Fundamental scaling constraints for equilibrium molecular computing

Erin Crawley [1,2,*,†] Qian-Ze Zhu [1,*,‡] and Michael P. Brenner [1,2,§]

[1] *School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts 02138, USA*

[2] *Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

(Received 25 November 2025; accepted 7 May 2026; published 6 August 2026)

Molecular computing promises massive parallelization to explore solution spaces, but so far practical implementations remain limited due to off-target binding and exponential proliferation of competing structures. Here, we investigate the theoretical limits of equilibrium self-assembly systems for solving computing problems, focusing on the directed Hamiltonian path problem (HPP) as a benchmark for NP-complete problems. The HPP is encoded via particles with directional lock-key patches, where self-assembled chains form candidate solution paths. We determine constraints on the required energy gap between on-target and off-target binding for an acyclic HPP to be encoded and solved. We simultaneously examine whether components with the required energy gap can be designed. Combining these results yields a phase diagram identifying regions where HPP problems are both solvable and designable. These results establish fundamental upper bounds on equilibrium molecular computation and highlight the necessity of nonequilibrium approaches for scalable molecular computing architectures.

DOI: 10.1103/x9sw-h1rh

I. INTRODUCTION

Frontiers of computing are expanding to tackle larger problems and broader application domains. Beyond the remarkable success of silicon-based computing, there is interest in how alternative physical systems can provide novel approaches to computation and information processing. While quantum computing has made tremendous progress in recent years, achieving improvements over classical approaches [1] and major advances in quantum error correction [2,3], there has been comparatively less progress in molecular computing. However, molecular computing has innate advantages: molecular components—such as self-assembling particles or interacting polymers—can collectively explore vast configuration spaces in parallel [4–7]. At the same time, biology itself offers an existence proof for molecular computers: biological systems demonstrate that collections of molecules paired with directed energy sources can solve complex computational problems.

Here, we consider the Directed Hamiltonian path problem (HPP) as a benchmark problem for molecular computing. The HPP is a canonical NP-complete problem that asks whether there exists a path in a directed graph that starts and ends at designated nodes and visits all nodes exactly once [8]. It is a well-posed problem where solving large instances is challenging for conventional silicon-based hardware.

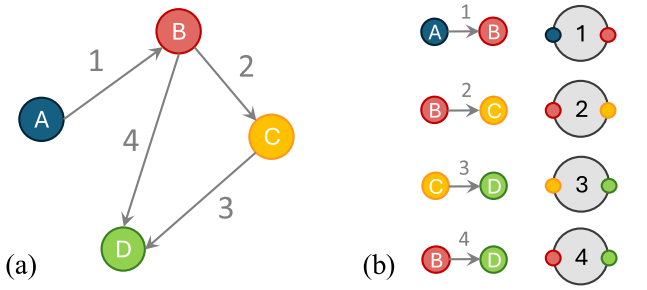
Leveraging molecular computers for NP-complete problems is an old idea; Adleman demonstrated that programmed DNA hybridization reactions could solve the HPP for a seven-node graph [9]. However, Adleman's idea has since stagnated beyond proof-of-principle demonstrations, owing to key scaling limitations [7,10]. As the problem size (N) increases, the solution space expands combinatorially, yielding an exponential number of competing structures. Thus, suppressing crosstalk and maintaining the requisite binding specificity becomes prohibitively difficult—a constraint that has been established quantitatively in the context of multicomponent DNA and colloidal self-assembly [10–14]. This fatally hinders reliable computations as the problem size increases ($N \to \infty$). Most work extending Adleman's initial ideas focuses on computation with biological molecules, including enzymes and proteins, which likewise suffer from poor binding specificity [15–19].

Yet, the key question is not whether *existing* biological molecules can solve difficult computational problems. Rather, it is whether there exists *any* molecular design for the underlying molecules that simultaneously circumvents the exponential growth of competing structures and avoids the crosstalk catastrophe that limits the computational ability of molecular systems.

Recent convergence of capabilities in synthetic biology, protein engineering, and nanoscale manufacturing now provide practical technologies to enable more refined approaches. In principle, such a design could operate in or out of equilibrium, with the only constraint being that it can be designed with existing components. As an example, one may modify the information carrying model with the binding characteristics of components [20–22]. Alternatively, the binding characteristics could change depending on the local environment of a bound particle [23–25].

Here, we consider a subset of this general problem and ask whether it is possible to design molecular components that can solve the HPP when the system is constrained to be in equilibrium. We focus on two competing requirements: (1) that a molecular system can encode and solve the target

*These authors contributed equally to this work.
†Contact author: erincrawley@g.harvard.edu
‡Contact author: qianzezhu@g.harvard.edu
§Contact author: brenner@seas.harvard.edu



Species	1: A → B	2: B → C	3: C → D	4: B → D
1: A → B	W	S	W	S
2: B → C	W	W	S	W
3: C → D	W	W	W	W
4: B → D	W	W	W	W

(c)

(d) $$Yield = \frac{Z_{target}}{Z_{all}} = \frac{1\,2\,3}{1\,2 \quad 1\,3 \quad 1\,3\,4 \cdots}$$

FIG. 1. Encoding the Hamiltonian Path Problem into a patchy particle system. (a) A directed graph with four nodes and four directed edges, including one Hamiltonian path $A \to B \to C \to D$ (edges $1 \to 2 \to 3$). (b) Each directed edge is represented as a unique particle species with two patches: the source and target nodes. (c) The interaction matrix encoding pairwise binding strengths between species. A strong interaction s is assigned if two edges are consecutively connected in the graph; otherwise a weak interaction w is used to model crosstalk. (d) The yield is defined as the partition function of the target structure (the Hamiltonian path) divided by the total partition function of all possible assembled chains.

problem, and (2) that the necessary interactions are designable with realistic components. While prior work has established that the required energy gap grows with the number of components, the encodability and designability constraints have not previously been treated simultaneously. We find that these simultaneous constraints are in tension, and we demonstrate that equilibrium assembly is only possible in a small sliver of the designable phase space.

II. SYSTEM SET UP AND OPTIMIZATION

In contrast to Adleman's pure DNA strand encoding of this problem, we encode the HPP as a self-assembling system of patchy particles: each edge in the directed graph corresponds to a unique particle species, and each particle has directional interacting patches corresponding to the source and target nodes of that edge [Figs. 1(a) and 1(b)]. Two particle species are designed to interact strongly if the target node of one matches the source node of the other, and weakly otherwise to model crosstalk, as shown in an example interaction matrix in Fig. 1(c).

Self-assembled linear chains of interacting "edge" particles represent candidate paths through the graph, including the possibility of node mismatch due to crosstalk. The Hamiltonian path for a graph with N nodes, if it exists, corresponds to a specific self-assembled sequence of $N-1$ particles with all strong interactions. In this paper, we restrict ourselves to *acyclic* graphs which contain no loops; results and discussion for general graphs (including cycles) are provided in Supplemental Material (SM) [26].

Specifically, in our molecular-computing framework, "solving" the HPP means designing a system whose equilibrium ensemble is concentrated on the correct assembled structure. The molecular system does not calculate the yield; it physically realizes it. Here, self-assembled chains represent candidate paths, and the Hamiltonian path is a particular chain of length $N-1$. Since the equilibrium abundance of each chain is proportional to its Boltzmann weight, an instance is solvable only if the equilibrium weight of the Hamiltonian path remains appreciable relative to the exponentially many competing assemblies.

We analyze the system in the grand canonical ensemble, where each species of edge particle i has an associated concentration c_i. Each pair of particles i, j interacts with Boltzmann weight $e^{-\beta E_{ij}}$, where E_{ij} is the interaction energy and $\beta = 1/(k_B T)$. The interaction energy matrix then takes the form

$$E_{ij} = \begin{cases} s, & \text{target node of } i \text{ equals source node of } j \\ w, & \text{otherwise,} \end{cases} \tag{1}$$

where $s < 0$ is the strong binding energy and $w < 0$ is the average off-target binding energy, satisfying $w > s$. The yield of the Hamiltonian path is given by the equilibrium fraction of the Hamiltonian path partition function as $Y_{\mathrm{HP}} = Z_{\mathrm{HP}}/Z_{\mathrm{all}}$, where Z_{all} is the total partition function summing over all possible structures. An illustration of the yield of the Hamiltonian path is shown in Fig. 1(d). Following [11], Z_{all} can be exactly calculated using a transfer matrix method, allowing us to express the yield as

$$Y_{\mathrm{HP}} = \frac{Z_{\mathrm{HP}}}{Z_{\mathrm{all}}} = \frac{(\prod_{i\in \mathrm{HP}} c_i) e^{-(N-2)\beta s}}{\sum_{i,j} c_i [(\mathbb{I} - \mathbf{B})^{-1}]_{ij}}, \tag{2}$$

where $\mathbf{B}$ is a matrix with entries $B_{ij} = e^{-\beta E_{ij}} c_j$. Given this parametrization of the yield, we can optimize over component concentrations to maximize the yield.

Before optimizing for concentrations, we note necessary constraints on the energy gap, $w - s$. For the yield to be maintained, the energetic contributions must outweigh the entropic ones for the desired structure. For the trivial case with no extraneous edges outside the Hamiltonian path, the analysis of heterogeneous one-dimensional chain self-assembly in [11] directly yields the bound: $w - s > 2\ \ln(N) k_B T$. As such, we parametrize the energy gap with two parameters, a and b, as

$$w - s = a\ \ln(N) + b. \tag{3}$$

In (3), $a \geqslant 2$ is the energy gap-scaling coefficient, which determines the strength of the energy gap $w - s$, in units of $k_B T / \ln(N)$, and $b \geqslant 0$ is an offset parameter. We will demonstrate below that solving general HPP instances will require even stronger constraints on the gap-scaling coefficient a. A key challenge in solving the HPP via the optimization framework in [11] is that the target structure—the Hamiltonian

path itself—is unknown. This precludes direct optimization of its yield as in Fig. 1(d), as the correct path cannot be explicitly enumerated in advance. Although the specific path is unknown, any Hamiltonian path must contain exactly $N-1$ edges in order to visit all N nodes once and only once. Therefore, we can maximize a surrogate objective: the total yield of all linear assemblies of length $N-1$. We refer to this quantity as the filtered yield. Physically, this corresponds to restricting attention to chains of the correct length, e.g., by a size-selection step. This filtered yield includes the desired Hamiltonian path as well as competing structures of the same length. Hence the ratio $Y_{\mathrm{HP}}/Y_{\mathrm{filtered}}$ measures how uniquely the Hamiltonian path is selected within the decoded sector. Mathematically, the filtered yield is given by

$$Y_{\mathrm{filtered}} = \frac{Z_{\mathrm{length}=N-1}}{Z_{\mathrm{all}}} = \frac{\sum_{i,j} c_i [\mathbf{B}^{N-2}]_{ij}}{\sum_{i,j} c_i [(\mathbb{I}-\mathbf{B})^{-1}]_{ij}}, \tag{4}$$

where $\mathbf{B}$ is defined as in (2). We then optimize the species concentrations $\{c_i\}$ to maximize the surrogate objective $\ln(Y_{\mathrm{filtered}})$, and use the resulting set of optimal $\{c_i^*\}$ to evaluate the true yield of the Hamiltonian path Y_{HP} given in (2). When $Y_{\mathrm{HP}} \approx Y_{\mathrm{filtered}}$, nearly all chains of length $N-1$ are the Hamiltonian path, so filtering by length is sufficient to recover the solution; when $Y_{\mathrm{HP}} \ll Y_{\mathrm{filtered}}$, the length-$(N-1)$ sector is dominated by competing chains and the problem is not reliably solvable by this equilibrium protocol.

We test and evaluate this method on randomly generated directed graphs. The graphs are generated as follows: (1) a single Hamiltonian path is constructed, in order to enable ground-truth yield evaluation; (2) additional extraneous edges are added with probability ρ_b; (3) these directed edges are only added to "shortcut" the Hamiltonian path, thus restricting the graphs to be acyclic (see an accompanying python notebook [27] for details). For general graphs with loops, there is an exponential proliferation of competing structures with all strong bonds (see additional results and discussion in the SM [26]). Under these restrictions, the expected number of edges in an acyclic graph with N nodes is given by

$$N_{\mathrm{edge}} = N - 1 + \frac{\rho_b}{2}(N-1)(N-2). \tag{5}$$

For each N ranging from 5 to 100 nodes, we randomly generate 200 graphs following the above process and perform the optimization over concentrations for each graph independently.

Figure 2(a) shows the yield behavior as a function of graph size N for a representative energy parameter choice $a = 6.0$, $b = 2.0$ [defined in (3)]. In the upper panel, we plot the median of the logarithmic filtered yield (blue) and the Hamiltonian path (HP) yield (red) over 200 graph instances. Individual points indicate yields for each realization, while the variability—defined as the difference between the 75th and 25th quartiles—is used to quantify fluctuations and is shown in the lower panel of Fig. 2(a). Beyond a critical threshold value of N, denoted by N_c, a sharp separation between filtered yield and HP yield emerges, with an abrupt drop in the HP yield. The variability also peaks sharply near the critical N_c, reflecting the coexistence of two nearly degenerate free-energy states—the ordered HP and disordered misbound configurations.

This critical transition becomes more apparent when examining the effect of varying the gap-scaling coefficient a, as shown in Fig. 2(b). To quantify the observed scaling, we collapse the yield curves using the ansatz:

$$\ln(Y_{\mathrm{HP}}) = a^{\alpha} F\left(\frac{N}{a^{\eta}}\right) \tag{6}$$

for some function F, with fitted exponents $\eta = 1.222$ ($R^2 = 1.000$) and $\alpha = -0.178$ ($R^2 = 0.485$), as shown in the inset of Fig. 2(c). The excellent collapse along the N/a^{η} axis confirms that the transition is governed by a scaling law, resembling a first-order phase transition.

We extract the critical size N_c for each a based on a yield threshold criterion: $\ln Y_{\mathrm{HP}} < -10$. This threshold is robust, as the HP yield decreases rapidly and consistently near the transition. We can also independently extract N_c as the point where the variability spikes. As shown in Fig. 2(c), the critical N_c values extracted from the yield drop (red) agree closely with those inferred from the peak of the variability fluctuation curves (blue), providing independent validation. These results are consistent with a first-order-like transition, with the HP yield serving as the order parameter. The system exhibits competing free-energy minima: an energetically favored ordered HP configuration and an exponentially large manifold of competing structures favored entropically. As N increases, the entropic manifold gains weight until, at $N = N_c$, the two basins become comparable.

To further understand this, in Fig. 2(d) we examine the critical energy gap $\Delta E_c = w - s = a \ln N_c + b$ versus N_c for varying ρ_b and b values. The critical energy gap required to reliably assemble the HP increases with both system size N and ρ_b. The parameter ρ_b controls the number of possible extraneous structures that can compete with the HP. All curves collapse by ρ_b, indicating that the density of extraneous edges sets the universal energy–entropy balance of this problem. We derive the total partition function of all length-$(N-1)$ chains in closed form via a generating function approach (see details in SM [26]), and show the asymptotic scaling to be

$$a(N_c) = \frac{\Lambda(\rho_{\mathrm{eff}})}{\ln N_c} N_c + O(1), \tag{7}$$

where $\Lambda(\rho) = 2\ln(1+\sqrt{\rho})$, and ρ_{eff} is an effective edge density that incorporates unequal particle concentrations, and reduces to ρ_b when component concentrations are equal. However, in the N range we probed here, the logarithmic correction is of order unity, so a simple linear fit effectively captures the data: the dashed line in Fig. 2(c) provides an excellent linear fit with $R^2 = 0.9983$. For simplicity, we use this linear approximation in the designability discussion that follows. On the other hand, (7) also implies that the critical energy gap, ΔE_c, scales linearly with N_c:

$$\Delta E_c \sim \Lambda(\rho_{\mathrm{eff}}) N_c. \tag{8}$$

This explains the near linear dependence observed in Fig. 2(d), with slopes dependent on ρ_b.

III. DESIGNABILITY OF INTERACTIONS

Now that we have a protocol for optimization without prior knowledge of the solution, we must ask whether the interac-

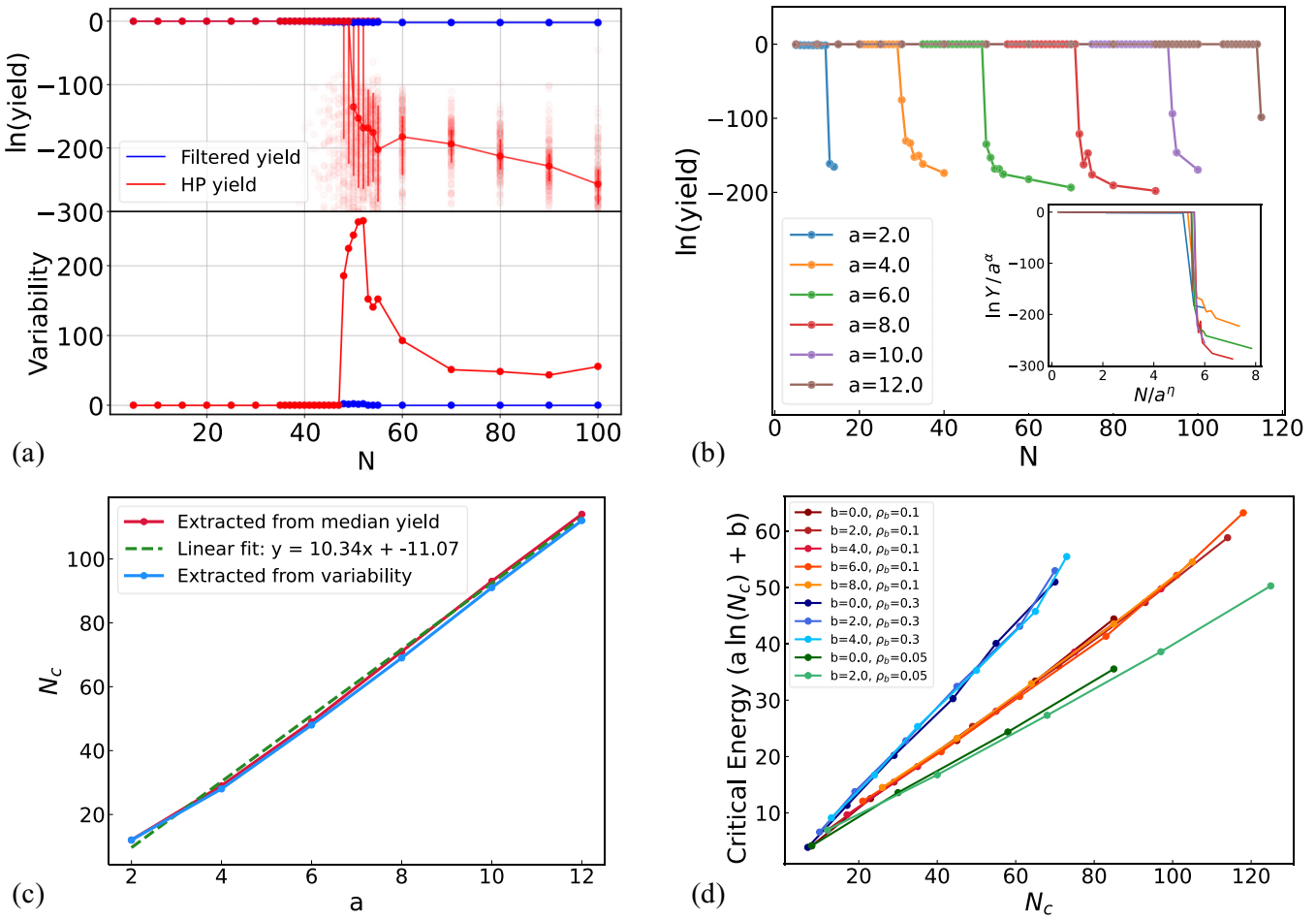


FIG. 2. Scaling of HP yield and phase transitionlike behavior. (a) Upper: Median log yield of filtered assemblies (blue) and Hamiltonian path assemblies (red) as a function of graph size N for $a = 6.0$, $b = 2.0$, $w = 0.0$, $\rho_b = 0.1$. Points show individual graph instances; error bars indicate interquartile range. Lower: Variability of log yields across graph instances for the same parameters. (b) HP yield curves as a function of graph size N for various gap-scaling coefficients a. Other conditions are the same: $b = 2.0$, $w = 0.0$, $\rho_b = 0.1$. Inset: Curves for different a collapse under the scaling form $\ln Y_{\mathrm{HP}} = a^{\alpha} F(N/a^{\eta})$. (c) Critical graph size N_c extracted from yield drop (red) and variability peaks (blue). Dashed line: best linear fit with $R^2 = 0.9983$. (d) Critical energy gap $w - s = a \ln N_c + b$ as a function of N_c.

tions required to encode and solve the HPP are designable, at least in principle. In order to encode and solve the HPP in a molecular system, one must be able to design N_{edge} distinct bindings for each of the graph edges while ensuring that the difference between on-target and off-target bindings, $w - s$, is sufficiently large.

We model potential bindings as a lock-key pair of chains, each of length L with components drawn from an alphabet of size A containing the "letters" $\{1, \ldots, A\}$. A diagram of this setup with locks denoted by $x^{(i)}$ and keys denoted by $y^{(i)}$ is shown in Fig. 3(a). Suppose that on-target bonds between matching letters have average energy $\epsilon_s < 0$, and off-target bonds have binding energy $\epsilon_w < 0$, with $|\epsilon_w| < |\epsilon_s|$. Then, the average strong binding for a chain of length L is $s = \epsilon_s L$ and the average weak binding energy is $w = \epsilon_s \bar{n}_w + \epsilon_w (L - \bar{n}_w)$, where $\bar{n}_w$ is the average number of matching letters between two off-target chains.

Thus, in order to encode the HPP into a system with alphabet size A, one must find N_{edge} bindings of some length L so that the energy gap between on- and off-target binding satisfies

$$w - s = \epsilon_s \bar{n}_w + \epsilon_w (L - \bar{n}_w) - \epsilon_s L \geqslant a \ln(N) + b. \tag{9}$$

To consider self-assembly processes over experimentally realistic timescales, we require that s is fixed below some threshold. With a fixed $|s|$, we must then require $\epsilon_s = s/L$, so that ϵ_s is smaller for larger L. A similar scheme is presented for magnetic systems in [28], with $s = -10 k_B T$. With fixed strong bonding s, the "encodability" bound (9) becomes

$$|s| \left(1 - \frac{\bar{n}_w}{L}\right) \left(1 - \frac{\epsilon_w}{\epsilon_s}\right) \geqslant a \ln(N) + b. \tag{10}$$

For a molecular system with the three design parameters of energy fraction $\frac{\epsilon_w}{\epsilon_s}$, overlap fraction $\frac{\bar{n}_w}{L}$, and the maximum allowed s, (10) yields the maximum allowed a and b values that can be realized for a given N.

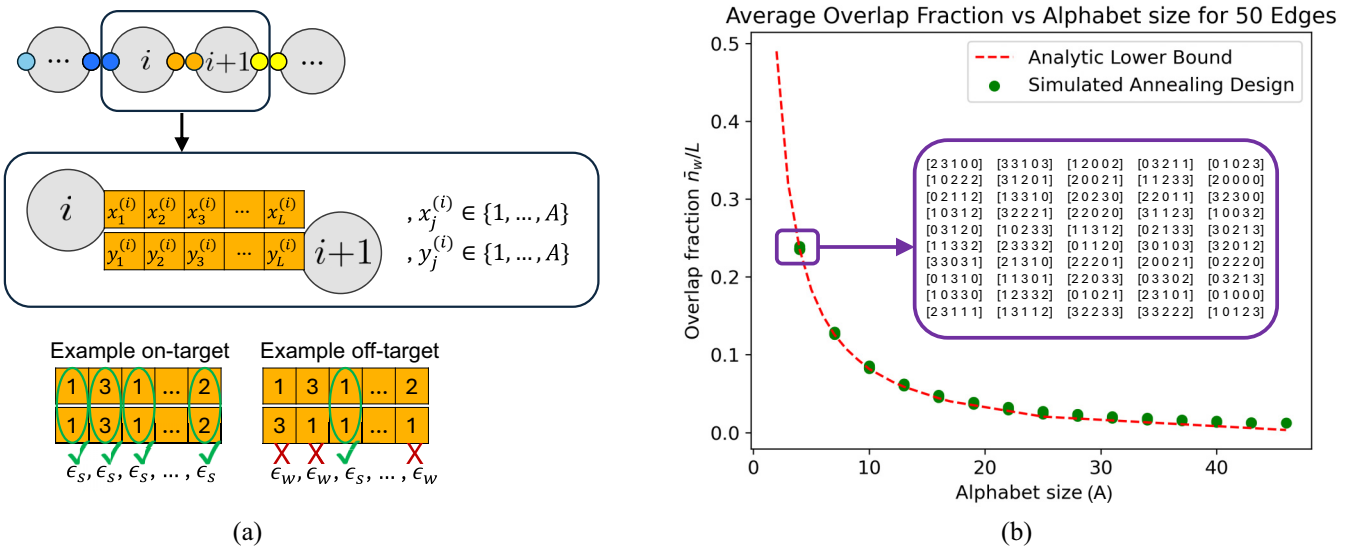


FIG. 3. Binding design scheme and agreement with analytic expression. (a) Binding sites between particles are modeled via a lock-key pair of chains of length L with components chosen from an alphabet of size A. On-target "letters" have an average energy of ϵ_s and off-target components an average energy of ϵ_w, to model crosstalk. Examples of an on-target binding pair and an off-target pair are shown, where $A = 3$ and $L = 5$. The on-target binding has total strength $s = \epsilon_s L$. (b) Overlap fraction $\frac{\bar{n}_w}{L}$ as a function of alphabet size, A. The analytic lower bound (11) and a numerical simulated annealing protocol show good agreement. An example of bindings for $A = 4$ and $L = 5$ generated via simulated annealing is shown.

A lower bound for the parameter $\frac{\bar{n}_w}{L}$ can be found in terms of the alphabet size A by (see SM [26] for details)

$$\frac{\bar{n}_w}{L} = \frac{(A-r)q^2 + r(q+1)^2 - N_{\text{edge}}}{N_{\text{edge}}(N_{\text{edge}}-1)}, \tag{11}$$

where $q = \lfloor N_{\text{edge}}/A \rfloor$ and $r = N_{\text{edge}} \bmod A$. When N_{edge} is divisible by A this simplifies to $\frac{\bar{n}_w}{L} = \frac{1}{N_{\text{edge}}-1}(\frac{N_{\text{edge}}}{A} - 1)$. Figure 3(b) shows good agreement between the analytic lower bound (11) and a numerical simulated annealing protocol. In the limit where $N_{\text{edge}} \gg A$, this lower bound becomes $\frac{\bar{n}_w}{L} \sim \frac{1}{A}$. This makes sense as given a particular "lock" component $x_k^{(i)}$, there is a probability of $\frac{1}{A}$ that an off-target "key" component, $y_k^{(j)}$, will match. Then, the total expected number of overlaps between off-target pairs $x^{(i)}$ and $y^{(j)}$ is $\bar{n}_w = \frac{L}{A}$.

IV. PHASE DIAGRAM FOR ENCODING AND SOLVING HPP PROBLEMS

Using the results of the previous two sections, we can now determine the largest graph size for an HPP to be encoded and solved by a given molecular system in equilibrium. By imposing equality in (10) and substituting the lower bound on $\frac{\bar{n}_w}{L}$ from (11), we obtain the maximal attainable values of a and b for any molecular system with alphabet size A and off-target energy fraction $\frac{\epsilon_w}{\epsilon_s}$. In Fig. 4, we provide a phase diagram showing regions where the design of certain components is possible and regions where the HPP is solvable. The intersection of these regions is where it is feasible that a given designed molecular system can encode and solve an HPP.

The phase diagram in Fig. 4 shows regions for an ideal molecular design, a synthetic magnetic panel system from [28], and a toy model of DNA. The ideal design features a very large alphabet size A and zero crosstalk between letters, $\frac{\epsilon_w}{\epsilon_s} = 0$, maximizing the attainable energy gap in (10). Even in this ideal case, however, the molecular system can solve the HPP only up to a finite graph size. For the representative parameters $s = 10$, $b = 0$, and $\rho_b = 0.1$, the maximum solvable graph size is $N = 19$.

In the magnetic system [28], sets of magnetic dipole panels were designed to minimize crosstalk between off-target panels. The phase diagram yields an estimate of the largest HPP that could be solved with their 12-panel system, treating the magnetic panels as the system's letters ($A = 12$). For the experimentally realized design with three dipoles per panel—corresponding to an off-target energy fraction of $\frac{\epsilon_w}{\epsilon_s} = 0.66$—the maximum designable and solvable graph problem has five nodes.

For our estimated DNA model, we use an alphabet size of $A = 4$, corresponding to the canonical base pairs. To estimate the off-target energy fraction ϵ_w/ϵ_s, we use the worst case overlap between off-target base pairs from [29], namely, Guanine-Guanine with binding energy -2.22 kcal/mol. On-target base pair energies are -12.1 kcal/mol for A–T and -21.0 kcal/mol for G–C [30]. Taking the latter as the representative on-target value, we obtain an estimate of $\epsilon_w/\epsilon_s \approx 0.1$ for a DNA-like system. We note that our DNA-like system is a toy model: it does not instantiate the actual base pair binding energies but instead fixes the total strong bonding energy s as well as the ratio ϵ_w/ϵ_s.

Table I lists further example systems with the largest possible solvable HPP problem for each case. Phase diagrams and additional results for these systems (along with the "Pacman" particles from [31]) can be found in SM [26].

In both the ideal case and in these specific examples, Fig. 4 demonstrates that for this problem, there is a fundamental tension between designability and solvability. With increasing

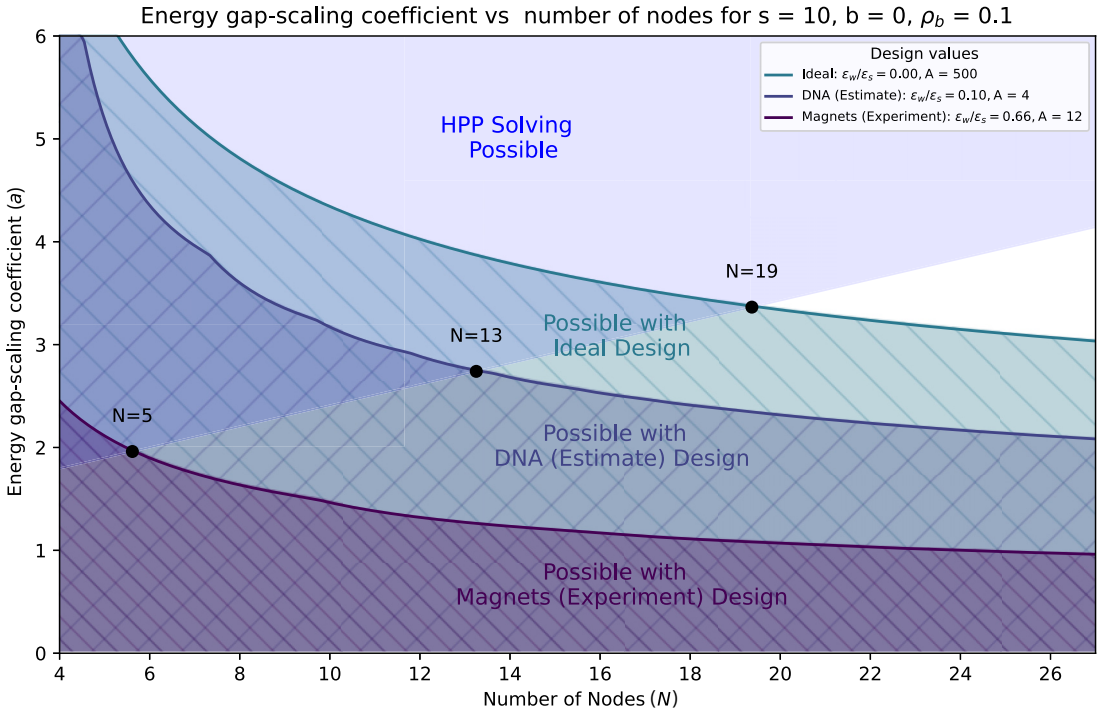


FIG. 4. Phase diagram for designable molecular systems that can solve HPP problems. Phase diagram for encoding and solving an HPP in equilibrium, with parameters $b = 0$, $s = 10$, $\rho_b = 0.1$. For the ideal case with large A and $\frac{\epsilon_w}{\epsilon_s} = 0$, the largest solvable graph has 19 nodes. Using a heuristic for DNA base pairs as the letters, with $A = 4$ and $\epsilon_w/\epsilon_s = 0.1$, the largest graph has 13 nodes. Using the magnetic dipole system from [28] as the "letters," with $A = 12$ and $\frac{\epsilon_w}{\epsilon_s} = 0.66$, the largest solvable graph has five nodes.

N, a larger number of components need to be designed, so the maximum energy gap between on- and off-target components (governed by a) decays. On the other hand, for the HPP to be encoded and solved, a must increase with N [as shown in Fig. 2(c)]. These competing constraints ensure that even in the ideal design case, there is a stopping point for the largest possible solvable problem in equilibrium. Moreover, this is a best case scenario bound: our model for component design assumes lock and key binding strings exactly overlap, and the method we outline for solving HPP problems by optimizing for the filtered yield holds only for acyclic graphs. The existence of a maximum possible problem size, even in this best case scenario, motivates that molecular computing in equilibrium cannot scale robustly enough to solve larger and more complex problems. Moreover, since the equilibrium yield is an upper bound on the yield of such a self-assembly protocol, the scaling limits found here are conservative: kinetic barriers and finite equilibration times can only further reduce the yield. We propose that to do this, one will need to include out-of-equilibrium protocols, such as imposing constraints on particle supply [10] or incorporating kinetic proofreading mechanisms [23,32,33].

ACKNOWLEDGMENTS

This work was supported by the NSF Award 2112085 for the AI Institute in Dynamic Systems and the Alfred P. Sloan Foundation under Grant No. G-2021-14198 and a Natural Sciences and Engineering Research Council of Canada (NSERC) PGS-D fellowship to E.C.

TABLE I. Design and HPP-solving parameters for several example systems and the maximum size N that can be attained.

	Design			HPP		
System	A	Mean ϵ_w/ϵ_s	$\lvert s\rvert$ (k_BT)	b	ρ_b	Max. N
Ideal system	∞	0	10	0	0.05	22
Ideal system	∞	0	10	0	0.1	19
Ideal system	∞	0	10	0	0.3	15
3-Dipole magnet panels ($b = 0$)	12	0.66	10	0	0.1	5
3-Dipole magnet panels ($b = 2$)	12	0.66	10	2	0.1	None
25-Dipole magnet panels ($b = 0$)	12	0.33	10	0	0.1	13
25-Dipole magnet panels ($b = 2$)	12	0.33	10	2	0.1	11
DNA (estimate, $\rho_b = 0.05$)	4	0.1	10	0	0.05	14
DNA (estimate, $\rho_b = 0.1$)	4	0.1	10	0	0.1	13

DATA AVAILABILITY

The data that support the findings of this article are openly available [27].

[1] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, R. Biswas, S. Boixo, F. G. Brandao, D. A. Buell, *et al.*, Quantum supremacy using a programmable superconducting processor, Nature (London) **574**, 505 (2019).

[2] Google Quantum AI, Suppressing quantum errors by scaling a surface code logical qubit, Nature (London) **614**, 676 (2023).

[3] D. Bluvstein, S. J. Evered, A. A. Geim, S. H. Li, H. Zhou, T. Manovitz, S. Ebadi, M. Cain, M. Kalinowski, D. Hangleiter, *et al.*, Logical quantum processor based on reconfigurable atom arrays, Nature (London) **626**, 58 (2024).

[4] J. H. Reif, Paradigms for biomolecular computation, in *First International Conference on Unconventional Models of Computation* (Springer Verlag, Singapore, 1998), pp. 72–93.

[5] M. H. Garzon and R. J. Deaton, Biomolecular computing and programming, IEEE Trans. Evol. Comput. **3**, 236 (2002).

[6] E. Winfree, DNA computing by self-assembly, in *2003 NAE Symposium on Frontiers of Engineering* (The National Academies Press, Washington, DC, 2004), pp. 105–117.

[7] Z. Ezziane, DNA computing: Applications and challenges, Nanotechnology **17**, R27 (2006).

[8] M. R. Garey and D. S. Johnson, *Computers and Intractability* (W. H. Freeman and Company, New York, 2002), Vol. 29.

[9] L. M. Adleman, Molecular computation of solutions to combinatorial problems, Science **266**, 1021 (1994).

[10] S. A. Kurtz, S. R. Mahaney, J. S. Royer, and J. Simon, Active transport in biological computing, in *DNA Based Computers II* (American Mathematical Society, Providence, RI, 1999), pp. 171–179.

[11] A. Murugan, J. Zou, and M. P. Brenner, Undesired usage and the robust self-assembly of heterogeneous structures, Nat. Commun. **6**, 6203 (2015).

[12] M. H. Huntley, A. Murugan, and M. P. Brenner, Information capacity of specific interactions, Proc. Natl. Acad. Sci. USA **113**, 5841 (2016).

[13] Z. Zeravcic, V. N. Manoharan, and M. P. Brenner, Size limits of self-assembled colloidal structures made using specific interactions, Proc. Natl. Acad. Sci. USA **111**, 15918 (2014).

[14] C. G. Evans and E. Winfree, Physical principles for DNA tile self-assembly, Chem. Soc. Rev. **46**, 3808 (2017).

[15] W. D. Smith, DNA computers in vitro and vivo, in *DNA Based Computers*, Series in Discrete Mathematics and Theoretical Computer Science (American Mathematical Society, Providence, RI, 1996), Vol. 27, pp. 121–185.

[16] S. Roweis, E. Winfree, R. Burgoyne, N. V. Chelyapov, M. F. Goodman, P. W. Rothemund, and L. M. Adleman, A sticker-based model for DNA computation, J. Comput. Biol. **5**, 615 (1998).

[17] D. Faulhammer, A. R. Cukras, R. J. Lipton, and L. F. Landweber, Molecular computation: RNA solutions to chess problems, Proc. Natl. Acad. Sci. USA **97**, 1385 (2000).

[18] H. Hug and R. Schuler, Strategies for the development of a peptide computer, Bioinformatics **17**, 364 (2001).

[19] R. Bar-Ziv, T. Tlusty, and A. Libchaber, Protein–DNA computation by stochastic assembly cascade, Proc. Natl. Acad. Sci. USA **99**, 11589 (2002).

[20] W. B. Rogers and V. N. Manoharan, Programming colloidal phase transitions with DNA strand displacement, Science **347**, 639 (2015).

[21] A. McMullen, S. Hilgenfeldt, and J. Brujic, DNA self-organization controls valence in programmable colloid design, Proc. Natl. Acad. Sci. USA **118**, e2112604118 (2021).

[22] E. M. King, C. X. Du, Q.-Z. Zhu, S. S. Schoenholz, and M. P. Brenner, Programming patchy particles for materials assembly design, Proc. Natl. Acad. Sci. USA **121**, e2311891121 (2024).

[23] Q.-Z. Zhu, C. X. Du, E. M. King, and M. P. Brenner, Proofreading mechanism for colloidal self-assembly, Phys. Rev. Res. **6**, L042057 (2024).

[24] J. Evans and P. Šulc, Designing 3D multicomponent self-assembling systems with signal-passing building blocks, J. Chem. Phys. **160**, 084902 (2024).

[25] J. Metson, Designing complex behaviors using transition-based allosteric self-assembly, Phys. Rev. Res. **7**, 033044 (2025).

[26] See Supplemental Material at http://link.aps.org/supplemental/10.1103/x9sw-h1rh for optimization analysis for graphs with loops, detailed derivations for analytical formulas, and phase diagrams for additional parameters.

[27] E. Crawley, Q.-Z. Zhu, and M. P. Brenner, GitHub, 2025, https://github.com/RosellaZ/equilibrium-molecular-computing.

[28] C. X. Du, H. A. Zhang, T. G. Pearson, J. Ng, P. L. McEuen, I. Cohen, and M. P. Brenner, Programming interactions in magnetic handshake materials, Soft Matter **18**, 6404 (2022).

[29] N. Peyret, P. A. Seneviratne, H. T. Allawi, and J. SantaLucia, Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches, Biochemistry **38**, 3468 (1999).

[30] I. K. Yanson, A. B. Teplitsky, and L. F. Sukhodub, Experimental studies of molecular interactions between nitrogen bases of nucleic acids, Biopolymers **18**, 1149 (1979).

[31] S. Sacanna, W. T. M. Irvine, P. M. Chaikin, and D. J. Pine, Lock and key colloids, Nature (London) **464**, 575 (2010).

[32] J. J. Hopfield, Kinetic proofreading: A new mechanism for reducing errors in biosynthetic processes requiring high specificity, Proc. Natl. Acad. Sci. USA **71**, 4135 (1974).

[33] Z. Liang, M. X. Lim, Q.-Z. Zhu, F. Mottes, J. Z. Kim, L. Guttieres, C. Smart, T. Pearson, C. X. Du, M. Brenner, *et al.*, Magnetic decoupling as a proofreading strategy for high-yield, time-efficient microscale self-assembly, Proc. Natl. Acad. Sci. USA **122**, e2502361122 (2025).