dual-rail logic[54, 24]; i.e., we assume a second input species $\bar{I}$ that is high when the input I is low and vice-versa. In principle, such dual rail $(\bar{I}, I)$ input can be created from a single input I by a fast reusable NOT-gate [54]. Note also that although the reactions involving I and $\bar{I}$ are not designed to be catalytic, these species have concentrations set directly (i.e. their dynamics are not coupled to any other species in the system).

To understand the mechanism, consider the response to a single step up in I shown in Fig.2.2c. Production of P requires both I and A to be present. While A is high in its resting state, P is produced only when input I steps up. However, turning on I also effectively suppresses A on a timescale,$\tau_{\mathrm{a}} = \frac{1}{k_1[I]+\lambda_1}$, thus ceasing production of P. The production of P will thus show a stereotyped profile for every step up in I provided the pulse width is considerably longer than τ_{a}. To emphasize this, we label this stereotyped flux leading to the production of P as G (depending on A and another species promoted by the input) and plot its dynamics in 2.2c. As shown in the figure, a subsequent step down in I that occurs more than a time τ_{a} after the step up has minimal impact on G because A is near zero. After the step down of I, A is restored back to its resting value on a timescale $1/\lambda_1$. Provided A has sufficient time to recover its initial value, each pulse of G will be identical and thus the total amount of P accumulated will report the number of pulses n without regard to time period

T or duty fraction δ.

Limits of operation: There are two critical requirements for the mechanism above to work. First, the pulse width $T\delta$ must be larger than the length of the stereotyped adaptive response τ_{a}, so that the stereotyped response is not interrupted by the step down in I. Thus, we require that

$$T\delta \gg \frac{1}{k_1[I] + \lambda_1}.$$

Second, the pulse off time $T(1-\delta)$ needs to be long enough so that A can be restored to its resting state before the next pulse in I comes, hence requiring

$$T(1-\delta) \gg 1/\lambda_1.$$

Taken together, we require $T \gg T_{min} = \sup(\{1/\lambda_1, \tau_{\mathrm{a}}\})$.

To verify that real chemical networks can operate in this kinetic regime, we designed a DNA strand displacement implementation of this scheme. As with all the DNA circuits we present here, our design process leaned on the reaction designs laid out earlier [55]; e.g., we introduce intermediate species to ensure that strands in different parts of the circuit have completely distinct sequences. A representation of the reaction network with waste products suppressed is shown in Fig 2.2b. Simulating this DNA strand displacement network using Visual DSD [9] software with realistic kinetic parameters, we find that the output is indeed sensitive to pulse number n but insensitive to duty fraction δ and time period T over a significant range.

Duty fraction decoder

We now develop a circuit needed to decode duty fraction, δ. The reaction network capable of decoding δ is deceptively simple in topology,

$$\mathrm{I} \xrightarrow{k} \mathrm{I} + \mathrm{P} \tag{2.4}$$

$$\mathrm{P} \xrightarrow{\lambda} \phi. \tag{2.5}$$

Species P is created by every pulse of I but P also decays with a time constant $1/\lambda$.

Limits of operation: Intuitively, such an output P effectively reports the exponential moving average of input species I over a time window $1/\lambda$. As shown in the supplementary information, this moving average, in principle, depends on all three features n, T, δ. However, the n dependence is exponentially suppressed if $\lambda n T \gg 1$. The T dependence is also weak if $\lambda T \ll 1$. Under these two limiting operations,

$$T \ll T_{max} = 1/\lambda, \quad n \gg n_{min} = 1/(\lambda T)$$

we find $\bar{P} \approx \frac{k}{\lambda}\delta$ depends only on the duty fraction δ.

Finally, note that the mean level of P is a good readout only if the size of the oscillations about $\bar{P}$ seen in Fig.2.3c are small. The size of such oscillations relative to the mean is given by $\frac{\Delta \bar{P}}{\bar{P}} = \lambda T(1-\delta)$ which is naturally small in the limits of operation defined above ($\lambda T \ll 1$ and $\delta \in [0, 1]$). The operational regime can be expanded by introducing intermediate species or shortening toeholds, effectively decreasing the value for λ (though this comes with the cost of increasing the value of $n_m in$).

To check whether real chemical systems can operate in the kinetic regime defined above, we implemented this scheme using DNA strand displacement reactions. We generated the network and equations using Visual DSD software [9] with realistic kinetic parameters and then simulated this system in MATLAB. The impact of varying n, T, δ on the output is

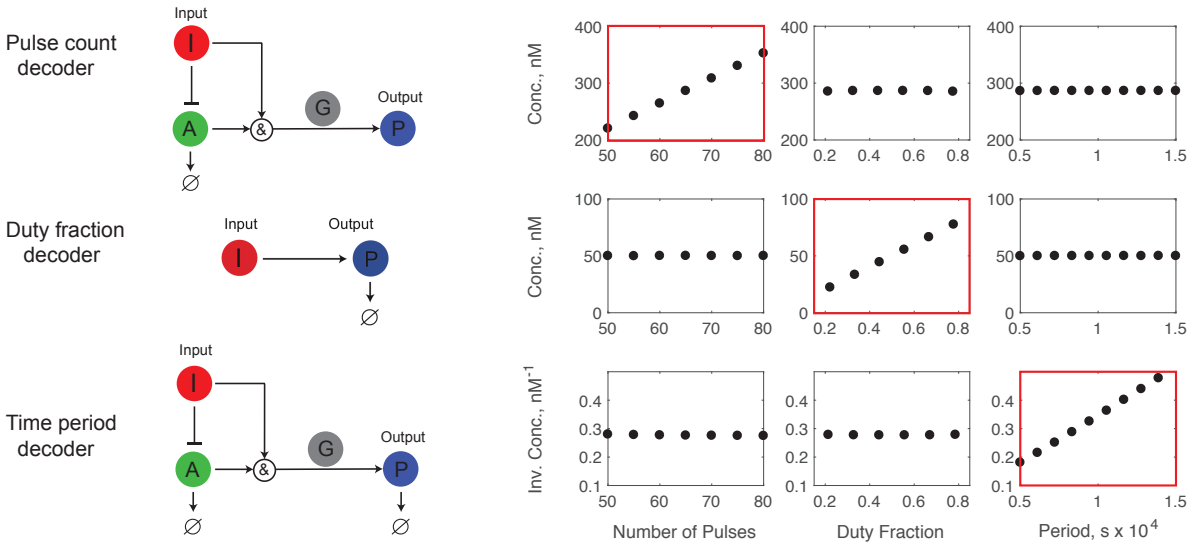

Figure 2.5: **Output of DNA-based decoders with realistic kinetic parameters on a library of temporal patterns that systematically vary in pulse number, duty fraction and time period.** Each decoder circuit's output changes significantly in response to changes in only one of the three features. For example, the duty fraction decoder changes by a factor of 4× as duty fraction changes from 0.2 to 0.8 but barely changes as time period changes from 5000 seconds to 15000 seconds. Limits of operation are consistent with formulas derived in the respective Results section.

shown in Fig. 2.3c-f. As desired, output P is only sensitive to changes in duty fraction δ and insensitive to changes in n, T.

Time period decoder

We can build a time period T decoder by modifying the adaptive circuit motif introduced above; we simply add a decay process for species P so that P reflects the moving average of G (as defined above) over a fixed timescale $1/\lambda_2$,

$$\mathrm{P} \xrightarrow{\lambda_2} \phi. \tag{2.6}$$

All other reactions are as shown in Eqn.A.1 - A.3. The analysis for this network is

nearly identical to that done for duty fraction. Note that the flux $G(t)$ possesses the same periodicity T as the input pulse train $I(t)$ but G has pulse width set by the width of the stereotyped adaptive response τ_a and independent of the pulse width of I. Thus, $G(t)$ has an effective duty fraction $\delta_{\text{eff}} = \tau_a/T$. If we tune kinetic parameters to the same regime from our duty fraction sensor, the output species P will then report δ_{eff} and is thus proportional to $1/T$ but independent of the δ and n of the input species I. Note that this sensor output technically measures 1/T (the frequency of the input) and could just as well be referred to as a frequency decoder.

Limits of operation: This network inherits kinetic constraints from both the pulse count decoder applied to the adaptive part of the network and the duty fraction decoder applied to P. The former restricts $T(1-\delta) \gg 1/\lambda_1$ as discussed for the pulse count decoder. The latter requires

$$T \ll 1/\lambda_2, \quad n \gg 1/(\lambda_2 T).$$

In this kinetic regime, output P depends strongly on T and only weakly on n, δ.

The DNA implementation is shown in Fig 2.4b. Simulations show that the output is insensitive to changes in δ, n but sensitive to changes in T.

Finally, we systematically tested all three decoders proposed here against a library of temporal patterns that vary in all three features (n, δ, T). Each decoder was implemented with DNA strand displacement with fixed kinetic rates. We see in Fig.2.5 that each decoder shows a much larger response to changes in its relevant feature than to the other features over a substantial dynamic range. Also, because these relationships are linear over the designed parameter range, a few simple control experiments would allow the extraction of single parameter fits for explicit quantitative relationships between the temporal features and the concentration of P. Thus, collectively, the three circuits can discriminate each member of the temporal pattern library. In Figure 5, we've used the following fixed parameters to generate each curve: pulse counting- $\delta = .5, T = 8000$ s, $n = 65$, duty fraction - $n = 50, T = 8000$

s, $\delta = .5$ and time period- $\delta = .5, n = 60, T = 8000$ s.

Discussion

The circuits introduced here exploit a weakness in the context of digital computation - internal transients - to naturally perform analog computation on temporally coded information. The performance of such analog 'computation using transients' is limited by the dynamic range over which timescales of transients can be tuned. Hence DNA strand displacement reactions are particularly suitable for such computation since their kinetics can be tuned over a large dynamic range through toehold sequence design [56, 57, 58, 59]. In addition to synthetic applications, our design principles (e.g., asymmetric rectified adaptation) are relevant to understanding how natural systems decode temporal patterns as well [51, 50, 52, 60]. While we have successfully decoded temporal features into the concentrations of DNA species, it is important to emphasize that the experimental measurement of the different temporal features would require different methods. Duty fraction and period should be measured while the input is presented so that the steady-state value can be read off. However, pulse count should be measured after the input has been presented (although it could be observed concurrently with the input to identify the arrival of each pulse).

The methods developed here can be combined with other developments in the programmable molecular technology field [61]. For example, a drug payload carried by a DNA origami pill [62] can be released only in those cells with a pulsatile pattern of the transcription factor NFkB that precedes a inflammatory response but not in cells with NFkB patterns that precede an adaptive immune response [5]. Similar *in situ* temporal computation using molecules can also help surveil complex ecosystems, such as the gut, where a future ecological collapse is often indicated by temporal precursors [63].

We have produced circuits that decode broadly relevant but *predetermined* temporal features. Going forward, it would be interesting to develop molecular circuits that can *learn*

relevant temporal features dynamically [64, 65, 66] as in machine learning approaches. In the learning paradigm, for example, a molecular circuit could be exposed to two classes of time-varying patterns during a 'training phase' (the temporal equivalent of cat and dog images in static pattern recognition). The circuit would determine which temporal features can best distinguish those two classes.

Supplemental Information

The supplemental information (see Appendix A), provides expanded DNA reaction networks along with explicit values for initial conditions and reaction rates used in simulation. Some expanded calculations are also presented to supplement the analysis of each circuit.

Acknowledgments

The authors thank Aaron Dinner, Anders Hansen, Kabir Husain, John Hopfield, Sidney Nagel and Michael Rust for useful discussions. We acknowledge NSF-MRSEC 1420709 for funding and the University of Chicago Research Computing Center for computing resources. AM is grateful to the Simons Foundation's Mathematical Modeling of Living Systems investigator program for support.

CHAPTER 3

PATTERN RECOGNITION IN THE NUCLEATION KINETICS OF NONEQUILIBRIUM SELF-ASSEMBLY

This chapter is reproduced from a collaborative manuscript draft. [1]

Abstract: Molecular processes can conceal complex information processing abilities while appearing to build structures, transform molecules and dissipate energy. Such information processing has been revealed in processes ranging from protein interactions and gene regulation to the growth of self-assembled structures. However, current examples are often limited to a sequence of digital operations on relatively low dimensional inputs. To understand the full scope of computational abilities that can be hidden in complex biomolecular environments, we must understand the potential for the collective dynamics of many promiscuously interacting molecules to carry out computation on high-dimensional inputs. Here we show that nucleation of multi-component structures can recognize high dimensional patterns in the concentrations of molecules in a manner reminiscent of Hopfield neural networks. We find that competitive nucleation in a soup of promiscuously interacting molecules can assemble distinct structures in response to subtle patterns in the relative concentrations of nearly a thousand molecular species. Using DNA hybridization-based self-assembly as a validation platform, we experimentally demonstrate such kinetically controlled self-assembly. This work reveals new computational abilities inherent to large promiscuous networks of molecular interactions.

In biology, a primary use of cellular information processing is to know how to make the

1. Author List: Constantine G. Evans, Jackson O'Brien (co-first author), Erik Winfree, Arvind Murugan. CGE, EW, and AM conceived the study. CGE and EW designed the molecules. CGE, JOB, EW, and AM designed the experiments and CGE, JOB, and AM performed the experiments. CGE wrote the code for stochastic, greedy nucleation rate estimation (Figure B.4) and the optimization of the pixel-to-tile map used in Figure 3.5. JOB performed the simulations demonstrating winner-take-all (Figure 3.4 d and B.7), the simulations showing chimeric growth for various tile set designs (Figure B.2), and the simulations demonstrating the capacity of the pixel-to-tile map (Figure B.9). All analyzed the data and wrote the manuscript.

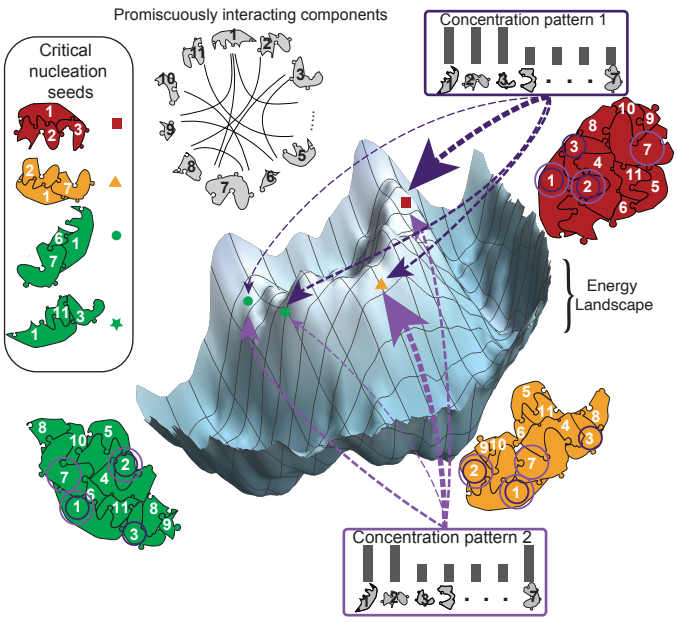

Figure 3.1: **Molecules with multiple stable assemblies can naturally recognize high-dimensional concentration patterns through nucleation kinetics.** Promiscuous interactions allow the same molecules to spatially co-localize in distinct combinations to form distinct structures. Then, nucleation (traversal of the barrier from monomers to the wells corresponding to finished structures) can select between structures by discriminating patterns in the concentrations of different molecules; e.g., high concentrations for species 1, 2, and 3 (pattern 1) selectively nucleates the red assembly wherein they are co-localized, but high concentrations for 1, 2, and 7 (pattern 2) kinetically selects for the orange assembly (thicker arrows indicate stronger nucleation). As in an associative neural network, this system can be naturally extended to recognize more classes of concentration patterns by adding promiscuous molecular interactions—e.g., through evolution or engineering—that co-localize the same species in new spatial combinations.

right molecular assemblies at the right time and in the right place. Signal transduction and gene regulatory networks [67] are often seen as providing the necessary information processing, while a distinct molecular assembly process builds the right structures in response to such a decision. Thus, information processing and structural assembly are often seen as distinct processes.

While the deterministic growth of geometrically-ordered structures has been previously shown to be capable of computation (algorithmic self-assembly [68]), our approach demonstrates a new class of neural network-like algorithms that exploit the inherently stochastic

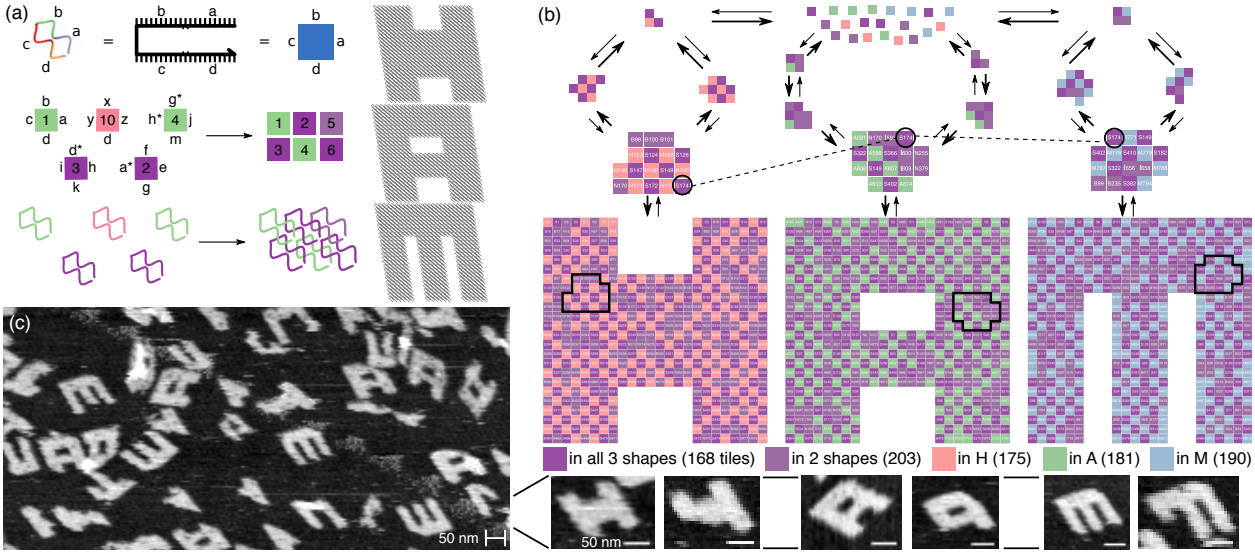

Figure 3.2: **A multifarious mixture of 917 molecular species can assemble three distinct structures from one set of molecules.** (a) 42-nucleotide DNA strands self-assemble into 2-d structures by forming bonds with four complementary strands in solution. The strands can be abstracted as square tiles with four different 10 or 11 nucleotide glues determined by strand sequence. (b) One pool of 917 tile types assembles into three distinct structures, H, A and M. While each tile occurs only once in each structure, the shared purple species re-occur in multiple shapes, in distinct spatial arrangements. (c) Annealing an equal mix of all tiles results in a mixture of fully and partially assembled H, A and M, imaged by atomic force microscope (AFM).

nature of nucleation. Theory has hinted that this was possible in abstract by analogies with neural network. Here, we experimentally demonstrate that a structural process, nucleation, hides within it a neural network-like capacity to decide what structure to build based on high-dimensional inputs, without any need for a separate regulatory module. We find that competitive nucleation between different potential assemblies of one set of molecules can recognize high-dimensional patterns in the concentrations of nearly a thousand assembly components (Figure 3.1).

To study the computational ability inherent to nucleation, we attempted to design a self-assembling system with DNA nanotechnology (Figure 3.2[a]). Prior works have demon-

strated self-assembly of structures with uniquely addressed or algorithmically chosen components that interact through Watson-Crick pairing [69, 70, 71, 72, 73]. We build on these ideas to create a molecular system capable of assembling multiple target structures from a shared set of promiscuously interacting components, i.e., a multifarious system [74]. We then study the pattern recognition capability inherent to the competitive nucleation between these structural polymorphs. Our multifarious soup of DNA tiles can assemble three different structures, H, A and M (Figure 3.2[b]). These structures co-localize the same shared set 'S' of 371 distinct tiles (purple tiles in Figure 3.2) in different arrangements.

To create such a multifarious soup, we design S tiles that do not directly bind each other. By then introducing three sets of interaction mediating-tiles, H, A and M for each of the desired structures H, A and M respectively, we can avoid constraints from Watson-Crick complementarity, allowing almost arbitrary interactions to be engineered between S tiles. For example, each interaction tile in H will bind four specific S tiles together in a way that reflects neighborhood constraints between shared S tiles in structure H. These H tiles are unique to structure H and do not occur in the assembled A or M structures. Tiles in a 1:1 stoichiometric mix of S+H, S+A or S+M will have no promiscuous interactions and will assemble H, A or M respectively, as with prior work on uniquely addressable structures [70]. But a 1:1:1:1 mix of S+H+A+M, henceforth called our SHAM mix, is capable of assembling three distinct structures. This additive construction is analogous to Hebbian learning of multiple memories in Hopfield neural networks [75, 74].

Extensive promiscuous interactions present in the SHAM mix could in principle lead to unplanned chimeric structures and aggregates, but these can be largely prevented by cooperative effects such as proofreading[76, 77]. Much like with neural networks, random arrangement of tiles ensures a degree of statistical proofreading inherent to multifarious structures[74]. We further modified tile placement to ensure proofreading while also maximizing the number of shared species (see Methods in Appendix B). The resulting design, in

Figure 3.2[b], has 168 tiles shared across all three structures and 203 additional tiles shared across a pair. We performed sequence design using tools from [72] to reduce unintended interactions and secondary structure.

To test whether proofreading was sufficient to combat promiscuity and to test the unbiased yield of different structures, we annealed all tiles at equal concentration in solution over 150 hours from 48°C to 46°C. Atomic force microscopy (AFM) revealed a roughly equal yield of all three structures (Figure 3.2[c]), despite the inequivalent thermodynamic energies of H, A, and M. These results indicate that yield is kinetically controlled and no one structure has intrinsically favorable kinetic pathways on the timescales probed here. Additionally, we did not observe significant chimeric structures or uncontrolled aggregation, indicating that proofreading was functioning as desired. A fraction of structures formed, however, were missing fragments at two specific corners, which could arise from asymmetric growth kinetics or lattice curvature [69].

In our system, kinetic pathways can be controlled by concentration patterns because the target structures differ in co-localization of tiles. We can model such selectivity by adapting classical nucleation theory [78]. The free energy of a structure of size N_{tiles} with B_{bonds} total bonds is $G(A) = \sum_{\text{bonds}} G_{\text{bond}} - \sum_{\text{tiles}} G_{\text{tiles}}$ where G_{bond} is the energy of each bond in units of RT and $G_{\text{tiles}} = \log c_i/c_0$ is the chemical potential (equivalently, translational entropy) of each tile i at concentration c_i. This free energy has competing contributions that scale with the area and perimeter and hence is maximized for structures of a critical size K. The formation of such critical nucleation seeds is often rate-limiting; once structures of this critical nucleus size are created, subsequent growth is faster and mostly 'downhill' in free energy. The nucleation rate can be computed using an Arrhenius-like approximation $\eta \sim e^{-G(A_s)}$, where $G(A_s)$ is the free energy of a critical nucleation seed s, and summing over all potential critical seeds s.

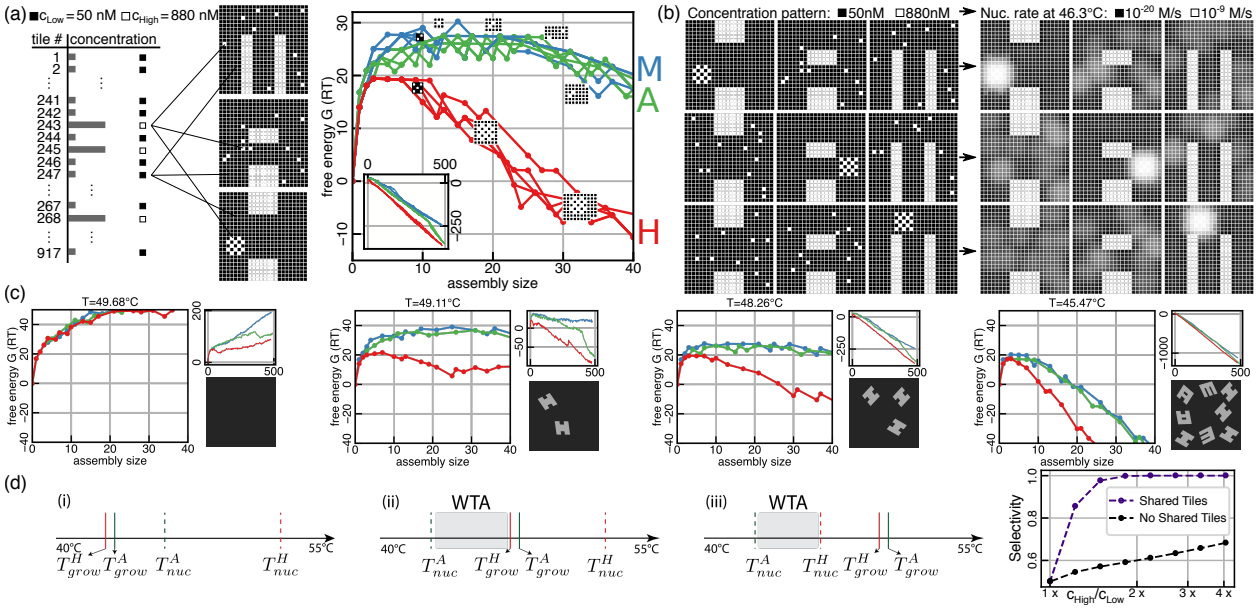

Figure 3.3: **Simulations and theory show that enhanced concentration of shared tiles selectively enhances nucleation rate of structures in which high concentration tiles are co-localized.** (a) A pattern enhancing the concentration of shared tiles co-localized in H but relatively dispersed in A and M. Nucleation pathways for H climb a lower nucleation barrier, with smaller critical nucleation seeds, as seen from a stochastic nucleation simulation. (b) Regions predicted to participate in nucleation by the simulation for three concentration patterns (lighter colors correspond to higher participation) (c) Varying temperature trades off selectivity and the scale of pattern recognition for speed of recognition, even with constant tile monomer concentrations. (d) With tile monomer depletion, if assembly occurs at temperatures $T < T^H_{nuc}$ and $T < T^H_{grow}$ allowing nucleation and growth of on-target structures (here, H) but too high $T > T^A_{nuc}$ for off-target nucleation (A, shown, or M), a winner-take-all (WTA) effect enhances selectivity compared to systems with no shared components (right) illustrated using simulations of a toy model, see Extended Figure B.7.

While such analyses are often applied to homogeneous crystals with uniform concentration $c_i = c$ of components through the structure, heterogeneous concentration patterns require a new kind of analysis. Identifying critical seeds is now more involved since they can be arbitrarily shaped, e.g. paying in perimeter to gain high concentration tiles. We implemented a stochastic sampling algorithm to estimate the nucleation rate of a structure with an uneven concentration patterns c_i, with results in Figure 3.3.

We can gain insight into how nucleation performs pattern recognition using the example

shown in Figure 3.3[a], where the concentrations of some shared tiles in the SHAM Mix have been enhanced. If the high concentration tiles are co-localized in structure H, such a pattern will lower kinetic barriers for the nucleation of H while maintaining high barriers for A and M since those same high concentration tiles are scattered across the structure in A and M. The area K over which co-localization promotes nucleation depends on the temperature and the typical concentration of high tiles and e.g., can be estimated from the size of critical seeds predicted by classical nucleation theory. The nucleation rate of a structure is high if it contains contiguous regions of area K with high average (log) concentration and is low otherwise. Since the area of critical seeds, and hence K, is generally larger at high temperatures, we expect a trade-off between speed and complexity of pattern recognition (Figure 3.3), with more subtle discrimination at higher temperatures (large K)—at the expense of slower experiments—and lower discriminatory power at lower temperatures (small K). See Figure 3.3[c].

To experimentally assess the kinetics of nucleation, we designed distinct fluorophore/quencher pairs on adjacent tiles in four locations on each shape, using tiles not shared between shapes (Figure 3.4). Each pair quenches when the local region of that specific structure assembles, providing a real-time indicator of nucleation or growth in different parts of a structure (Extended Figure B.3).

To experimentally characterize the basis of selectivity, we systematically varied the location of a 5×5 checkerboard pattern of high concentration tiles—a 'flag'—in each of the three shapes, through 37 total locations. We enhanced concentrations of only shared tiles in the SHAM Mix, and thus did not create additional thermodynamic bias towards any one structure. We ramped the temperature down slowly, from 48°C to 46°C, the expected range of melting temperatures for structures, to provide robustness to potential sequence or regional variations in precise nucleation temperatures.

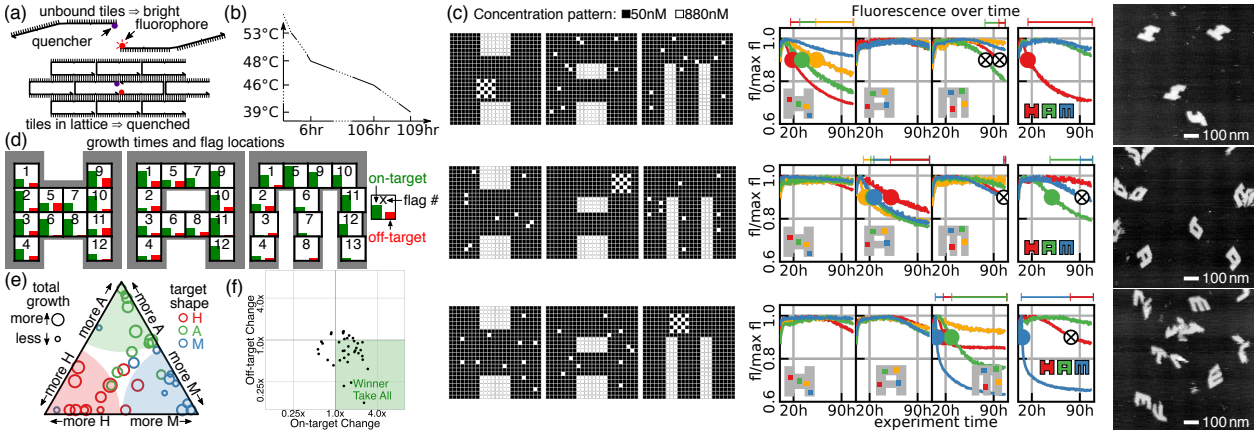

Figure 3.4: **Enhanced concentrations of shared tiles can selectively nucleate structures in which those tiles are co-localized.** (a) We incorporated four different fluorophore/quencher pairs on adjacent tiles in four locations on each structure, allowing the choice of any four locations as labels: quenching of a label indicates growth of that local region on that structure. (b) Samples were annealed with the temperature protocol shown here, spending a majority of the experiment between 48 and 46 °C. (c) We prepared 37 different patterns of concentration (three shown here) that had 12 or 13 shared tiles of increased concentration (16.6x) in checkerboard pattern in a particular 5×5 location in H, A, or M. Fluorophore quenching in multiple samples, and AFM imaging, showed that many patterns resulted in selective nucleation of the shape with the checkerboard pattern; results for each location are summarized in (d) and (e). (f) Patterns enhance nucleation on the shape with localized high concentrations while simultaneously decreasing nucleation of the other two shapes for most experiments. Here, values are given relative to the fractional quenching observed for an equimolar SHAM Mix, described in Extended Figure B.6.

As shown in Figure 3.4[c], when the pattern localizes high concentration species in H, the fluorophore in the expected nucleation region of H quenched first and rapidly; we then observed a delayed drop in the fluorophore signals from other parts of the same structure, indicating growth. Fluorophores on other structures do not show a corresponding quench, indicating selective nucleation and growth of one structure out of three, based on concentration patterns. The summary in Figure 3.4[d] shows that nucleation does usually occur earlier in structures where the high concentration tiles are localized than in competing structures where those tiles are scattered. We then imaged resulting samples using an AFM to confirm that fluorophore quenching corresponded to selective self-assembly.

In multifarious systems, we expect enhanced selectivity because of a competitive suppres-

sion of nucleation. Let T^H_{nuc}, T^H_{grow} be temperatures above which nucleation and growth of H are insignificant on experimental timescales (and similarly for A, M). If $T^A_{nuc} < T^H_{grow}, T^H_{nuc}$, we can use an annealing protocol that spends sufficient time in a temperature range $T^A_{nuc} < T < T^H_{grow}, T^H_{nuc}$ where the on-target H can nucleate and grow out before the off-target A can nucleate at all (gray region in Figure 3.3[d](ii),(iii)). We then expect a winner-take-all (WTA) effect in which the assembly of H depletes shared tiles S and thus actively suppress nucleation of A. As shown in Figure 3.4[f], we see evidence for such a WTA effect in most experiments (see also Extended Figure B.6). Such a winner-take-all effect can enhance the effect of nominal differences in nucleation kinetics at a fixed temperature and deserves further investigation.

Our work thus far shows the space of all concentration patterns $\mathbf{C} = \mathcal{R}^n$ is composed of regions that result in the selective assembly of each of H, A and M respectively. These regions may be separated by decision boundaries or by regions where selectivity is low. In this way, selective nucleation can be seen as solving a particular pattern recognition problem based on which tiles are co-localized in the designed structures.

We then asked if our system could solve an arbitrary image classification problem like that in Figure 3.5[a]. Here, each image corresponds to a specific concentration pattern if the grayscale value at each pixel position (i, j) is interpreted as the concentration of a specific tile $f(i, j)$. Note that images in one class share no more resemblance than images across classes, e.g., class H is Hodgkin, Hopfield, a Horse etc. In this way, the number of distinct images per class (6 in the experiments presented below) tests the flexibility of decision surfaces inherent to this self-assembling molecular system as a classifier.

We found that such arbitrary pattern recognition problems can be potentially solved by our tile system by optimizing the pixel-to-tile map $f(i, j)$. We used an optimization algorithm (see Methods in Appendix B) on $f(i, j)$ that sought to map high concentration pixels in each

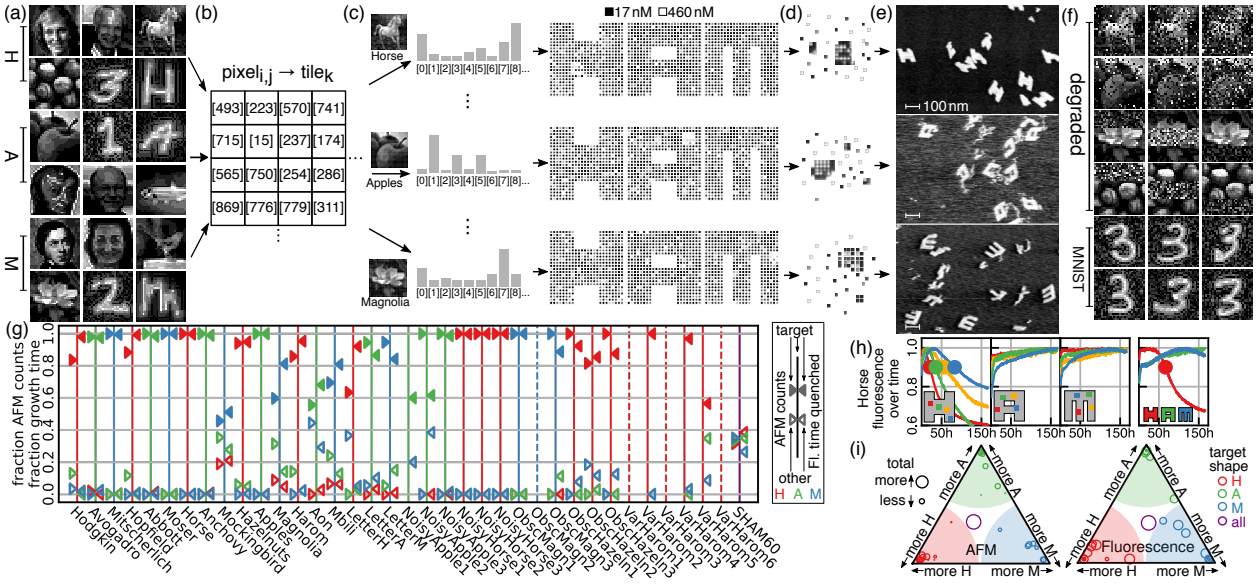

Figure 3.5: **The distinct co-localization patterns in the three structures are expressive enough to classify up to 18 images, even after distortion, where each pixel value is mapped to a tile concentration.** (a)-(e) Training: Selecting 18 images (30×30 pixels), we computationally searched for one pixel-to-tile assignment that would map each image to a concentration pattern with high concentrations localized in the shape associated with the image and distributed in other shapes; (f) For testing, we also considered 18 distorted variations of these images, using the same assignment. (g) In all trained images, and most distorted images, both fluorescence and AFM results showed that the associated concentration pattern for each image resulted in selective nucleation of the correct corresponding shape. The results across all 36 pattern recognition experiments for both AFM imaging and fluorescence monitoring are summarized, averaged over different samples with different fluorophore configurations. Arrows indicate relative fraction of quenching time or number of shapes counted in AFM images. Line color indicates desired shape for the pattern (purple: uniform SHAM Mix at 60 nM) in each sample. Dashed lines - no significant quenching or shapes seen under AFM. (h) Sample fluorescence traces and AFM images for the 'Horse' concentration pattern. (i) Same data as (g); here, proximity to triangle corners indicate relative fractions for different shapes.

image (e.g., Moser) to co-localized tiles in the corresponding on-target structure (here, M) to enhance nucleation while mapping those same pixels to scattered tiles in undesired structures (here, A and H). Note that this map $f(i, j)$ is simultaneously optimized for all images and not independently for each image. Hence no map $f(i, j)$ might be able to perfectly satisfy all the above requirements simultaneously for all images in all classes; we explore the limits of this classifier through simulations in Extended Figure B.9.

In experiments, we then enhanced concentrations of tiles in the SHAM Mix in accordance to each of the 18 training images (using the optimized $f(i, j)$) and annealed each of the 18 mixes with the same ramp protocol as earlier. We found that the 18 training images yielded highly selective correct nucleation, as verified by AFM imaging and real-time fluorescence quenching (Figure 3.5 and Extended Figure B.8).

We also tested 12 degraded images and 6 alternate handwriting images (Figure 3.5[b]). We find successful pattern recognition for random speckle distortions, and all but one partly obscured image. The ability to recognize distorted images, not present in a training set, is a critical aspect of learning in neural networks since it tests the ability to generalize. A given neural network architecture can be naturally robust to certain families of distortions (e.g., convolution networks can handle translation) but not others (e.g., dilation). By nature, our self-assembling system does not allow for robustness to, say, translations but random uncorrelated pixel flips are easily corrected by the cooperative nature of nucleation.

Discussion

We have demonstrated high dimensional pattern recognition by exploiting competitive nucleation in a system of molecules with multiple crystal structures i.e., crystal polymorphism, first described by Mitscherlich [79]. Indeed, many scientific and industrial applications rely on biasing nucleation towards one of the crystal polymorphs by tuning annealing protocols. Our work updates these classic ideas for heterogeneous crystals where the number of distinct components is of the size of the crystal itself[80, 81]. This new heterogeneous context introduces novel elements such as pattern recognition and winner-take-all nucleation.

More broadly, our work demonstrates kinetic control of non-equilibrium multi-component self-assembly that can be exploited even in single target self-assembly. Such kinetic control can enhance yields by avoiding kinetic traps [80] and ensuring that nucleation occurs at a user-specified location. Similar kinetic control has been seen in polymer folding, e.g., in

synthetic DNA origami [82] and natural co-transcriptional folding of RNA [83].

Unlike many prior relationships between physics and computation, the computational ability of our system relies on the inherently stochastic nature of nucleation; monomers must make many unsuccessful attempts at forming a critical seed in both on- and off-target structures, with repeated backtracking in order to realize complex pattern recognition. If instead nucleation were greedy and deterministic, always taking steps to locally minimize free energy without backtracking, the system would not necessarily discover the minimal nucleation barrier for complex patterns, limiting the computational capacity. In this way, the system presented here, in mechanism, resembles stochastic local search algorithms used, e.g., for boolean satisfiability problems. In fact, trade-offs inherent to stochastic algorithms emerge from physical constraints in our work; e.g., at higher temperatures, nucleation is slower but critical nucleation seeds are expected to be larger. Consequently, the pattern recognition problems solved are more complex than in faster experiments carried out at lower temperatures[84]; see Extended Figure B.9. Finally, the nucleation-driven pattern recognition here can potentially be combined with growth-driven abilities, e.g., the nucleation of 2-d facets in 3-d structures.

The pattern recognition functionality here is much like that of an associative neural network [75], recognizing subtle, high-dimensional patterns even in the presence of particular forms of noise. Further, much like Hopfield's model of associative memory, the molecular neural network presented here is naturally expandable with potentially learnable interactions. Our system can be expanded to assemble a fourth structure without modifying any existing tiles by simply adding a fourth set of unique tiles that mediate new interactions between shared tiles. This additive procedure resembles the procedure for adding stable states in Hopfield associative memory via superposition of the corresponding interactions for each desired state , especially in continuous attractor networks with spatial structure [85, 86, 87]. If the interaction mediating unique tiles in our system can be enzymatically created in

response to environmental inputs or conditions our molecular neural network can potentially learn new pattern recognition behaviors from presented examples [88, 89, 90].

Finally, though our experimental system exploits DNA in non-biological setting, the principles behind our work hold lessons for biology. Molecular biology is often presented as a story of intentional specific interactions between numerous molecular agents. As a consequence, promiscuity is seen a deleterious perturbation that degrades performance. However, in other contexts, including neural networks, many promiscuous interactions are known to collectively provide a function if the interactions are cooperative [67]. Our work adds to a small but growing list of examples [91, 92, 93] of functionality that exploits molecular promiscuity and cannot be understood as perturbations of conventional picture with specific interactions.

Acknowledgments

We thank Michael Brenner, Aaron Dinner, David Doty, Deborah K. Fygensen, Stanislas Leibler, Richard M. Murray, Lulu Qian, Paul W. K. Rothemund, Chris Thachuk, Grigory Tikhomirov, Damien Woods, and Zorana Zeravcic. We thank Lulu Qian (use of FastScan AFM and experimental advice), Paul W. K. Rothemund (experimental advice), Richard M. Murray (use of Echo), Deborah K. Fygensen (advice on fluorescence monitoring of nucleation with qPCR), Damien Woods (advice on sequence design and nucleation theory), Chris Thachuk (advice on Echo), Grigory Tikhomirov Zorana Zeravcic, Michael Brenner, Stanislas Leibler, Aaron Dinner (nucleation discussions)

This work was supported by National Science Foundation grant NSF CCF-1317694, the Evans Foundation for Molecular Medicine, European Research Council grant 772766, Science Foundation Ireland grant 18/ERCS/5746, and the Carver Mead New Adventures Fund. JOB, AM were partially supported by the University of Chicago Materials Research Science and Engineering Center, which is funded by National Science Foundation under award number

DMR-2011854. AM acknowledges support from the Simons Foundation.

CHAPTER 4

CONTROL OF NUCLEATION THROUGH CONCENTRATION PATTERNS IN HETEROGENEOUS SELF-ASSEMBLY

Adapted from work being prepared for publication [1]

Abstract Heterogeneous self-assembly, where many diverse components spontaneously aggregate into a large complex, is an important process in many systems. The complex dependence on assembly component concentrations and binding energies allows for interesting modulation of nucleation rates from a broad array of control parameters, but the calculation of this rate is a difficult problem. Here, we use rare-event sampling to estimate nucleation rates in such systems to elucidate the pattern-discriminatory capabilities of such systems. We show that even small changes in the spatial arrangement of a concentration pattern can lead to large changes in nucleation rates for heterogeneous systems, and explore the complex interplay between length-scales, time-scales, and discrimination in these systems. Our approach provides mechanistic insights into the sculpting of kinetic pathways and control in out-of-equilibrium self-assembly systems using spatial concentration patterns.

While heterogeneous self-assembly is a general phenomenon not limited to any one particular system, we use a lattice of single-stranded DNA tiles (SSTs) [69] as our model system because of the feasibility of experiments. In simulation, we assume mono-energetic bonds for designed interactions and assume off-target bonds have no energy contribution. Kinetic simulations are carried out in XGrow [94]. As in Chapter 3, we model the energy of assemblies as simply dependent upon the concentrations of the incorporated species and the number of bonds formed, and we refer to the dimensionless bond strength throughout this chapter as G_{se}.

1. Author List: Constantine G. Evans, Jackson O'Brien (co-first author) and Arvind Murugan. CGE and JOB performed simulations. JOB wrote the text.

In order to effectively analyze the pattern discrimination intrinsic in these heterogeneous self-assembly systems, we require a method to efficiently calculate nucleation rates. In general, this is a difficult problem because the system is out-of-equilibrium, the transition from monomers to completed assemblies can take large numbers of individual attachment/detachment reactions, and rate calculations must be repeated for every change in temperature or tile concentration.

For systems with a large mismatch in time-scale between individual reaction steps and reactions of interest, estimating event frequency directly from dynamics is extremely inefficient. So called "rare events" where stochastic fluctuations allow traversal of a free energy barrier are an important phenomenon in a large variety of systems. As a result, many rare event sampling algorithms have been developed to efficiently sample such processes. Here, we employ Forward Flux Sampling [95, 96, 97, 98] In short, this method estimates transition rates between two states of interest by ratcheting and cacheing trajectories between these two states according to a user-defined, strictly-increasing order parameter, λ. Here, our states of interest are un-assembled monomers (A) and a fully grown out lattice (B), and the number of components in an assembly is our order parameter of choice. States are initialized as monomers, and simulations are allowed to run until a chosen value of the order parameter, λ_0 is achieved. When this occurs, those states are cached. By repeating this process many times, a dimensionful flux, $\Phi_{\text{mons},0}$ out of the initial basin can be computed. Analogous to a ratchet, we begin simulations where the previous progressive simulations left off (i.e. trajectories that reached from A to λ_0), using a random cached state with order parameter λ_0 with uniform sampling . Now, instead of a flux, it is only necessary to keep track of a forward probability, $p_{0,1}$ of reaching the next chosen value $\lambda_1 > \lambda_0$ before melting back to monomers and cacheing these states similar to before. If we keep track of the lineage relationships between our sampled states, this method naturally provides trajectory sampling for successful excursions from A to B, allowing mechanistic probing of the transition.

By computing this probability for a series of increasing λ values, the transition rate can be estimated using the product of those conditional probabilities and the initial flux, see 4.1c.

$$k_{AB} = \Phi_{A,0} \prod_{i=0}^{n-1} p_{i,i+1}$$

with $\lambda_n \equiv B$. Importantly, nowhere in this procedure do we rely on assumptions about equilibrium or bias the microscopic dynamics.

By running many trajectories initialized in a particular assembly, we can estimate the probability of reaching state A or state B first; we call the probability of reaching state B first the committor probability, p_{commit}. States without a tendency toward state A or state B are of particular interest as they capture - we refer to such states with $p_{\text{commit}} \sim .5$ as *critical seeds* (also often referred to collectively as the "transition state ensemble" [98]).

There are several hyperparameters that must be assigned in the process - e.g. the number of λ-surfaces used and the number of trajectories to sample at each surface. Here, we use λ-surfaces separated by a single monomer addition, and we launch a minimum of 1000 trajectories from each λ-surface and increase this number to achieve a constant variance for our p_f estimate.

Results

To demonstrate the subtle dependence of nucleation rates on concentration patterns, we began with four patterns that are all within precisely 4 concentration swaps (i.e. swapping the concentration between two different monomers), shown in order of increasing nucleation rate in 4.2a. While it is quite intuitive that a tighter grouping of the high concentration (here, 20× from a base 50 nM concentration) species will lead to more favorable initial growth, it is impressive that these relatively small changes to the patterns can lead to more than 4 orders of magnitude of change in nucleation rate at the selected temperature.

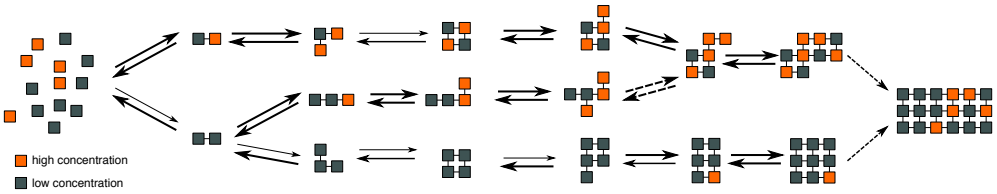

Figure 4.1: **Nucleation pathways and rate-limiting steps in multi-component self-assembly can change dramatically with small changes in the pattern of unequal concentrations of components.**. The rate-limiting nucleation step of self-assembly is set by a balance of bulk and boundary effects, and is the lowest of the highest free energy state along each possible assembly pathway (a min-max problem). With uneven concentrations, the energy landscape is rugged and critical seeds can take on unusual shapes due to the favorable kinetic attachments of high concentration components that are co-localized on the structure. While the min-max problem on a rugged landscape is a computationally difficult problem, we can determine dominant nucleation pathways and thus rate by rare-event trajectory sampling methods.

Further, *a priori* it is difficult to guess whether the "checker" or "cross" pattern might be more effective, and yet these patterns differ by more than an order of magnitude in their nucleation rate. Panels 4.2b-d help build up a mechanistic explanation for the large difference observed for the "checker" and "cross" patterns. Energy trajectories in 4.2b display that at early assembly sizes, the "cross" trajectories are typically at lower energies. However, for larger assemblies, eventually the purple "checker" trajectories typically reach lower energies earlier than the green "cross" trajectories. This trend is emphasized by the two bold example trajectories, and the selected assemblies suggest this comes from the early incorporation of almost exclusively high concentration species for the "cross" pattern leads to an irregular geometric structure with a large perimeter-to-area ratio which results in relatively high energy assemblies even after many monomers have been incorporated. On the other hand, the checkerboard isn't able to include as many high concentration species initially, but the regular grid of high concentration species effects a roughly square structure.

This trend is made a bit more precise in the forward probability plots shown in 4.2c. As expected, the enhanced attachment rates of the contiguous high concentration species in the "cross" relatively enhance that pattern's forward probability until size 5. For larger

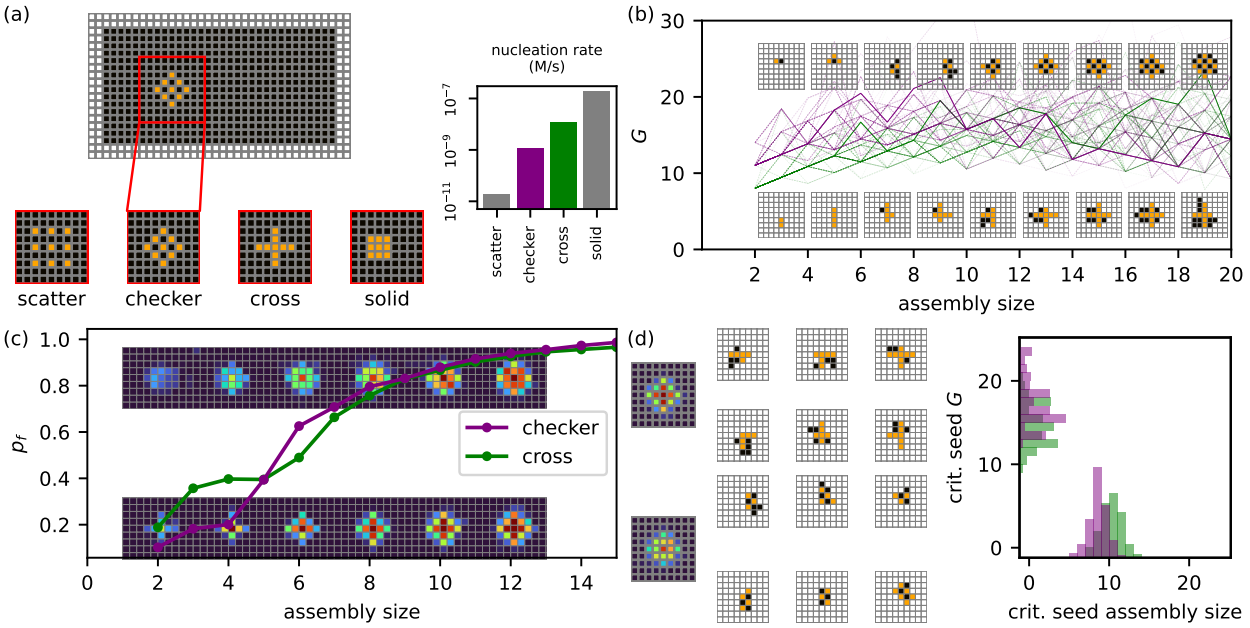

Figure 4.2: **Subtle changes in concentration patterns lead to significantly altered kinetic pathways and overall nucleation rates** (a) Here we compare four patterns with the same number of high (20x) concentration components (shown in orange) on a 16×32 grid at a fixed $G_{se} = 5.2$: the cut outs below show the same 9×9 red-outlined region. Each pattern differs from adjacent patterns only in the location of 4 high concentration species which have only been shifted by one tile on the lattice. Yet each small change causes nucleation rates to differ by a factor of ~ 10, as shown in the bar graph on the right. (b) Energy along sampled trajectories for checkerboard pattern (orange) and cross pattern (blue). (inset: configurations along representative trajectories (bold).) (c) The forward probability p_f describes the probability assemblies of a particular size will evolve to a structure of the next size (usually one tile larger) without dissociating back to monomers. Heat maps shown along the top (checkerboard) and bottom (cross) indicate the probability a particular tile is included in an assembly of a given size. Up until size 5, the cross pattern has a significantly enhanced p_f because of the 5 contiguous high concentration tiles located at the center of the cross. This initial advantage more than compensates for lower p_f for the cross at later assembly stages. (d) Critical seeds along sampled trajectories for the cross pattern (top) and the checkerboard pattern (bottom). (left) Composite critical seed heat maps show the probability a particular tile is included in the sampled critical seeds. (middle) 6 example critical seed assemblies are shown for each pattern. (right) Histogram of energy and size of critical seeds. Critical seeds for cross are larger,on average, than those for the checkerboard but have lower energy because seeds for the cross can readily accommodate a contiguous region of high concentration tiles.

structures, the "checker" pattern has a small but visible advantage. The heatmaps along the bottom show that the central contiguous region of 5 high concentration species at the center of the cross has a very high probability of inclusion, but the "checker" heatmaps along the top demonstrate a higher probability of a square structure. The critical seeds shown in 4.2d show that despite the delayed barrier for the "cross" pattern (its critical seeds are on average larger than for the "checker" pattern), the average height of this barrier is smaller and ultimately results in more favorable nucleation kinetics for the "cross".

From intuitive simple models such as classical nucleation theory, we know that at higher temperatures our system will have correspondingly larger critical seeds (i.e. assemblies will require more correctly attached monomers before further additions become favorable). To explore the consequences of this changing scale on concentration pattern dependent nucleation rates, we constructed a library of 100 random patterns with exactly 9 high concentration species randomly distributed within a $k \times k$ square for five different k values (i.e. 500 total patterns). See Figure 4.3a for two examples for $k = 8$. For this library of patterns, increasing G_{se} (corresponding to decreasing temperature) in general increasing nucleation rates as shown in Figure 4.3b,top. Additionally, patterns with a smaller length scale tend to nucleate fastest at all temperatures as expected. Interestingly, the relative spread of these average values is largest at intermediate G_{se} values. At high G_{se} (low temperature), we see all 5 curves coalesce within an order of magnitude (in fact, $k = 7$ and $k = 8$ become indistinguishable in mean for $G_{se} > 5.5$) whereas near $G_{se} = 5.2$ the average nucleation rates spanned roughly 4 orders of magnitude. Even for a given k, the distribution about the mean for the 100 nucleation rates varies dramatically over G_{se} as shown in Figure 4.3b, bottom for $k = 5$, reaching a maximum range of nucleation rates around $G_{se} = 5.2$

To better understand the discriminatory power of nucleation, we selected two patterns for $k = 8$ (the same patterns shown in Figure 4.3b). Figure 4.3c shows the ratio of these two nucleation rates over several different G_{se} (ranging from 5.0 to 6.1) - however, we use

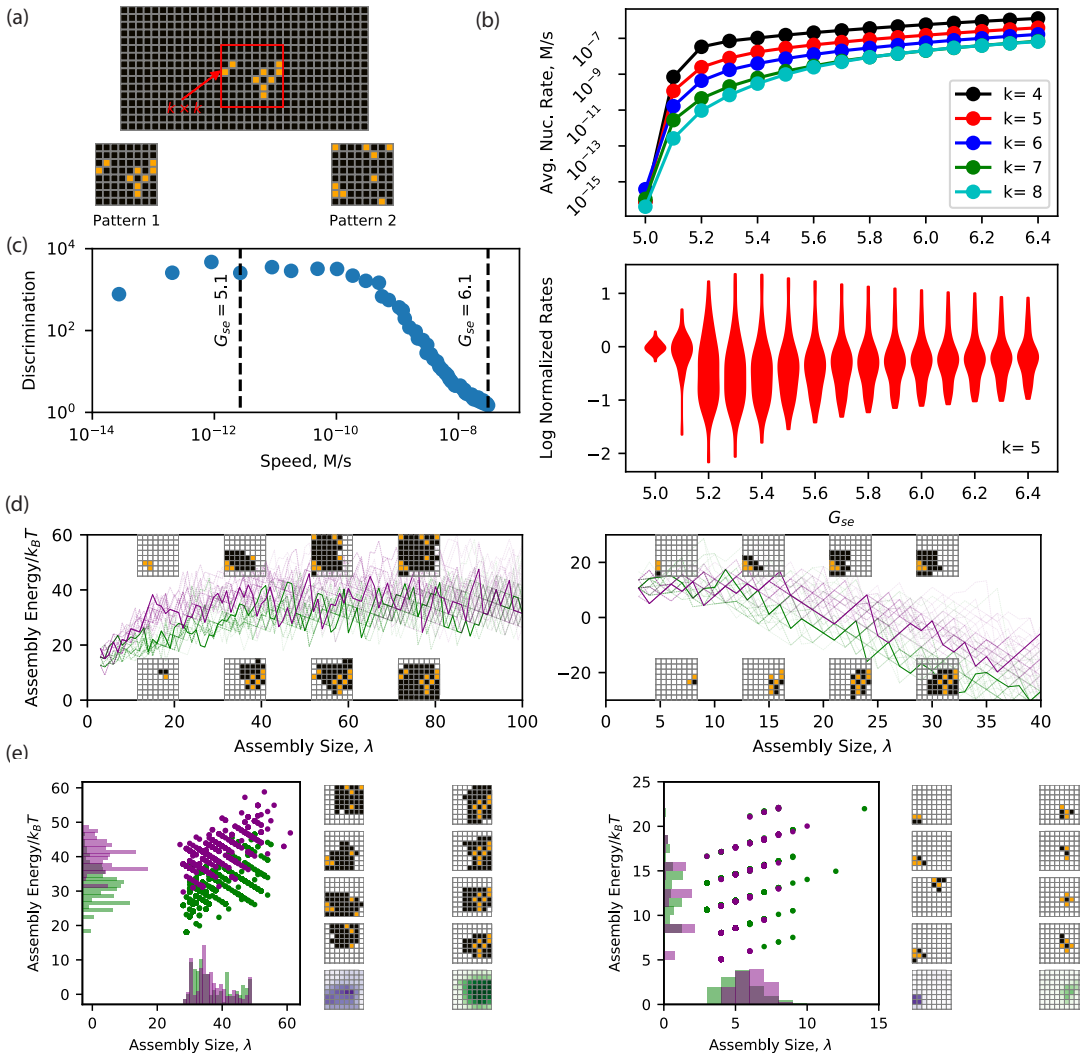

Figure 4.3: **Temperature sets a trade-off between speed and length-scale resolution.** (a) Random patterns with a fixed number of high (20x) concentration components (shown in orange) were generated such that the high concentration species fit inside a $k \times k$ square. The cut outs below show 2 such examples for $k = 8$. (b) (top) Average nucleation rates increase for higher G_{se} and for smaller k localization. (bottom) Violin plots for $k = 5$ show the distributions of these rates reaches a maximal functional range at intermediate G_{se} ($\log_{10}$ of ratio to mean rate for each G_{se}). (c) Discrimination (the ratio of nucleation rates from Pattern 1 to Pattern 2 from panel (a)) decreases as nucleation rates increases. (d) Energy trajectories of assembly at the G_{se} values shown with dashed lines in (c) ($G_{se} = 5.1$ on top, $G_{se} = 6.1$ on bottom)(Pattern 1 in green and Pattern 2 in purple). Example growth states for a particular trajectory (bold) are shown along the top for Pattern 2 and the bottom for Pattern 1 (e) (left) For 4000 sampled critical seeds, a scatter plot of size against energy for both patterns at $G_{se} = 5.1$ is shown (Pattern 1 in green, Pattern 2 in purple). Example critical seeds are shown for the different patterns at this G_{se} (Pattern 1 on the right column). The purple and green heat maps at the bottom of each column indicate the probability each tile is included in a critical seed. (right) Same as the left but for $G_{se} = 6.1$.

the nucleation rate of pattern 1 as the x-axis to emphasize the changing discrimination can be thought of as a consequence of the speed of the computation. At low $G_{se} = 5.1$ and slow nucleation, these two patterns differ by a factor of $\sim$ 5,000 but are within a factor of ~ 2 by $G_{se} = 6.1$.

There are several perspectives to understand this change in discrimination, but it is perhaps most intuitive to look at the length scale and associated energy barrier for these different G_{se} values. Figure 4.3d and e show sample trajectories and sample critical seeds for both patterns at the two G_{se} values of interest. Almost all critical seeds and maximal energy states along the trajectories occur for assembly sizes larger than $\sim 6 \times 6$ for $G_{se} = 5.1$. For the critical seeds, the source pattern is visually obvious, and the energy histograms are clearly separated by several dimensionless energy units showing the kinetic barriers are well differentiated for the two patterns. In contrast, at large G_{se} the energy barrier generally occurs for assembly sizes less than 10, and the critical seeds and critical seed energy histograms are difficult to distinguish. From this example, it is clear that the system's ability to distinguish different concentration patterns is strongly dependent upon the temperature (and correspondingly the speed) at which the self-assembly is occurring. As a final note, there is another trade-off at high temperatures where the critical seed size becomes much larger than the $k \times k$ pattern and the discrimination between such patterns decreases (though for the parameter choices here the timescale becomes infeasible for experiments).

As a final demonstrative example, we built upon this intuition about changing nucleation scale leading to very different relative nucleation rates to design a pair of patterns whose nucleation rates invert as G_{se} is increased. Figure 4.4a shows the patterns - both contain exactly 16 high concentrations species. In one case, the high concentration species form the perimeter of a 5×5 square (Outline) and in the other there is a $3x3$ solid region of high concentration species with 7 other high concentration species dispersed toward the edges of the structure (Solid). At low G_{se} and large critical seeds, the Outline is more favorable by

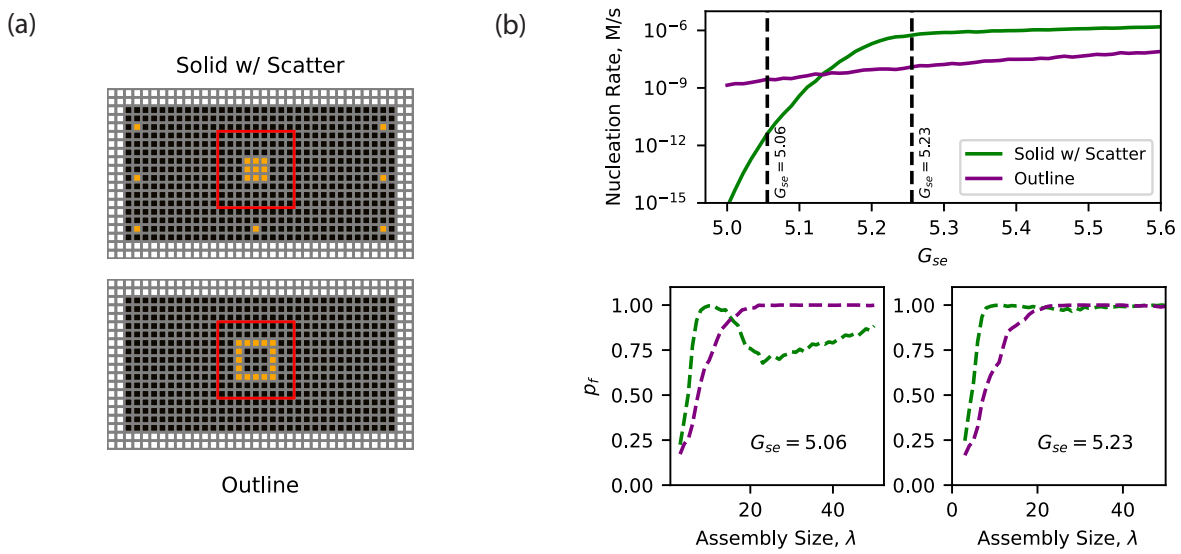

Figure 4.4: **Interplay between temperature and nucleation scale can lead to significant inversions of relative nucleation rates.** (a) Two patterns with an equal number of high concentration (20×) species are shown. (b) (top) Nucleation rates for both patterns are shown as a function of G_{se}. Two G_{se} values of interest are denoted with vertical lines, and the corresponding critical seeds are shown in panel (c). (bottom) Forward probabilities as a function of λ show that at low G_{se} the Solid w/ Scatter pattern faces a significant barrier after growing out the high concentration region in comparison with the Outline pattern. However, at larger G_{se}, the high concentration region in the Solid w/ Scatter pattern is sufficient to allow favorable growth to larger structures, and it reaches this favorable growth more quickly than for the Outline pattern.

several orders of magnitude as shown in 4.4b. Even though the initial $3x3$ structure forms more favorably for the Solid than the initial attachments for the Outline, the structure size is still insufficient for favorable attachment of low concentration species for low G_{se} and p_f decreases from $\lambda \sim 9$ until $\lambda \sim 25$. However, at high G_{se} the initial favorable attachments are sufficient to cross the energy barrier and p_f strictly increases, leading to Solid nucleating orders of magnitude faster.

Discussion

This work establishes the analog control of heterogeneous self-assembly through concentration patterns in the assembly components. By changing the concentration of just a few

components, the nucleation rate can be modulated over several orders of magnitude.

Similar to the classification problem presented in Chapter 3, a high-dimensional input is processed into a simple output (here, a single number) autonomously by the physical interactions between the assembly components. In fact, this computational method largely extends immediately to multifarious systems - however, at current this method ignores depletion effects which were clearly experimentally relevant for the system from Chapter 3 and can also play an important role even in uniquely addressed structures as unproductive kinetic trap states could use up monomer components without ever generating complete assemblies. This generalization of is left to future work.

CHAPTER 5

DISCUSSION AND OUTLOOK

This thesis builds upon our understanding of natural computing systems, expanding on the applications for these systems and elucidating a new computational paradigm surrounding the ubiquitous physical process of nucleation. In terms of immediate practical applications, these proof-of-principal systems could pave the way for context dependent therapeutics triggered either by a temporal signature or correlations between concentrations of a several unique components. Perhaps more importantly, this body of work broadens the current understanding of the tools at the disposal of molecular systems to recognize patterns, and my hope is that this will ultimately lead to the discovery and understanding of novel, biologically-relevant computational phenomena.

APPENDIX A

SUPPLEMENTARY INFORMATION: TEMPORAL PATTERN RECOGNITION THROUGH ANALOG MOLECULAR COMPUTATION

Methods

We first formulated abstract chemical networks with the desired feature detecting properties. Then, largely following the design principles laid out in [55], we implemented these chemical reaction networks as DNA strand displacement reactions. All strand displacement circuits are designed within Microsoft's Visual DSD software described in [9] using default kinetic parameters and concentrations $\in [.05 \text{ nM}, 10 \ \mu\text{M}]$. Then, by adapting the MATLAB code generated within this program, we exposed these circuits to the pulsatile inputs defined in the main text, defined by their duty fraction δ, number of pulses n, and period T. All results shown are from deterministic simulations without leak reactions.

Detailed Chemical Networks

All reactions shown utilize the default kinetic parameters within the Visual DSD software (3×10^{-4} $\frac{1}{\text{nM s}}$ bind, .1226 $\frac{1}{\text{s}}$ unbind corresponding to toe-holds with 4-6 nucleotides [99]). Different reaction rates were achieved by selecting appropriate initial concentrations. The dynamic range we require ($\sim$ 5 orders of magnitude) is also achievable through toehold design [99]. Species with specified initial concentrations are outlined in bold and their values are given in accompanying tables. All species whose initial concentrations are specified and do not have explicit time dependence displayed in the main text are held at their initial concentrations throughout all simulations. In the tables of concentrations below, parenthetical values indicate high and low oscillatory values in the time-varying input.

Pulse Counting

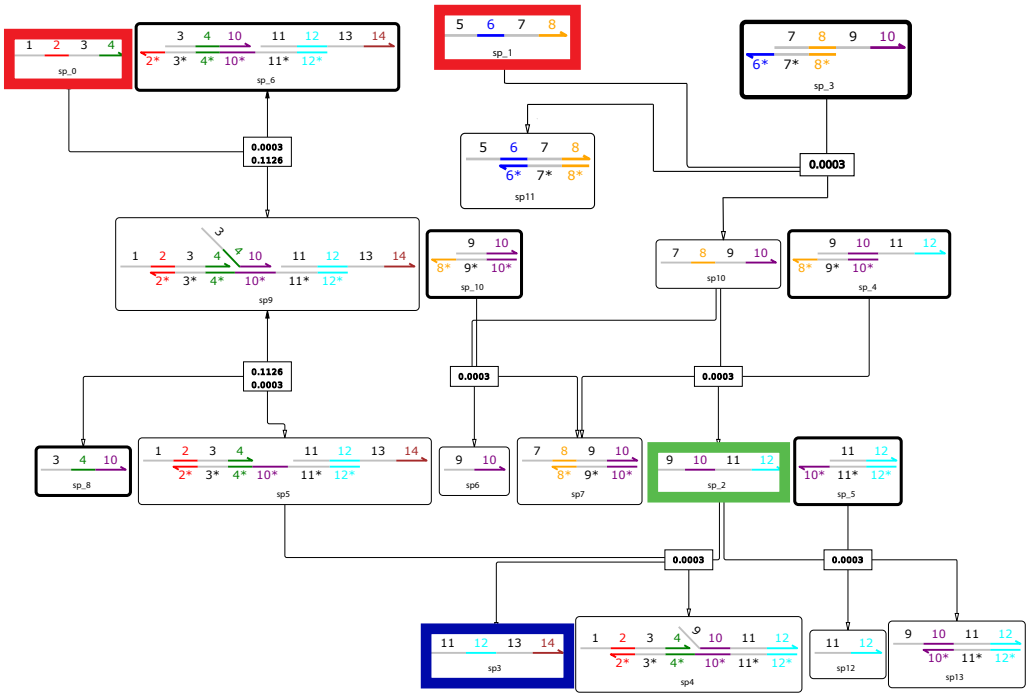

Figure A.1: **Full version of DNA strand displacement reaction network for pulse counting decoder including waste products.** The two species in red had their dynamics directly modulated to the parameters of the input series, with sp_1 (referred to as $\bar{I}$ in the main text) pulsing exactly out of phase with sp_0 (I in the main text). Graphs and labels are generated automatically within the Visual DSD software [9]. See Table A.1 for a list of initial concentrations. Note that the flux G plotted in Figure 2 is defined $G \equiv \sqrt{\mathrm{sp}_2 * \mathrm{sp}_5}$.

Table A.1: Pulse Counter Initial Conditions

Species	Initial Conc. (nM)
sp_0	1 (0)
sp_1	100 (0)
sp_2	50
sp_3	10000
sp_4	10000
sp_5	10
sp_6	10000
sp_8	10000
sp_10	10000

Duty Fraction

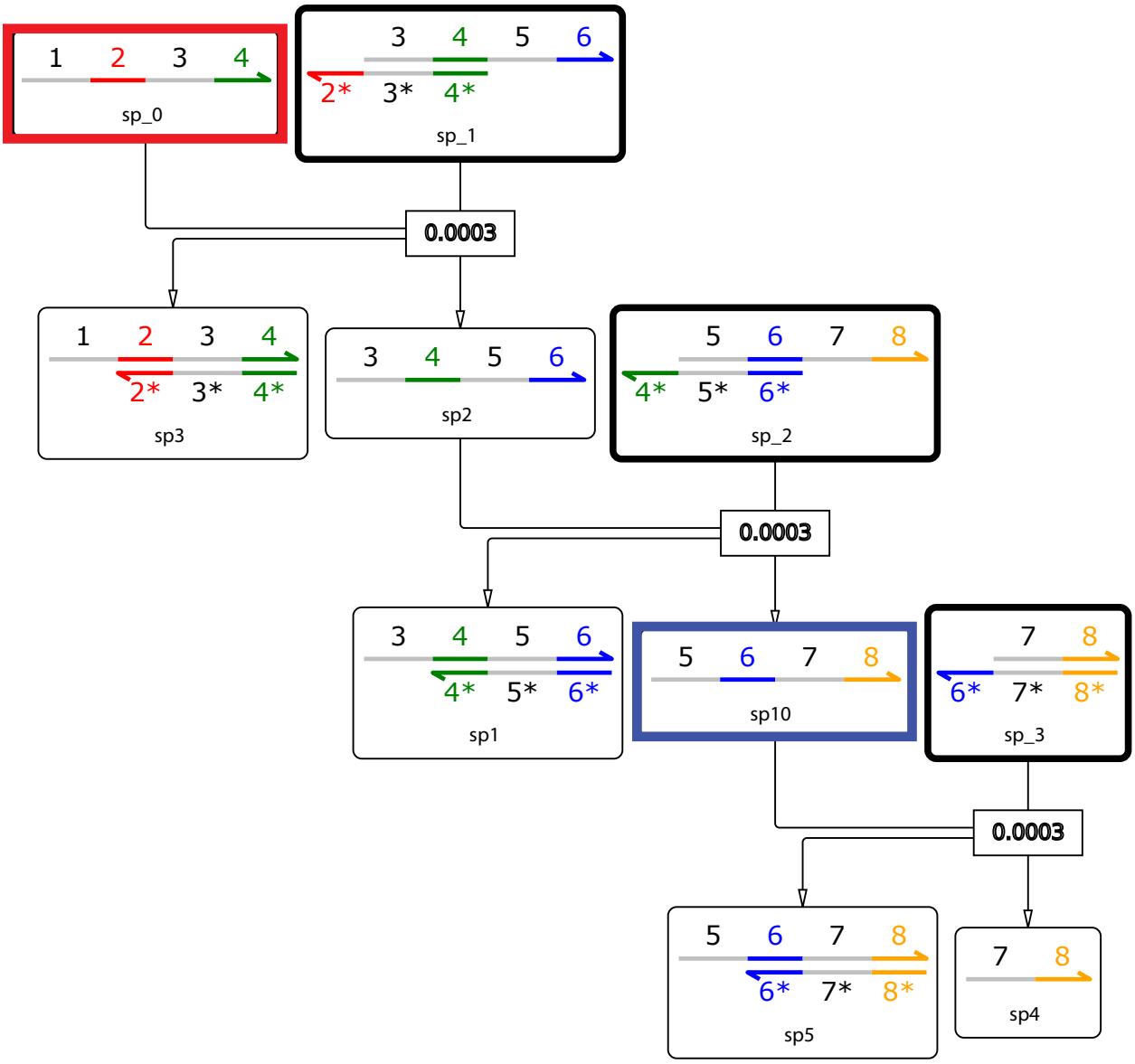

Figure A.2: **Full version of DNA strand displacement reaction network for duty cycle decoding, including waste products.** This circuit effectively takes the moving average of the dynamics of sp_0 and reports it in sp10. See main text Figure 3 for analysis and Figure 5 for performance. Initial concentrations of bolded species are listed in Table A.2.

Table A.2: Duty Fraction Decoder Initial Conditions

Species	Initial Conc. (nM)
sp_0	1(0)
sp_1	100
sp_2	100
sp_3	10

Period Detecting

Table A.3: Period Decoder Initial Conditions

Species	Initial Conc. (nM)
sp_0	1 (0)
sp_1	100 (0)
sp_2	50
sp_3	10000
sp_4	10000
sp_5	10
sp_6	10000
sp_8	10000
sp_10	10000
sp_11	.1375

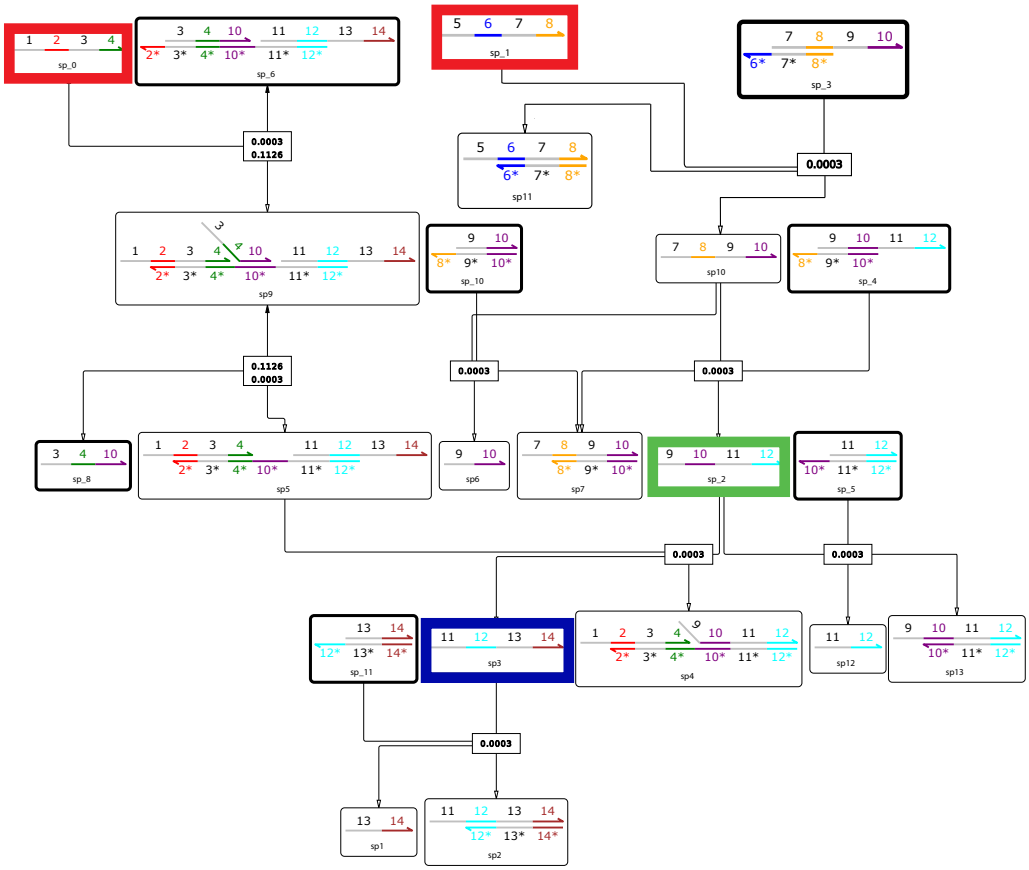

Figure A.3: **Full version of DNA strand displacement reaction network for time period decoding, including waste products.** By taking the moving average of the incoming flux $G \equiv \sqrt{\mathrm{sp}_2 * \mathrm{sp}_5}$, this circuit decodes the period of *sp*_0. Initial concentrations of bolded species are reported in Table A.3.

Supplemental Calculations

Pulse Counter

$$\mathrm{I} + \mathrm{A} \xrightarrow{k_1} \mathrm{P} \tag{A.1}$$

$$\mathrm{A} \xrightarrow{\lambda_1} \phi \tag{A.2}$$

$$\bar{\mathrm{I}} \xrightarrow{k_2} \mathrm{A} \tag{A.3}$$

$$50 \tag{A.4}$$

To supplement the analysis presented in the main text, we present the system of corresponding ordinary differential equations governing the evolution of A and P, presented in Equation A.6.

$$\dot{A}(t) = k_2\bar{I}(t) - k_1A(t)I(t) - \lambda_1A(t) \tag{A.5}$$

$$\dot{P}(t) = k_1A(t)I(t) \tag{A.6}$$

The discussion of timescales relies on exponential decays and exponential approaches to steady state exhibited by these differential equation in transitioning between the two states $I = C, \bar{I} = 0$ and $I = 0, \bar{I} = C$ where C corresponds to the finite value presented in Table A.1 for sp_0 and sp10.

When I is turned on, P starts being produced since P requires both I and A to be present (Note that A has a resting state at a high concentration $\frac{k_2C}{\lambda_1}$ when the input I is off). However, turning on I also causes exponential decay of A from $\frac{k_2C}{\lambda_1}$ to 0 on a timescale $\frac{1}{k_1C+\lambda_1}$.Consequently, the term promoting P decays after a short transient. Thanks to the simple form of these equations, we can compute $P(t)$ analytically

$$P(t) = \frac{C^2k_1k_2}{\lambda_1}\frac{(1 - e^{-(\lambda_1+k_1C)t})}{k_1C + \lambda_1} \tag{A.7}$$

which exhibits the dependence of P on the degradation timescale of A, $\tau_{\mathrm{a}} = \frac{1}{\lambda_1+k_1C}$.

Note that any pulse of width,

$$T\delta \gg \tau_{\mathrm{a}}$$

will produce a stereotyped exponential decay profile (i.e. independent of T and δ) for P. This restriction provides one limit on the window of operation for pulse counters.

Finally, once I switches off, A returns to its steady state $\frac{k_2C}{\lambda_1}$ at a timescale $\frac{1}{\lambda_1}$. The

pulse needs to be off for long enough

$$T(1-\delta) \gg \frac{1}{\lambda_1},$$

so that A can be restored to its original state between every pulse. This condition ensures each step up of P will have the same value (the ODE for P only depends on the values of I and A) and can serve as a proxy for pulse number n.

Within the regime defined by the two inequalities, we find that the circuit presented is able to count pulses independent of duty fraction or time period.

Duty Fraction Decoder

The duty fraction decoder is given by the simple network,

$$\mathrm{I} \xrightarrow{k} \mathrm{I} + \mathrm{P} \tag{A.8}$$

$$\mathrm{P} \xrightarrow{\lambda} \phi \tag{A.9}$$

and consequently, $P(t)$ is governed by,

$$\dot{P}(t) = kc(t) - \lambda P(t) \tag{A.10}$$

where the input concentration $I = c(t)$ is taken to vary as $c(t)$. The solution to the above equation can be written in terms of a exponentially decaying kernel,

$$P(t) \;=\; k \int_{-\infty}^{t} c(t') e^{-\lambda(t-t')} dt'. \tag{A.11}$$

P(t) will be at its smallest value at the front edge of a pulse; this value, after $n-1$ pulses

can be computed from the above,

$$P(t=(n-1)T) \;=\; \frac{k}{\lambda}\frac{(e^{T\lambda\delta}-1)(1-e^{-(n-1)T\lambda})}{e^{T\lambda}-1} \tag{A.12}$$

P(t) will be largest right at the end of the pulse; this value, after n pulses can be computed to be,

$$P(t=(n-1+\delta)T) \;=\; \frac{k}{\lambda}\frac{(e^{T\lambda\delta}-1)(e^{(1-\delta)\lambda T}-e^{-(n-1+\delta)T\lambda})}{e^{T\lambda}-1}. \tag{A.13}$$

We consider the average of these two quantities as representing the readout value of P. Adding and simplifying yields

$$\bar{P} \;=\left(\frac{e^{\lambda T\delta}-1}{2(e^{\lambda T}-1)}\right)\;(1+e^{\lambda T(1-\delta)}-e^{-\lambda T(n-1)}-e^{-\lambda T(\delta+(n-1))}). \tag{A.14}$$

In the limit of

$$n \gg 1/(\lambda T), \quad T \ll 1/\lambda,$$

$\bar{P} \approx \frac{k\delta}{\lambda}$ is proportional to the duty fraction δ but independent of n and T.

Finally, we can consider the difference between the maximal and minimum values of $P(t)$ as a measure of the variation away from this readout. Taking the difference of Equations A.13 and A.12 yields

$$\Delta P \;=\left(\frac{e^{\lambda T\delta}-1}{(e^{\lambda T}-1)}\right)\;(-1+e^{\lambda T(1-\delta)}+e^{-\lambda T(n-1)}-e^{-\lambda T(\delta+(n-1))}). \tag{A.15}$$

For large n and small λT, this approximates to $\Delta P \approx \frac{k\delta}{\lambda}\lambda T(1-\delta) \approx \bar{P}\lambda T(1-\delta)$. Thus, in the limit defined above, $T \ll \frac{1}{\lambda}$, the fractional variation about $\bar{P}$ defined as $\frac{\Delta P}{\bar{P}}$ is small, and $\bar{P}$ is a reliable readout for the duty fraction δ.

APPENDIX B

SUPPLEMENTARY INFORMATION: PATTERN RECOGNITION IN THE NUCLEATION KINETICS OF NON-EQUILIBRIUM SELF-ASSEMBLY

Methods

Multifarious DNA tile system design

Prior theoretical proposals[74, 100] for multifarious mixtures require each component to accept multiple strongly binding partners at each binding site. However, in DNA tile assembly, each binding site can usually only bind its Watson-Crick complement, and not an arbitrary set of other domains. Hence, we used an alternate approach: we assumed three structures made of entirely unique tiles and ran a merging algorithm that attempted to reuse tiles in a roughly checkerboard pattern across structures. The algorithm accepted merges if consequences for unintentional binding between other tiles were minimal. After determining the abstracted shared and tile layout in this way, we designed DNA sequences reflecting this layout.

The three target shapes were drawn on a 24×24 single-stranded tile (SST) molecular canvas[70], at an abstract level without sequences, with each tile in each shape initially a unique tile with glues assigned to bind only to that tile's neighbors within its shape. Edges of the shapes were an exception, instead using a special null glue throughout. In total, this initial design had 2,706 binding domains, and 1,456 tiles. The three shapes were then processed through a 'merging' algorithm that attempted to reuse the same tiles in different shapes. Randomly choosing two tiles in two different shapes (with null glues on the sames sides of each tile, if any), each step of the algorithm attempted to make the two tiles identical by reusing the same four glues in both, and propagating the changed glues to the

complementary glues on the neighbors of each tile (e.g., Extended Figure B.1[a]). To avoid having these changes create undesired growth pathways, for example, allowing chimera of multiple shapes, the modified set was checked after each trial merging step for two criteria from algorithmic self-assembly (Extended Figure B.1[b]). The self-healing criterion requires that, for any correct subassembly of any shape, while attachments of the wrong tile for a particular location may take place by one bond, only the correct tile can attach by two or more bonds[101]. The second-order sensitivity criterion requires that, for any correct subassembly of any shape, if an incorrect attachment by one bond takes place, the incorrectly attached tile will not create a neighborhood where an additional incorrect tile can attach by two bonds, and thus the initial error will be likely to fall off[77]. These two criteria, which are trivially satisfied when every tile and bond is unique to a particular location, continue to be satisfied after each merging step. Thus, we ensure that there is at least a minimum barrier to continued incorrect growth in a regime where tile attachment by two or more bonds is favorable, and attachment by one bond is unfavorable, which is the case close to the melting temperature of most DNA tile assembly systems[102].

The algorithm repeatedly merged tiles that satisfied the two criteria until no further acceptable merges were possible. As each merge could affect the acceptability of later merges by changing the glues around each tile, in order to guide the algorithm toward series merges more likely to be compatible, the algorithm was initially restricted to considering an alternating 'checkerboard' subset of tiles, which, apart from edges, were likely to be merge-able. After exhausting possible merges of these subsets, the algorithm then attempted merges using all tiles in the system. After repeating this algorithm multiple times, and selecting the system with the smallest number of tiles, the final resulting system had 698 binding domains and 917 tiles, with 371 of tiles shared between at least two shapes (Extended Figure B.1[c]).

After the assignment of abstract binding domains to each tile by the merging algorithm, the sequences for the binding domains, and thus tiles themselves, were generated using the

sequence design software of Woods and Doty et al[72]. Tiles used a standard SST design, with alternating 10 and 11 nt binding domains, designed to have similar binding strengths. Null binding domains on the edges of shapes, not intended to bind to any other tiles, were assigned poly-T sequences.

Models of nucleation

To model the dependence of the nucleation rates of the three shapes on patterns of unequal concentration, we developed a simple nucleation model based on the stochastic generation of possible nucleation pathways and critical nuclei. The model estimates nucleation rates by analyzing stochastic paths generated in a greedy manner by making single-tile additions starting from a particular monomer in the system. At each step, all favorable attachments are added and then an unfavorable attachment is performed with probability weighted by the relative free-energy differences of the available tile attachment positions. When multiple favorable attachments are available, the most favorable attachment is made deterministically. This procedure is repeated for many paths over all possible initial positions within the shape considered, and the barrier (highest free energy state visited in "growing" a full structure) is recorded for each path. A nucleation rate is estimated by assuming an equilibrium occupation of this barrier state (Arrenhius' approximation [78]) and summing over the kinetics of the available attachments from this state. See Extended Figure B.4 for a detailed discussion. The approximations here could be improved by running fully reversible simulations, e.g., using Forward Flux Sampling [95], discussed further in Chapter 4.

Fluorophore labels, DNA synthesis, and growth

Sites for fluorophore and quencher modifications were chosen to avoid edges, modify only unshared tiles, and provide a reasonable distribution of locations on each shape. Fluorophores were chosen for spectral compatibility and temperature stability[103]. ROX,

ATTO550, and ATTO647N were paired with Iowa Black RQ, and FAM was paired with Iowa Black FQ. Both fluorophore and quencher modifications were made on the 5' ends of tiles; to sufficiently co-localize fluorophores and quenchers, one tile in the label pair used a reversed orientation (Figure 3.4[a]).

Tiles without fluorophore or quencher modifications were ordered unpurified (desalted) and prenormalized to 400 µM in TE buffer (Integrated DNA Technologies). Strands with fluorophore or quencher modifications were ordered HPLC-purified and normalized to 100 µM. Individual tiles were mixed, in the concentration patterns used for experiments, using an Echo 525 acoustic liquid handler (Beckman Coulter). Samples used TEMg buffer (TE buffer with 12.5 mM $MgCl_2$). Flag experiments used a 50 nM base concentration of unenhanced tiles, and an 880 nM concentration of enhanced concentration tiles, while pattern recognition experiments had tiles with concentrations between 16.6 nM to 450 nM in pattern recognition experiments, quantized into ten discrete values to simplify mixing and conserve material. Fluorophore and quencher-modified tile locations always had tiles mixed at the lowest concentration used in the experiment.

For flag experiments, and pattern recognition of trained images, four samples were prepared per concentration pattern: one sample for each shape with all four fluorophore labels on only that shape, to monitor growth of multiple regions on each shape, and an additional sample with one fluorophore on each shape: ROX, ATTO550 ('five'), and ATTO647N ('six') on H, A, and M respectively. For pattern recognition of test images, for experiment size reasons, only the lattermost sample was prepared.

Samples were grown in an mx3005p quantitative PCR (qPCR) machine (Agilent), in order to provide a program of controlled temperature over time while monitoring fluorescence. Growth protocols began with a ramp from 71°C to 53°C over 40 minutes to ensure all potentially preexisting complexes were melted, and then a slower ramp from 53°C to an initial growth temperature at 1°C per hour. At this point, three different protocols were used.

For constant temperature flag growth experiments, the growth temperature was 47°C, and this was held for 51 hours. For temperature ramp flag growth, the initial growth temperature was 48°C, which was reduced over 100 hours to 46°C. For pattern recognition, a ramp from 48°C to 45°C over 150 hours was used. For constant temperature experiments, fluorescence readings were taken every 12 minutes, and for other experiments, every 30 minutes. After the growth period, temperature was lowered to 39°C at 1°C per 26 minutes. See Extended Figure B.5 and B.8 for temperature protocols plotted as a function of time. Once protocols were finished, samples were stored at room temperature until ready for AFM imaging.

AFM imaging was performed using a FastScan AFM (Bruker) in fluid tapping mode. To achieve better images, two techniques were combined: sample warming to prevent nonspecific clumping of structures, and washing with Na-added buffer to prevent smaller material, such as unbound, single DNA tile strands, from adhering to the mica surface. Each sample was diluted 50x into TEMg buffer with an added 100 mM NaCl, then warmed to approximately 40°C for 15 minutes. 50µL of the sample mix was deposited on mica, then left for two minutes. As much liquid as possible was pipetted off of the mica and discarded, then immediately replaced with added-Na buffer again, and mixed by pipetting up and down. This washing process of buffer removal and addition was repeated twice with added-Na buffer, then once with TEMg buffer to remove remaining Na, before imaging was performed in TEMg buffer. As adhesion of DNA to mica is dependent upon the ratio of monovalent and divalent cations in the imaging buffer, this process was meant to ensure that unbound tiles were washed away during the washing process where Na and Mg were present, while imaging itself took place with only Mg, so that the lattice structures would be more strongly adhered to the surface, resulting in better image quality.

Fluorescence and AFM data analysis

Fluorophore signals are known to be affected by partial assemblies and extraneous factors such as the total amount of single stranded DNA in solution and sequence [103], which prevents quantitative interpretation of absolute fluorescence levels.For this reason, the fluorescence of each fluorophore was normalized to the maximum raw fluorescence value of that fluorophore in that particular sample. The *time* at which the fluorescence signal was lowered by 10% was then used as a measure of fluorophore label quenching. The duration between the point of 10% quenching and the end of the growth segment of the experiment was defined as the "growth time" for that fluorophore label, which was defined as 0 in the event of quenching never reaching 10%. For concentration patterns with four samples with different fluorophore arrangements, the total growth time of a shape was defined as the average of the growth time of the five total fluorophore labels on the shape across the four samples (four in the shape-specific sample, and one in the each-shape sample), while for concentration patterns with only one sample, the growth time of the corresponding fluorophore label was used.

For flag experiments, AFM imaging was done only for qualitative confirmation of the selective nucleation and growth indicated by fluorescence results. For pattern recognition and equal-concentration experiments, however, shapes in AFM images were uniformly quantified. At least one sample of each of the patterns had three 5×5 µm images taken under comparable conditions. The sample corresponding with each image was blinded, and structures were counted independently by each of the four authors, classifying structures as either "nearly complete" or "clearly identifiable" examples of each of the three shapes. For the purposes of analysing pattern-dependent nucleation and growth, no clear distinction between the number of nearly complete and clearly identifiable shapes was found, and so the two categories were summed. Counts were averaged across the three images, then averaged across the counts of the four authors, to obtain a count per shape per 25 μm^2 region for each pattern.

To measure the selectivity of patterns, the fraction of on-target shape growth time, and AFM counts, compared to the sum of shape growth times and AFM counts, was used. The total growth times, and total AFM counts, of the on-target shapes were used to measure overall shape growth.

Pattern recognition training

Images for pattern recognition were selected from several sources, rescaled to 30×30, discretized to 10 grayscale values, and adjusted so that the number of pixels with each value was consistent across all images (see SI for details). Pixel values were converted to concentrations using an exponential formula, $c = (1\text{ nM})e^{3p \log 3}$, where p is a pixel value between 0 and 1. The intention of the numbers used was to make the average tile concentration 60 nM for each image. As each image had 900 pixels and there are 917 tiles in the system, 17 tiles did not have their concentrations set by any pixel; these tile concentrations were uniformly set to the lowest concentration, and the assignment of these tiles was used to ensure that fluorophore label locations did not vary in concentration.

The tile-pixel assignment was optimized through a simple hill-climbing algorithm, starting from a random assignment, where random modifications to the assignment map are attempted at each step and accepted if the move increases the efficacy of the map. This efficacy was quantified through a heuristic function that accounts for relative nucleation rates, location of nucleation sites (with emphasis given to locations that succeeded in the flag experiments, see Figure 3.4, and satisfaction of constraints related to the fluorescent reporters. Because the nucleation algorithm described above is costly, a simplistic model of nucleation based upon the Boltzmann-weighted sum of concentrations over a $k \times k$ window swept over each structure (similar to the model employed in [84]) was used to evaluate relative nucleation rates for a majority of the optimization steps. The more detailed but computationally costly model described above was then employed for an additional several hours in hopes

of improving the mapping. The window-based nucleation (along with all constraints about nucleation location and fluorescent reporters) is employed to explore the capacity of this map training procedure in Extended Figure B.9.

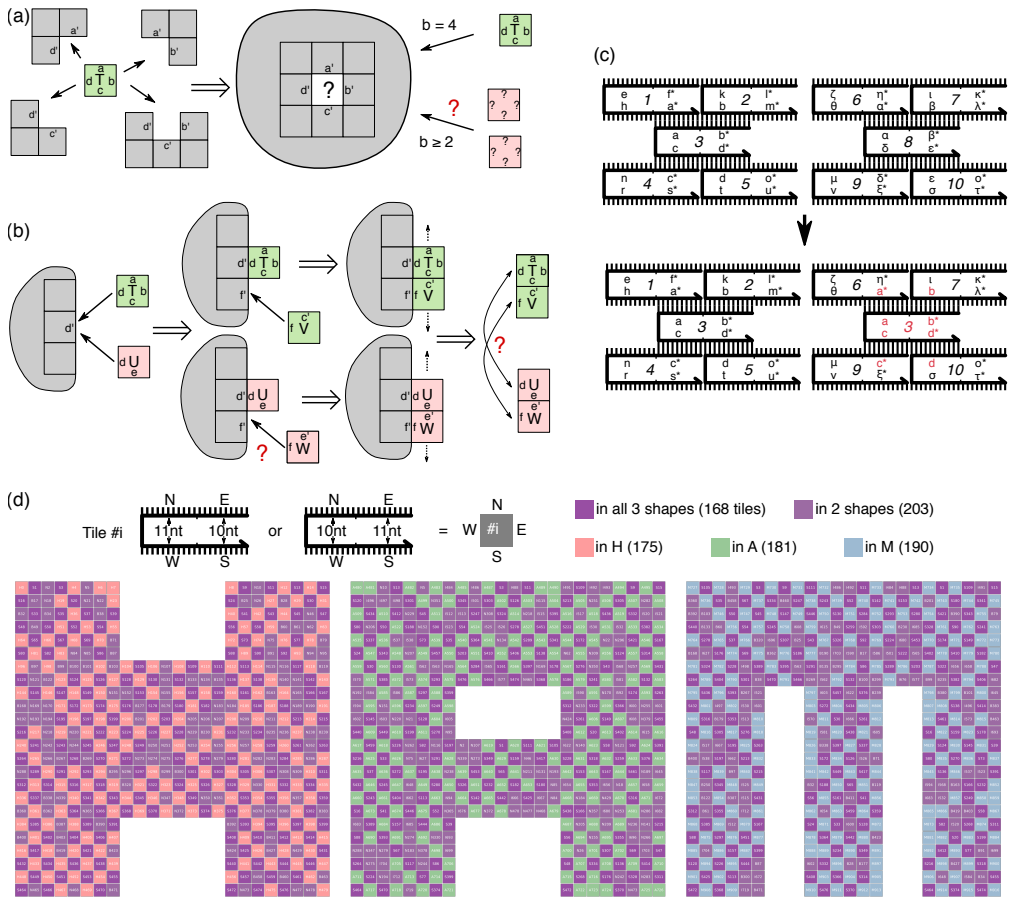

Figure B.1: **Proofreading tile set design and tile assignment map. (a)** Our systems are designed to grow in a regime where a tile attaching by at least two bonds is favorable, but a tile attaching by one bond is not ("threshold 2"). Motivated by self-healing tile systems [101], we seek a tile set where no correct partial assembly should ever allow an undesired tile to attach by two or more bonds, though undesired attachments by one bond are allowed, such that any favorable attachment to a partial assembly will be correct. **(b)** In addition to tiles attaching favorably by 2 bonds to growing facets, new facets in the system will only be created by tiles attaching unfavorably by one bond, and then being stabilized by further, favorable growth. At a site where tile T would correctly attach by one bond, a tile U might be able to attach incorrectly by the same bond. T would correctly be stabilized by the subsequent attachment of V by two bonds, but U might be able to be stabilized as well if there is a tile W that can attach to it and shares the same glue as V. Thus, if for every pair of tiles that can bind to each other (eg, $T + V$), there is no other pair of binding tiles (eg, $U + W$) that share two glues on the same edges of the tiles, then any tile that attaches by one bond to an assembly will either be the correct tile, or will not allow a subsequent attachment, and will likely detach. This is equivalent to "second-order sensitivity" with all directions treated as inputs, functioning as a as a form of "proofreading" [77, 76]. We created a multifarious tile system by first starting with three shapes constructed entirely of unique tiles, then repeatedly attempting **(c)** to "merge" tiles in different shapes by constraining the sequences of their domains to be identical, and checking whether each merge of two tiles results in a tile system that does not have any tile pairs violating criteria in (a) and (b). From multiple trials of the merging process, we selected the smallest result, **(d)**, containing 917 tiles. Tiles in the system were designed with the single-stranded tile (SST) motif in [69], with two alternating tiles motifs of 11 nt and 12 nt domains.

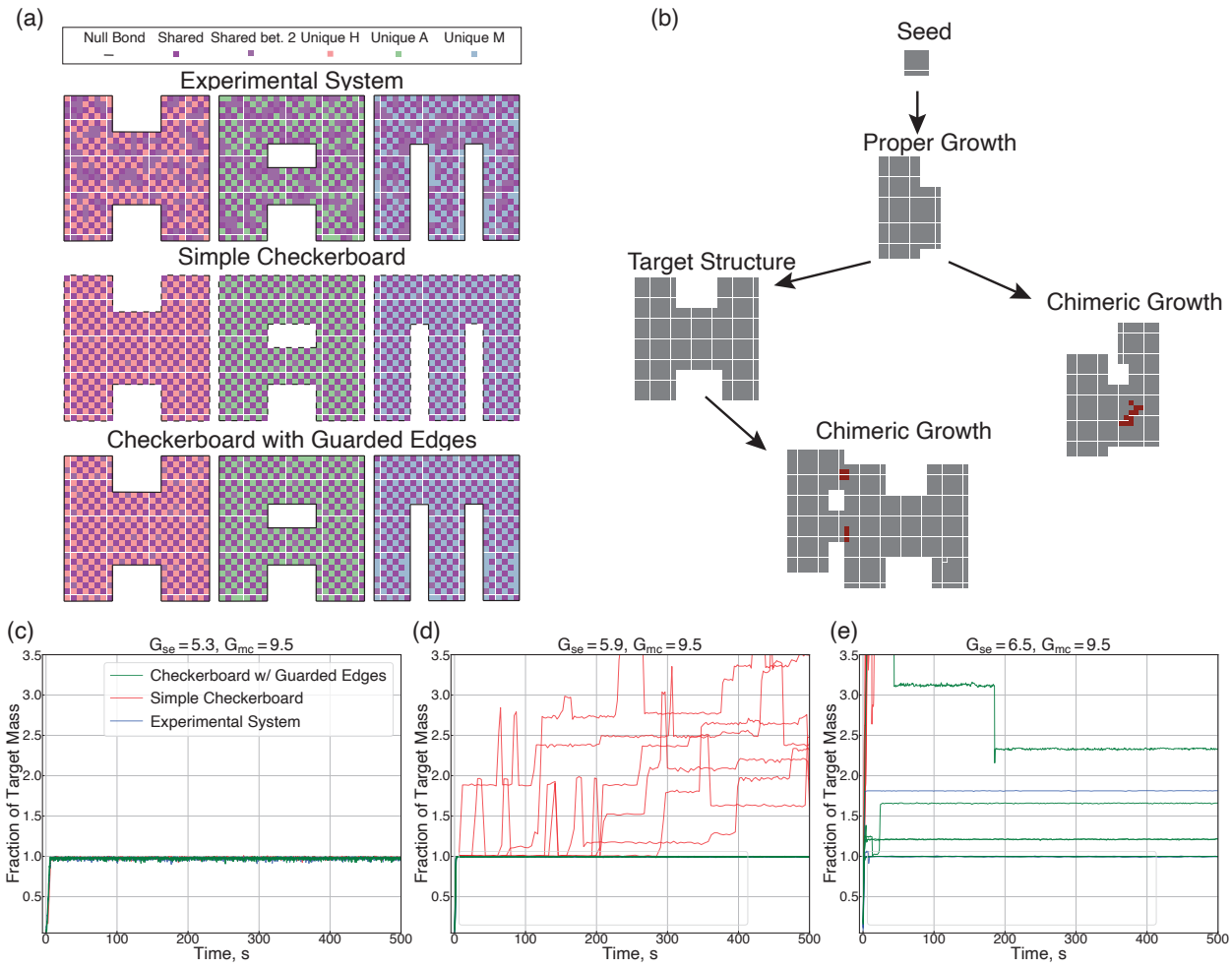

Figure B.2: **Suppression of chimeric growth through tile set design.** We contrast assembly errors in three distinct tile sets: **(a, top)** the proofreading tile set with an inert boundary used in experiments (described in Figure 3.2); **(a, middle)** a simple checkerboard tile set with a strictly alternating shared and unique tile pattern for each shape, where unique tiles can be seen as mediating different interactions between shared tiles; and **(a, bottom)** an edge-guarded checkerboard in which we additionally enforce inert bonds around each shape's perimeter. **(b)** Schematic shows the two distinct kinds of chimeric structures (e.g., part-H, part-A) seen in simulation due to promiscuous interactions; chimeric structures can grow either before full assembly of the target structure or emerge spontaneously from the edge of a properly formed structure. Chimeras like those illustrated along the lower path are held together by just a few bonds and will quickly break apart (species with unintended bonds shown in red): this corresponds to the observed sharp drops in the mass trajectories of (c)-(e). **(c)-(e)** For each tile set, we performed kinetic growth simulations, starting from a pre-formed 5×5 seed, for a single seed localized on H, were performed using XGrow (with chunk fission) [94] with uniform tile concentrations corresponding to $G_{mc} = 9.5$, and varying bond energies relative to k_BT, G_{se}. The size of the assembly (in units of the size of the fully formed H) is shown as a function of time. **(c)** For small $G_{se} = 5.3$ (i.e., high temperature, slow growth), no chimeras are observed on the simulated timescales for any tile set. **(d)** For intermediate $G_{se} = 5.9$ (all 6 checkerboard trajectories still result in chimeras, while no errors are observed on the timescale probed for the guarded checkerboard or proofreading-satisfying tile set. **(e)**: Large $G_{se} = 6.6$, (i.e., low temperatures, fast growth) leads to chimeras with all 3 tile sets; chimeras are seen in all runs for checkerboard structures (red traces), 4 of the 6 runs for guarded checkerboard structures (green traces) and 1 of the 6 runs for proofreading-satisfying structures form chimeras). Simulations were carried out with model detailed in Extended Figure B.4 [102] with $\alpha = 0$.

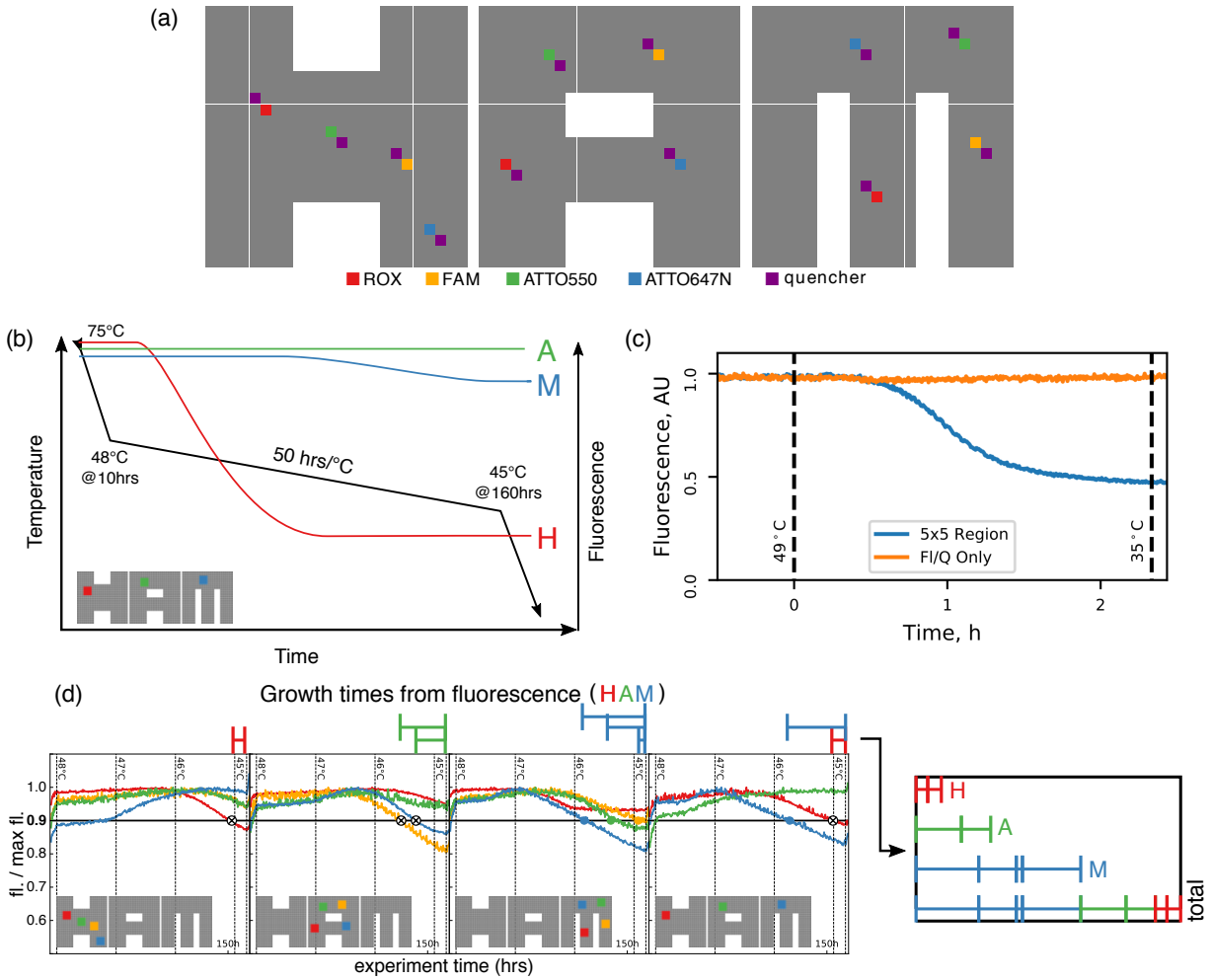

Figure B.3: **Fluorophore quenching as a measure of nucleation and growth.** **(a)** shows positions and types of all fluorophore/quencher pairs available for use; only one of each type of fluorophore can be used in a single sample; four selections of fluorophores were used in different samples. **(b)** shows expected behavior of fluorophore labels on shapes as one shape nucleates and grows. **(c)** shows fluorescence data for non-quenching (fluorophore tile only, orange) and quenching (5×5 lattice around fluorophore and quencher tiles, blue) controls for the ATTO647N fluorophore/quencher pair on A. Here, the temperature ramps linearly from 49°C to 35°C at a rate of $0.1 \frac{^\circ\text{C}}{\text{min}}$, with all tiles at 50 nM, and each sample has its fluorescence normalized to its maximum value independently. **(d)** shows an example of fluorescence growth time measurements. Each fluorophore signal, in each sample, is independently normalized to its maximum value during the experiment, and the time between the point where the signal goes below 0.9 ("10% quenching") and the end of the experiment is measured ("growth time"). These times are then summed for all fluorophores, in all four samples, on each shape, resulting in a growth time for each shape, and, when normalized to the sum of all growth times, a relative growth time for each shape. See Methods above for fluorophore details.

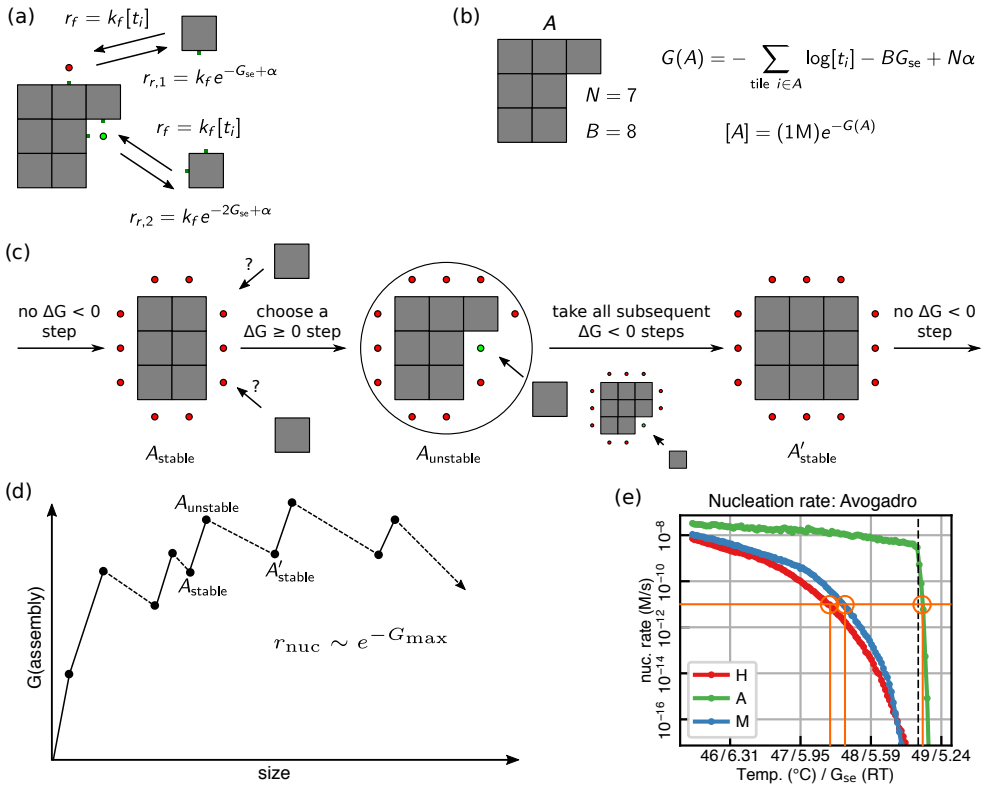

Figure B.4: **Stochastic Greedy Nucleation Model, based on repeated stochastic simulations.** The **(a)** frequently-used kinetic Tile Assembly Model (kTAM) [68, 102] has rates for tile attachment and detachment events based on tile and assembly diffusion and total binding strength of correct attachments a tile can make at a lattice site. These rates can be used **(b)** to derive a free energy for any tile assembly in a system, and, assuming fixed monomer concentrations, an equilibrium concentration for any assembly. [104] showed that the equilibrium concentration of the highest-energy assembly along a nucleation trajectory under this assumption provides an upper bound for nucleation rate through that trajectory, with or without fixed monomer concentrations. However, in a large system, considering all possible intermediate assemblies and all pathways, including many that are extremely unlikely, would be infeasible. Thus, we developed the Stochastic Greedy Nucleation Model to generate stochastically-chosen paths of tile attachments. Starting from a single tile (chosen with probability proportional to relative concentration), **(c)** whenever the assembly is in a state A_{stable} where there is no tile attachment that would be favorable (have $\Delta G < 0$), one of the possible unfavorable (with $\Delta G \geq 0$) attachments is stochastically chosen, resulting in a higher-G state A_{unstable}. Then, all subsequent possible $\Delta G < 0$ attachments are made, resulting in the next A'_{stable} state; for our system of unique tiles for each site in the lattice, this sequence of favorable steps has a unique resulting assembly. The process repeats until all tiles in a shape are attached, which results **(d)** in a trajectory with a maximum-G assembly that can be used to bound the rate of nucleation through that particular trajectory. By using this process to collect many trajectories, and then repeating the entire process for each of the three shapes in the system, we can estimate nucleation rates dependent upon temperature **(e)**, with the assumption that tile monomer concentrations do not deplete, and that the trajectories found are a reasonable representation of likely trajectories. For temperature ramps, we determined a reference temperature for each shape at which the model predicted a nucleation rate fast enough to exceed some threshold (orange line), and used this to compare to fluorescence results in Extended Figure B.5(d) and B.8(b).

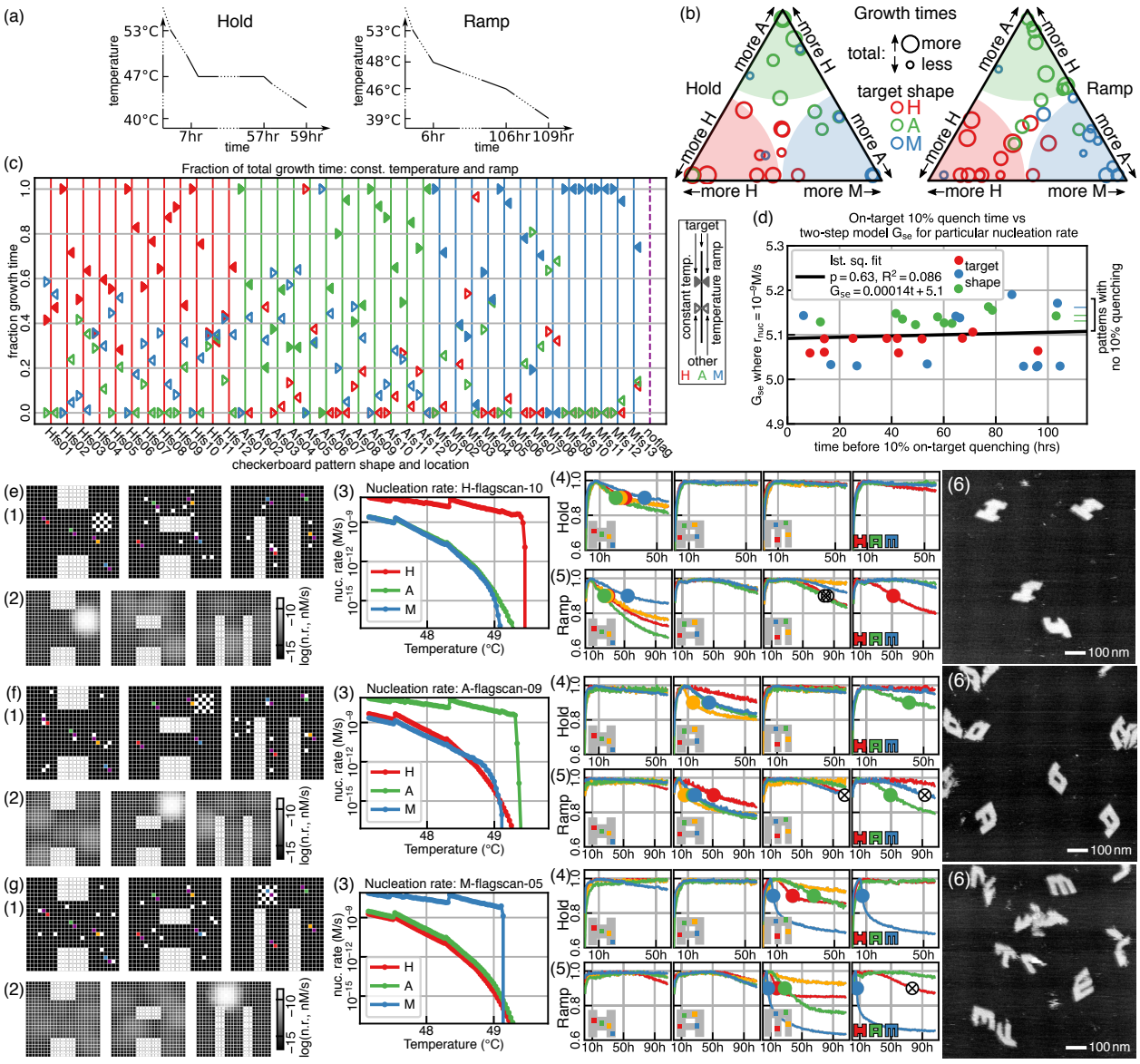

Figure B.5: **Nucleation and growth with 'flag' patterns of enhanced concentration.** 36 different concentration patterns with enhanced concentrations of shared tiles in 5×5 regions were prepared, each with four different standard sets of fluorophores in four samples, and grown using two temperature protocols **(a)**: a ramp focusing on 48°C to 46°C over 100 hours, and a hold at 47°C. Using growth times as described in Extended Figure B.3, fluorescence data for many samples in both experiments showed preference for the desired shapes **(b, c)**, but with considerable variation in selectivity and total amount of growth. No statistically significant correlation was found between the nucleation model prediction for temperature of on-target nucleation and the time of on-target shape quenching in the temperature ramp experiment **(d)**. **(e-g)** show details of three patterns, with concentration patterns (1), weighted critical nucleus free energy starting from particular tiles (2), nucleation-model-estimated nucleation rates (3), temperature hold (4) and temperature ramp (5) experiment fluorescence results, and (6) AFM images from the temperature hold experiments.

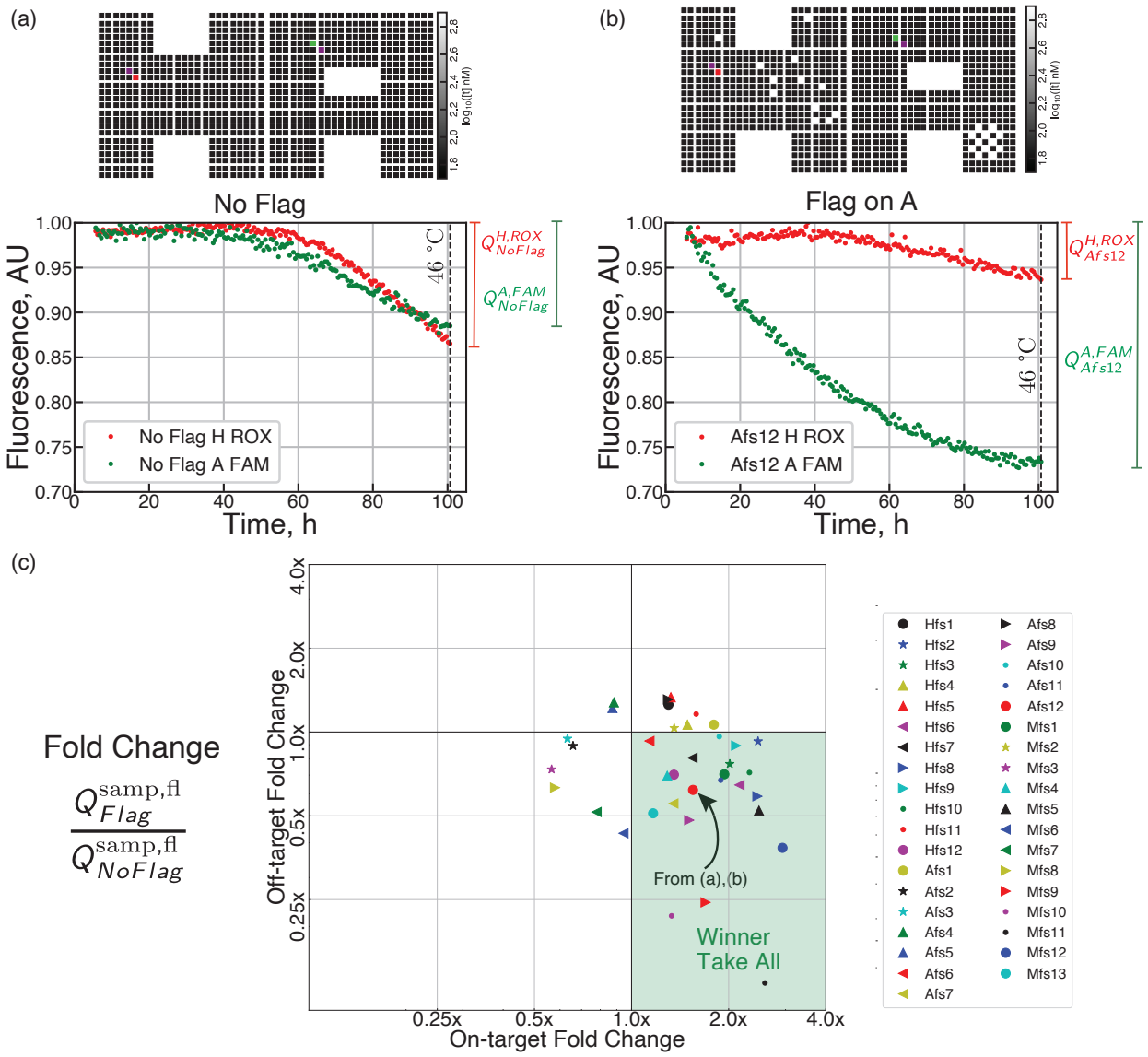

Figure B.6: **Evidence of winner-take-all in flag experiments (a)** Normalized fluorescence for ROX fluorophore in the H sample and the FAM fluorophore in the A sample for a uniform concentration pattern (i.e. no flag). We define the quenching Q for each fluorophore as the difference between the maximum value (always 1 due to normalization) and the value of the fluorescence when the ramp protocol reaches 46 °C. **(b)** Same as in (a) but for the Afs12 concentration pattern . **(c)** Same data as 3.4 (f) with the data points labeled by their corresponding concentration patterns. The horizontal axis values are calculated by averaging the fold changes of the 5 on-target fluorophores (4 in the sample with the fluorophores on the flagged structure and 1 fluorophore from the '3' sample) while the vertical axis is calculated by averaging over the 10 off-target fluorophores (8 from the samples with fluorophores on the non-flagged shape and the 2 non-corresponding fluorophores from the '3' sample).

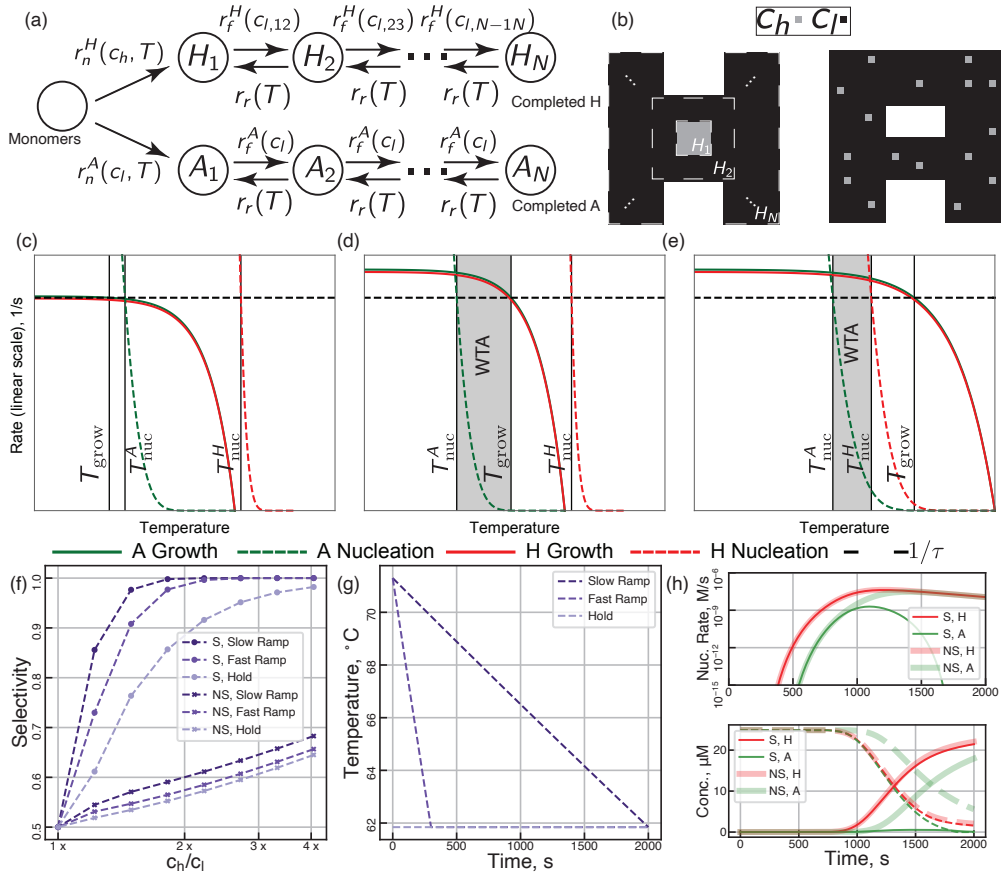

Figure B.7: **Simple model illustrates winner-take-all dynamics. (a)** A 2N+1 state ($N = 50$) master equation models the evolution from monomers along two one-dimensional chains of states corresponding to the growth of two distinct structures (kTAM rates are used for attachment/detachment, see Extended Figure B.4) after a non-reversible nucleation step (Arrhenius' barrier crossing formula assumed). **(b)** There are high c_h and low c_l concentrations initially present in the system, with H assumed to have a region of localized high concentration species. Nucleation of H is assumed in a single location, and thus it's states correspond to growth of specific locations in H. In A, states correspond to fragments of a particular size with random composition, leading to distinct depletion dynamics. Species can be totally shared between the two structures (S) or totally unique in each structure (NS). **(c)-(e)** Extended schematics from main text Fig. 3d showing nucleation rates as a function of temperature (assuming fixed concentrations). Here, nucleation and growth temperatures are defined when rate curves become equal to inverse experimental timescale, $1/\tau$ (where the nucleation rates have been scaled by the low concentration in the system to match units). When H structures are able to nucleate and grow at temperatures before A nucleation becomes significant, winner-take-all dynamics are possible with many H's forming and depleting shared tiles, effectively decreasing $T^{\mathrm{A}}_{\mathrm{nuc}}$. **(f)** Selectivity, quantified as the fraction of structure mass in the H branch relative to the total mass in the H and A branches at the final time point, is plotted for the three temperature protocols shown in **(g)** for systems with shared species (S, dots) and systems with no shared species (NS, x's). **(h), top** Nucleation rates for the slow ramp temperature protocol are shown over the duration of the simulation. Note that when the components are shared between H and A, depletion keeps the nucleation rate of the A structure small even at low temperatures. **(h), bottom** The concentrations in state H_N and A_N (solid lines) and the average values of c_l in A and H (dashed lines) are shown for the slow ramp protocol.

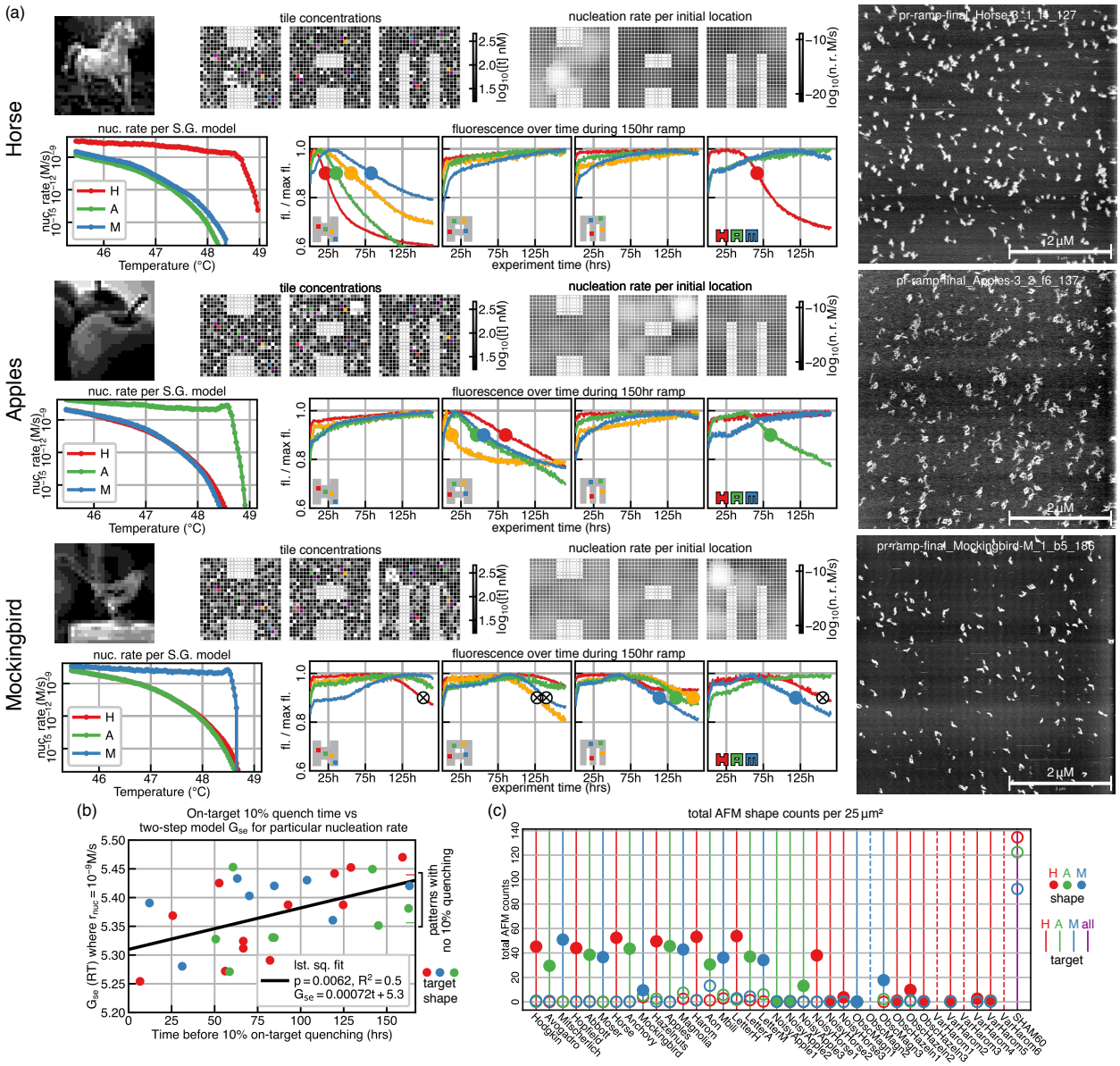

Figure B.8: **Classification of images viewed as concentration patterns** 36 different concentration patterns, derived from a mapping of 36 grayscale images, were run using a ramp between 48°C to 45°C over approximately 150 hours. **(a)** Three pattern examples, with source image, concentration pattern, nucleation model nucleation rate starting from particular tiles, nucleation model nucleation rates, fluorescence results, and AFM images. **(b)** Across all patterns there was some correlation between the on-target nucleation temperature predicted by the nucleation model and on-target shape quenching time. **(c)**Total AFM shape counts for each sample.

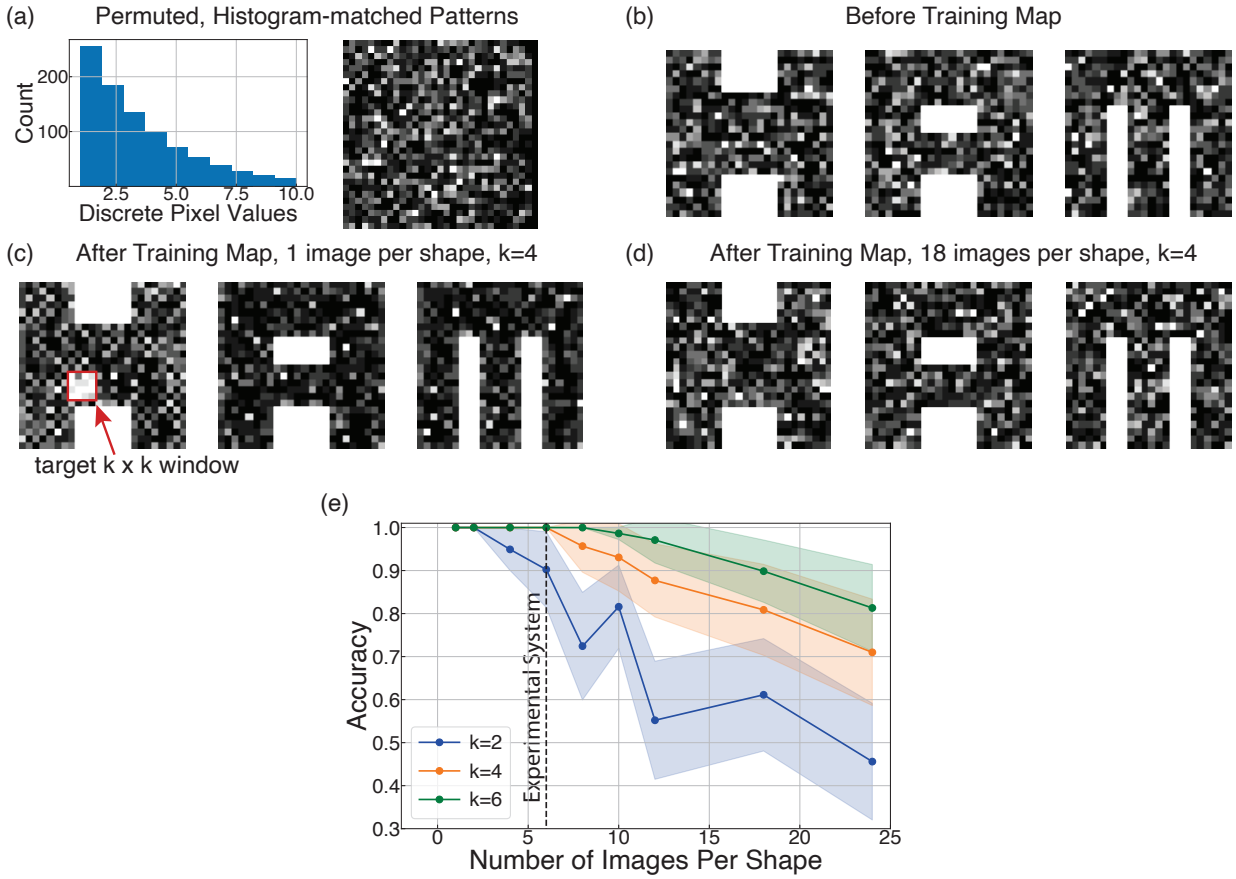

Figure B.9: **Pattern recognition capacity** To analyze the pattern-recognition capabilities of the designed tile set, a map-training algorithm was run for increasingly larger sets of random images. **(a)** Random images are generated by permuting a list of 900 pixel values (with matching histograms). As in experiments, patterns correspond to 30x30 images with the same number of tiles at each of 10 possible grayscale values. **(b)** Before training, bright pixels (i.e., high concentrations) are randomly assigned to tiles across the three shapes. **(c)** The map training algorithm attempts to create regions of localized high concentrations on a length scale k on the targeted shape (here, H) while dispersing any high concentration tiles on the off-target shapes to prevent spurious nucleation. Note this nucleation center could occur at many different spatial locations on the H **(d)** When the number of images becomes large (18 images per shape shown here), the training algorithm is still able to create a scale k nucleation site on the target shape (H in this example), but the concentration of high concentration tiles on the desired shape is less apparent visually. **(e)** As the number of images in the set increases, the performance of the training algorithm decreases. For larger k, the pixel-tile map can exploit higher-order correlations and can thus accommodate more images. Accuracy is calculated by taking the ratio of the nucleation rate of the target shape over the nucleation rates of all three shapes (calculated using the nucleation model shown in Extended Figure B.4 at $G_{se} = 5.4$ which roughly corresponds to a temperature of 48.6 °C). Here, the shaded areas around each line correspond to the standard deviation of the accuracy contribution over the images in each set.

REFERENCES

[1] Y. LeCun et al. "Backpropagation Applied to Handwritten Zip Code Recognition". In: *Neural Computation* 1.4 (1989), pp. 541–551. DOI: 10.1162/neco.1989.1.4.541.

[2] A. Morales-Esteban et al. "Pattern recognition to forecast seismic time series". In: *Expert Systems with Applications* 37.12 (2010), pp. 8333–8342.

[3] Mark P. Mattson. "Superior pattern processing is the essence of the evolved human brain". In: *Frontiers in Neuroscience* 8 (2014), p. 265.

[4] Julius Adler. "Chemotaxis in Bacteria". In: *Science* 153.3737 (1966), pp. 708–716.

[5] Jeremy E Purvis and Galit Lahav. "Encoding and decoding cellular information through signaling dynamics". In: *Cell* 152.5 (Feb. 2013), pp. 945–956. ISSN: 0092-8674.

[6] Anders S Hansen and Erin K O'Shea. "Encoding four gene expression programs in the activation dynamics of a single transcription factor". en. In: *Curr. Biol.* 26.7 (Apr. 2016), R269–71. ISSN: 0960-9822, 1879-0445. DOI: 10.1016/j.cub.2016.02.058.

[7] LM Adleman. "Molecular computation of solutions to combinatorial problems". In: *Science* 266.5187 (1994), pp. 1021–1024.

[8] Jocelyn Y Kishi et al. "Programmable autonomous synthesis of single-stranded DNA". en. In: *Nat. Chem.* 10.2 (Feb. 2018), pp. 155–164. ISSN: 1755-4330, 1755-4349. DOI: 10.1038/nchem.2872.

[9] Matthew R Lakin et al. "Visual DSD: a design and analysis tool for DNA strand displacement systems". en. In: *Bioinformatics* 27.22 (Nov. 2011), pp. 3211–3213. ISSN: 1367-4803, 1367-4811. DOI: 10.1093/bioinformatics/btr543.

[10] Shiping Song et al. “Aptamer-based biosensors”. In: *Trends Analyt. Chem.* 27.2 (Feb. 2008), pp. 108–117. ISSN: 0165-9936. DOI: 10.1016/j.trac.2007.12.004.

[11] Souvik Modi et al. “A DNA nanomachine that maps spatial and temporal pH changes inside living cells”. en. In: *Nat. Nanotechnol.* 4.5 (May 2009), pp. 325–330. ISSN: 1748-3387, 1748-3395. DOI: 10.1038/nnano.2009.83.

[12] Ursula E Spichiger-Keller. *Chemical Sensors and Biosensors for Medical and Biological Applications*. en. John Wiley and Sons, Nov. 2008. ISBN: 9783527612260.

[13] Xi Chen et al. “Stacking nonenzymatic circuits for high signal gain”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 110.14 (Apr. 2013), pp. 5386–5391. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1222807110.

[14] Nan Hao and Erin K O’Shea. “Signal-dependent dynamics of transcription factor translocation controls gene expression”. en. In: *Nat. Struct. Mol. Biol.* 19.1 (Dec. 2011), pp. 31–39. ISSN: 1545-9993, 1545-9985. DOI: 10.1038/nsmb.2192.

[15] Anders S Hansen and Erin K O’Shea. “Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression”. en. In: *Mol. Syst. Biol.* 9 (Nov. 2013), p. 704. ISSN: 1744-4292. DOI: 10.1038/msb.2013.56.

[16] Erica Del Grosso et al. “Enzyme-Operated DNA-Based Nanodevices”. en. In: *Nano Lett.* 15.12 (Dec. 2015), pp. 8407–8411. ISSN: 1530-6984, 1530-6992. DOI: 10.1021/acs.nanolett.5b04566.

[17] Jongmin Kim and Erik Winfree. “Synthetic in vitro transcriptional oscillators”. en. In: *Mol. Syst. Biol.* 7 (Feb. 2011), p. 465. ISSN: 1744-4292. DOI: 10.1038/msb.2010.119.

[18] Nathaniel Roquet et al. “Synthetic recombinase-based state machines in living cells”. In: *Science* 353.6297 (2016). ISSN: 0036-8075. DOI: 10.1126/science.aad8559.

eprint: http://science.sciencemag.org/content/353/6297/aad8559.full.pdf. URL: http://science.sciencemag.org/content/353/6297/aad8559.

[19] Weixin Tang and David R. Liu. "Rewritable multi-event analog recording in bacterial and mammalian cells". In: *Science* 360.6385 (2018). ISSN: 0036-8075. DOI: 10.1126/science.aap8992.

[20] Ravi U. Sheth et al. "Multiplex recording of cellular events over time on CRISPR biological tape". In: *Science* 358.6369 (2017), pp. 1457–1461. ISSN: 0036-8075. DOI: 10.1126/science.aao0958. eprint: http://science.sciencemag.org/content/358/6369/1457.full.pdf. URL: http://science.sciencemag.org/content/358/6369/1457.

[21] Victoria Hsiao et al. "A population-based temporal logic gate for timing and recording chemical events". In: *Molecular Systems Biology* 12.5 (2016). DOI: 10.15252/msb.20156663. eprint: http://msb.embopress.org/content/12/5/869.full.pdf. URL: http://msb.embopress.org/content/12/5/869.

[22] Ari E. Friedland et al. "Synthetic Gene Networks That Count". In: *Science* 324.5931 (2009), pp. 1199–1202. ISSN: 0036-8075. DOI: 10.1126/science.1172005. eprint: http://science.sciencemag.org/content/324/5931/1199.full.pdf. URL: http://science.sciencemag.org/content/324/5931/1199.

[23] L Qian and E Winfree. "A simple DNA gate motif for synthesizing large-scale circuits". In: *DNA Computing* (Jan. 2009), pp. 70–89.

[24] L Qian and E Winfree. "Scaling Up Digital Circuit Computation with DNA Strand Displacement Cascades". In: *Science* 332.6034 (June 2011), pp. 1196–1201.

[25] Lulu Qian, Erik Winfree, and Jehoshua Bruck. "Neural network computation with DNA strand displacement cascades". en. In: *Nature* 475.7356 (July 2011), pp. 368–372. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature10262.

[26] Kevin M Cherry and Lulu Qian. "Scaling up molecular pattern recognition with DNA-based winner-take-all neural networks". en. In: *Nature* 559.7714 (July 2018), pp. 370–376. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/s41586-018-0289-6.

[27] R Sarpeshkar. "Analog synthetic biology". en. In: *Philos. Trans. A Math. Phys. Eng. Sci.* 372.2012 (Mar. 2014), p. 20130110. ISSN: 1364-503X. DOI: 10.1098/rsta.2013.0110.

[28] R Sarpeshkar. "Analog versus digital: extrapolating from electronics to neurobiology". en. In: *Neural Comput.* 10.7 (Oct. 1998), pp. 1601–1638. ISSN: 0899-7667.

[29] Ramiz Daniel et al. "Synthetic analog computation in living cells". en. In: *Nature* 497.7451 (May 2013), pp. 619–623. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature12148.

[30] Tianqi Song et al. "Design and Analysis of Compact DNA Strand Displacement Circuits for Analog Computation Using Autocatalytic Amplifiers". en. In: *ACS Synth. Biol.* 7.1 (Jan. 2018), pp. 46–53. ISSN: 2161-5063. DOI: 10.1021/acssynbio.6b00390.

[31] Tianqi Song et al. "Analog Computation by DNA Strand Displacement Circuits". In: *ACS Synth. Biol.* 5.8 (Aug. 2016), pp. 898–912. DOI: 10.1021/acssynbio.6b00144.

[32] Joshua Fern et al. "DNA Strand-Displacement Timer Circuits". In: *ACS Synth. Biol.* 6.2 (Feb. 2017), pp. 190–193. DOI: 10.1021/acssynbio.6b00170.

[33] Boyan Yordanov et al. “Computational design of nucleic acid feedback control circuits”. en. In: *ACS Synth. Biol.* 3.8 (Aug. 2014), pp. 600–616. ISSN: 2161-5063. DOI: 10.1021/sb400169s.

[34] David Doty. “Timing in chemical reaction networks”. In: *CoRR* abs/1304.0872 (2013). arXiv: 1304.0872. URL: http://arxiv.org/abs/1304.0872.

[35] Deepak K Agrawal, Elisa Franco, and Rebecca Schulman. “A self-regulating biomolecular comparator for processing oscillatory signals”. en. In: *J. R. Soc. Interface* 12.111 (Oct. 2015), p. 20150586. ISSN: 1742-5689, 1742-5662. DOI: 10.1098/rsif.2015.0586.

[36] Anthony J Genot, Jonathan Bath, and Andrew J Turberfield. “Combinatorial displacement of DNA strands: application to matrix multiplication and weighted sums”. en. In: *Angew. Chem. Int. Ed Engl.* 52.4 (Jan. 2013), pp. 1189–1192. ISSN: 1433-7851, 1521-3773. DOI: 10.1002/anie.201206201.

[37] Muppirala Viswa Virinchi, Abhishek Behera, and Manoj Gopalkrishnan. “A Stochastic Molecular Scheme for an Artificial Cell to Infer Its Environment from Partial Observations”. In: *DNA Computing and Molecular Programming*. Springer International Publishing, 2017, pp. 82–97. DOI: 10.1007/978-3-319-66799-7_6.

[38] Muppirala Viswa Virinchi, Abhishek Behera, and Manoj Gopalkrishnan. “A reaction network scheme which implements the EM algorithm”. In: (Apr. 2018). arXiv: 1804.09062 [cs.ET].

[39] Anton S. Zadorin et al. “Synthesis of Programmable Reaction-Diffusion Fronts Using DNA Catalyzers”. In: *Phys. Rev. Lett.* 114 (6 Feb. 2015), p. 068301. DOI: 10.1103/PhysRevLett.114.068301. URL: https://link.aps.org/doi/10.1103/PhysRevLett.114.068301.

[40] Dominic Scalise, Nisita Dutta, and Rebecca Schulman. "DNA Strand Buffers". en. In: *J. Am. Chem. Soc.* 140.38 (Sept. 2018), pp. 12069–12076. ISSN: 0002-7863, 1520-5126. DOI: 10.1021/jacs.8b05373.

[41] Christian Cuba Samaniego and Elisa Franco. "A Robust Molecular Network Motif for Period-Doubling Devices". en. In: *ACS Synth. Biol.* 7.1 (Jan. 2018), pp. 75–85. ISSN: 2161-5063. DOI: 10.1021/acssynbio.7b00222.

[42] K. Oishi and E. Klavins. "Biomolecular implementation of linear I/O systems". In: *IET Systems Biology* 5.4 (July 2011), pp. 252–260. ISSN: 1751-8849. DOI: 10.1049/iet-syb.2010.0056.

[43] Niranjan Srinivas et al. "Enzyme-free nucleic acid dynamical systems". In: *Science* 358.6369 (2017). ISSN: 0036-8075. DOI: 10.1126/science.aal2052. eprint: http://science.sciencemag.org/content/358/6369/eaal2052.full.pdf. URL: http://science.sciencemag.org/content/358/6369/eaal2052.

[44] Adrien Padirac, Teruo Fujii, and Yannick Rondelez. "Bottom-up construction of in vitro switchable memories". In: *Proceedings of the National Academy of Sciences* 109.47 (2012), E3212–E3220. ISSN: 0027-8424. DOI: 10.1073/pnas.1212069109.

[45] A Graves, A r. Mohamed, and G Hinton. "Speech recognition with deep recurrent neural networks". In: *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*. May 2013, pp. 6645–6649. DOI: 10.1109/ICASSP.2013.6638947.

[46] Yuhai Tu. "Quantitative Modeling of Bacterial Chemotaxis: Signal Amplification and Accurate Adaptation". In: *Annu. Rev. Biophys.* 42 (Jan. 2013), pp. 337–359. ISSN: 1936-122X.

[47] James E. Ferrell. "Perfect and Near-Perfect Adaptation in Cell Signaling". In: *Cell Systems* 2.2 (2016), pp. 62–67. ISSN: 2405-4712. DOI:

https://doi.org/10.1016/j.cels.2016.02.006. URL: http://www.sciencedirect.com/science/article/pii/S2405471216300369.

[48] Steven M. Chirieleison et al. "Pattern transformation with DNA circuits". In: *Nature Chemistry* 5.12 (Sept. 2013), pp. 1000–1005. ISSN: 1755-4349. DOI: 10.1038/nchem.1764. URL: http://dx.doi.org/10.1038/nchem.1764.

[49] N Barkai and S Leibler. "Robustness in simple biochemical networks". en. In: *Nature* 387.6636 (June 1997), pp. 913–917. ISSN: 0028-0836. DOI: 10.1038/43199.

[50] Amir Mitchell, Ping Wei, and Wendell A Lim. "Oscillatory stress stimulation uncovers an Achilles' heel of the yeast MAPK signaling network". en. In: *Science* 350.6266 (Dec. 2015), pp. 1379–1383. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.aab0892.

[51] Benoit Sorre et al. "Encoding of temporal signals by the TGF-β pathway and implications for embryonic patterning". en. In: *Dev. Cell* 30.3 (Aug. 2014), pp. 334–342. ISSN: 1534-5807, 1878-1551. DOI: 10.1016/j.devcel.2014.05.022.

[52] Carolyn Zhang et al. "Processing Oscillatory Signals by Incoherent Feedforward Loops". en. In: *PLoS Comput. Biol.* 12.9 (Sept. 2016), e1005101. ISSN: 1553-734X, 1553-7358. DOI: 10.1371/journal.pcbi.1005101.

[53] S Mangan, A Zaslaver, and U Alon. "The Coherent Feedforward Loop Serves as a Sign-sensitive Delay Element in Transcription Networks". In: *Journal of Molecular Biology* 334.2 (2003), pp. 197–204. ISSN: 0022-2836. DOI: https://doi.org/10.1016/j.jmb.2003.09.049. URL: http://www.sciencedirect.com/science/article/pii/S0022283603012038.

[54] Erik Winfree Lulu Qian. "A Simple DNA Gate Motif for Synthesizing Large-Scale Circuits". In: *Journal of The Royal Society Interface* 8.62 (2011), pp. 1281–1297.

[55] Erik Winfree David Soloveichik Georg Seelig. "DNA as a universal substrate for chemical kinetics". In: *Proceedings of the National Academy of Sciences* 107.12 (Mar. 2010), pp. 5393–5398.

[56] David Soloveichik, Georg Seelig, and Erik Winfree. "DNA as a universal substrate for chemical kinetics". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 107.12 (Mar. 2010), pp. 5393–5398. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.0909380107.

[57] David Yu Zhang and Georg Seelig. "Dynamic DNA nanotechnology using strand-displacement reactions". In: *Nat. Chem.* 3.2 (Jan. 2011), pp. 103–113. ISSN: 1755-4330.

[58] Robert R F Machinek et al. "Programmable energy landscapes for kinetic control of DNA strand displacement". en. In: *Nat. Commun.* 5 (Nov. 2014), p. 5324. ISSN: 2041-1723. DOI: 10.1038/ncomms6324.

[59] Thomas E. Ouldridge. "DNA nanotechnology: understanding and optimisation through simulation". In: *Molecular Physics* 113.1 (2015), pp. 1–15. DOI: 10.1080/00268976.2014.975293. eprint: https://doi.org/10.1080/00268976.2014.975293. URL: https://doi.org/10.1080/00268976.2014.975293.

[60] Davit A. Potoyan et al. "Molecular stripping in the NF-ÎºB/IÎºB/DNA genetic regulatory network". In: *Proceedings of the National Academy of Sciences* 113.1 (2016), pp. 110–115. ISSN: 0027-8424. DOI: 10.1073/pnas.1520483112. eprint: http://www.pnas.org/content/113/1/110.full.pdf. URL: http://www.pnas.org/content/113/1/110.

[61] Dirk Benzinger and Mustafa Khammash. "Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation". en. In:

Nat. Commun. 9.1 (Aug. 2018), p. 3521. ISSN: 2041-1723. DOI: 10.1038/s41467-018-05882-2.

[62] Veikko Linko, Ari Ora, and Mauri A. Kostiainen. "DNA Nanostructures as Smart Drug-Delivery Vehicles and Molecular Devices". In: *Trends in Biotechnology* 33.10 (2015), pp. 586–594. ISSN: 0167-7799. DOI: https://doi.org/10.1016/j.tibtech.2015.08.001. URL: http://www.sciencedirect.com/science/article/pii/S0167779915001614.

[63] Lei Dai et al. "Generic Indicators for Loss of Resilience Before a Tipping Point Leading to Population Collapse". In: *Science* 336.6085 (2012), pp. 1175–1177. ISSN: 0036-8075. DOI: 10.1126/science.1219805. eprint: http://science.sciencemag.org/content/336/6085/1175.full.pdf. URL: http://science.sciencemag.org/content/336/6085/1175.

[64] Matthew R Lakin and Darko Stefanovic. "Supervised Learning in Adaptive DNA Strand Displacement Networks". en. In: *ACS Synth. Biol.* 5.8 (Aug. 2016), pp. 885–897. ISSN: 2161-5063. DOI: 10.1021/acssynbio.6b00009.

[65] Matthew R Lakin et al. "Design of a biochemical circuit motif for learning linear functions". en. In: *J. R. Soc. Interface* 11.101 (Dec. 2014), p. 20140902. ISSN: 1742-5689, 1742-5662.

[66] William Poole et al. "Chemical Boltzmann Machines". In: *DNA Computing and Molecular Programming*. Springer International Publishing, 2017, pp. 210–231. DOI: 10.1007/978-3-319-66799-7_14.

[67] D Bray. "Protein molecules as computational elements in living cells". In: *Nature* 376.6538 (Jan. 1995), pp. 307–312. ISSN: 0028-0836.

[68] Erik Winfree. *Simulations of Computing by Self-Assembly*. Tech. rep. CaltechCSTR:1998.22. Pasadena, CA, 1998. DOI: 10.7907/Z9TB14X7.

[69] Peng Yin et al. “Programming DNA Tube Circumferences”. In: *Science* 321.5890 (Aug. 8, 2008), pp. 824–826. ISSN: 0036-8075, 1095-9203. DOI: 10.1126/science.1157312. URL: http://science.sciencemag.org/content/321/5890/824 (visited on 03/31/2019).

[70] Bryan Wei, Mingjie Dai, and Peng Yin. “Complex shapes self-assembled from single-stranded DNA tiles”. In: *Nature* 485.7400 (May 2012), pp. 623–626. ISSN: 1476-4687. DOI: 10.1038/nature11075. URL: https://www.nature.com/articles/nature11075 (visited on 03/09/2019).

[71] Yonggang Ke et al. “Three-Dimensional Structures Self-Assembled from DNA Bricks”. In: *Science* 338.6111 (2012), pp. 1177–1183. ISSN: 0036-8075. DOI: 10.1126/science.1227268. eprint: https://science.sciencemag.org/content/338/6111/1177.full.pdf. URL: https://science.sciencemag.org/content/338/6111/1177.

[72] Damien Woods et al. “Diverse and robust molecular algorithms using reprogrammable DNA self-assembly”. In: *Nature* 567.7748 (Mar. 2019), p. 366. ISSN: 1476-4687. DOI: 10.1038/s41586-019-1014-9. URL: https://www.nature.com/articles/s41586-019-1014-9 (visited on 03/26/2019).

[73] E. Winfree et al. “Design and Self-Assembly of Two-Dimensional DNA Crystals”. In: *Nature* 394 (1998), pp. 539–544.

[74] Arvind Murugan et al. “Multifarious assembly mixtures: Systems allowing retrieval of diverse stored structures”. In: *Proceedings of the National Academy of Sciences* 112.1 (Jan. 2015), pp. 54–59.

[75] J J Hopfield. “Neural networks and physical systems with emergent collective computational abilities”. en. In: *Proc. Natl. Acad. Sci. U. S. A.* 79.8 (Apr. 1982), pp. 2554–2558. ISSN: 0027-8424. DOI: 10.1073/pnas.79.8.2554.

[76] Erik Winfree and Renat Bekbolatov. “Proofreading Tile Sets: Error Correction for Algorithmic Self-Assembly”. In: *DNA Computing*. Ed. by Junghuei Chen and John Reif. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2004, pp. 126–144. ISBN: 978-3-540-24628-2. DOI: 10.1007/978-3-540-24628-2_13.

[77] Constantine G. Evans and Erik Winfree. “Optimizing Tile Set Size While Preserving Proofreading with a DNA Self-assembly Compiler”. In: *DNA Computing and Molecular Programming*. Ed. by David Doty and Hendrik Dietz. Lecture Notes in Computer Science. Springer International Publishing, 2018, pp. 37–54. ISBN: 978-3-030-00030-1. DOI: 10.1007/978-3-030-00030-1_3.

[78] Daan Frenkel and Beren Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, 2002. DOI: 10.1016/B978-0-12-267351-1.X5000-7.

[79] Eilhard Mitscherlich. “Sur la relation qui existe entre la forme cristalline et les proportions chimiques. I. Mémoire sur les arsenates et les phosphates”. In: *Ann. Chim. Phys* 19 (1821), pp. 350–419. URL: https://gallica.bnf.fr/ark:/12148/bpt6k6571078m/f356.item.

[80] Arvind Murugan, James Zou, and Michael P Brenner. “Undesired usage and the robust self-assembly of heterogeneous structures”. In: *Nat. Commun.* 6 (Feb. 2015), p. 6203.

[81] William M Jacobs, Aleks Reinhardt, and Daan Frenkel. “Communication: Theoretical prediction of free-energy landscapes for complex self-assembly”. In: *J. Chem. Phys.* 142.2 (Jan. 2015), p. 021101. ISSN: 0021-9606.

[82] Katherine E Dunn et al. “Guiding the folding pathway of DNA origami”. en. In: *Nature* 525.7567 (Sept. 2015), pp. 82–86. ISSN: 0028-0836, 1476-4687. DOI: 10.1038/nature14860.

[83] Tao Pan and Tobin Sosnick. “RNA folding during transcription”. en. In: *Annu. Rev. Biophys. Biomol. Struct.* 35 (2006), pp. 161–175. ISSN: 1056-8700. DOI: 10.1146/annurev.biophys.35.040405.102053.

[84] Weishun Zhong, David J. Schwab, and Arvind Murugan. “Associative Pattern Recognition Through Macro-molecular Self-Assembly”. In: *Journal of Statistical Physics* 167.3 (May 1, 2017), pp. 806–826. ISSN: 1572-9613. DOI: 10.1007/s10955-017-1774-2. URL: https://doi.org/10.1007/s10955-017-1774-2 (visited on 04/03/2019).

[85] Edvard I Moser, May-Britt Moser, and Bruce L McNaughton. “Spatial representation in the hippocampal formation: a history”. en. In: *Nat. Neurosci.* 20.11 (Oct. 2017), pp. 1448–1464. ISSN: 1097-6256, 1546-1726. DOI: 10.1038/nn.4653.

[86] F P Battaglia and A Treves. “Attractor neural networks storing multiple space representations: A model for hippocampal place fields”. In: *Phys. Rev. E* 58.6 (Dec. 1998), pp. 7738–7753. DOI: 10.1103/PhysRevE.58.7738.

[87] Weishun Zhong, David J Schwab, and Arvind Murugan. “Associative Pattern Recognition Through Macro-molecular Self-Assembly”. en. In: *J. Stat. Phys.* 167.3-4 (May 2017), pp. 806–826. ISSN: 0022-4715, 1572-9613. DOI: 10.1007/s10955-017-1774-2.

[88] EB Baum. “Building an associative memory vastly larger than the brain”. In: *Science* 268.5210 (1995), pp. 583–585. ISSN: 0036-8075. DOI: 10.1126/science.7725109. eprint:

https://science.sciencemag.org/content/268/5210/583.full.pdf. URL: https://science.sciencemag.org/content/268/5210/583.

[89] Allen P Mills, Bernard Yurke, and Philip M Platzman. "Article for analog vector algebra computation". In: *Biosystems* 52.1 (1999), pp. 175–180. ISSN: 0303-2647. DOI: https://doi.org/10.1016/S0303-2647(99)00044-1. URL: http://www.sciencedirect.com/science/article/pii/S0303264799000441.

[90] Matthew R. Lakin and Darko Stefanovic. "Supervised Learning in Adaptive DNA Strand Displacement Networks". In: *ACS Synthetic Biology* 5.8 (2016). PMID: 27111037, pp. 885–897. DOI: 10.1021/acssynbio.6b00009. eprint: https://doi.org/10.1021/acssynbio.6b00009. URL: https://doi.org/10.1021/acssynbio.6b00009.

[91] Ang Gao et al. "Evolution of weak cooperative interactions for biological specificity". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 115.47 (Nov. 2018), E11053–E11060. ISSN: 0027-8424, 1091-6490. DOI: 10.1073/pnas.1815912115.

[92] Yaron E Antebi et al. "Combinatorial Signal Perception in the BMP Pathway". en. In: *Cell* 170.6 (Sept. 2017), 1184–1196.e24. ISSN: 0092-8674, 1097-4172. DOI: 10.1016/j.cell.2017.08.015.

[93] Elissa A Hallem and John R Carlson. "Coding of odors by a receptor repertoire". en. In: *Cell* 125.1 (Apr. 2006), pp. 143–160. ISSN: 0092-8674. DOI: 10.1016/j.cell.2006.01.050.

[94] *The Xgrow Simulator*. URL: http://dna.caltech.edu/Xgrow/.

[95] Rosalind J Allen, Patrick B Warren, and Pieter Rein Ten Wolde. "Sampling rare switching events in biochemical networks". en. In: *Phys. Rev. Lett.* 94.1 (Jan. 2005), p. 018104. ISSN: 0031-9007. DOI: 10.1103/PhysRevLett.94.018104.

[96] Rosalind J. Allen, Daan Frenkel, and Pieter Rein ten Wolde. "Simulating rare events in equilibrium or nonequilibrium stochastic systems". In: *The Journal of Chemical Physics* 124.2 (2006), p. 024102.

[97] Sarwar Hussain and Amir Haji-Akbari. "Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook". In: *The Journal of Chemical Physics* 152.6 (2020), p. 060901.

[98] Rosalind J Allen, Chantal Valeriani, and Pieter Rein ten Wolde. "Forward flux sampling for rare event simulations". In: *Journal of Physics: Condensed Matter* 21.46 (Oct. 2009), p. 463102.

[99] David Yu Zhang and Erik Winfree. "Control of DNA Strand Displacement Kinetics Using Toehold Exchange". In: *Journal of the American Chemical Society* 131.47 (2009). PMID: 19894722, pp. 17303–17314. DOI: 10.1021/ja906987s. eprint: https://doi.org/10.1021/ja906987s. URL: https://doi.org/10.1021/ja906987s.

[100] L Zhou, H J Su, and C E Castro. "Programmable motion of DNA origami mechanisms". In: *Proceedings of the ...* Jan. 2015.

[101] Erik Winfree. "Self-healing Tile Sets". In: *Nanotechnology: Science and Computation*. Ed. by Junghuei Chen, Nataša Jonoska, and Grzegorz Rozenberg. Natural Computing Series. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 55–78. ISBN: 978-3-540-30296-4. DOI: 10.1007/3-540-30296-4_4. URL: https://doi.org/10.1007/3-540-30296-4_4 (visited on 03/08/2019).

[102] Constantine G. Evans and Erik Winfree. "Physical principles for DNA tile self-assembly". In: *Chemical Society Reviews* 46.12 (2017), pp. 3808–3829. ISSN: 0306-0012, 1460-4744. DOI: 10.1039/C6CS00745G. URL: http://xlink.rsc.org/?DOI=C6CS00745G (visited on 02/25/2019).

[103] Yong You, Andrey V. Tataurov, and Richard Owczarzy. "Measuring Thermodynamic Details of DNA Hybridization Using Fluorescence". In: *Biopolymers* 95.7 (2011), pp. 472–486. ISSN: 1097-0282. DOI: 10.1002/bip.21615. URL: https://www.onlinelibrary.wiley.com/doi/abs/10.1002/bip.21615 (visited on 03/18/2020).

[104] R. Schulman and E. Winfree. "Programmable Control of Nucleation for Algorithmic Self-Assembly". In: *SIAM Journal on Computing* 39.4 (Dec. 4, 2009), pp. 1581–1616. ISSN: 0097-5397. DOI: 10.1137/070680266. URL: https://epubs.siam.org/doi/abs/10.1137/070680266 (visited on 05/02/2019).