Phase transition of parallelizability in assembly systems

Ikumi Kobayashi and Shin-ichi Sasa
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

(Received 2 February 2023; accepted 8 June 2023; published 5 July 2023)

We propose a phase transition on the feasibility of efficient parallel assembly. By introducing the parallel efficiency that measures how efficiently the parallel assembly works, the parallelizable phase is defined by its positive value. The parallelizable-unparallelizable transition is then identified by the nonanalytic change in the parallel efficiency from a positive value to zero. We present two analyzable models to demonstrate this phase transition in the limit of infinite system size.

DOI: 10.1103/PhysRevResearch.5.033004

I. INTRODUCTION

In industry and applied science, there are often situations where a large number of parts are assembled to make a complex product. Typical examples are automotive assembly and polymer synthesis. Recently, the advancement of nanotechnology [1] has made it possible to assemble nanoscale objects into desired structures [2–5]. Furthermore, control of polymer sequences [6,7] and assembly of colloidal particles [5,8–12] have been vigorously studied.

We use the term parallelization to describe simultaneously assembling subunits and then combining them to complete the final product. Parallelization increases assembling efficiency. The concept of parallelization has been studied in computer science [13,14]. Some problems can be efficiently solved by parallel computing, whereas others cannot [15]. Inspired by these studies, we explore analogous concepts in physical assembly work. Specifically, we aim to determine under what conditions efficient parallel assembly is feasible.

The feasibility of parallelization qualitatively changes the time required for assembly. When the number of parts L becomes large, the L dependence of the number of parallel steps d required for assembly is crucial. For example, when assembling hundreds to thousands of parts, the assembly time is drastically different depending on whether $d = O(\log L)$ or $d = O(L^{\alpha})$.

The feasibility of parallel assembly can be clarified by introducing the parallel efficiency η. The parallel efficiency η is defined as the ratio of the minimum number of steps $\log_2 L$ required to assemble L parts to the actual number of steps d taken for assembly. For instance, imagine the assembly of a 2-mer is achieved by combining two monomers, followed by the combination of 2-mers to form 4-mers, and so on. Under such a fully parallelized assembly, we have $d = \log_2 L$ and $\eta = 1$. Conversely, in the case of sequential assembly where components are added one by one, we have $d = L - 1$ and $\eta = \log_2 L/(L-1)$, which goes to zero in the limit of $L \to \infty$.

In this paper, we propose a phase transition called *parallelizability transition*, where parallelizable and unparallelizable phases are characterized by parallel efficiency. That is, when a system parameter is continuously changed, parallel efficiency exhibits a transition from a positive value to zero in the limit of infinite system size. We demonstrate this phase transition by presenting two analyzable models. In the first model, the quenched combinability model, one-dimensional chains are assembled in the smallest number of parallel steps. In the second model, the ANP model, a final product is assembled through random bonding reactions. We introduce the parallel efficiency to measure how efficiently the parallel assembly works. Then, we exactly show that both models exhibit the parallelizability transition.

II. SETUP OF THE QUENCHED COMBINABILITY MODEL

Let us consider the assembly work of connecting L different parts to create a single chain. An external operator tries to perform the most efficient parallel assembly possible. However, the components do not always fit together. Which pairs of states can be combined is predetermined and does not change during the assembly process. The quenched combinability model idealizes such a situation.

We consider the assembly of one-dimensional chains of length L. To precisely specify the geometric structure of the states, we use graph theory notations and terminologies. We denote as $G_{i,j}$ a path graph in which vertices from i to j are connected in order:

$$\begin{aligned} G_{i,j} &= (V, E), \\ V &= \{i, i+1, ..., j\}, \\ E &= \{(v, v+1) \mid v = i, i+1, ..., j-1\}. \end{aligned} \tag{1}$$

See Fig. 1 for the illustration of $G_{i,j}$. The final product G is a path graph of length L, i.e.,

$$G = G_{1,L}. \tag{2}$$

FIG. 1. We denote as $G_{i,j}$ a path graph in which vertices from i to j are connected in order. For simplicity, we assume that there are edges between touching vertices and omit describing the edges

The set of possible states S is the entirety of the connected subgraphs of G, i.e.,

$$S = \{G_{i,j} \mid 1 \leqslant i \leqslant j \leqslant L\}. \tag{3}$$

The initial parts set M is the set of states with a single vertex, i.e.,

$$M = \{G_{i,i} \mid 1 \leqslant i \leqslant L\}. \tag{4}$$

The case $L = 7$ is shown in Fig. 2. Each state $s \in S$ is either active (filled circles in Fig. 2) or inactive (open circles in Fig. 2). This active-inactive distinction represents the bonding properties of the state with other states. That is, active states can always combine with other states, while inactive–inactive pairs can combine with probability p. Note that randomness is introduced as a quenched disorder. We determine the set of allowed bondings $\hat{R}$ probabilistically according to the following rules:

For each tuple i, j, k $(1 \leqslant i \leqslant j < k \leqslant L)$,

(1) if either $G_{i,j}$ or $G_{j+1,k}$ is active, $(G_{i,j}, G_{j+1,k}) \in \hat{R}$ with probability 1;

(2) if both $G_{i,j}$ and $G_{j+1,k}$ are inactive, $(G_{i,j}, G_{j+1,k}) \in \hat{R}$ with probability p.

These rules are illustrated in Fig. 3. Let us assume that the activity is carried over to the postbonding state. That is, the product in case 1 is active and the product in case 2 is inactive. In addition, we suppose that only $G_{1,1}$ is active in M. From the initial condition that only $G_{1,1}$ is active in M and the propagation rule of the active state, we have

$$G_{i,j} = \begin{cases} \text{active} & (i = 1) \\ \text{inactive} & (i \neq 1). \end{cases} \tag{5}$$

Let us introduce the parallel efficiency to measure how well the parallelization is working in this system. We call diagrams like Fig. 4 assembly pathways. In the assembly pathway shown in Fig. 4, a chain of length $L = 7$ is assembled with four parallel steps. The number of parallel steps means the maximum distance from the upmost states to the bottom state. We denote by $d(T)$ the number of parallel steps of the assembly pathway T.

Let $\hat{d}$ be the least number of parallel steps required to assemble the final product G, i.e.,

$$\hat{d} := \min_{T \in \hat{U}_G} \{d(T)\}, \tag{6}$$

where $\hat{U}_G$ is the set of all realizable assembly pathways generating G. Note that not all assembly pathways are necessarily realizable. For example, if $(G_{5,5}, G_{6,7}) \notin \hat{R}$, the assembly pathway in Fig. 4 is unrealizable. Because the set of allowed bondings $\hat{R}$ is probabilistically determined, $\hat{d}$ is also a random variable.

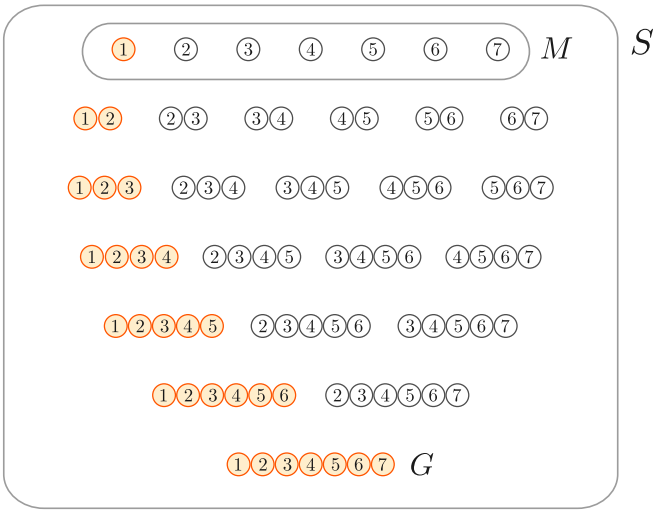

FIG. 2. All elements of S in the case $L = 7$. The final product G is shown at the bottom. The set of possible states S is the entirety of the connected subgraphs of G. The initial parts set M is shown at the top.

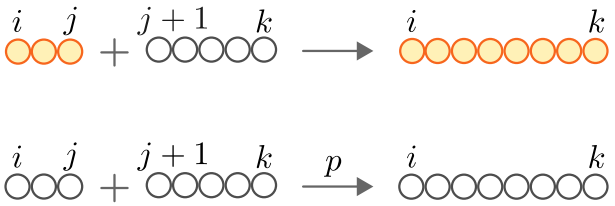

FIG. 3. Schematic of the decision procedure for the set of possible bondings $\hat{R}$. For each pair of states $(G_{i,j}, G_{j+1,k})$, (i) if either $G_{i,j}$ or $G_{j+1,k}$ is active $(i = 1)$, $(G_{i,j}, G_{j+1,k}) \in \hat{R}$ with probability 1; (ii) if both $G_{i,j}$ and $G_{j+1,k}$ are inactive $(i > 1)$, $(G_{i,j}, G_{j+1,k}) \in \hat{R}$ with probability p. By making this determination for every pair $(G_{i,j}, G_{j+1,k})$ of graphs, we determine the set $\hat{R}$ probabilistically.

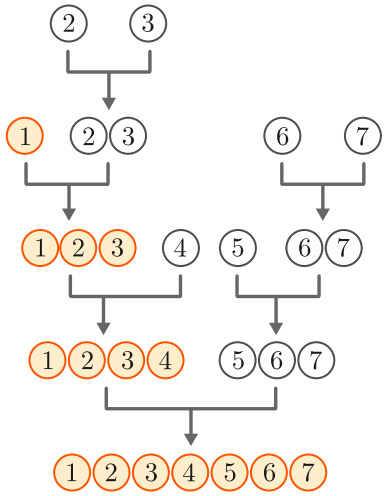

FIG. 4. Example of an assembly pathway generating $G_{1,7}$. In this case, $d(T) = 4$. Assembly pathways are binary trees that express the building process of a state. The product is placed at the bottom. The elements of M are placed at the top.

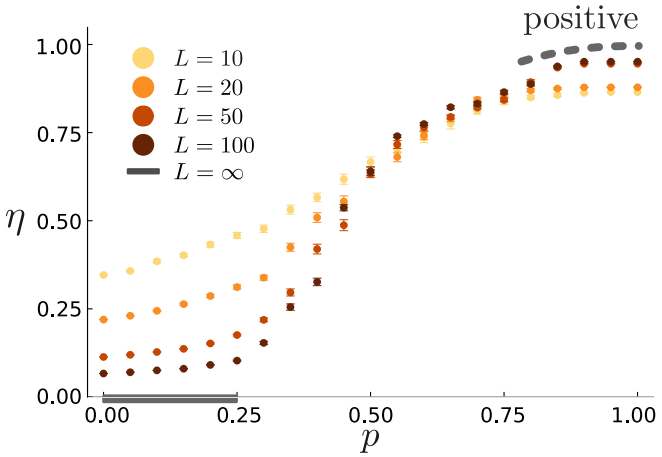

FIG. 5. Simulation results for $L = 10$, 20, 50, and 100 in the quenched combinability model. For each p, we took 100 samples and calculated the mean and standard deviation of η. The schematic of parallel efficiency η in the limit of infinite system size is overlaid. The dashed line represents qualitative behavior. A rigorous analysis shows that the analyticity of η is broken at a point p_c satisfying $1/4 \leqslant p_c < 3/4$.

The parallel efficiency η is defined as

$$\eta := \frac{\log_2 L}{\langle \hat{d} \rangle}, \tag{7}$$

where L is the total number of vertices and $\langle \hat{d} \rangle$ is the average minimum number of parallel steps. We can use the quantity η to measure the efficiency of parallel assembly in this system because it satisfies the following two properties. First, η satisfies the normalization condition $0 \leqslant \eta \leqslant 1$. Second, we can determine the feasibility of efficient parallel assembly in the limit of infinite system size by checking whether $\lim_{L\to\infty} \eta$ is positive or zero. If the assembly process is sufficiently parallelized and $\langle \hat{d} \rangle \sim \log L$, then $\lim_{L\to\infty} \eta$ is positive. In contrast, if the parallelization breaks down and $\langle \hat{d} \rangle$ grows faster than $\log L$, then $\lim_{L\to\infty} \eta$ becomes zero.

III. RESULTS FOR THE QUENCHED COMBINABILITY MODEL

We first display the numerical results in Fig. 5. Although the data suggest the existence of a phase transition, it is quite difficult to perform the numerical calculation for a larger system. Nevertheless, we have a rigorous proof for the existence of the phase transition (see also Appendix) when L becomes infinite. We can show

$$\lim_{L\to\infty} \eta = 0 \tag{8}$$

for $0 \leqslant p < 1/4$, and

$$\lim_{L\to\infty} \eta \neq 0 \tag{9}$$

for $3/4 \leqslant p \leqslant 1$. Thus, the analyticity of η is broken at a point p_c satisfying $1/4 \leqslant p_c < 3/4$. This result is illustrated in Fig. 5. The breaking of the analyticity of η allows us to identify the parallelizable and unparallelizable phases without arbitrariness. In other words, the region of p satisfying $\lim_{L\to\infty} \eta \neq 0$ is identified as a parallelizable phase and the region of p satisfying $\lim_{L\to\infty} \eta = 0$ as an unparallelizable phase. Here, we briefly outline the proof of Eqs. (8) and (9). A rigorous proof is given in Appendix. The essence of the proof of Eq. (8) which represents the unparallelizable phase is that when p is small, it is difficult to construct large, inactive states. The number of assembly pathways generating a state with n vertices is given by the Catalan number, which asymptotically increases as 4^n. The realization probability of each assembly pathway is p^{n-1}. From the balance of these two factors, it can be understood that when p is smaller than $1/4$, there is a high possibility that there are no ways to construct large inactive states. Therefore, when p is smaller than $1/4$, it is necessary to connect inactive states one by one using active states, and it is found that the assembly cannot be completed in $O(\log L)$ steps. The obtained upper limit of the parameter of the unparallelizable phase, $1/4$, originated from the asymptotic behavior of the Catalan number.

The essence of the proof of Eq. (9) which represents the parallelizable phase is that when p is large, there is a high possibility of the existence of relatively unbiased assembly pathways. For example, considering an assembly pathway with no bias such that both children of a state with n vertices have $n/2$ vertices, we have $d = \log_2 L$. In contrast, considering a very biased assembly pathway such that the children of a state with n vertices have $n-1$ and 1 vertices, we have $d = L - 1$. In the proof of Eq. (9), it is shown by mathematical induction on the number of vertices that when p is greater than $3/4$, at least one relatively unbiased assembly pathway is likely to exist. Therefore, it is found that when p is greater than $3/4$, there is a high possibility that there is at least one assembly pathway with $d = O(\log L)$. In contrast to the case of the parallelizable phase, the obtained lower bound, $3/4$, does not have a clear origin. It is possible to slightly improve the lower bound by increasing the number of the base cases of the induction.

IV. SETUP OF THE ANP MODEL

Instead of optimizing the assembly pathways, we study typical pathways of stochastic evolution of assembly in the second model. This model is interpreted as a mean-field version of the first model. To simplify the analysis, inactive states in the quenched combinability model are further classified into neutral states and passive states in the ANP model. That is, each component takes one of three states: active, neutral, or passive. In the quenched combinability model, once it is determined that two states s_1 and s_2 cannot combine, they will never combine during the assembly process. The ANP model incorporates this effect by assuming that passive states never combine with each other.

The stochastic assembly rule is as follows. Initially, there are one active and $(L-1)$ neutral components. The assembly of the components proceeds in a repetition of the following two steps: (i) randomly pair two components as possible [16]; (ii) for each pair, perform the following bonding reaction: an active component can bond with any other component, a neutral component can bond with inactive components with probability p', and a passive component cannot bond with passive components. If a neutral component fails to bond, the

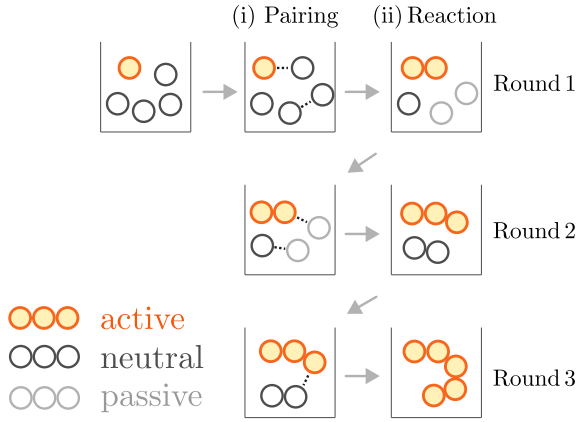

FIG. 6. Schematic of the assembly of five parts in three rounds ($L = 5, \hat{d}' = 3$). Initially, there are one active and $(L-1)$ neutral components. A single round consists of the following two steps: (i) Randomly pair two components as possible, (ii) For each pair, perform the reactions shown in Eqs. (10)-(14). The quantity $\hat{d}'$ is the number of rounds until there is only one component.

component becomes passive. We define the set of operations (i) and (ii) as a single round.

The ANP model is expressed symbolically by denoting active, neutral, and passive components as A, N, and P, respectively. Operation (ii) is then written as the following set of chemical reactions:

$$\mathrm{A} + \mathrm{N} \to \mathrm{A}, \tag{10}$$

$$\mathrm{A} + \mathrm{P} \to \mathrm{A}, \tag{11}$$

$$\mathrm{N} + \mathrm{N} \to \begin{cases} \mathrm{N} & \text{with probability } p' \\ \mathrm{P} + \mathrm{P} & \text{with probability } 1 - p', \end{cases} \tag{12}$$

$$\mathrm{N} + \mathrm{P} \to \begin{cases} \mathrm{N} & \text{with probability } p' \\ \mathrm{P} + \mathrm{P} & \text{with probability } 1 - p' \end{cases} \tag{13}$$

$$\mathrm{P} + \mathrm{P} \to \mathrm{P} + \mathrm{P}. \tag{14}$$

This procedure is illustrated in Fig. 6. The active state is necessary to ensure that the assembly process can always be executed. Even if all states become passive, the assembly process can be completed by the reaction represented by Eq. (11). We measure the number of rounds $\hat{d}'$ until all parts are connected. The assembly of five parts in three rounds is shown in Fig. 6. The bonding reactions occur probabilistically, and the number of rounds required to connect all the parts varies from trial to trial. Therefore, $\hat{d}'$ is a random variable.

We introduce the parallel efficiency η' to characterize the feasibility of the efficient parallel assembly. The parallel efficiency η' is defined as

$$\eta' := \frac{\log_2 L}{\langle \hat{d}' \rangle}, \tag{15}$$

where L is the number of parts and $\langle \hat{d}' \rangle$ is the average number of required rounds. In a way similar to the first model, η' satisfies the following two properties. First, η' satisfies the normalization condition $0 \leqslant \eta' \leqslant 1$. Second, we can determine the feasibility of efficient parallel assembly by checking whether η' is positive or zero in the limit $L \to \infty$.

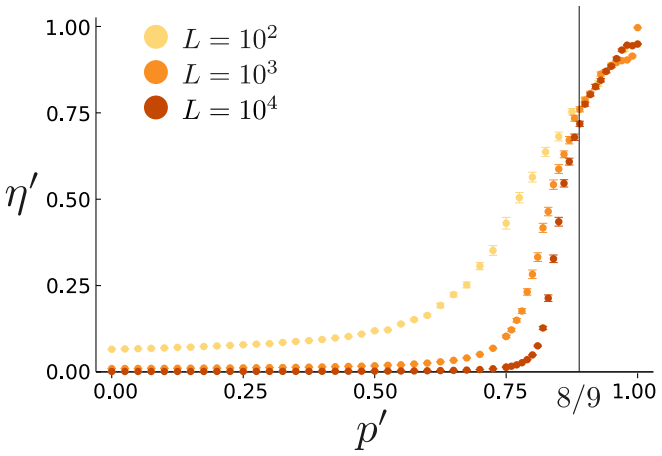

FIG. 7. Simulation results for the ANP model. For each p', we took 100 samples and calculated the mean and standard deviation of η'. The data points approach the line $p' = 8/9$ as L increases.

V. RESULTS FOR THE ANP MODEL

The simulation results are shown in Fig. 7. These graphs suggest that a discontinuous transition exists at a point p'_c. Indeed, for this model, we prove that the parallel efficiency η' has a discontinuous transition when L becomes infinite. Quantitatively, we show

$$\lim_{L\to\infty} \eta' = 0 \tag{16}$$

for $0 \leqslant p' < 8/9$, and

$$\lim_{L\to\infty} \eta' \neq 0 \tag{17}$$

for $8/9 < p' \leqslant 1$. This implies that the parallel efficiency η' is discontinuous at $p'_c = 8/9$. The proof is the following. Let a_n, b_n, and c_n be the population of A, N, and P in the nth round, respectively. The total population of components is

$$T_n = a_n + b_n + c_n. \tag{18}$$

Because the population of A does not change through the reactions, $a_n = 1$ always holds. Assuming that L is large enough, we analyze the behavior of the expected values of b_n, c_n, and T_n ignoring terms of $O(1)$. Let $\langle \hat{S}_n \rangle$ be the expected number of P − P pairs in the nth round. We obtain

$$\langle \hat{S}_n \rangle \simeq \frac{T_n}{2}\left(\frac{c_n}{T_n}\right)^2, \tag{19}$$

where the symbol $\simeq$ represents an approximation ignoring the terms of $O(1)$ [17]. For each pair described by Eq. (12) or Eq. (13), N is generated with probability p'. Therefore, we obtain

$$\begin{aligned} b_{n+1} &= \left(\frac{T_n}{2} - \langle \hat{S}_n \rangle\right) p' \\ &\simeq \frac{T_n}{2}\left(1 - \left(\frac{c_n}{T_n}\right)^2\right) p'. \end{aligned} \tag{20}$$

Because the number of components is halved in the reaction that produces A or N, we obtain

$$T_{n+1} = T_n - (a_{n+1} + b_{n+1}) \simeq T_n - b_{n+1}. \tag{21}$$

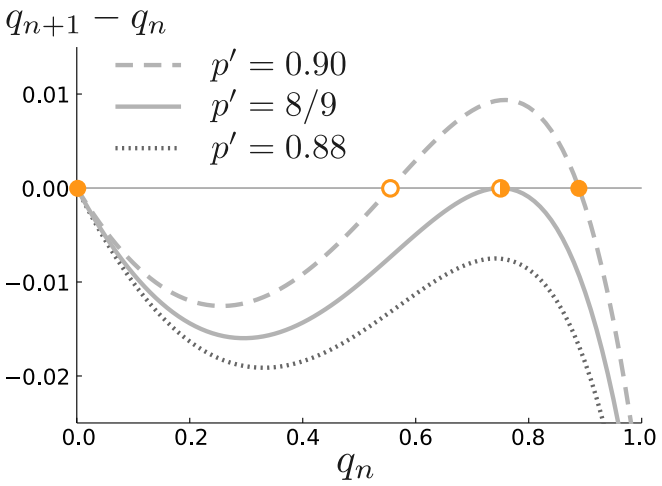

FIG. 8. Saddle node bifurcation exhibited by the discrete dynamical system of Eq. (25). When p' is less than $8/9$, only $q_n = 0$ is a stable fixed point. When p' is greater than $8/9$, a new stable fixed point and an unstable fixed point appear in $0 < q_n < 1$.

By setting

$$q_n := 1 - \left(\frac{c_n}{T_n}\right)^2, \tag{22}$$

we obtain

$$b_{n+1} \simeq \frac{T_n}{2} q_n p', \tag{23}$$

$$T_{n+1} \simeq \left(1 - \frac{q_n p'}{2}\right) T_n. \tag{24}$$

Substituting Eqs. (18), (23), and (24) into Eq. (22), we obtain

$$q_{n+1} \simeq \frac{q_n p' \left(1 - \frac{3}{4} q_n p'\right)}{\left(1 - \frac{1}{2} q_n p'\right)^2}. \tag{25}$$

The discrete dynamical system given by Eq. (25) exhibits a saddle node bifurcation as parameter p' changes. This bifurcation structure is shown in Fig. 8. When p' is less than $8/9$, only $q_n = 0$ is a stable fixed point. The assembly cannot be completed with $\hat{d}' = O(\log L)$ because $T_{n+1}/T_n \simeq 1$. As a result, η' becomes zero in this case. In contrast, when p' is greater than $8/9$, a new stable fixed point appears in $0 < q_n < 1$. The total number of components decreases exponentially because $0 < T_{n+1}/T_n < 1$. As a result, $\hat{d}' = O(\log L)$ and η' takes a positive value in this case.

VI. DISCUSSION

The two models presented in this paper characterize in-principle and realistic parallelizability, respectively. In the quenched combinability model, the pathway with the least number of parallel steps is chosen after considering all possible assembly pathways. Therefore, η characterizes whether efficient parallel assembly is feasible, in principle. The equivalent situation would be bottom-up manufacturing of industrial products, where the manufacturing process is well-designed and optimized in advance. In the ANP model, L parts are randomly paired and combined, and the number of rounds until all parts are connected is measured. Therefore, η' characterizes whether efficient parallel assembly is realistically possible. The equivalent situation would be chemical synthesis, where the molecules randomly collide.

The parallel efficiency defined in this paper is related to the complexity of molecular structures. The minimum number of parallel steps $\hat{d}$ is essentially the same as the molecular assembly index defined in the literature [18–20]. It may be possible to extend this study to classify the complexity of molecules using parallel efficiency.

The model analyzed in this paper can be considered a variation of several known models. By excluding the single active unit, the quenched combinability model can be considered a form of directed percolation in $(1 + 1)$ dimensions [21]. Therefore, findings in directed percolation may be used to estimate the transition point in this model. The ANP model can be related to the cluster merging process described by the Smoluchowski equation [22]. In this process, the number of clusters exhibits exponential decay, which corresponds to the parallelizable phase in this paper. The ANP model can also be viewed as one special case of stochastic chemical reactions or reaction-diffusion systems [23,24]. One such model is the activated random walk, which consists of active particles A and sleeping particles S [25]. Discussing parallelizability in general chemical reaction systems is an important future task.

We present possible future directions. As a practical direction, this study can apply to actual industrial production processes. Parallelizable-unparallelizable transition would emerge in connection to the success rate p of each process during the assembly of complex structures. Applying the method of this study may make it possible to calculate the threshold success rate of the elementary process to achieve efficient parallel assembly.

As a theoretical direction, this paper could lead to methods of classifying chemical reaction systems using parallel efficiency. Chemical reaction systems are classified according to the number of steady states or the number of conserved quantities [26]. Extending this study may make it possible to add another axis [parallelizable or unparallelizable] to the classification of chemical reaction systems.

The model discussed here may be realized in chemical reaction systems. In organic synthetic chemistry, chemical reactions such as living radical polymerization [27,28] and multicomponent reactions [29,30] are studied. Such reaction systems could correspond directly to the model analyzed in this paper.

VII. CONCLUSION

In this paper, we proposed a phase transition on the feasibility of efficient parallel assembly. We demonstrated the parallelizable-unparallelizable transition through two models. We can consider some extensions of the models. For example, the quenched combinability model assumes that all reaction probabilities between inactive states are p, but this could be extended to depend on the internal composition to resemble a real chemical reaction. It is an important future task to extend the model to make the theory more easily comparable to real experiments.

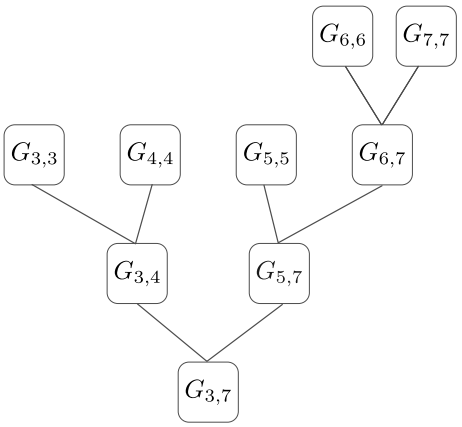

FIG. 9. Example of an assembly pathway T generating a state $s = G_{3,7}$. Each vertex of T is an element of S, that is, $G_{i,j}$. The root of T is the state s. For each vertex $G_{i,j}$ of T, the children are $G_{i,k}$ and $G_{k+1,j}$ $(i \leqslant k \leqslant j-1)$. Every leaf of T is an element of M, that is, $G_{i,i}$. In this case, $d(T) = 3$.

ACKNOWLEDGMENTS

We thank Ryohei Kakuchi, Masato Itami, Tomohiro Tanogami, and Yusuke Yanagisawa for fruitful discussions. The present study was supported by JSPS KAKENHI Grants No. JP19H05795, No. JP20K20425, and No. JP22H01144.

APPENDIX: PROOF OF THE RESULTS

In this Appendix, we prove Eqs. (8) and (9). Before we begin the proof, we define the assembly pathway more formally.

An assembly pathway T generating a state $s \in S$ is a binary tree that satisfies the following four conditions:

(1) Each vertex of T is an element of S.

(2) The root of T is the state s.

(3) For each vertex $G_{i,j}$ of T, the children are $G_{i,k}$ and $G_{k+1,j}$ $(i \leqslant k \leqslant j-1)$.

(4) Every leaf of T is an element of M.

The number of parallel steps $d(T)$ is the height of the tree T [15]. An example of an assembly pathway T generating a state $s = G_{3,7}$ is shown in Fig. 9. Note that assembly pathways are defined not only for G but also for every state $s \in S$. We call an assembly pathway T generating an inactive state an *inactive subtree*.

Let two children of a state s be C_s^1 and C_s^2. Such vertices C_s^1 and C_s^2 having the same parent are called siblings. We then introduce the following terms:

(1) An assembly pathway T is realizable if all siblings (C_s^1, C_s^2) in T are included in $\hat{R}$

(2) An assembly pathway T is unrealizable if there exist siblings (C_s^1, C_s^2) in T that are not included in $\hat{R}$.

1. Unparallelizable phase

We show the proof of Eq. (8). When $p = 0$, $\lim_{L\to\infty} \eta = 0$ is trivial because $\langle \hat{d} \rangle = L - 1$ holds. Thus, we consider the case $0 < p < 1/4$.

a. Number of assembly pathways

Let a_n be the total number of assembly pathways of a path graph with n vertices. Focusing on the last step, we obtain the following recurrence relation:

$$a_1 = 1,$$
$$a_n = \sum_{i=1}^{n-1} a_i a_{n-i} \quad (n \geqslant 2). \tag{A1}$$

This is the same as the recurrence relation that defines the Catalan number. Using the general terms of Catalan numbers [31], we obtain

$$a_n = \frac{(2(n-1))!}{n!(n-1)!}. \tag{A2}$$

This number has the following upper bound:

$$a_n \leqslant 4^{n-1} \quad (n \geqslant 1). \tag{A3}$$

Proof. We prove Eq. (A3) by mathematical induction.

Base case: In the case $n = 1$, $a_n \leqslant 4^{n-1}$ is true because $a_1 = 4^0 = 1$.

Induction step: Assuming that $a_n \leqslant 4^{n-1}$ holds $(n = 1, 2, 3, \ldots)$, we obtain

$$\begin{aligned} a_{n+1} &= \frac{(2n)!}{n!(n+1)!} \\ &= \frac{2n(2n-1)}{n(n+1)} \times a_n \\ &= \left(4 - \frac{6}{n+1}\right) \times a_n < 4 \times 4^{n-1} = 4^n. \end{aligned} \tag{A4}$$

By mathematical induction, we have proved Eq. (A3). ■

b. Number of inactive subtrees

Let $\mathcal{T}_m$ be the set of all inactive subtrees generating inactive states of size m. The number of the size m inactive states is $(L - m)$. For each of them, there are

$$a_m = \frac{(2(m-1))!}{m!(m-1)!} \tag{A5}$$

inactive subtrees (see 1 a). Thus, we obtain

$$|\mathcal{T}_m| = (L - m)a_m. \tag{A6}$$

c. Evaluation of $\hat{d}$

Let A_T be a stochastic event that an assembly pathway T is realizable. Then, the following proposition holds.

Proposition 1.

$$\bigwedge_{m=n}^{L-1} \bigwedge_{T\in\mathcal{T}_m} \overline{A_T} \quad \Rightarrow \quad \hat{d} \geqslant \frac{L-1}{n-1} \quad (2 \leqslant n \leqslant L-1) \tag{A7}$$

Proof. As shown in Fig. 10, any assembly pathway T generating G is decomposed into the sequential bonding of inactive subtrees to active states. Let K be the number of inactive subtrees in T. Let m_k be the size of the state generated by the kth inactive subtree (see Fig. 10). Because the premise of Proposition 1 means that there is no way to generate an

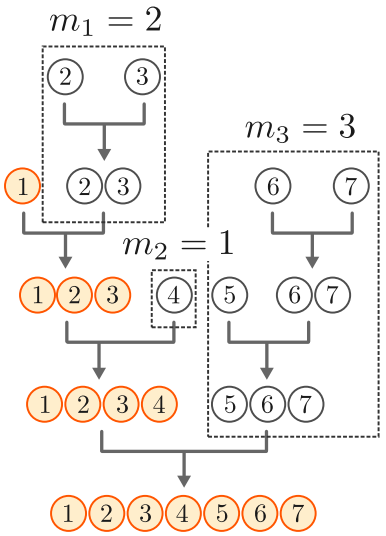

FIG. 10. Any assembly pathway T generating G is decomposed into the sequential bonding of inactive subtrees (dashed rectangles) to active states (filled circles). This diagram shows the case $K = 3$ and $d(T) = 4$.

inactive state with more than n vertices, $m_k \leqslant n-1$ holds. Therefore, we obtain

$$L - 1 = \sum_{k=1}^{K} m_k \leqslant (n-1)K. \tag{A8}$$

Because the distance from the root G to $G_{1,1}$ is K, $d(T) \geqslant K$ also holds. Therefore, we obtain

$$d(T) \geqslant \frac{L-1}{n-1}. \tag{A9}$$

Because Eq. (A9) holds for any assembly pathway T, we obtain

$$\hat{d} = \min_{T \in \hat{U}_G} \{d(T)\} \geqslant \frac{L-1}{n-1}. \tag{A10}$$

■

Assembling an inactive state with m vertices requires $(m-1)$ bondings, which are independently realized with probability p. Therefore, for any assembly pathway $T \in \mathcal{T}_m$,

$$\Pr(A_T) = p^{m-1}, \tag{A11}$$

where $\Pr(A)$ represents the probability that stochastic event A occurs.

d. Evaluation of the probability

Using the above preparation, we evaluate the probability as

$$\begin{aligned}
\Pr\left(\hat{d} \geqslant \frac{L-1}{n-1}\right) &\geqslant \Pr\left(\bigwedge_{m=n}^{L-1} \bigwedge_{T \in \mathcal{T}_m} \overline{A_T}\right) \\
&= \Pr\left(\overline{\bigvee_{m=n}^{L-1} \bigvee_{T \in \mathcal{T}_m} A_T}\right) \\
&= 1 - \Pr\left(\bigvee_{m=n}^{L-1} \bigvee_{T \in \mathcal{T}_m} A_T\right) \\
&\geqslant 1 - \sum_{m=n}^{L-1} \sum_{T \in \mathcal{T}_m} \Pr(A_T) \\
&= 1 - \sum_{m=n}^{L-1} |\mathcal{T}_m| p^{m-1}.
\end{aligned} \tag{A12}$$

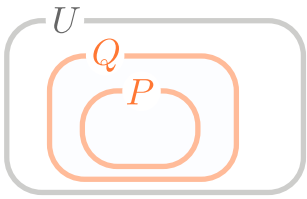

FIG. 11. Diagram of the relationship between propositions and sets. $\Pr(\mathrm{Q}) \geqslant \Pr(\mathrm{P})$ holds when $\mathrm{P} \Rightarrow \mathrm{Q}$ holds.

Note that

$$\Pr(\mathrm{Q}) \geqslant \Pr(\mathrm{P}) \tag{A13}$$

holds when

$$\mathrm{P} \Rightarrow \mathrm{Q} \tag{A14}$$

holds (see Fig. 11). We used this relation and Proposition 1 in the first line. Furthermore, we used de Morgan's rule in the second line and Boole's inequality in the fourth line. Boole's inequality, also known as the union bound, is an inequality given by

$$\Pr\left(\bigvee_{i=1}^{n} E_i\right) \leqslant \sum_{i=1}^{n} \Pr(E_i), \tag{A15}$$

where E_i $(i = 1, 2, 3, ..., n)$ represent arbitrary events which may not be independent. We further evaluate the sum as

$$\begin{aligned}
\sum_{m=n}^{L-1} |\mathcal{T}_m| p^{m-1} &= \sum_{m=n}^{L-1} (L-m) a_m p^{m-1} \\
&\leqslant \sum_{m=n}^{L-1} (L-m)(4p)^{m-1} \\
&< \sum_{m=n}^{\infty} L(4p)^{m-1} \\
&= \frac{L(4p)^{n-1}}{1-4p},
\end{aligned} \tag{A16}$$

where we used Eq. (A3) (see Appendix 1 a) in the second line and $p < 1/4$ in the fourth line.

e. Evaluation of $\langle \hat{d} \rangle$ and η

Let us define the integer

$$n_0 := 2 + \left\lceil \log_{4p}\left(\frac{1-4p}{L}\right) \right\rceil. \tag{A17}$$

Then, we obtain

$$\begin{aligned}\langle \hat{d} \rangle &\geqslant \Pr\left(\hat{d} \geqslant \frac{L-1}{n_0-1}\right) \times \frac{L-1}{n_0-1} \\ &> \left(1 - \frac{L(4p)^{n_0-1}}{1-4p}\right) \times \frac{L-1}{n_0-1} \\ &\geqslant (1-4p) \times \frac{L-1}{n_0-1}, \end{aligned} \tag{A18}$$

where we used Markov's inequality in the first line, Eqs. (A12) and (A16) in the second line, and Eq. (A17) and $\lceil x \rceil \geqslant x$ in the third line. Markov's inequality is an inequality given by

$$\Pr(|\hat{X}| \geqslant a) \leqslant \frac{\langle |\hat{X}| \rangle}{a}, \tag{A19}$$

where $\hat{X}$ is a stochastic variable and $a > 0$.

Using Eq. (A18), we evaluate η as

$$\begin{aligned}\eta &\coloneqq \frac{\log_2 L}{\langle \hat{d} \rangle} \\ &\leqslant \frac{(n_0-1)\log_2 L}{(1-4p)(L-1)} \\ &< \frac{\left[2 + \log_{4p}\left(\frac{1-4p}{L}\right)\right]\log_2 L}{(1-4p)(L-1)}. \end{aligned} \tag{A20}$$

We thus obtain

$$\lim_{L\to\infty} \eta = 0 \tag{A21}$$

for $0 \leqslant p < 1/4$.

2. Parallelizable phase

We show the proof of Eq. (9).

a. A strategy that enables logarithmic height assembly

We introduce another term:

An assembly pathway T is **α-splitting** if, for any bonding process $s_i + s_j \to s_k$ in T [17],

$$\min\{|s_i|, |s_j|\} \geqslant \lceil \alpha |s_k| \rceil, \tag{A22}$$

where $|s|$ represents the number of vertices of the state s.

Proposition 2. If T is α-splitting,

$$d(T) \leqslant \frac{\log L}{-\log(1-\alpha)} \tag{A23}$$

holds.

Proof. From Eq. (A22) and $|s_i| + |s_j| = |s_k|$, we obtain

$$\max\{|s_i|, |s_j|\} \leqslant |s_k| - \lceil \alpha |s_k| \rceil \leqslant (1-\alpha)|s_k|. \tag{A24}$$

Using this inequality repeatedly, we obtain

$$1 \leqslant (1-\alpha)^{d(T)}|G| = (1-\alpha)^{d(T)} L, \tag{A25}$$

which is equivalent to Eq. (A23). ■

b. Probability that this strategy is not available

Let B_s be the stochastic event that all α-splitting assembly pathways generating a state s are unrealizable. Let L_m be the subgraph consisting of the leftmost m vertices of state s, and $R_{|s|-m}$ be the subgraph consisting of the rightmost $|s| - m$ vertices of state s (see Fig. 12). Focusing on the final step, we obtain

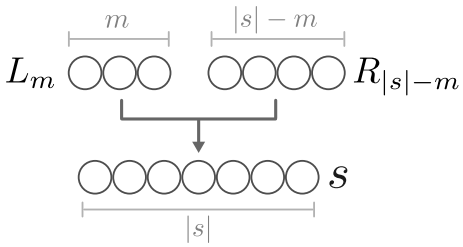

FIG. 12. Let L_m be the subgraph consisting of the leftmost m vertices of state s, and $R_{|s|-m}$ be the subgraph consisting of the rightmost $|s| - m$ vertices of state s.

$$\begin{aligned} B_s \quad \Leftrightarrow \quad & \text{For all } m \ (\lceil \alpha |s| \rceil \leqslant m \leqslant |s| - \lceil \alpha |s| \rceil), \\ & (L_m, R_{|s|-m}) \notin \hat{R} \quad \vee \\ & \times \{(L_m, R_{|s|-m}) \in \hat{R} \ \wedge (B_{L_m} \vee B_{R_{|s|-m}})\}. \end{aligned} \tag{A26}$$

Then, we obtain

$$\begin{aligned}\Pr(B_s) &= \prod_{m=\lceil \alpha |s| \rceil}^{|s|-\lceil \alpha |s| \rceil} [(1-\tilde{p}) + \tilde{p} \Pr(B_{L_m} \vee B_{R_{|s|-m}})] \\ &\leqslant \prod_{m=\lceil \alpha |s| \rceil}^{|s|-\lceil \alpha |s| \rceil} \{(1-\tilde{p}) + \tilde{p}[\Pr(B_{L_m}) + \Pr(B_{R_{|s|-m}})]\}, \end{aligned} \tag{A27}$$

where we defined $\tilde{p}$ as

$$\tilde{p} = \begin{cases} p & (L_m \text{ is inactive}) \\ 1 & (L_m \text{ is active}). \end{cases} \tag{A28}$$

In the second line, we used Boole's inequality.

Let us define

$$Q_n \coloneqq \max_{s \in S_n}\{\Pr(B_s)\}, \tag{A29}$$

where $S_n = \{s \in S \mid |s| = n\}$. Then, we obtain the following recursive inequalities for Q_n:

$$\begin{aligned} &Q_1 = 0, \\ &Q_n \leqslant \prod_{m=\lceil \alpha n \rceil}^{n-\lceil \alpha n \rceil} ((1-\tilde{p}) + \tilde{p}(Q_m + Q_{n-m})) \quad (n \geqslant 2). \end{aligned} \tag{A30}$$

Proof. Let $s^* \in S_n$ be the state for which $\Pr(B_s)$ is maximum. Applying Eq. (A27) to this s^*, we obtain

$$\begin{aligned} Q_n &= \Pr(B_{s^*}) \\ &\leqslant \prod_{m=\lceil \alpha n \rceil}^{n-\lceil \alpha n \rceil} \{(1-\tilde{p}) + \tilde{p}[\Pr(B_{L_m}) + \Pr(B_{R_{n-m}})]\} \\ &\leqslant \prod_{m=\lceil \alpha n \rceil}^{n-\lceil \alpha n \rceil} [(1-\tilde{p}) + \tilde{p}(Q_m + Q_{n-m})], \end{aligned} \tag{A31}$$

where we used Eq. (A29) in the third line. ■

c. Evaluation of Q_n

Proposition 3. If $p \geqslant 3/4$ and $\alpha = 1/6$, $Q_n \leqslant 1/4$ for all $n(= 1, 2, 3, ...)$.

Proof. We prove Proposition 3 by mathematical induction.

Base case: Because $\alpha = 1/6$, $\lceil \alpha n \rceil = 1$ for $n = 2, 3$. Then, we obtain

$$Q_2 \leqslant (1-\tilde{p}) + \tilde{p} \cdot 2Q_1 = 1 - \tilde{p}$$
$$Q_3 \leqslant ((1-\tilde{p}) + \tilde{p}(Q_1 + Q_2))^2 \leqslant (1-\tilde{p}^2)^2. \tag{A32}$$

Using $\tilde{p} \geqslant p \geqslant 3/4$, we obtain $Q_2 \leqslant 1/4$ and $Q_3 \leqslant 49/256 < 1/4$. $Q_1 = 0 < 1/4$ also holds trivially.

Induction step: Assume that $Q_1, Q_2, ..., Q_{n-1} \leqslant 1/4$ holds $(n = 4, 5, 6, ...)$. From Eq. (A30), we obtain

$$\begin{aligned} Q_n &\leqslant \prod_{m=\lceil \alpha n \rceil}^{n-\lceil \alpha n \rceil} \left((1-\tilde{p}) + 2\tilde{p} \cdot \frac{1}{4} \right) \\ &= \prod_{m=\lceil \alpha n \rceil}^{n-\lceil \alpha n \rceil} \left(1 - \frac{\tilde{p}}{2} \right) \\ &\leqslant \left(1 - \frac{p}{2} \right)^{n-2\lceil \alpha n \rceil + 1} \\ &\leqslant \left(1 - \frac{p}{2} \right)^{3} \\ &\leqslant \frac{125}{512} < \frac{1}{4}, \end{aligned} \tag{A33}$$

where we used the induction hypothesis in the first line, $\tilde{p} \geqslant p$ in the third line, $n - 2\lceil \alpha n \rceil + 1 \geqslant 3$ $(n \geqslant 4)$ in the fourth line, and $p \geqslant 3/4$ in the fifth line.

By mathematical induction, we have proved Proposition 3. ∎

Substituting $Q_n \leqslant 1/4$ into Eq. (A30) and using $\tilde{p} \geqslant p$, $\lceil x \rceil < x + 1$, we obtain the evaluation of Q_n:

$$Q_n \leqslant \left(1 - \frac{p}{2} \right)^{n-2\lceil \alpha n \rceil + 1} < \left(1 - \frac{p}{2} \right)^{(1-2\alpha)n-1}. \tag{A34}$$

d. Evaluation of $\langle \hat{d} \rangle$ and η

We evaluate $\langle \hat{d} \rangle$ by separately considering the following two cases:

(1) There exists a realizable 1/6-splitting assembly pathway of G; that is, B_G is false.

(2) There exists no such assembly pathway of G; that is, B_G is true.

In the first case, we can use $d \leqslant \log L/(-\log(1-\alpha))$ through Proposition 2, where α is set to 1/6 to simplify the appearance. Even in the second case, we can use the inequality $d < L$, which always holds.

We then obtain

$$\begin{aligned} \langle \hat{d} \rangle &< \Pr(\overline{B_G}) \frac{\log L}{-\log(1-\alpha)} + \Pr(B_G) L \\ &\leqslant \frac{\log L}{-\log(1-\alpha)} + Q_L L \\ &< \frac{\log L}{-\log(1-\alpha)} + L\left(1 - \frac{p}{2}\right)^{(1-2\alpha)L-1}, \end{aligned} \tag{A35}$$

where we used $\Pr(\overline{B_G}) \leqslant 1$ and Eq. (A29) in the second line and Eq. (A34) in the third line.

We then evaluate η using Eq. (A35) as

$$\begin{aligned} \eta &:= \frac{\log_2 L}{\langle \hat{d} \rangle} \\ &> \frac{\log_2 L}{\frac{\log L}{-\log(1-\alpha)} + L\left(1 - \frac{p}{2}\right)^{(1-2\alpha)L-1}}, \end{aligned} \tag{A36}$$

where the right side goes to $-\log_2(1-\alpha)$ in the limit $L \to \infty$. Substituting $\alpha = 1/6$ into the result, we obtain

$$\lim_{L\to\infty} \eta \neq 0 \tag{A37}$$

for $3/4 \leqslant p \leqslant 1$.

[1] C. P. Poole, Jr. and F. J. Owens, *Introduction to Nanotechnology* (John Wiley & Sons, New York, 2003).

[2] K. J. Bishop, C. E. Wilmer, S. Soh, and B. A. Grzybowski, Nanoscale forces and their uses in self-assembly, Small **5**, 1600 (2009).

[3] D. Nykypanchuk, M. M. Maye, D. Van Der Lelie, and O. Gang, DNA-guided crystallization of colloidal nanoparticles, Nature (London) **451**, 549 (2008).

[4] S. Koh, Strategies for controlled placement of nanoscale building blocks, Nanoscale Res. Lett. **2**, 519 (2007).

[5] B. A. Grzybowski, K. Fitzner, J. Paczesny, and S. Granick, From dynamic self-assembly to networked chemical systems, Chem. Soc. Rev. **46**, 5647 (2017).

[6] N. Badi and J.-F. Lutz, Sequence control in polymer synthesis, Chem. Soc. Rev. **38**, 3383 (2009).

[7] J. K. Szymanski, Y. M. Abul-Haija, and L. Cronin, Exploring strategies to bias sequence in natural and synthetic oligomers and polymers, Acc. Chem. Res. **51**, 649 (2018).

[8] Q. Chen, S. C. Bae, and S. Granick, Directed self-assembly of a colloidal kagome lattice, Nature (London) **469**, 381 (2011).

[9] O. D. Velev and K. H. Bhatt, On-chip micromanipulation and assembly of colloidal particles by electric fields, Soft Matter **2**, 738 (2006).

[10] J. J. Juárez and M. A. Bevan, Feedback controlled colloidal self-assembly, Adv. Funct. Mater. **22**, 3833 (2012).

[11] X. Tang, B. Rupp, Y. Yang, T. D. Edwards, M. A. Grover, and M. A. Bevan, Optimal feedback controlled assembly of perfect crystals, ACS Nano **10**, 6791 (2016).

[12] F. Li, D. P. Josephson, and A. Stein, Colloidal assembly: The road from particles to colloidal molecules and crystals, Angew. Chem. Int. Ed. **50**, 360 (2011).

[13] R. Greenlaw, H. J. Hoover, and W. L. Ruzzo, *Limits to Parallel Computation: P-Completeness Theory* (Oxford University Press on Demand, New York, 1995).

[14] S. Arora and B. Barak, *Computational Complexity: A Modern Approach* (Cambridge University Press, Cambridge, 2009).

[15] This is called the NC versus P problem in computational complexity theory.

[16] When the total number of components is odd, leave the extra one and do nothing until the next round.

[17] Consider a random variable $\hat{\sigma}_i$ such that $\hat{\sigma}_i = 1$ when the ith pair is P − P and $\hat{\sigma}_i = 0$ otherwise. Then $\langle \hat{S}_n \rangle \simeq \langle \sum_{i=1}^{T_n/2} \hat{\sigma}_i \rangle = T_n/2 \cdot \langle \hat{\sigma}_i \rangle$, and we obtain Eq. (19).

[18] S. M. Marshall, C. Mathis, E. Carrick, G. Keenan, G. J. Cooper, H. Graham, M. Craven, P. S. Gromski, D. G. Moore, S. Walker *et al.*, Identifying molecules as biosignatures with assembly theory and mass spectrometry, Nat. Commun. **12**, 3033 (2021).

[19] S. M. Marshall, A. R. Murray, and L. Cronin, A probabilistic framework for identifying biosignatures using Pathway Complexity, Philos. Trans. R. Soc. A **375**, 20160342 (2017).

[20] S. M. Marshall, D. Moore, A. R. Murray, S. I. Walker, and L. Cronin, Quantifying the pathways to life using assembly spaces, arXiv:1907.04649.

[21] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, Adv. Phys. **49**, 815 (2000).

[22] R. M. Ziff, E. M. Hendriks, and M. H. Ernst, Critical Properties for Gelation: A Kinetic Approach, Phys. Rev. Lett. **49**, 593 (1982).

[23] N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, New York, 1992), Vol. 1.

[24] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 1985), Vol. 3.

[25] L. Levine and V. Silvestri, How far do activated random walkers spread from a single source? J. Stat. Phys. **185**, 18 (2021).

[26] M. Feinberg, *Foundations of Chemical Reaction Network Theory* (Springer, New York, 2019).

[27] G. Moad, E. Rizzardo, and S. H. Thang, Toward living radical polymerization, Acc. Chem. Res. **41**, 1133 (2008).

[28] W. A. Braunecker and K. Matyjaszewski, Controlled/living radical polymerization: Features, developments, and perspectives, Prog. Polym. Sci. **32**, 93 (2007).

[29] R. Kakuchi, Multicomponent reactions in polymer synthesis, Angew. Chem. Int. Ed. **53**, 46 (2014).

[30] R. Kakuchi, The dawn of polymer chemistry based on multicomponent reactions, Polym J **51**, 945 (2019).

[31] T. Koshy, *Catalan Numbers with Applications* (Oxford University Press, Oxford, 2008).