Design Principles for Fast and Efficient Self-Assembly Processes

Florian M. Gartner[1] and Erwin Frey[1,2,*]

[1]*Arnold Sommerfeld Center for Theoretical Physics (ASC) and Center for NanoScience (CeNS),
Department of Physics, Ludwig-Maximilians-Universität München,
Theresienstraße 37, D-80333 Munich, Germany*
[2]*Max Planck School Matter to Life, Hofgartenstraße 8, D-80539 Munich, Germany*

Self-assembly is a fundamental concept in biology and of significant interest to nanotechnology. Significant progress has been made in characterizing and controlling the properties of the resulting structures, both experimentally and theoretically. However, much less is known about kinetic constraints and determinants of dynamical properties like time efficiency, although these constraints can become severe limiting factors of self-assembly processes. Here, we investigate how the time efficiency and other dynamical properties of reversible self-assembly depend on the morphology (shape) of the building blocks for systems in which the binding energy between the constituents is large. As paradigmatic examples, we stochastically simulate the self-assembly of constituents with triangular, square, and hexagonal morphology into two-dimensional structures of a specified size. We find that the constituents' morphology critically determines the assembly time and how it scales with the size of the target structure. Our analysis reveals three key structural parameters defined by the morphology: the nucleation size and attachment order, which describe the effective order of the chemical reactions by which clusters nucleate and grow, respectively, and the growth exponent, which determines how the growth rate of an emerging structure scales with its size. Using this characterization, we formulate an effective theory of the self-assembly kinetics, which we show exhibits an inherent scale invariance. This allows us to identify general scaling laws that describe the minimal assembly time as a function of the size of the target structure. We show how these insights on the kinetics of self-assembly processes can be used to design assembly schemes that could significantly increase the time efficiency and robustness of artificial self-assembly processes.

I. INTRODUCTION

Self-assembly is an essential concept in biology that explains the formation of large functional structures from smaller subunits and thereby "bridges the gap" between different length scales [1,2]. Artificial self-assembly has recently attracted considerable interest, promising numerous applications in nanotechnology, medicine, and biology [2–6]. Typically, the formation of artificial structures [4], as well as many biological structures such as virus capsids [7,8] and micelles [9], is promoted by weak and reversible interactions between subunits [10–18]. Consequently, single bonds between subunits are usually unstable, and a large number of mutual interactions is required before an emerging larger structure becomes stabilized as a whole. In this way, kinetic traps are eluded that would arise if too many stable intermediates are formed that cannot be completed with the available amount of resources [17,19–23]. While there are various strategies to evade kinetic traps even in irreversible self-assembly processes [24–26], here we focus exclusively on reversible reaction kinetics, which plays a pivotal role in nanotechnology and biology [13,15,27].

Reversible self-assembly (in thermal equilibrium systems) can be viewed as a process in which the system minimizes its free energy and thereby reaches a state in which a large fraction of the monomers are tied to complete structures [16,22,28,29]. This notion motivates using ensemble theory from statistical physics to characterize the final steady state [18,20,30–32]. Indeed, equilibrium methods have been applied successfully to a broad class of self-assembling systems [28,33–36] and—when applicable—represent powerful and convenient tools to characterize the resulting final structure and the yield.

However, equilibrium methods have a clear limitation in that they do not provide insights into the dynamic aspects
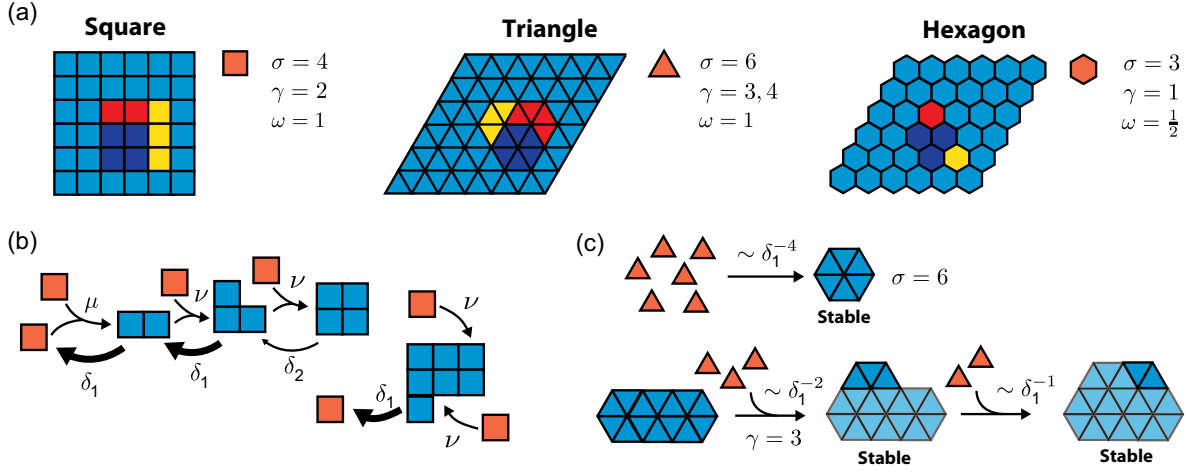
*frey@lmu.de

FIG. 1. Illustration of the model. (a) We consider the self-assembly of two-dimensional structures of volume (total number of constituents) S from square-, triangle-, or hexagon-shaped building blocks and ask for the time the assembly processes take to reach substantial yields. The configurations shaded in dark blue, red, and yellow illustrate the first intermediate assembly states with enhanced stability (wherein each particle has at least two bonds). Typical assembly pathways consecutively pass through these (or equivalent) intermediate states. The dynamics can, thus, be effectively described by the three parameters σ, γ, and ω, which are listed for the three cases and denote the *nucleation size*, *attachment order*, and *growth exponent*, respectively [see (c) and main text Secs. III A and IV]. (b) Self-assembly dynamics is modeled as follows (illustrated for the system with square-shaped monomers): Two monomers dimerize at rate μ along any of their edges. Subsequently, monomers attach at rate ν to any free neighboring site until a square consisting of S monomers with edge length $L = \sqrt{S}$ ($L = \sqrt{S/2}$ in the case of triangular monomers) is completed as illustrated in (a), assuming periodic boundary conditions between top and bottom and left and right. At the same time, particles can also detach from a given structure at rate $\delta_n \sim \exp[-nE_B/k_BT]$, with n denoting the number of bonds between the particle and the structure and E_B the binding energy per bond. (c) Coarse-grained effective dynamics (illustrated for the system with triangle-shaped monomers): Assembly starts with forming a stable nucleus [dark blue shaded states in (a)], which requires a minimum number of σ particles (nucleation size). The effective nucleation rate scales as $1/\delta_1^{\sigma-2}$, since $\sigma - 2$ unstable states must be traversed to form the nucleus. Analogously, attachment processes are initiated by an effective reaction of the order of γ (attachment order), after which a series of reactions of smaller (subleading) order may follow. As a particularity of triangular monomers, the first attachment reaction right after nucleation has attachment order $\gamma = 4$, while all subsequent attachment reactions have a smaller attachment order $\gamma = 3$.

of self-assembly processes. In particular, they do not inform about the time the system requires to achieve a specified yield. However, understanding the kinetics of self-assembly processes is important, as experiments clearly show that the time required to complete self-assembly processes can become a severe obstacle, especially when the target structures are large [6]. Hence, it is indispensable for the advancement of the field to gain a deeper understanding of those factors that control kinetic properties, like the time efficiency of self-assembly.

An essential feature of self-assembling systems is the morphology (shape) of the building blocks. For example, several icosahedral virus capsids are assembled from triangular or trapezoidal capsid proteins [37]. Carboxysomes—another example of biological structures with icosahedral symmetry—emerge from pentagon- and hexagon-shaped building blocks [38]. In DNA nanotechnology, simple regular shapes are typically used like squares in algorithmic self-assembly [35] or triangles to build artificial capsids [6]. The shape of the constituents might play an important role, as it determines how the subunits connect with each other when forming higher-order assemblies. Does this "connectivity" and, thus, the

shape of the subunits crucially influence the speed of self-assembly? Do the typical morphologies of the above examples support efficient self-assembly, or could constituents with another morphology assemble even faster?

We use a generic conceptual model for reversible self-assembly to address these basic questions. As paradigmatic examples, we perform stochastic simulations of systems composed of monomers with triangular, square, and hexagonal shapes that self-assemble into two-dimensional structures of a specified size [Fig. 1(a)]. As an essential and informative characteristic, we thereby analyze the "time complexity," describing how the minimally required assembly time scales with the size of the target structure. A similar concept was applied previously by us to quantify the time efficiency of different (reversible and irreversible) self-assembly scenarios disregarding effects arising due to the morphology of the constituents [26]. Here, we show that the time efficiency and its scaling as a function of structure size (*time complexity*) for reversible self-assembly processes depend strongly on the morphology of the constituents. Using an effective kinetic model, we determine analytic estimates for these scaling laws describing how the assembly time scales with the size of the target

structure and the detachment rate. These findings suggest that the morphology of the building blocks is a major determinant of the time efficiency of self-assembly processes and implies various possibilities to enhance the time efficiency of artificial self-assembly: For example, by enforcing a hierarchical assembly step, monomers can be induced to first form higher-order assemblies with a more suitable morphology, which subsequently assembles significantly faster than the original monomers. We simulate such assembly schemes explicitly and show that, in this way, the assembly time can be decreased by up to several orders of magnitude. This is an important result, because it shows that a putatively unimportant detail, like the precise shape of the building blocks, can critically influence the time efficiency and even its scaling behavior in nonequilibrium processes.

This paper is structured as follows: In Sec. II, we introduce the mathematical model and the central concepts. In Sec. III, we discuss the phenomenology of the self-assembly dynamics for the case in which the binding energy between the subunits is large, and, thus, only leading-order detachment reactions need to be considered (strong binding limit of self-assembly). As a major result, we show that the required assembly time exhibits an interesting scaling behavior as a function of the target structure size, depending on the morphology of the building blocks. In Sec. IV, we formulate an effective mathematical model for self-assembly in the strong binding limit, and in Sec. V, we show that this theory reveals an underlying scale invariance, thereby explaining the observed scaling behavior. Finally, in Sec. VI, we summarize the results and discuss their implications and possible applications in biology and nanotechnology.

II. MODEL DESCRIPTION

We consider the self-assembly of N identical particles (building blocks, monomers) of triangular, square, and hexagonal shapes into two-dimensional square-shaped structures of size (number of constituents) S assuming chemical reaction kinetics in a well-mixed fluid environment [Fig. 1(a)]. We assume that during the self-assembly process monomers from a well-mixed, dilute solution bind to sparse clusters. These conditions typically apply to a broad class of biological and experimental self-assembly systems like, e.g., virus capsid assembly for various virus species [39], self-assembly of artificial DNA nanobricks [40], and self-assembly of various cellular organelles [41]. Our model does not cover systems in which spatial degrees of freedom or scaffolding play a significant role (e.g., as studied in Refs. [17,42,43]) or where monomers organize into spatial aggregates before forming the final products [44–46]. Furthermore, we assume that all binding reactions are specific and result in correctly assembled intermediates. This excludes systems with a significant rate for incorrect bonding [47,48] or where particles have multiple different

binding partners (e.g., see Ref. [49]); albeit we expect that the model can be generalized in future work to include these additional effects. To simplify the analysis, we assume that the final structures have periodic boundary conditions; i.e., they form closed tori. The boundary conditions, representing the geometry of the target structures, induce a mechanism for self-limiting growth of the clusters [32]. Below, we show that our results apply identically to other kinds of boundary conditions and even to systems with unlimited cluster growth. In the case of periodic boundary conditions, it can be shown mathematically (cf. Appendix J 1) that, in the limit of large particle numbers, it is irrelevant for the assembly dynamics whether the particles are heterogeneous and bind only with specific neighbors ("information-rich" systems [23,40,50–52]) or whether all particles are identical [27] and bind indistinctly with each other [26,53]. For convenience, throughout the main text, we therefore assume homogeneous systems in which the particles are indistinguishable. In Appendix J, we show that our results apply equally well to heterogeneous (information-rich) systems if the reaction rate constants and concentrations of all species are the same, and we discuss several special cases (e.g., where the concentrations or reaction rates of the species differ strongly). In both the homogeneous and the heterogeneous case, we do not consider the possibility of incorrect binding of the monomers, e.g., binding in a wrong orientation or with the wrong binding partners. The initial concentration of monomers is denoted by $C = N/V$, with V being the reaction volume.

The self-assembly dynamics of the model are defined by the following reaction kinetics [Fig. 1(b)]: Any two building blocks (monomers) can bind with rate μ along any edge to form a dimer, which can subsequently grow at rate ν by further attachment of monomers at any free neighboring site. We formally distinguish the dimerization rate from the attachment rate, because cooperative binding effects might disfavor the dimerization of monomers over the attachment of monomers to larger structures in biological or experimental systems. However, we generally set $\mu = \nu$ in the simulations. Following the assumptions of classical or ideal aggregation theory [32], we assume that structures grow only by the attachment of single monomers, while interactions among larger oligomers are neglected [54,55]. This assumption is generally justified, since the concentration of monomers is usually much larger than those of oligomers. However, we assess below to which extent ideal aggregation theory is a reasonable assumption specifically for the different particle morphologies considered here.

Furthermore, to account for the reversibility of the binding reactions, it is assumed that single monomers detach from existing structures at a rate δ_n that decreases exponentially with the number n of bonds that need to be broken (Arrhenius' law for diffusion-limited processes [56,57]):

$$\delta_n = A \exp[-nE_B/k_BT]. \tag{1}$$

Here, E_B is the binding energy per bond formed between the building blocks, and A is a constant that sets the unbinding rate's overall scale. The building blocks of triangular, square, and hexagonal shapes differ in the maximum number z_{bond} of bonds they can form with other building blocks: $z_{\text{bond}} = 3$, 4, and 6, respectively. In particular, if complete structures are to be stable, the binding energy must be large enough so that the detachment rate $\delta_{z_{\text{bond}}}$ is sufficiently small.

The initial number N of monomer determines the maximum number N/S of complete structures that can be formed. The yield at time t is defined as the actual number of complete structures divided by their maximum number. In other words, it represents the fraction of resources bound into complete structures. Furthermore, to quantify the time efficiency of the self-assembly processes, we define the assembly time T_{90}, which measures the time at which the yield first surpasses a value of 90%. In particular, we investigate how the minimal assembly time T_{90}^{min} that can be achieved by tuning the experimental control parameters depends on the structure size S, and we quantify its asymptotic scaling behavior.

The dynamics of the assembly process can be controlled by changing the ratio between the frequencies of detachment and attachment events, i.e., by changing the detachment rates relative to the overall scale of the reaction kinetics: $\tilde{\delta}_n := \delta_n/(C\nu)$. Assuming detailed balance, the control parameters can also be written as $\tilde{\delta}_n \sim \exp[\Delta G_n/k_B T]$, where ΔG_n is the free energy difference between an n-fold bound subunit and a free subunit. Experimentally, this implies various possibilities to control the assembly process. For example, it can be controlled thermodynamically by changing parameters that affect the free energy difference ΔG_n, like temperature, binding energy, pH value, or salt concentration, in addition to changing the monomer concentration C. Additional possibilities to control the assembly process kinetically arise, assuming that the monomers have different internal states, among which they switch under energy consumption, whereby detailed balance may be broken [26,58,59]. In this kinetic approach [60], therefore, we describe the system's behavior in its most general form in terms of the dimensionless control parameters $\tilde{\delta}_n = \delta_n/(C\nu)$. Similarly, the resulting assembly time is measured relative to the reactive timescale in units $(C\nu)^{-1}$. We simulate the stochastic dynamics of the system using Gillespie's algorithm [61]; see Appendix A for details on how the simulations were implemented.

III. SCALING LAWS

In systems with sufficiently strong bonds between the building blocks, $E_B \gg k_B T$, the detachment rate δ_1 is significantly larger than the higher-order detachment rates δ_n with $n \geq 2$. The resulting separation of timescales implies that cluster configurations where each particle forms at least two bonds with neighboring particles
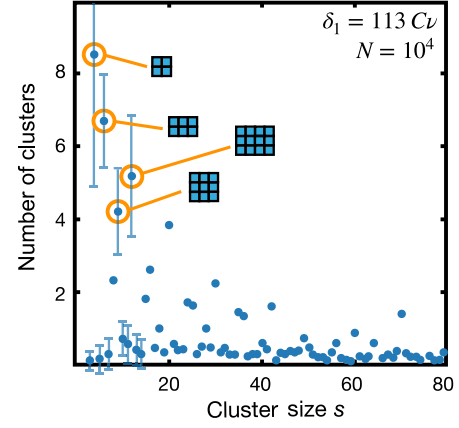

FIG. 2. Typical cluster size distribution of a system with square-shaped monomers at a fixed, early time point ($T = 0.1 \cdot T_{90}$) averaged over 100 independent simulation runs with target structure size $S = 81$. Parameters: $E_B = 10 k_B T$ and $A = 2.5 \times 10^6 C\nu$. Error bars show the sample standard deviation for a few selected cluster sizes. Structure sizes that allow for stable configurations are significantly more frequent on average than those which do not allow for a stable configuration. A few of these stable configurations are illustrated explicitly. In an effective theory, one can exploit this feature and use a reduced state space that considers only the most likely connection pathways between these stable configurations; see Sec. IV.

are significantly more stable than configurations in which particles are connected with only one bond. Consequently, the self-assembly process typically proceeds through a path that leads through these intermediate states of enhanced stability. Hallmarks of these dynamics can be inferred from the typical distribution of cluster sizes in the assembly process. For example, Fig. 2 shows the cluster size distribution for square-shaped monomers at a fixed time point, averaged over 100 independent runs of the stochastic simulation. We observe that cluster sizes that allow for configurations of enhanced stability are significantly more frequent on average than those that do not (similar results based on molecular dynamics simulations are reported in Ref. [15]; Refs. [51,52] furthermore demonstrate that the prevalence of particular assembly pathways can be shown with calculations of the free energy landscape). This suggests that the detachment rate δ_1 plays a key role in the assembly dynamics, as it determines the prevalence of intermediate configurations and controls the assembly pathway.

Throughout the rest of this paper, we consider the limit where the higher-order detachment rates δ_n with $n \geq 2$ are so small that they effectively do not affect the assembly dynamics on the relevant timescales and can, thus, be neglected. This limit can also be understood as the *leading order* in an expansion in which one truncates the sequence of detachment rates $\delta_1 \ll \delta_2 \ll \delta_3 \ldots$ at successively higher order. The idea is that—also for intermediate or small binding energies—the leading-order effect already

describes essential features of the kinetics of self-assembly processes. In a follow-up paper, we will investigate the diametric case of low binding energies and analyze the resulting effect of higher-order detachment rates on the self-assembly dynamics. Note that, if $\delta_n = 0$ for $n \geq 2$, configurations in which each particle has two bonds cannot disassemble anymore; hence, the assembly process contains irreversible steps. We refer to this case, which is the subject of this paper, also as the strong binding limit of self-assembly dynamics.

For a given binding energy E_B, a characteristic timescale $\tau_{\mathrm{irr}} \approx \delta_2^{-1} = \delta_1^{-1} e^{\beta E_B}$ with $\beta = 1/k_B T$ can be defined, below which higher-order detachment processes are negligible, and the dynamics will be accurately described by the strong binding limit. Conversely, for any given timescale τ, one can specify a minimal binding energy $\beta E_{\mathrm{irr}} \approx \log(\tau \delta_1)$, so that on timescales smaller than τ the system behaves effectively as described by the strong binding limit.

A. Predominant assembly pathways

In the strong binding limit, one can characterize the self-assembly pathways as follows: The assembly process begins with the formation of a nucleus, which requires a minimum number of σ building blocks to achieve a stable configuration where each building block has two binding partners. The *nucleation size* σ depends on the morphology of the building blocks [Figs. 1(a) and 1(c)]. Subsequent attachment processes proceed in two steps, as illustrated in Fig. 1(c): First, a leading-order reaction involving γ building blocks must occur, where we refer to γ as the *attachment order*, to reach a new stable configuration. Again, the morphology of the building blocks determines the value of the attachment order [62]. Second, for larger clusters, this leading-order attachment event may be followed by a series of lower-order reactions where the cluster traverses further stable configurations until another leading-order γ reaction is required. The probability of all the above processes is proportional to $\tilde{\delta}_1^{-u}$, where u denotes the number of unstable states passed through before the next stable configuration is reached. This is the basis for the effective kinetic theory that we derive in Sec. IV. Note that, during the nucleation process, the system passes through $u_{\mathrm{nuc}} = \sigma - 2$ unstable states, while in a leading attachment reaction, $u_{\mathrm{att}} = \gamma - 1$ unstable states are passed through. Therefore, if $\sigma > \gamma + 1$, the effective nucleation rate can be slowed down relative to the growth rate of existing structures by tuning the control parameter $\tilde{\delta}_1$. Since a reduced nucleation rate allows for more structures to get completed, the yield increases as a function of $\tilde{\delta}_1$, as we discuss in the next section.

We remark that the concept of morphology is more general than merely the shape of the monomers. For example, we discuss in Sec. VI how square-shaped monomers with six instead of four binding sites can have the same morphological parameters σ and γ as hexagonal
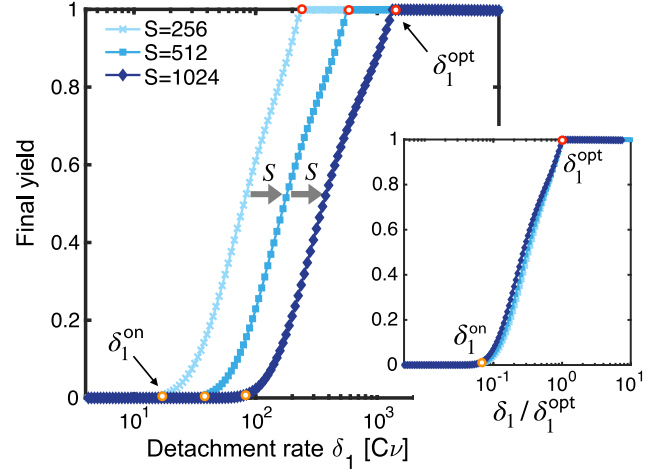

FIG. 3. Final yield in the strong binding limit as a function of the detachment rate $\tilde{\delta}_1$ (δ_1 in units of $C\nu$) for different target structure sizes S as indicated in the graph. Data points represent averages over ten independent runs of the stochastic simulation performed with square-shaped monomers and particle numbers $N = 200S$. By increasing $\tilde{\delta}_1$, the final yield increases from 0 to a perfect value of 1. Larger target structure sizes require a larger value of $\tilde{\delta}_1$ to achieve a specified yield. Inset: Rescaling of the detachment rate $\tilde{\delta}_1$ by the optimal detachment rate $\tilde{\delta}_1^{\mathrm{opt}}(S)$ leads to a data collapse of the yield curves.

monomers [cf. Fig. 8(a)]. Furthermore, icosahedral virus capsids often self-assemble from triangle-shaped monomers, but the assembly process starts with the formation of a pentamer ($\sigma = 5$) instead of a hexamer ($\sigma = 6$) as in our case [1,37]. Hence, the parameters σ and γ do not uniquely relate with either the shape of the monomers or their number of binding sites. Nevertheless, in the context of our specific system we often use the terms "shape" and "morphology" interchangeably.

B. Scaling of the yield curves

For reaction kinetics where the detachment rates are $\delta_n = 0$ for $n \geq 2$, stable intermediate configurations no longer decay, and the assembly process, thus, contains irreversible steps (strong binding limit). Consequently, the system eventually reaches an absorbing state where all monomers are bound into complete structures or stable intermediate states. The self-assembly process can, therefore, be uniquely characterized by its final yield, which is found to be a monotonic function of $\tilde{\delta}_1$. Our stochastic simulations reveal the following characteristic features (Fig. 3): If the ratio $\tilde{\delta}_1$ is small, structures nucleate very quickly, and there are kinetic traps, resulting in a poor final yield. Increasing $\tilde{\delta}_1$ slows down nucleation relative to cluster growth (provided $\sigma > \gamma + 1$), thereby increasing the number of structures that get completed but also the time required to reach the final state.
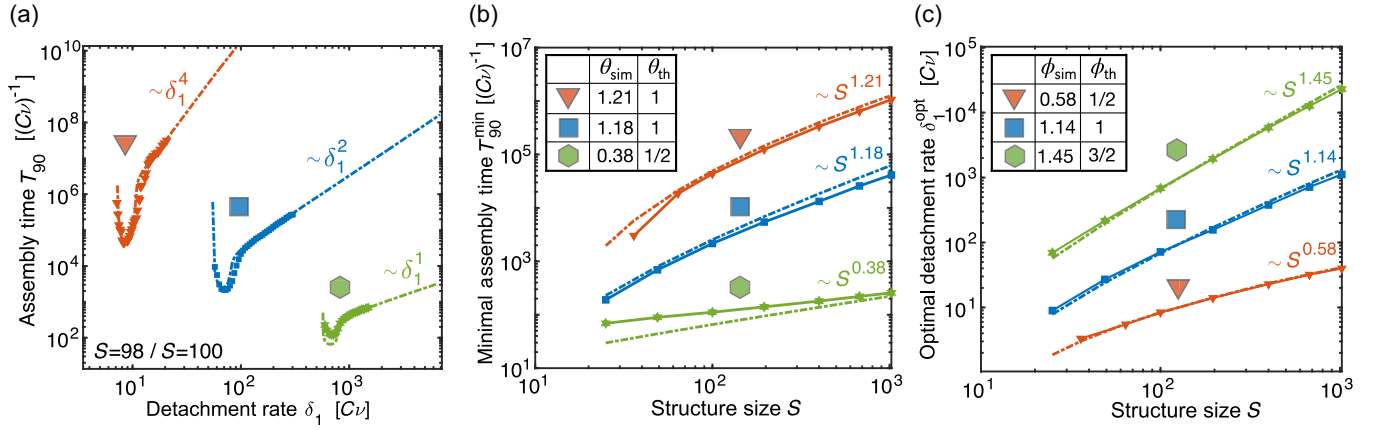
FIG. 4.   Assembly time and optimal detachment parameter in the strong binding limit. (a) The time T_{90} required to achieve 90% yield in the stochastic simulations (markers) is plotted against the detachment rate $\tilde{\delta}_1$ (δ_1 in units of $C\nu$) for systems with triangle- (red), square- (blue), and hexagon-shaped (green) building blocks for a fixed target structure size $S = 98$ (triangle-shaped building blocks) or $S = 100$ (square- and hexagon-shaped building blocks). [Note that, in the case of triangle-shaped building blocks, there is no regular structure consisting of $S = 100$ particles since regular structures consist of $2L^2$ (with an integer L) triangles; compare Fig. 1(a). Hence, we choose a structure size $S = 98$ ($L = 7$), which is as close as possible to the structure size $S = 100$ for square and hexagonal monomers.] Simulations are performed with particle numbers $N = 1000S$ (square- and hexagon-shaped building blocks) and $N = 200S$ (triangle-shaped building blocks). Dash-dotted lines represent the prediction of the effective theory obtained by numerically integrating Eq. (4). For sufficiently large detachment rates, the assembly time scales proportionally to $\tilde{\delta}_1^{\sigma-2}$, while the assembly time diverges in the limit of small detachment rates. The value of $\tilde{\delta}_1$ at which the assembly time attains its minimum, $T_{90}^{\min}$, defines the optimal detachment rate $\tilde{\delta}_1^{\mathrm{opt}}$. The minimal assembly time $T_{90}^{\min}$ (b) and the optimal detachment rate $\tilde{\delta}_1^{\mathrm{opt}}$ (c) inferred from the stochastic simulations (markers and solid lines) and effective theory (dash-dotted lines) are plotted against the size S of the target structure. They both exhibit approximate power-law dependencies for all three morphologies of the building blocks. Tables show the scaling exponents θ_{sim} and ϕ_{sim} inferred from the last three data points of the stochastic simulation in comparison with their theoretical asymptotic values θ_{th} and ϕ_{th} derived by the mathematical analysis; see Eqs. (14) and (17).

The optimal values of the detachment rate, $\tilde{\delta}_1^{\mathrm{opt}}(S)$, with which one can achieve a 90% yield in the smallest amount of time, are shown as open red circles on the yield curves in Fig. 3 for different target sizes S. The optimal values happen to approximately coincide with the values of $\tilde{\delta}_1$ for which 100% yield is reached for the first time. Furthermore, the optimal values are an increasing function of the target structure size S. Remarkably, rescaling the detachment rate $\tilde{\delta}_1$ by the optimal detachment rate $\tilde{\delta}_1^{\mathrm{opt}}$ leads to a data collapse of the yield curves (inset in Fig. 3). In particular, this implies that $\tilde{\delta}_1^{\mathrm{opt}}(S)$ scales identically with the structure size S as the onset rate for finite yield $\tilde{\delta}_1^{\mathrm{on}}(S)$, i.e., the value of the control parameter at which the final yield is larger than zero for the first time (shown as orange circles in Fig. 3). Using this scaling equivalence allows us to simplify the analysis in Sec. V A, since it permits us to work in a regime in which the yield is negligible. Further below, in Sec. V B, we prove the validity of the phenomenological scaling law for the yield curves based on a scale invariance of the underlying reaction kinetics.

C. Scaling laws for the optimal detachment rate and the minimal assembly time

The assembly time T_{90} as a function of the detachment rate $\tilde{\delta}_1$ in the strong binding limit is displayed in Fig. 4(a)

for the different morphologies of the building blocks. The curves exhibit a characteristic shape: For small $\tilde{\delta}_1$, the assembly time shows a U shape with a pronounced minimum, followed by a power-law scaling $T_{90} \sim \tilde{\delta}_1^{\sigma-2}$ as $\tilde{\delta}_1$ increases. This power law results from the fact that, for large $\tilde{\delta}_1$, the formation of the nucleus is the rate-limiting step in the assembly process and that $\sigma - 2$ unstable states must be traversed to form a nucleus (cf. Fig. 10). In contrast, the regime of the U shape is characterized by simultaneous structure assembly: Heuristically, if the detachment rate is close to its optimal value, the timescale of nucleation is slightly smaller than the timescale of growth of an entire structure. Therefore, the required number of nuclei can form initially within a short time interval while the concentration of monomers is still large. Since the rate of nucleation depends strongly on the monomer concentration, the assembly time is very sensitive to variations in $\tilde{\delta}_1$ in this regime, resulting in the typical U shape of the T_{90} curves with a pronounced minimum.

Remarkably, the minimum assembly times for the three different morphologies differ by almost 3 orders of magnitude. Even more strikingly, the minimum assembly times $T_{90}^{\min}$ at the optimal detachment rates $\tilde{\delta}_1^{\mathrm{opt}}(S)$ behave differently as a function of the structure size S [Fig. 4(b)]: For triangle- and square-shaped building blocks, the minimum

assembly time increases approximately as a power law in the structure size with exponent $\theta \approx 1.2$ (for $S \gg 1$), while for hexagon-shaped building blocks, it increases with a much smaller exponent $\theta \approx 0.4$. Similarly, the optimal detachment rate shows approximate power-law dependencies on the target structure size S with characteristic exponents ϕ for each particle morphology [Fig. 4(c)]. We refer to the (asymptotic) exponents θ and ϕ as *time complexity* and *control parameter exponent*, respectively [26]. Differences in the time complexity exponents imply that the disparities in the assembly times of the various monomer morphologies become ever more pronounced with the increasing size of the target structure. We, thus, conclude that, for the self-assembly of large objects, the morphology of the building blocks plays a crucial role, as it decisively determines the time efficiency of their self-assembly process.

Please note that for the minima depicted in Fig. 4(a) to be not affected by higher-order detachment rates, a minimal binding energy $E_B^{\mathrm{irr}}/(k_B T) \approx \log\left(T_{90}^{\min}\delta_1^{\mathrm{opt}}\right)$ between 10 and 12 is required as a sufficient condition. These values appear plausible, even though typical values for the binding energies between subunits, for example, in virus capsid assembly, are reported to be still a bit smaller [17,20,48]. We, thus, expect that the minimal assembly time in typical biological and experimental systems is reasonably well described by the approximation that neglects higher-order detachment rates.

The assembly time can alternatively be characterized as a function of the total average binding energy of a completed structure, $E_{\mathrm{tot}} = \frac{1}{2}Sz_{\mathrm{bond}}E_B$, taking into account the total number z_{bond} of bonds per particle. Using $E_B/(k_B T) = \log(A/\delta_1)$ and $\tilde{\delta}_1 \approx S^{\phi_{\mathrm{th}}}$, we find that the *optimal* total binding energies of structures formed from triangle-, square-, and hexagon-shaped monomers are generally different but become identical if $\tilde{A} \approx S^{2.5}$ ($E_{\mathrm{tot}}^{\mathrm{opt}} \approx 3S\log S$). Upon crossing this manifold, the total binding energies for the different monomer morphologies change their relative ordering. Thus, it is important to note that system-specific stability requirements can affect the choice of the optimal particle morphology.

Adjoined time complexity exponents. The time complexity exponent θ above describes the scaling of the assembly time in units of the reactive timescale $(C\nu)^{-1}$. Thus, if the monomer concentration C or binding rate ν are changed rather than the binding energy E_B or temperature T, the change in timescale must be taken into account as well. For example, assuming that binding energy and temperature are held constant and only C or ν are changed to optimize the control parameter $\tilde{\delta}_1$ for a given structure size S, the minimal assembly time (in physical units) scales as $\sim S^{\theta+\phi} := S^{\tilde{\theta}}$. We call $\tilde{\theta} = \theta + \phi$ the *adjoined* time complexity exponent. Table I displays the values of the adjoined exponents for the different

TABLE I. Adjoined time complexity exponents, $\tilde{\theta}_{\mathrm{sim}} = \theta_{\mathrm{sim}} + \phi_{\mathrm{sim}}$, measured with the stochastic simulation (cf. Fig. 4; for $400 \leq S \leq 1032$) together with their theoretical asymptotic estimates $\tilde{\theta}_{\mathrm{th}}$ for the different monomer morphologies.

	$\tilde{\theta}_{\mathrm{sim}}$	$\tilde{\theta}_{\mathrm{th}}$
Triangle	1.79	3/2
Square	2.32	2
Hexagon	1.83	2

particle morphologies. The adjoined exponents are larger than the time complexity exponents, indicating that self-assembly in the strong binding limit is more efficient if C and ν are kept constant (optimally at their maximum values), and the binding energy or temperature are varied (the product $C\nu$ then fixes the minimal achievable assembly time in a specific system). We further deepen this discussion in Sec. V E. Also note that the adjoined exponents rank the particle morphologies differently concerning their efficiency, with triangular and hexagonal monomers having the smallest adjoined exponents in the limit of large structure sizes.

IV. EFFECTIVE KINETIC THEORY

To understand how the kinetic parameters and the morphology of the building blocks affect the assembly process and, in particular, how the power-law dependence of the assembly time on the structure size emerges, we formulate an effective kinetic theory [60] for the concentrations $c_s(t)$ of clusters of size s and the concentration of monomers $m(t) := c_1(t)$. To accomplish this, we leverage the fact that, in the strong binding limit, the assembly process typically follows a particular path that proceeds through a sequence of stable intermediate states, as shown in Fig. 2 for square-shaped building blocks. In general, the larger the number of unstable states that must be traversed to reach a new stable configuration, the less likely the particular path will be. Therefore, in formulating the effective theory, we can restrict ourselves to considering only the most likely assembly pathways that connect stable configurations via the smallest number of unstable intermediate states. The predominance of these assembly pathways could also be retraced with calculations of the free energy landscape of all cluster sizes [51,52].

A. Nucleation process

The assembly process starts with the formation of a stable nucleus consisting of σ monomers [Fig. 1(a)]. Its formation can be viewed as a one-step Markov process with intermediate cluster states $1 \leq s < \sigma$ and the final nucleus as an absorbing state σ; see the illustration for hexagon-shaped monomers in Fig. 5. In the limit where the backward rates are large compared to the forward rates, which
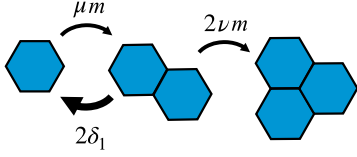
FIG. 5. The nucleation process (illustrated for hexagon-shaped monomers) corresponds to a one-step Markov process with the nucleus as the absorbing state. The forward rates μm and $2\nu m$ correspond to dimer formation and the subsequent assembly step, respectively, where the numerical prefactor in the second step accounts for the multiplicity of the process. Note that the forward rates are proportional to the monomer concentration m, considering the monomers as an unlimited reservoir. The backward rate equals $2\delta_1$, as either of the two monomers can detach. Hence, in the effective rate constant $\bar{\mu}$ for the nucleation process, Eq. (2), the combinatorial prefactors of the forward and backward reactions cancel out exactly. The same holds for the more complex nucleation processes for particles with square and triangular morphology.

here corresponds to $\delta_1 \gg \nu m$, the effective rate for the transition from a monomer to a nucleus is simply given by the ratio of the product of the forward and backward rates (cf. Appendix B):

$$r_{1 \to \sigma} = \mu \left(\frac{\nu}{\delta_1} \right)^{\sigma - 2} m^{\sigma} \equiv \bar{\mu} m^{\sigma}. \qquad (2)$$

In a kinetic theory, this translates to an effective loss term $-\sigma \bar{\mu} m^{\sigma}$ in the rate equation for the monomer concentration m with the effective rate for a reaction of the order of σ given by $\bar{\mu} = \mu(\nu/\delta_1)^{\sigma-2}$. The corresponding gain term in the rate equation for the nuclei concentration c_{σ} is $\bar{\mu} m^{\sigma}$.

B. Effective assembly process

After the formation of the nucleus, a minimum number γ of monomers must attach to the nucleus before the next stable cluster configuration is reached [Fig. 1(c)]. Thus, as in the nucleation process, one has a one-step Markov chain with the cluster of size $\gamma + \sigma$ representing an absorbing state. The corresponding effective rate, again under the assumption $\delta_1 \gg \nu m$, is given by (cf. Appendix B)

$$r_{\sigma \to \sigma + \gamma} = p_\sigma q_\sigma \nu \left(\frac{\nu}{\delta_1} \right)^{\gamma - 1} m^{\gamma} \equiv p_\sigma q_\sigma \bar{\nu} m^{\gamma}. \qquad (3)$$

Here, p_σ denotes the perimeter of the nucleus, accounting for the fact that the set of γ monomers can attach at any point along the nucleus boundary. Furthermore, q_σ is a numeric factor that accounts for the multiplicity of the seed-forming reaction pathways [e.g., for square-shaped monomers $q_\sigma = 1$, and triangle-shaped monomers $q_\sigma = 4$; cf. Appendix B and Fig. 11(a) therein]. The effective kinetic parameter for this reaction of the order of γ is given by $\bar{\nu} = \nu(\nu/\delta_1)^{\gamma-1}$.

In general, all of the following assembly steps depend on the morphology of the cluster and not just on its size s.
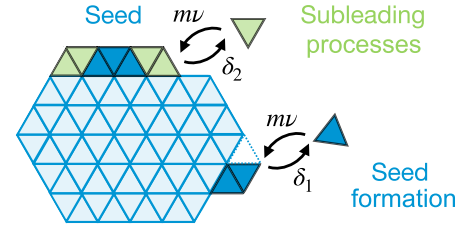

FIG. 6. Illustration of the assembly steps for large clusters (illustrated for triangle-shaped monomers). Cluster growth after nucleation proceeds in two steps for assembly processes with attachment order $\gamma > 1$ (here, $\gamma = 3$): The formation of a seed of size γ at one edge of the cluster (rate limiting slow process with an effective rate $p_s q_s \bar{\nu} m^{\gamma}$) is followed by a sequence of subleading fast processes that complete the edge very quickly (domino effect).

However, in the spirit of a mean-field approximation, we disregard the specific cluster morphology and consider only typical (large) clusters, as illustrated in Fig. 6 for triangle-shaped monomers.

In the case of hexagon-shaped monomers, clusters can grow by the attachment of a single monomer without having to go through unstable configurations. Hence, the attachment order for hexagon-shaped monomers is $\gamma = 1$, and clusters simply grow by monomer attachment at a rate $r_{s \to s+1} = p_s q_s \nu m$ with $q_s = 1$. We approximate the perimeter of the cluster as $p_s = a\sqrt{s}$ independent of the cluster morphology, where a can be estimated by considering a typical cluster or regarded as a parameter to be determined numerically by fitting to the results of stochastic simulations (cf. Appendix B). Hence, in the effective rate equation for c_s (with $s > \sigma$), there is a gain term of the form $f_s \nu m c_s$ with the combinatorial factor $f_s = p_s q_s = a s^{\omega}$ with $\omega = 1/2$ [63]. We call ω the *growth exponent* for cluster growth.

In contrast, for square- and triangle-shaped monomers, cluster growth proceeds in two steps (Fig. 6): First, a leading-order slow reaction involving γ monomers must occur to arrive at a new stable cluster configuration. This *seed formation process* occurs along one of the edges of the polygon at a rate given by Eq. (3), where the nucleus size σ is replaced by the cluster size s. For large s, the numeric factor q_s, which accounts for the degeneracy of the seed-forming reaction, is approximately a constant given by $q_s = 2$ for square-shaped monomers and $q_s = 3$ for triangles; see Appendix B with Fig. 11(b). Once such a seed is formed, it can further grow in a series of reactions of lower effective order (order 1 in the case of square-shaped monomers and order 2 in the case of triangle-shaped monomers). Since the corresponding rates are large compared to the rate of seed formation, the leading-order process is followed with high probability—in a kind of "domino effect"—by a series of such lower-order processes until the entire edge of the polygon is filled up. Subsequent growth then requires another seed formation process along any polygon edge. Since the seed formation process is rate limiting, the total rate $r_{s \to s+l_s}$ of filling up the edge of a

polygon of length l_s approximately equals the rate of seed formation $r^{\text{seed}}_{s \to s+\gamma}$.

To analyze systems with different monomer morphologies on equal footing, we now formulate the growth dynamics in terms of an effective rate at which the cluster grows by one unit. The corresponding *effective monomer attachment rate* is given by $\tilde{r}_{s \to s+1} = l_s r_{s \to s+l_s}$, where $l_s \sim \sqrt{s}$ is a typical edge length of the cluster. Thus, in an effective rate equation for the concentration c_s of clusters of size s, the gain term has the form $f_s \bar{\nu} m^\gamma c_s$, where $f_s = q_s p_s l_s \approx a s^\omega$ with a growth exponent $\omega = 1$. This takes into account that the multiplicity of the reactions in the growth process is proportional to the product of the typical perimeter $p_s \sim \sqrt{s}$, the multiplicity factor q_s for the seed-forming reaction, and the edge length $l_s \sim \sqrt{s}$ of a cluster. Similar to hexagon-shaped monomers, for large s, the prefactor a is approximately a constant that depends only on the morphology of the building blocks [64]. It can be determined numerically by fitting it to the results of the stochastic simulations or estimated by combinatorial arguments (cf. Appendix B).

C. Effective kinetic model

Taken together, one obtains the following set of *rate equations* for the effective kinetic model:

$$\partial_t m(t) = -\sigma \bar{\mu} m^\sigma(t) - \bar{\nu} m^\gamma(t) \sum_{s=\sigma}^{S-1} f_s c_s(t), \quad (4a)$$

$$\partial_t c_\sigma(t) = \bar{\mu} m^\sigma(t) - \bar{\nu} m^\gamma(t) f_\sigma c_\sigma(t), \quad (4b)$$

$$\partial_t c_s(t) = \bar{\nu} m^\gamma(t) [f_{s-1} c_{s-1}(t) - f_s c_s(t)]. \quad (4c)$$

Here, to describe the cluster growth, we use the effective one-particle growth rates derived in the previous paragraph. For the combinatorial factor f_s, we use the asymptotic expression for the entire domain $\sigma \le s < S$:

$$f_s = a s^\omega, \quad (5)$$

with the prefactor a determined numerically for each monomer morphology (cf. Appendix B) [65]; for completed structures (absorbing state) $s = S$, we set $f_S = 0$. Equation (4a) accounts for the loss of monomers due to nucleation events with the effective rate $\bar{\mu}$ and attachment of monomers to incomplete complexes with the effective rate $\bar{\nu}$. Similarly, Eq. (4b) describes the gain and loss of stable nuclei due to nucleation events and the initial growth of stable nuclei, respectively. Note that, for triangle-shaped monomers, the first assembly step following nucleation has a larger attachment order $\gamma = 4$ than all subsequent assembly steps where $\gamma = 3$. This is accounted for explicitly in the implementation of the effective model for triangle-shaped monomers (cf. Appendix B), as we find that it significantly improves the agreement with the results

from the stochastic simulations. Finally, Eq. (4c) specifies the cluster dynamics with a gain due to the growth of smaller complexes and loss due to the growth of complexes of the same size s. As shown in Fig. 4, we find excellent agreement between the effective kinetic theory and the stochastic simulations for all particle morphologies.

In conclusion, the effective kinetic theory identifies the three parameters σ (nucleation size), γ (attachment order), and ω (growth exponent) as the major determinants of self-assembly kinetics. Since the above kinetic approach is general, we suppose that it can accurately describe the self-assembly of a broad class of systems, including, for example, three-dimensional structures.

V. SCALING THEORY

The numerically observed power laws for the optimal detachment rate $\tilde{\delta}_1^{\text{opt}}$ and the minimum assembly time T_{90}^{min} as a function of the structure size S suggest that the self-assembly dynamics exhibit some underlying scale invariance (Fig. 4). In this section, we analyze the kinetic theory for cluster growth in the strong binding limit, Eq. (4), using several complementary approaches that offer crucial insights into the nature of the dynamics. We limit our presentation to discussing the main ideas, leaving the formal mathematical analysis to Appendixes C and D.

A. Onset criterion for finite yield

In the strong binding limit, the yield curves obtained from the stochastic simulations exhibit scaling behavior (cf. the data collapse in Fig. 3)

$$Y(\tilde{\delta}_1, S) = \hat{Y}[\tilde{\delta}_1 / \tilde{\delta}_1^{\text{opt}}(S)]. \quad (6)$$

This means that all detachment rate-defining characteristic features of the yield curves, such as the threshold rate for finite yield $\tilde{\delta}_1^{\text{on}}$, must show the same scaling behavior with structure size S as the optimal detachment rate $\tilde{\delta}_1^{\text{opt}}$. Because of this equivalence in scaling behavior, it is sufficient to analyze the self-assembly dynamics for a vanishingly small yield. This considerably simplifies the rate equations Eq. (4), which we do first in the following. Using the asymptotic form for the combinatorial factor f_s, the monomer dynamics [Eq. (4a)] can be written as

$$\partial_t m(t) = -\sigma \bar{\mu} m^\sigma - a \bar{\nu} \langle s^\omega \rangle m^\gamma K, \quad (7)$$

where $K := \sum_{s=\sigma}^{S-1} c_s$ is the total concentration of incomplete clusters and $\langle s^\omega \rangle$ the ω moment of the cluster size distribution c_s. As long as the yield is zero, no clusters are completed, and, thus, the number of incomplete clusters grows with each nucleation event:

$$\partial_t K(t) = \bar{\mu} m^\sigma. \quad (8)$$

These equations for the monomers and incomplete clusters are not closed, because they require information about the moments $\langle s^\omega \rangle$ of the cluster size distribution. However, in the limiting case of vanishingly low yield considered here, where no appreciable amount of finished structure is present, these equations can be greatly simplified by taking advantage of the fact that all bound monomers are contained in unfinished clusters:

$$C - m = K\langle s \rangle, \tag{9}$$

where $\langle s \rangle$ is the average size of a cluster. For a growth exponent $\omega = 1$ (valid for triangle- and square-shaped monomers), this immediately leads to a closed set of equations; otherwise, for general ω, we use a mean-field-like "factorization approximation" $\langle s^\omega \rangle \approx \langle s \rangle^\omega$ to close the equations.

To determine when the system displays finite yield, one must specify under which conditions the cluster size distribution c_s can develop a finite weight at fully formed clusters of size S. The dominant factor in the dynamics of the cluster size distribution is drift due to (irreversible) cluster growth, which implies that (at the mean-field level) the largest cluster s_{max} obeys the growth law

$$\partial_t s_{max}(t) = a\bar\nu s_{max}^\omega m^\gamma, \tag{10}$$

corresponding to cluster growth by forming a seed of size γ at one of the cluster's edges, followed by a domino effect filling up the cluster edge as described in Sec. IV.

The set of equations, Eqs. (7)–(10), exhibits scale invariance, as shown in Appendix C. In particular, one finds that the maximal cluster size obeys the scaling form

$$s_{max}(t) = \eta^y \tilde s_{max}(a\bar\nu C^\gamma \eta^z t) \tag{11}$$

with $\eta = (\bar\mu/a\bar\nu)C^{\sigma-\gamma-1}$ and the scaling exponents

$$z = \frac{1-\omega}{2-\omega} \quad \text{and} \quad y = -\frac{1}{2-\omega}. \tag{12}$$

The onset condition for finite yield, $S = s_{max}(\infty)$, can then be written as a power law using the definitions of the effective rates $\bar\mu$ and $\bar\nu$ in Eqs. (2) and (3):

$$\tilde\delta_1^{on} \sim S^\phi, \tag{13}$$

with the *control parameter exponent*

$$\phi = \frac{2-\omega}{\sigma-\gamma-1}. \tag{14}$$

Furthermore, from Eq. (11), we read off that the typical timescale of the assembly process scales as

$$\tau_{assem} \sim (a\bar\nu C^\gamma \eta^z)^{-1} \sim \tilde\delta_1^\alpha, \tag{15}$$

where $\alpha = \gamma - 1 + z(\sigma - \gamma - 1)$. By using the scaling of the onset detachment rate $\tilde\delta_1^{on}$ with target size S, Eq. (13), one thus obtains

$$\tau_{assem} \sim S^{\alpha\phi} := S^\theta \tag{16}$$

with the *time complexity exponent* given by

$$\theta = \frac{(1-\omega)\sigma + \gamma + 2\omega - 3}{\sigma - \gamma - 1}. \tag{17}$$

Strictly speaking, τ_{assem} corresponds to the timescale for forming the first complete structure starting from an existing nucleus. However, we are actually interested in the minimal assembly time T_{90}^{min} required to achieve 90% yield. To see that T_{90}^{min} scales identically as τ_{assem}, we first determine what the rate-limiting step in the assembly process is by comparing τ_{assem} with the typical time required to form a nucleus:

$$\tau_{nuc} = (\bar\mu C^\sigma)^{-1} \sim \tilde\delta_1^\beta, \tag{18}$$

where $\beta = \sigma - 2$. For all the particle morphologies considered here, one has $\beta > \alpha$ [66], implying that nucleation is the rate-limiting step of the self-assembly process. Using the scaling equivalence $\tilde\delta_1^{opt} \sim \tilde\delta_1^{on} \sim S^\phi$ and that nucleation is rate limiting, the minimal assembly time (for 90% yield), T_{90}^{min}, can be approximated by the time required for $0.9N/S$ nucleation events to happen, each of which occurs at an effective rate $\sim \bar\mu_{opt} C^\sigma$. This results in the scaling law

$$T_{90}^{min} \sim \frac{C}{S} \cdot \frac{1}{\bar\mu_{opt} C^\sigma} \sim S^\theta \tag{19}$$

with θ identical to Eq. (17). In conclusion, the time to assemble the first structure and the minimal time to reach 90% yield show the same scaling behavior with target size S.

The theoretical estimates of the exponents for the three different particle morphologies are shown in the tables in Figs. 4(b) and 4(c) compared with the values inferred from stochastic simulations. The theoretical exponents generally provide good estimates for the numerically determined exponents. The remaining differences can be attributed to the asymptotic scaling regime not being fully achieved for the simulated system sizes and the actual growth exponent ω deviating slightly from its estimated value [67].

For a given ω, the smallest possible value for the time complexity exponent according to Eq. (17) is obtained for $\gamma = 1$, giving $\theta = 1 - \omega$. In contrast, if $\gamma > 1$, θ decreases steadily with increasing nucleation size σ, asymptotically reaching the minimum $1 - \omega$ in the limit $\sigma \to \infty$. Realistically, however, this limit cannot be reached, because, for very large σ, the description by our effective model breaks down, since δ_1^{opt} can no longer be assumed to

be large compared to the reaction rate $C\nu$. In summary, this suggests that a minimum attachment order $\gamma = 1$ is key to achieving a minimal time complexity exponent θ and, thus, maximal time efficiency for the self-assembly of large structures in the strong binding limit.

B. Scale invariance of the self-assembly dynamics

The theory in the strong binding limit can be shown to exhibit general scale invariance, regardless of the assumption of a small yield. We outline the main arguments leading to this scale invariance below while leaving the mathematical details to Appendix D. This scale invariance explains the collapse of the yield curves observed in Fig. 3 and has further essential consequences, as discussed in this subsection.

We use a hydrodynamic approximation to see that the self-assembly dynamics shows scale invariance and write Eq. (4) as an effective advection equation for the cluster size distribution $c(s)$, as proposed by previous work [19,68,69]. We consider $c(s)$ as a continuous function of a real variable $s \in [\sigma, S]$ and Taylor expand the first term in Eq. (4c) to first order in ∂_s to obtain the advection equation

$$\partial_t c(s, t) = -a\bar{\nu} m^\gamma \partial_s [s^\omega c(s)]. \tag{20}$$

This equation must be completed with the appropriate boundary conditions:

$$a\bar{\nu} m^\gamma(t)\sigma^\omega c(\sigma, t) = \bar{\mu} m^\sigma(t), \tag{21}$$

accounting for the influx of complexes due to nucleation, and an absorbing boundary condition $c(S, t) = 0$ for the complexes that get completed. In the equation for the monomer concentration, Eq. (4a), we neglect the loss term due to nucleation, since monomer loss due to attachment processes plays a much more significant role whenever there is significant growth of the structures (if there is substantial yield, roughly S/σ times as many monomers are consumed by attachment reactions as compared to nucleation reactions). Approximating the sum in Eq. (4a) by an integral, we thus obtain for the concentration of the monomers

$$\partial_t m(t) = -a\bar{\nu} m^\gamma \int_\sigma^S s^\omega c(s) ds, \tag{22}$$

with initial condition $m(0) = C$. As the last step, we approximate the lower limit of the integral σ by a small target-size-dependent value ϵS; i.e., we neglect a fraction $(\epsilon S - \sigma)/S \approx \epsilon$ of all assembly steps, where ϵ can, in principle, be arbitrarily small. This step is helpful, as it allows us to perform a variable transformation in the integral, thus making the integral, including its boundaries, independent of S. If $\omega > 1$, which can be the case for the self-assembly of three-dimensional structures

(cf. Appendix D 2), the approximation of the lower integral boundary can become inaccurate, because the growth of small clusters from size σ to ϵS defines the predominant timescale in the growth process of a cluster. In this case, we therefore approximate the upper boundary of the integral instead; see Appendix D 2 for details.

With these approximations [70], it can be shown with a scaling ansatz that the monomer concentration and complex concentration as functions of both s and t, the parameter δ_1, and the target structure size S obey the scaling forms

$$m(t, \delta_1, S) = C\hat{m}(S^{-\theta}\tilde{t}, S^{-\phi}\tilde{\delta}_1) \tag{23}$$

and

$$c(s, t, \delta_1, S) = CS^{-2}\hat{c}(S^{-1}s, S^{-\theta}\tilde{t}, S^{-\phi}\tilde{\delta}_1), \tag{24}$$

with $\tilde{t} = C\nu t$ and ϕ and θ identical to Eqs. (14) and (17) as obtained in Sec. V A. Moreover, the condition to obtain a yield Y translates directly into a condition on the scaling functions $\hat{m}$ and $\hat{c}$:

$$\hat{m}(S^{-\theta}\tilde{t}, S^{-\phi}\tilde{\delta}_1) + \int_\epsilon^1 x\hat{c}(x, S^{-\theta}\tilde{t}, S^{-\phi}\tilde{\delta}_1)dx = 1 - Y, \tag{25}$$

where we use that the yield being Y is equivalent to the number of resources that remain in the system, i.e., that have not yet been absorbed by the right boundary, equals $(1 - Y)C$. This shows that the yield obeys the scaling form

$$Y(t, \delta_1, S) = \hat{Y}(S^{-\theta}\tilde{t}, S^{-\phi}\tilde{\delta}_1), \tag{26}$$

implying that all detachment rate-defining characteristics of the yield curve scale identically ($\sim S^\phi$) as functions of the structure size. For example, the threshold rate to obtain a fixed yield Y, $\delta_1^{\text{on}}(Y)$, scales identically as the optimal rate $\delta_1^{\text{opt}}(Z)$ to obtain a yield Z in the minimal amount of time, explaining the collapse of the yield curves in Fig. 3. Furthermore, by solving Eq. (26) for the time argument, we find that the time T_Y required to achieve a yield $Y \in (0, 1)$ as a function of δ_1 obeys

$$T_Y(\delta_1, S) = (C\nu)^{-1}S^\theta\hat{T}_Y(S^{-\phi}\tilde{\delta}_1) \tag{27}$$

whenever the function has a finite value. This shows that the same scaling behavior that we observe for the optimal detachment rate and minimal assembly time in Figs. 4(b) and 4(c), in fact, applies identically to the entire T_Y curve.

Robustness of T_{90}^{min}. The fact that the T_{90} curves in Fig. 4(a) maintain their shape when increasing S allows us to uniquely characterize the sensitivity of the minimal assembly time as a function of the control parameter. Figure 4(a) shows that the morphology strongly influences how sensitively the minimal assembly time depends on $\tilde{\delta}_1$:

For example, for the assembly time to differ from the minimal assembly time by at most one order of magnitude, the parameter $\tilde{\delta}_1$ may vary only within a relative range of 40% in the case of triangular monomers but 340% for hexagonal monomers. These values correspond to respective ranges of the binding energies of $\pm 0.18 k_B T$ and $\pm 1.0 k_B T$. This suggests that hexagonal monomers not only self-assemble more efficiently, they also require less fine-tuning of the control parameter to achieve their maximal efficiency. In general, we expect that the nucleation size σ crucially determines the sensitivity of the minimal assembly time, since σ impacts both the depth of the U shape and the scaling of T_{90} in the limit of large $\tilde{\delta}_1$ [cf. Fig. 4(a)]. Small nucleation sizes, therefore, seem favorable to reduce the sensitivity.

Dimerization barrier. Note that, due to the boundary condition Eq. (21), the scaling functions Eqs. (23) and (24) still depend on the ratio $\tilde{\mu} := \mu/\nu$ between the dimerization and the binding rate (dimerization barrier). In biological systems, due to cooperative or allosteric binding effects, the dimerization rate μ can be significantly smaller than the rate of attachment ν of monomers to larger clusters [1,71,72], resulting in $\tilde{\mu} \ll 1$. In Appendix D 1, we show that, in an analogous way, scaling functions for the parameter $\tilde{\mu} := \mu/\nu$ can be obtained. A fixed value of $\tilde{\mu}$ does not affect the scaling behavior of the assembly time in dependence on the target structure size. However, we find that, for square- and triangle-shaped monomers (generally for systems with $\gamma > 1$), the minimal assembly time decreases proportionally with $\tilde{\mu}$ (cf. Appendix D 1). Furthermore, one can also determine the time complexity exponent for a scenario in which the assembly process is controlled entirely through $\tilde{\mu}$ (while $\tilde{\delta}_1$ is kept constant). For this scenario, we obtain the same time complexity exponent $1 - \omega$ as when changing $\tilde{\delta}_1$ in the case of hexagonal monomers (cf. Appendix D 1 for a further discussion). However, in experiments, it is presumably considerably simpler to control $\tilde{\delta}_1$ (via temperature, monomer concentration, etc.) rather than $\tilde{\mu}$, which requires control over allosteric effects and, hence, a particular molecular design of the building blocks.

C. Heuristic onset criterion for high yield

The results from the scaling analysis of the rate equations can also be understood heuristically based on the following scaling arguments for the onset of high yield. To achieve a high yield of target structures of size S, in total, about S times as many attachments (cluster growth) events as nucleation events must occur:

$$\nu_{\text{tot}} = S \cdot \mu_{\text{tot}}, \tag{28}$$

or, in other words, during the assembly process, any structure formed by a nucleation event must also have a significant probability of completion. The total nucleation rate is given by [cf. Eq. (8)] $\mu_{\text{tot}} = \bar{\mu} m^\sigma$, and the total growth rate can be estimated as [cf. Eq. (7)]

$$\nu_{\text{tot}} = a\bar{\nu}\langle s^\omega \rangle m^\gamma K, \tag{29}$$

where $\langle s^\omega \rangle := \int_\sigma^S s^\omega ds/(S - \sigma) \approx S^\omega/(\omega + 1)$ is the average growth rate of a cluster. Note that, during the assembly process, most clusters undergo all assembly steps to obtain a high yield, so it is reasonable to consider $\langle s^\omega \rangle$ as proportional to S^ω, as calculated here. Finally, since a finite fraction of the K clusters formed should become finished structures, one has $K \sim C/S$. Using the scaling of the effective kinetic parameters $\bar{\mu}$ and $\bar{\nu}$ with the detachment rate, one finds

$$S = \frac{\nu_{\text{tot}}}{\mu_{\text{tot}}} \sim \tilde{\delta}_1^{\sigma - \gamma - 1} S^{\omega - 1}, \tag{30}$$

where we assume that the monomer density is of the order of C and, hence, independent of the structure size S. Therefore, to achieve high yield, the detachment rate must scale as $\tilde{\delta}_1^{\text{full}} \sim S^\phi$, where ϕ is identical to Eq. (14). Thus, both $\tilde{\delta}_1^{\text{full}}$ and $\tilde{\delta}_1^{\text{on}}$ (cf. Sec. V A) scale in the same way with the structure size, providing further confirmation of the scaling law for yield curves.

D. Slow nucleation limit

Interestingly, the assembly time T_{90} can be calculated analytically when considering the extreme case of very slow nucleation that applies when $\tilde{\delta}_1 \gg \tilde{\delta}_1^{\text{opt}}$. To see this, one needs to realize that, in this regime, the concentration of unfinished structures is low at any time, so one can assume that each cluster is completed before the next one is nucleated. Each nucleation event then reduces the number of monomers by S units; hence, the evolution of the monomer concentration obeys $\partial_t m = -S\bar{\mu} m^\sigma$. Separating variables and integrating m from C to $(1 - Y)C$, where $Y \in (0, 1)$ is the desired yield, and t from 0 to T_Y, one obtains

$$T_Y = Q_Y \frac{C}{S} \frac{1}{\bar{\mu} C^\sigma} \tag{31}$$

with

$$Q_Y := \frac{(1 - Y)^{1 - \sigma} - 1}{\sigma - 1} \tag{32}$$

and $\bar{\mu} \sim \delta_1^{\sigma - 2}$ [cf. Eq. (2)]. From a comparison with the numerical data [see Fig. 16(a) in Appendix F], we find that this analytic result matches the assembly time exactly in the regime where it scales as $T_{90} \sim \tilde{\delta}_1^{\sigma - 2}$.

Note that T_Y has the same functional form as Eq. (19), confirming the asymptotic scaling of T_{90} (Fig. 4) and the

validity of the general scaling law for the assembly time given by Eq. (27). Furthermore, the analytic result, Eq. (31), reveals a strong dependence of the time efficiency in the strong binding limit on the nucleation size σ: Not only does σ determine the dependence on the detachment rate ($T_{90} \sim \tilde{\delta}_1^{\sigma-2}$), but it also strongly affects the prefactor Q_Y. For example, to achieve 90% yield, one has $Q_{0.9} = 10^5$ for triangle-shaped monomers but only $Q_{0.9} = 333$ for square-shaped monomers. Hence, this explains why square-shaped monomers assemble more efficiently than triangle-shaped monomers in the strong binding limit (see Fig. 4), although both systems exhibit the same time complexity exponent $\theta = 1$. Note also that, by decreasing the demanded yield Y from 0.9 to 0.5, the assembly time for triangular monomers in the limit of large $\tilde{\delta}_1$ decreases by a factor of $Q_{0.9}/Q_{0.5} = 3225$. In contrast, for square and hexagonal monomers, it decreases only by a factor of 143 and 33, respectively. Thus, we expect that differences in the assembly time between the different particle morphologies tend to become less pronounced at a lower yield (note, however, that the yield Y does not affect the scaling of T_{90} with the target structure size S). Or, to put it another way, increasing the yield in systems with large σ is more expensive in terms of time efficiency than in systems with small σ.

E. Interpretation of the scaling results

By exploiting the scale invariance of the dynamic equations in Secs. V A and V B, we show that the assembly time and detachment rate in the strong binding limit exhibit a characteristic scaling behavior as functions of the target structure size. This scaling behavior is characterized by the exponents θ (time complexity exponent) [Eq. (17)] or $\tilde{\theta}$ (adjoined time complexity exponent) (cf. Sec. III C) and ϕ (control parameter exponent) [Eq. (14)], which themselves are functions of the three structural parameters σ, γ, and ω determined by the morphology of the building blocks. The different time complexity exponents associated with distinct monomer morphologies lead to considerable differences in the corresponding assembly times for large target structure sizes. How can we understand intuitively why monomers with different morphologies assemble at different speeds? To answer this question, first of all, we have to distinguish two experimental scenarios: In the first scenario, we assume that the monomer concentration and attachment rate are constant and the detachment rate δ_1 is changed (for example, by changing binding energy or temperature) to tune $\tilde{\delta}_1 = \delta_1/(C\nu)$. Conversely, in the second scenario, the monomer concentration (or attachment rate) is varied while δ_1 is kept constant. The scaling of the assembly time in the second scenario is, therefore, described by the adjoined time complexity exponents; see Sec. III C. We start by discussing the first scenario,

in which the reactive timescale $C\nu$ remains constant and δ_1 is varied.

Here, the heuristic argument from Sec. V C provides a useful interpretation for why different particle morphologies assemble at vastly different speeds. Intuitively speaking, monomers with certain types of shapes assemble faster than others because their particular morphologies allow for a reduction of the nucleation rate (by decreasing δ_1) while the overall cluster growth rate remains unaffected by changing δ_1. Hexagon-shaped monomers are an example of such a system: To form a nucleus of size $\sigma = 3$, the system has to pass through an unstable state which makes the rate of nucleus formation sensitive to variations in the detachment rate δ_1; see Fig. 5. In contrast, once a nucleus has formed, monomers can always attach in a way so that they immediately have two binding partners, whereby the growth process is independent of δ_1 (in Refs. [51,52], this is reflected by the finding that the free-energy gradient of large clusters is generally much larger if the subunits have higher coordination numbers). Since a high yield of the assembly process in the strong binding limit requires the ratio between the overall nucleation and cluster growth rate to be small [cf. Eq. (30)], the process can, thus, be controlled efficiently by tuning δ_1. This means that only a comparatively small change in the nucleation timescale (upon changing δ_1) is necessary for the ratio to decline and the yield to grow.

In contrast, if even only a tiny fraction of the growth events following nucleation require a transition through an unstable state (e.g., for square- and triangle-shaped monomers), these slow steps eventually dominate the timescale of cluster growth, which, thus, also becomes dependent on δ_1. Therefore, the nucleation rate must now be decreased more significantly compared to the case in which the growth rate is independent of δ_1 to achieve the same ratio between nucleation and growth rate [Eq. (30)]. Thus, with nucleation being the rate-limiting step (cf. Sec. V A), this makes the overall process less time efficient.

Hence, as exemplified by the system with hexagon-shaped monomers, a minimal attachment order $\gamma = 1$ is key to achieving maximal time efficiency in the scenario where the binding energy or temperature are varied. In Sec. VI, we discuss several ways particle morphologies can be designed to assemble with $\gamma = 1$ and, generally, how this can be applied to increase the efficiency of artificial self-assembly processes.

But what happens if instead of the detachment rate δ_1 the monomer concentration C (or binding rate ν) is varied to tune the ratio $\tilde{\delta}_1$? In this case, with increasing target structure size, the monomer concentration must be further reduced to enhance $\tilde{\delta}_1$. This necessarily increases the timescale of all binding reactions and, thus, decreases time efficiency. Since the nucleation size σ is smallest for hexagonal monomers, the monomer concentration must be reduced even more than for square and triangular

building blocks to achieve the required reduction in the effective nucleation rate, as expressed by the magnitude of the respective control parameter exponents ϕ [cf. Fig. 4(c)]. Table I displays the adjoined time complexity exponents $\tilde{\theta} = \theta + \phi$, which describe the scaling of the assembly time with the target structure size for this scenario. Hence, assuming that the binding energy per bond for triangular and hexagonal monomers is the same and the monomer concentration is optimized, triangular monomers will self-assemble into large structures faster than hexagonal monomers. Irrespective of the morphology of the monomers, however, the first scenario (with the monomer concentration being constant and chosen as large as possible) is more efficient than the second (since $\theta < \tilde{\theta}$). In other words, for hexagonal monomers to self-assemble significantly faster than square or triangular monomers, the binding energy per bond needs to be considerably lower for hexagon-shaped monomers (note, however, that the total binding energy per particle might still be larger; see Sec. III C). Hence, in our discussion in Sec. VI, we generally relate to the first scenario. In the next section, we discuss a slightly different system where hexagonal particles self-assemble more efficiently, even in both experimental scenarios.

F. Dependence on the morphology of the target structure

We have discussed in detail how the assembly time depends on the morphology of the building blocks. Here, we briefly demonstrate that the assembly time also depends on the morphology of the target structure. Specifically, so far, we have considered the formation of bulk structures that extend symmetrically into both dimensions. An important application of self-assembly in nanotechnology is the formation of elongated tubes (cylinders) [73,74]. Therefore, we simulate the self-assembly of tubes with a fixed circumference of six monomers and a variable length of $S/6$ monomers as an example of structures that extend unequally in both dimensions. Again, we assume periodic boundary conditions in both dimensions so that structures stop growing once they reach the target size S (the case of unlimited cluster growth is analyzed in Appendix I, showing that the scaling properties are indeed identical). What time complexity exponent do we expect for the self-assembly of tubes? The key difference to the growth of symmetric target structures lies in the fact that, once the tubes close along the short dimension, they keep growing at a constant growth rate g_s, since the perimeter of the boundary remains constant. The growth rate g_s can easily be calculated with the same formalism as in Sec. IV as we show explicitly in Appendix B 6. In particular, the growth of large clusters is characterized by a growth exponent $\omega = 0$. Hence, Eq. (17) yields the time complexity exponents $\theta = 3$ for square- and triangle-shaped monomers and $\theta = 1$ for hexagonal monomers, respectively. The time
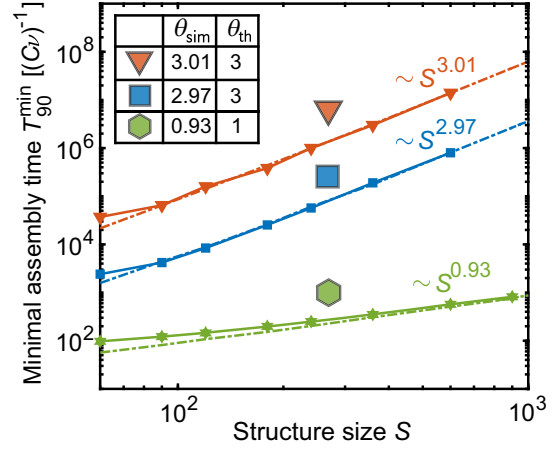

FIG. 7. Time efficiency of the self-assembly of tubes. The minimal assembly time $T_{90}^{\min}$ for the self-assembly of two-dimensional tubes (cylinders) inferred from the stochastic simulations (markers and solid lines) and effective theory (dash-dotted lines) (cf. Appendix B 6) is plotted against the size S of the target structure. The tubes have a fixed width (circumference) of six monomers and a variable height of $S/6$ monomers (with periodic boundary conditions in both dimensions). Simulations are performed with particle numbers $N = 500S$. The minimal assembly time is determined by averaging the assembly times over ten independent runs for each different value of δ_1 and subsequently taking the minimum of the averages. The table shows the time complexity exponent θ_{sim} inferred from the last three data points of the stochastic simulation in comparison with the theoretical value θ_{th} according to Eq. (17) with growth exponent $\omega = 0$. Predictions of the effective model agree excellently with the simulation data. The fact that the exponents are very different from those in Fig. 4(b) shows that also the morphology of the target structure crucially influences self-assembly efficiency.

complexity exponents measured from stochastic simulations precisely coincide with the theoretical values; see the table in Fig. 7. Furthermore, the assembly time is predicted very accurately by the effective model, with the growth rates modified accordingly (cf. Appendix B 6); see Fig. 7. Because of the large discrepancies of the time complexity exponents, hexagonal monomers self-assemble into large tubes significantly more efficiently than square or triangular monomers. Furthermore, the adjoined time complexity exponents (cf. Sec. III C) $\tilde{\theta} = 3$, 4, and 5 for hexagonal, triangular, and square monomers, respectively, are quite different in the case of tubes. This suggests that hexagon-shaped monomers are favorable also in the scenario where the monomer concentration is varied to find the optimal control parameter value.

This demonstrates that the time complexity also depends considerably on the morphology of the target structure. Our scaling approach, however, is very general and allows us to readily predict the scaling exponents if the growth exponent ω is correctly identified for the respective target structures. Our results suggest that,

also in the case of two-dimensional tubes, self-assembly is significantly more efficient if hexagonal rather than square or triangular monomers are used.

G. Robustness of the scaling behavior

To demonstrate the relevance and robustness of the scaling results, we investigate various differing scenarios and modifications of the model. The following briefly summarizes these investigations and their findings. Detailed discussions and explanations of the following analyses can be found in Appendixes E–J.

1. Cluster interactions

In the above model and analysis, we make the simplifying assumption that clusters do not interact with each other and grow only by attachment of single monomers (ideality assumption). In Appendix E, we discuss in detail whether and how cluster interactions could affect the assembly time and how this affects our results. We find that the potential role of cluster interactions depends strongly on the morphology of the particles as well. In any case, however, the assembly time is not or only slightly affected by cluster interactions, showing that our results derived under the ideality assumption are robust.

2. Lower yield thresholds

Furthermore, in Appendix F, we show that our scaling results are invariant to the threshold for the yield we use to define the assembly time. Specifically, we verify that the time T_{50} required to achieve 50% yield scales identically as the time T_{90} to achieve 90% yield.

3. Seeded systems

In Appendix G, we show that our scaling results are also robust to varying initial conditions. Specifically, we show that the scaling behavior is the same if we consider a "seeded" system in which we initially put in a number $\sim N/S$ of stable nuclei (seeds), as is often done in experiments [35,75]. Generally, we find that the minimal assembly time increases significantly due to seeding if $\gamma > 1$ (square- and triangle-shaped monomers) but remains constant if $\gamma = 1$ (hexagonal monomers) [cf. Appendix G and Fig. 17(b)]. In any case, however, we find that the scaling of the assembly time as a function of target structure size remains invariant if the system is seeded initially with a number $\sim N/S$ of nuclei [cf. Fig. 17(c)]. This can be easily explained with our scaling analysis, since the new initial condition $c_\sigma(t=0) \sim S^{-1}$ is consistent with the scale invariance described by the scaling form for the cluster size density, $c(s, t, S) \sim S^{-2}\hat{c}(S^{-1}s, S^{-\theta}\tilde{t})$ [cf. Eq. (24)]; see Appendix G for details. Hence, all our scaling results remain valid for seeded systems.

4. Annealing

Annealing is another frequently used experimental technique in which temperature continuously decreases during self-assembly. The idea is to compensate for the decreasing attachment rate resulting from monomer consumption with a simultaneous reduction of the detachment rate. In this way, the ratio between the frequencies of detachment events (rate δ_1) and attachment events (rate νm) can be kept at a constant level during self-assembly. In Appendix H, we study the behavior of self-assembly under a perfect annealing protocol, in which the ratio between attachment and detachment frequencies remains exactly constant during the process. We find that annealing enhances the time efficiency of self-assembly, but the time complexity exponent remains the same. In the framework of the effective theory, annealing is described with a time-dependent detachment rate $\delta_1(t) \sim m(t)$. Since the monomer concentration does not scale with S [cf. Eq. (23)], the scaling analysis can be performed in precisely the same way, and the invariance of the scaling exponents becomes evident.

5. Unlimited cluster growth

Furthermore, in Appendix I, we study the scaling properties of systems in which cluster growth is not self-limiting (i.e., there are no defined target structures) but clusters can grow indefinitely. In nature and nanotechnology, there are plentiful examples of such kinds of systems where structure growth is in principal unlimited [32,41]. Describing infinite cluster growth with an effective model as in Sec. V B requires us only to replace the upper limit of cluster sizes in the integral in Eq. (22) by infinity and to discard the absorbing boundary at $s = S$. Clearly, these modifications do not affect the scaling analysis. We define the yield in systems with unlimited cluster growth simply as the fraction of resources bound in clusters of a size $s \geq S$. Thereby, Eq. (25) immediately shows the scale invariance of this yield metric [Eq. (26)] and the corresponding assembly time [Eq. (27)]. Alternative definitions of the assembly time are possible as well (cf. Appendix I). Hence, our scaling results transfer directly to systems with unlimited cluster growth [76]. In Appendix I, we furthermore characterize the resulting final cluster size distribution in systems with unlimited cluster growth.

6. Heterogeneous systems

In the main text, we focus exclusively on homogeneous systems in which all particles are identical. In nature and nanotechnology, there are plentiful examples of heterogeneous systems that self-assemble from distinct species of constituents [40,41,50]. Heterogeneous designs can furthermore reduce assembly errors [47] and provide a viable way to limit the growth of structures. Hence, heterogeneous systems represent a suitable testing ground to verify our

scaling results experimentally. Therefore, it is important to show that our results apply likewise to heterogeneous systems. To test the validity of our results for heterogeneous systems, in Appendix J 1, we first consider the case of ideal heterogeneous systems, where distinct species bind only with their specific neighbors (see Fig. 19), but the reaction rates and concentrations of all species are the same. We show mathematically that the ideal heterogeneous system decouples into S identical homogeneous systems (one for each species). Thus, the ideal heterogeneous system is equivalent to a homogeneous system provided particle numbers are sufficiently large to suppress fluctuations between species' concentrations [53]. Departing from this idealized case, in Appendixes J 2 and J 3, we also consider systems with random heterogeneous binding rates as well as clusters that have explicit boundaries (rather than periodic boundary conditions). In both cases, self-assembly requires slightly longer, but the time complexity exponent is identical as in the ideal heterogeneous or homogeneous case. We furthermore study scenarios in which four neighboring species are provided in significantly larger concentrations (Appendix J 5) or the respective detachment rates of these same four species are strongly reduced (Appendix J 4). Both scenarios enhance the rate of nucleation of these specified four species and, thus, affect the assembly dynamics in similar terms as seeding discussed above. In particular, the minimal assembly time drastically increases as a result of these counterproductive measures (cf. Fig. 21). In contrast, a viable way to improve time efficiency is by enhancing the cluster growth rate (rather than the nucleation rate). In heterogeneous systems, this can be done particularly well through hierarchical self-assembly as we discuss in detail in the next section (see also Appendix J 6). Taken together, we find that heterogeneous systems behave very similarly to homogeneous systems, and our results explain and describe their properties in an analogous way.

7. Higher-order reversibility

Besides the assumption of an ideal self-assembly process, the second basic assumption we made in our analysis is that the binding energy is large, and, hence, all higher-order detachment rates δ_2, δ_3, etc., vanish (strong binding limit). In a follow-up paper, we will relax this assumption and investigate how higher-order detachment processes affect the assembly dynamics in cases where the binding energy is low. In this context, we will show that our effective theory for the strong binding limit can be generalized to incorporate the effect of higher-order detachment processes. We demonstrate that the minimum assembly time and its scaling behavior remain widely unaffected by higher-order detachment processes in a broad parameter range of structure sizes and binding energies. Hence, our scaling results for the strong binding limit remain valid and useful even beyond the regime where the binding energy is very

large and all higher-order detachment processes are strictly negligible.

VI. SUMMARY AND APPLICATIONS

In the following, we provide a concise overview of our findings and demonstrate how these insights contribute to a better understanding of biological self-assembly processes. Furthermore, we explore their potential applications in optimizing the efficiency of artificial self-assembly processes.

A. Summary

In this work, we consider the strong binding limit of self-assembly in well-mixed systems, which applies if the binding energy is large so that only the detachment of particles with a single bond from a cluster is relevant to the dynamics. On the other hand, detachment processes of particles with multiple bonds (higher-order detachment processes) can be neglected on the relevant timescales. We find that, in this regime, the morphology of the building blocks crucially determines the dynamic properties of the self-assembly processes, most importantly, their time efficiency. Three structural parameters essentially determine the self-assembly dynamics: the nucleation size and attachment order, which describe the effective order of reactions by which clusters nucleate and grow, respectively, and the growth exponent, which determines how the growth rate scales with the structure size. Importantly, we show that the kinetic rate equations exhibit an inherent scale invariance that allows us to derive scaling laws for the minimum required self-assembly time (time complexity exponent) and the optimal control parameter value (control parameter exponent) in dependence of the target structure size S. Both exponents are determined in terms of the three structural parameters alone. The time complexity exponent describes an important constraint for the time efficiency of self-assembly processes for large structure sizes. From its dependence on the structural parameters, it can be seen that particles with different morphologies can self-assemble into large structures at vastly different speeds.

Specifically, we find that hexagon-shaped monomers, which have an attachment order $\gamma = 1$, can self-assemble particularly fast and exhibit a low time complexity exponent. An attachment order $\gamma = 1$ means that clusters grow after nucleation by attachment of individual monomers (without having to go through a set of unstable states). This results in fast cluster growth and efficient self-assembly provided nucleation is sufficiently retarded. On the other hand, if the attachment order is larger than one (as in the case of square- and triangle-shaped monomers), i.e., if the growth processes are effectively few-particle reactions, a large nucleation size is favorable for the time complexity exponent to remain small. A large nucleation size, however, makes the time efficiency sensitive to variations in the control parameter and, thus, necessitates enhanced
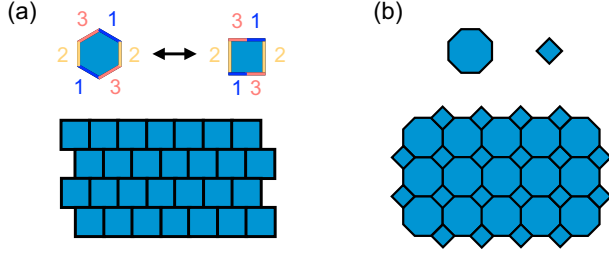
FIG. 8. Examples for morphologies that self-assemble efficiently in the strong binding limit. Squares with six instead of four binding sites form structures topologically equivalent to those formed by hexagons (cf. Fig. 1) (i.e., the squares in the finished structure can be smoothly deformed into hexagons without breaking any bonds). The same color and number indicate matching bonds. Hence, self-assembly of the squares is described by the same kinetic parameters ($\sigma = 3$, $\gamma = 1$, and $\omega \approx 1/2$) as for hexagons and exhibits the same minimal time complexity exponent $\theta \approx 1/2$. (b) Using two or more different types of constituents (here, octagons and squares) offers additional possibilities for designing efficient assembly schemes. For the example illustrated, again $\sigma = 3$, $\gamma = 1$, and $\omega \approx 1/2$ and, hence, $\theta \approx 1/2$ as for hexagons.

fine-tuning [cf. Eq. (31)]. Furthermore, the self-assembly time increases more strongly as a function of the target yield [cf. Eq. (32)].

We, thus, conclude that the particles' morphology strongly impacts the time efficiency of self-assembly processes in the strong binding limit. In particular, particles with an attachment order $\gamma = 1$ (like hexagon-shaped monomers) are found to self-assemble very efficiently. What do these results imply for nanotechnology and for our understanding of biology? In the following, we outline some ideas in this regard.

B. Increasing self-assembly efficiency by exploiting morphology effects

Self-assembly processes should be designed with specific characteristics to achieve both speed and resource efficiency. Our research indicates that particles that assemble with an attachment order $\gamma = 1$, such as hexagon-shaped monomers, possess a minimal time complexity exponent. This implies that these systems show rapid and efficient self-assembly of large structures.

More generally, one may now ask which other particle morphologies can be identified that have an attachment order $\gamma = 1$ and, thus, exhibit a minimal time complexity exponent. To give an idea for an answer, we first note that it is actually not only the shape of the monomers that determines the kinetically relevant parameters (like σ, γ, and ω) but that the concept of morphology also includes the number and positions of their binding sites. For example, the same kinetically relevant parameters as for hexagonal monomers can also be obtained with square-shaped monomers, provided each monomer has six instead of four
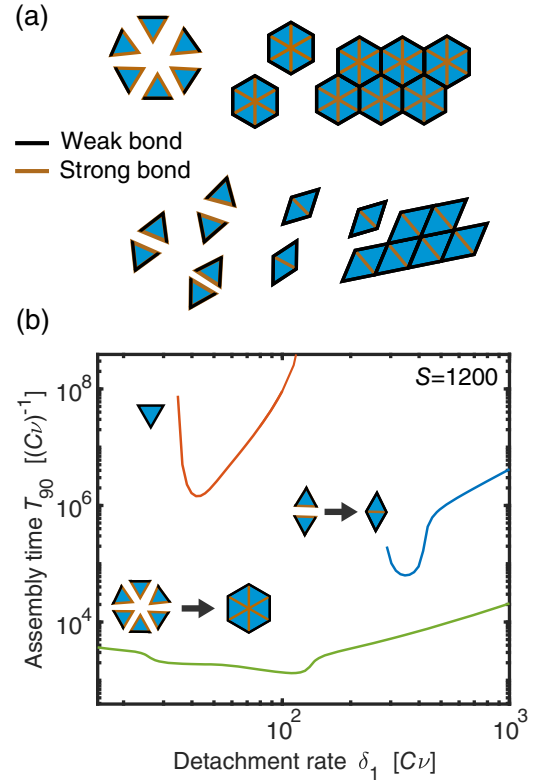

FIG. 9. Hierarchical self-assembly as a way to design optimized assembly schemes. Hierarchically organizing the assembly process can increase self-assembly efficiency significantly by optimizing the morphology of intermediate assembly products. (a) To induce triangle-shaped monomers to assemble hierarchically, bonds between edges indicated in brown are assumed to be more stable than bonds between black edges. As a result of this variation of the binding strengths, monomers first form hexamers (hexagons) or dimers (rhomboids), respectively, which subsequently assemble into the structures more efficiently than the original triangles; cf. Fig. 4. (b) The assembly time T_{90} is plotted against the detachment rate $\tilde{\delta}_1$ for the system with triangle-shaped monomers in comparison with the two hierarchical assembly schemes illustrated in (a) for a target structure of size $S = 1200$. Simulations are performed with the effective model (see Appendix B), assuming a constant ratio $r_h = 10^{-4}$ between the detachment rates of strong and weak bonds. Since, according to our findings, the self-assembly of particles with a hexagonal morphology is remarkably efficient, the assembly time can, thus, be reduced by 2–3 orders of magnitude over a broad range of the control parameter $\tilde{\delta}_1$.

binding sites, and the rows assemble "on gap" as shown in Fig. 8(a). The morphology of these squares is *topologically equivalent* to that of hexagons, meaning that, by deforming the building blocks while maintaining the adjacency relations with their neighbors, the resulting lattice of squares can be deformed into that of hexagons. Topologically equivalent morphologies, thus, have the same parameters σ, γ, and ω. Hence, morphologies that

are topologically equivalent to that of hexagons have the same minimal time complexity exponent.

A vast number of additional possibilities to achieve $\gamma = 1$ arises if two or more different kinds of building blocks are used to form the structures. One such example, in which structures are built from squares and octagons, is depicted in Fig. 8(b), but the number of possibilities here is sheer unlimited. Using particles with such morphologies can optimize the time and resource efficiency of their self-assembly processes.

Hierarchical self-assembly. Furthermore, the morphology effect can be used to optimize self-assembly experiments by designing highly time-efficient hierarchical self-assembly schemes. For instance, Fig. 9 shows two possibilities to significantly enhance the self-assembly efficiency of triangle-shaped monomers by inducing them to form higher-order constituents with a more favorable morphology through a hierarchical assembly step. Specifically, this is achieved by making certain bonds more stable than others [cf. Fig. 9(a)], enabling the monomers to first form squares or hexagons, which subsequently assemble into the final structures much more efficiently than the original triangles. We investigate this hierarchical scenario by modifying our effective kinetic theory (cf. Appendix B) and find that both a huge reduction in assembly time and increased robustness to parameter variations can be achieved. Specifically, for structures of size $S = 1200$, a reduction of the minimum assembly time by 3 orders of magnitude can be achieved, while at the same time the assembly time only weakly depends on the detachment rate (increasing robustness); see Fig. 9(b). Please note that the triangle-shaped building blocks do not assemble hierarchically *per se* but that they must be made to do so by discriminating the bond strengths along their edges. Therefore, hierarchical self-assembly provides a simple way to benefit from this morphology effect, thereby significantly increasing the efficiency and robustness of artificial self-assembly.

C. Virus and carboxysome capsid assembly

Our results might furthermore contribute to a better understanding of the kinetics of biological self-assembly processes. For example, several icosahedral virus capsids self-assemble from triangle-shaped capsid proteins [1,37], but the mechanisms underlying their self-assembly are still largely unknown. Our results suggest that a hierarchical step might be a possible strategy to allow large capsids to assemble time efficiently. However, the icosahedral symmetry of the spherical virus capsids poses an additional challenge to the self-assembly process: At precisely 12 sites, the capsid proteins must form pentamers, while everywhere else, they form hexamers. Only in this way can closed capsids form, whose geometry self-limits their own growth [32]. It has, therefore, been suggested that large virus capsids can assemble only in

the presence of a scaffold [43] that coordinates their formation. Thus, our results might not apply directly in this case as a consequence of the scaffolding. Otherwise, assuming that icosahedral shells assemble without a scaffold by forming a pentameric nucleus with nucleation size $\sigma = 5$, Eq. (17) would predict a time complexity exponent of $\theta = 2$, implying that self-assembly of large icosahedral capsids would be even significantly more inefficient than self-assembly of the planar triangle-shaped system ($\theta = 1$). Hence, besides structural aspects, enhancement of time efficiency could be another essential reason for the requirement of a scaffold in forming large icosahedral viruses.

In contrast, alpha-carboxysomes, which also form large icosahedral structures, are shown to assemble also in the absence of a scaffold and are composed of proteins with hexameric and pentameric quaternary structure (CsoS1ABC and CsoS4AB) [38]. Hence, carboxysome capsids represent the icosahedral analog of the hexagonal particle system studied here, being characterized by the same parameters σ, γ, and ω. This particular morphology of the constituents could be an essential factor for such huge structures to assemble efficiently even in the absence of a scaffold. Similar hierarchical assembly pathways, in which triangle-shaped capsid proteins first form pentameric and hexameric constituents, have also been confirmed for certain classes of smaller viruses [77], like Picorna viruses [78], the Brome Mosaic Virus [79], human Papillomavirus [80] or Cowpea Chlorotic Mottle Virus [81], all of which are also known to assemble without a scaffold. These insights might be particularly relevant for experiments that try to mimic the self-assembly of artificial capsids for biotechnological applications [6]. Taken together, this suggests that natural self-assembly processes have been optimized concerning their time efficiency by exploiting the dependence on the morphology of the building blocks. It is plausible that these adaptations occurred during evolution, since, according to our analysis, relatively small modifications (like alterations of individual bond strengths) could be enough to reduce the assembly time and increase the robustness of the self-assembly process significantly.

D. Conclusion and outlook

Our study suggests that understanding the kinetic properties of self-organizing processes is crucial for properly understanding these phenomena in biology and for the experimental realization of efficient self-organizing systems. Specifying structural determinants and conditions under which high yield is achieved in self-assembly is probably insufficient, because the time it takes could be extremely long.

Specifically, we have demonstrated that the time required to achieve high yield for large target structures can vary by several orders of magnitude depending on the

morphology of the building blocks. Therefore, in this work, we described a general mathematical framework that allows us to analyze and understand the kinetics of self-assembly processes in well-mixed systems better. In particular, we found that the morphology of the constituents is an essential kinetic determinant that strongly impacts self-assembly efficiency. We discussed how these insights could be used in nanotechnology to significantly enhance the efficiency and robustness of artificial self-assembly. Furthermore, we showed how this might impact our understanding of biological self-assembly phenomena.

In this paper, we considered the strong binding limit of self-assembly, which applies if the binding energy between the constituents is large. In a subsequent paper, we will extend our analysis to systems with low binding energy and investigate how higher-order detachment processes affect the assembly dynamics.

In our model and analysis, we assumed a well-mixed self-assembly system in which monomers from a dilute solution bind to sparse clusters. As discussed in the context of virus capsid assembly, however, some systems first form dense disordered aggregates (e.g., via phase separation or by binding to a scaffold like DNA) [42,43,45]. The disordered aggregates subsequently transit to an ordered state. Our well-mixed model does not describe such a scenario, and, hence, our time complexity results might not apply in this case. Importantly, such a spatial accumulation could significantly enhance self-assembly efficiency by increasing the local concentration of resources. Unspecific binding and local aggregation of the monomers might, therefore, crucially influence the time efficiency in spatial systems and could undermine the role of the monomer morphology. An interesting question to ask is whether the transition from the disordered to the ordered state can be described by a similar formalism as discussed in this manuscript, and, independent of that, what the time complexity of such spatial systems will be. In particular, it would be interesting to see how the time complexity of spatial systems compares to the time complexity of the well-mixed systems considered here.

Furthermore, in the current model, we disregard any possibilities of misbinding or assembly errors due to defects. The results of grand canonical ensemble simulations [51,52] suggest that systems with fast cluster growth tend to be more prone to defects and assembly errors. It would, therefore, be interesting to add reasonable error rates to the dynamic model and see whether this leads to additional constraints regarding the optimal morphology of the subunits. Finally, it would be interesting to see whether our theory can be extended to the case of an open system, i.e., when the total number of particles is not fixed but increases through a steady influx of monomers [82], and how this affects the kinetic behavior of the system.

APPENDIX A: STOCHASTIC SIMULATION

The simulations are performed using Gillespie's stochastic algorithm [61]. In the simulation, complexes are represented as boolean arrays of size S, which contain ones at sites occupied by a particle and zeros otherwise. For simplicity, one can imagine the structures as two-dimensional arrays. However, internally, all data structures are represented as one-dimensional memory sequences, and two-dimensional subscript indices must be converted to linear indices. The neighboring or adjacency relations between the sites in the two-dimensional array thereby determine the morphology and are defined as follows: In the square system, site (i, j) is adjacent to sites $(i \pm 1 \bmod L, j)$ and $(i, j \pm 1 \bmod L)$, where $L = \sqrt{S}$ is the linear extension of the target structure and by taking the modulo periodic boundary conditions are implied. Accordingly, in the hexagonal system, site (i, j) is adjacent to the six sites $(i \pm 1 \bmod L, j)$, $(i, j \pm 1 \bmod L)$, and $(i \pm 1 \bmod L, j \mp 1 \bmod L)$, and in the triangle-shaped system, site (i, j) has three neighbors: $(i \pm 1 \bmod L, j)$ and $(i, j + 1 \bmod L)$ if j is even or $(i, j - 1 \bmod L)$ if j is odd [see Fig. 1(a) for the logic behind these definitions].

When a dimer forms, such a boolean array is reserved for the complex, and two arbitrary neighboring sites are chosen and set to 1 to represent the initial dimer. Because of the periodic boundary conditions, which two neighboring positions are chosen is irrelevant, because the structure is translationally invariant. Subsequently, each unoccupied site in the complex with at least one occupied neighbor is occupied by a monomer at rate νn, where n is the number of monomers in the system (monomer attachment). The reaction rate ν is typically set to 1. Similarly, an occupied site becomes empty again with rate δ_i, where i denotes the number of occupied neighboring sites (monomer detachment). Each attachment (detachment) event decreases (increases) the number n of monomers in the system. In this way, the simulation respects all possible configurations of clusters that can emerge. By counting the number M of complete structures, i.e., structures with S occupied sites, the yield is calculated as yield $= (MS/N)$.
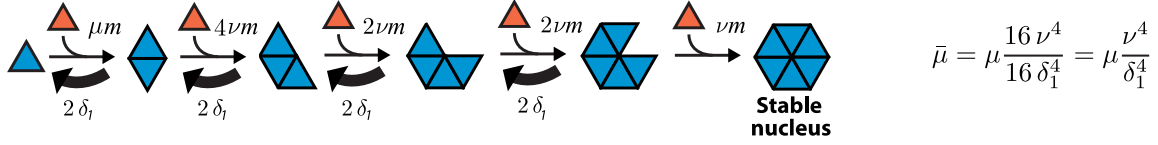
FIG. 10. Illustration of the nucleation process for triangle-shaped monomers. Starting from a monomer, the nucleus is formed by subsequent attachment of monomers with growth rates $g_i \sim \nu m$ indicated in the graph. The detachment of monomers counteracts this with the shrinkage (detachment) rate $d_i = 2\delta_1$. The prefactors account for the degeneracy of attachment or detachment reactions leading to equivalent configurations. In the limiting case of large detachment rates, the effective nucleation rate $\bar{\mu}$ is given by the product of the forward rates divided by the product of the reverse rates.

It is important to optimize the code for efficiency, because, since the detachment rate δ_1 is typically much larger than the reaction rate $N\nu$, many Gillespie steps are generally needed until a yield of 90% is reached even for intermediate particle numbers N (typically, we simulate the system with N between $100S$ and $1000S$ so that a maximum number of 100–1000 target structures is built). In particular, the simulation of the triangle-shaped system is computationally expensive due to the comparatively large number of intermediate steps between two stable configurations with the triangle-shaped morphology and the longer time spans required to be simulated to reach 90% yield. Partly, several billion Gillespie steps are needed for a single run to complete. The computational cost of the simulation strongly increases with the size of the target structure, because both the detachment rate δ_1 and the required simulation time T_{90} increase with S. Furthermore, the particle number N should typically be increased with S to keep the number of assembled structures constant. Hence, to simulate also large system sizes up to a size of $S = 1000$, the efficiency of the simulation is crucial. By associating additional data structures with the complexes that allow us to choose and update attachment and detachment events efficiently, our simulation written in C++ is able to perform more than one million Gillespie steps per second on a 3,1 GHz CPU. The C++ code of the simulation is available online.

APPENDIX B: EFFECTIVE KINETIC THEORY IN THE STRONG BINDING LIMIT

In this appendix, we show how, in the strong binding limit, the self-assembly dynamics can be approximately described by effective rate equations for the cluster size distribution $c_s(t)$; cf. Eq. (4). Our analysis assumes that the dynamics can be reduced to the most probable self-assembly paths, i.e., those reaction paths that traverse the minimum number of unstable configurations to move from one stable cluster configuration to the next.

1. Effective nucleation process

The self-assembly process starts with the formation of a stable nucleus, which corresponds to the formation of a cluster of σ monomers, as illustrated in Fig. 10 for triangle-shaped monomers. This stochastic process is a one-step Markov chain with σ states and the stable nucleus as the absorbing state. The effective forward (growth) rates g_i and backward (shrinkage) rates d_i are indicated in Fig. 10 for triangle-shaped and in Fig. 5 for hexagon-shaped monomers. Let $p_i(t)$ be the probability for the Markov chain to be in a state with i monomers. Then the set of master equations for all states $1 \leq i < \sigma$ reads

$$\partial_t p_i(t) = g_{i-1}p_{i-1} - d_i p_i - g_i p_i + d_{i+1}p_{i+1}, \quad \text{(B1)}$$

where $g_0 = 0$ and $d_1 = 0$ since $i = 1$ is the initial state. For the absorbing state, one has

$$\partial_t p_\sigma = g_{\sigma-1}p_{\sigma-1}. \quad \text{(B2)}$$

In the limiting case, where the decay rates d_i are large compared to the growth rates g_i, Eq. (B1) can be assumed to be stationary ($\partial_t p_i = 0$), leading to

$$p_{\sigma-1} = \frac{g_1 g_2 \ldots g_{\sigma-2}}{d_2 d_3 \ldots d_{\sigma-1}} p_1. \quad \text{(B3)}$$

This results in an effective equation for the nucleation process $1 \to \sigma$:

$$\partial_t p_\sigma(t) = r_{1 \to \sigma} p_1(t) \quad \text{(B4)}$$

with the effective nucleation rate given by the product of the forward (growth) and backward (shrinkage) rates

$$r_{1 \to \sigma} = \frac{g_1 g_2 \ldots g_{\sigma-1}}{d_2 d_3 \ldots d_{\sigma-1}}. \quad \text{(B5)}$$

For the three different monomer morphologies studied here, we find that the numerical prefactors cancel (compare Fig. 10), and one obtains

$$r_{1 \to \sigma} = \bar{\mu} m^\sigma, \quad \text{(B6)}$$

with the effective kinetic parameter $\bar{\mu}$ for the nucleation process of the order of σ given by

$$\bar{\mu} = \mu \left(\frac{\nu}{\delta_1}\right)^{\sigma-2}. \quad \text{(B7)}$$

Henceforth, we refer to this kinetic parameter simply as an *effective nucleation rate*.

2. Effective assembly processes after nucleation

As discussed in Sec. IV, there are typically two different types of processes by which clusters grow: the formation of a seed of size γ and the subsequent growth of the seed by a series of lower-order reactions. We refer to these processes as leading and subleading processes, respectively. For illustration, refer to Fig. 12, where a representative example of an assembly process for square-shaped building blocks in the early stages of cluster growth is shown.

A *leading-order* (seed formation) process requiring the sequential attachment of γ monomers typically occurs for cluster configuration with "completed (or smooth)" edges, i.e., the rectangles shown in Fig. 12. Using the same reasoning in terms of a Markov chain as for the nucleation process, the effective rate for seed formation reads

$$r_{s \to s+\gamma} = p_s q_s \nu \left(\frac{\nu}{\delta_1}\right)^{\gamma-1} m^\gamma \equiv p_s q_s \bar{\nu} m^\gamma, \quad \text{(B8)}$$

where p_s denotes the perimeter, or rather the number of available binding sites, of a cluster of size s. The additional factor q_s is a combinatorial factor that indicates the number of equivalent ways through which the remaining $\gamma - 1$ monomers can complete the seed after the first monomer has been attached. For example, in the case of square-shaped monomers, the second particle forming the seed can attach to either side of the first one unless the first particle attaches directly to a corner. Hence, for large clusters, one has $q_s = 2$. For smaller clusters, however, the average value of q is smaller; in the case of a nucleus, it is $q_\sigma = 1$. For triangle-shaped monomers, the equivalent assembly pathways by which a seed can be formed are illustrated in Fig. 11 for both small clusters (nuclei) and larger clusters. Each reaction pathway contributes with a weight that is calculated by Eq. (B5), and the sum of these weighted pathways (times a factor of 2, since the seed can grow to both sides) yields the combinatorial factors $q_\sigma = 4$ and $q_s = 3$, respectively. The value of the perimeter p_s generally depends on the cluster's shape. For simplicity, we generally assume that clusters have a regular shape. For example, in the case of square-shaped monomers, assuming that a typical cluster has a quadratic shape, we obtain $p_s = 4\sqrt{s}$. However, only the scaling of the multiplicity factor $p_s q_s \sim \sqrt{s}$ for typical large clusters is relevant for the following analysis.

These *slow* processes of leading order (seed formation along the edges) are usually followed by a sequence of *fast* processes of lower order, referred to as a domino effect; cf. Figs. 1(c) and 6. These processes typically fill up the entire row that is initiated by the seed and continue until the process can proceed only by another reaction of leading
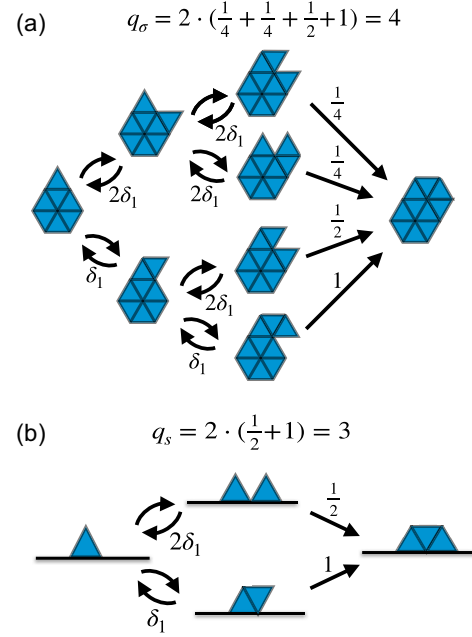

FIG. 11. Derivation of the multiplicity factor q_s for seed formation for (a) nuclei and (b) larger clusters with triangle-shaped monomers. The diagram depicts all equivalent assembly pathways by which a seed can be completed after the first seed-forming monomer has attached. The rate for all forward steps in the diagram is νm, while the rates for the backward steps are given by δ_1 or $2\delta_1$, depending on whether there are one or two monomers with a single bond. The total weight of each path (indicated above the rightmost arrows) is calculated by multiplying all backward rates along the respective path, i.e., by applying Eq. (B5) to the path. The multiplicity factor q_s corresponds to the sum of the weights of all contributing paths. Note that q_s has an additional factor of 2, because one must additionally account for the analogous pathways in which the seed forms to the left-hand side of the first monomer instead of to the right.

order γ. In the case of square-shaped monomers, these lower-order reactions have order 1 and happen at a rate $2\nu m$ or νm, depending on whether they can extend the seed in both directions or only in one (cf. Fig. 12). Analogously, in the case of triangle-shaped monomers, the effective order of these reactions is two, and the effective rates are given by $4\nu(\nu/\delta_1)m^2$ or $2\nu(\nu/\delta_1)m^2$, respectively, with an additional factor of 2 accounting for the two distinct but equivalent ways of adding two monomers to the seed. Since we assume $\delta_1 \gg \nu m$ (strong binding limit), the corresponding reaction rate for both morphologies of the building blocks is much larger than that of the leading-order reaction [cf. Eq. (B8)], which, therefore, defines the rate-limiting steps.

The total number of monomers attaching by a leading-order process and the subsequent domino effect is $l_s \sim \sqrt{s}$, where l_s is the typical edge length of a cluster of size s [83]. Since the seed formation process is rate limiting, the effective total rate between two leading-order events approximately equals the rate of seed formation
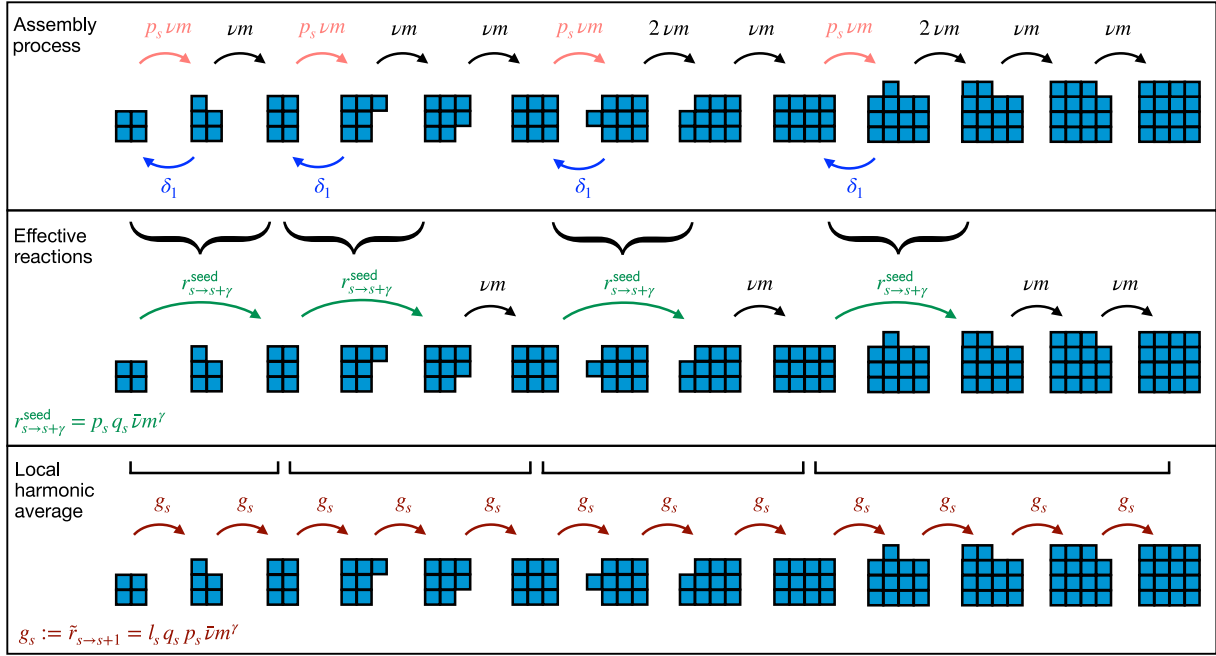
FIG. 12.   Effective assembly process in the strong binding limit, illustrated for square-shaped monomers. Top: typical assembly path leading from a stable nucleus of size $\sigma = 4$ to a structure of size $s = 16$, with the corresponding transition rates between the states, indicated in the graph (assuming $\delta_n = 0$ for $n \leq 2$). The prefactor p_s denotes the average perimeter (number of binding sites for the seed) of a cluster of size s. The effective theory is derived in two steps illustrated by the following. Middle: First, we describe transitions passing through unstable states as effective few-particle reactions of the order of γ (here, for square-shaped monomers, $\gamma = 2$) with rate $r_{s\to s+\gamma}^{\text{seed}}$. This is possible, since the backward rate δ_1 is large compared to the forward rates (cf. Appendix B). This way, we "integrate out" the fast detachment rate δ_1 from the process. Bottom: In the second step, we introduce one-step forward rates g_s, which are continuous functions of the structure size s. These average one-step rates are calculated by dividing the number of particles added per row by the total time it takes to add the entire row. In other words, this corresponds to taking a moving harmonic average of the rates of the leading-order reactions and the subsequent subleading-order reactions as indicated by the square brackets. Since the leading-order reactions are significantly slower than the subleading-order reactions, the latter can be assumed as infinitely fast. In this way, we derive a one-step Poisson process with transition rates that vary continuously as a function of the cluster size. This enables a convenient mathematical analysis and allows us to describe systems with different particle morphologies on equal footing.

$$r_{s\to s+l_s} \approx r_{s\to s+\gamma}^{\text{seed}}. \qquad (B9)$$

The alternating sequence of leading- and subleading-order processes is illustrated in the middle in Fig. 12. To arrive at a set of rate equations that have the same form for all the different morphologies of the monomers, we reduce them to an effective average growth rate g_s at which a cluster grows by one monomer unit:

$$g_s := \tilde{r}_{s\to s+1} = l_s \cdot p_s q_s \bar{\nu} m^\gamma. \qquad (B10)$$

Assuming that both the typical perimeter p_s and edge length l_s of a cluster scale as $\sqrt{s}$ for $s \gg 1$, one finds for large clusters

$$g_s = f_s \bar{\nu} m^\gamma, \qquad (B11)$$

with $f_s = l_s p_s q_s := as$. The factor a is a numerical prefactor that can be estimated (see below) by considering a typical large cluster or determined by fitting the assembly

time to data from the stochastic simulation for a fixed target structure size S.

For assembly processes where the attachment order is $\gamma = 1$, like in the case of hexagon-shaped monomers, there are no subleading processes, and, hence, the combinatorial factor f_s scales as $f_s \sim p_s \sim \sqrt{s}$.

Taking the results for all monomer morphologies together, we assume that the asymptotic form of the combinatorial factor can be written as

$$f_s = as^\omega, \qquad (B12)$$

where we call ω the growth exponent, which is $\omega = 1/2$ for $\gamma = 1$ (hexagonal building blocks) and $\omega = 1$ for $\gamma > 1$ (triangular and square building blocks).

3. Numerical prefactor of the multiplicity factor

The numerical prefactor a in the multiplicity factor f_s of the growth rate, Eq. (B12), affects the minimal assembly time and the optimal detachment rate. Still, it does not
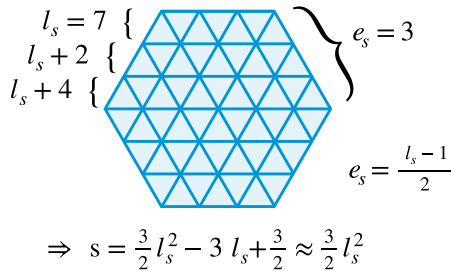
TABLE II. Table showing the multiplicity factors q_s for large cluster sizes s, as well as the theoretically estimated and fitted prefactors a_{th} and a_{fit}, respectively, for the different particle morphologies. The theoretical values a_{th} are estimated assuming a large cluster in the form of a square (square-shaped monomers) or a regular hexagon (triangle-shaped monomers). Small clusters typically have smaller values for q_s and a, which explains why the best fit is obtained with a smaller value a_{fit} compared to a_{th}. Clusters assembling from hexagon-shaped monomers have a rough interface, whereby estimating the prefactor a (and the growth exponent ω) in this case is more involved.

	Square	Triangle	Hexagon
q_s	2	3	1
a_{th}	8	6	$\cdots$
a_{fit}	5.3	4	2.3

affect their scaling, as shown in the next section. We determine the corresponding values of a by comparing the optimal replacement rate predicted by theory with stochastic simulation for target structures of size $S = 100$; see Table II, where the fitted values of a are compared with their theoretical estimates. Alternatively, the parameter a can be estimated by assuming typical cluster morphologies, as we briefly discuss in the following.

For systems with square-shaped monomers, assuming that typical clusters have the shape of a square, the cluster boundary is $p_s = 4\sqrt{s}$, and the number of particles attached per attachment sequence corresponds to one-quarter of the boundary, $l_s = \sqrt{s}$, implying $a = 8$ and $\omega = 1$ for large clusters. The best fit is obtained with a slightly smaller prefactor of $a = 5.3$, probably because small clusters have a smaller value $a \approx 4$ since $q_s \approx 1$ if s is small.

In the case of triangle-shaped monomers, clusters typically have a hexagonal morphology. Assuming a cluster has a regular hexagonal shape, its edge length e_s is given by $e_s = \sqrt{s/6}$; cf. Fig. 13. The perimeter is $p_s = 6e_s$, and

$$l_s = 7 \ \{$$
$$l_s + 2 \ \{$$
$$l_s + 4 \ \{$$
$$e_s = 3$$
$$e_s = \frac{l_s - 1}{2}$$
$$\Rightarrow \quad s = \frac{3}{2}l_s^2 - 3\,l_s + \frac{3}{2} \approx \frac{3}{2}l_s^2$$

FIG. 13. Regular hexagonal cluster consisting of $s = 54$ triangular building blocks to illustrate the relation between e_s (number of binding sites per edge), l_s (number of particles forming an edge), and the cluster size s. The cluster size s equates to $s = 2\sum_{k=0}^{e_s-1}(l_s + 2k)$, which can be calculated to yield the relation shown in the figure. In the limit of large s, one thus finds $s \approx \frac{3}{2}l_s^2$.

the total number of particles attached in an attachment sequence (domino effect) is $l_s = 2e_s + 1$ (because an edge of length e consists of $2e_s + 1$ triangles; see Fig. 13). Thus, $p_s \cdot l_s \approx 2s$, and together with $q_s = 3$ we obtain $a \approx 6$ for large clusters. In contrast, for small clusters, where the seed tends to form at the corner of an edge, one has $q_s \approx 1.5$ and, thus, approximately $a \approx 3$. The best fit between the estimated values for large and small clusters is obtained for $a = 4$.

There are no lower-order reactions for systems with hexagon-shaped monomers, because $\gamma = 1$. Consequently, the clusters typically have a rough interface. This makes the calculation of a (and the growth exponent ω) more involved than in the other two cases, and we refrain from making an estimate. The best numerical fit we obtain with a relatively small prefactor $a = 2.3$, which probably accounts for the fact that, to leading order, monomers bind only to sites at which they have two neighbors, which amounts to only a fraction of all boundary sites.

4. Assembly following nucleation for triangle-shaped monomers

For hexagon-shaped and square-shaped monomers, we use the asymptotic expression $f_s = as^\omega$ with the above numerically estimated values to compare the results of the effective kinetic theory with the stochastic simulations. However, the system with triangle-shaped monomers is special, because the first binding reaction after the formation of the nucleus has a higher attachment order $\gamma = 4$ than all subsequent binding reactions, which have $\gamma = 3$ [cf. Fig. 1(a)]. As it turns out, the fit to the stochastic simulation is significantly improved (in the region around the minimum at $\tilde{\delta}_1^{opt}$) if the higher order of the initial attachment reaction is taken into account. Specifically, by accounting for all equivalent assembly pathways that lead from the nucleus to the next stable configuration, the effective rate of the initial attachment reaction is found to be

$$r_{\sigma \to \sigma+4} = 24\bar{\nu}'c_\sigma m^4 \tag{B13}$$

with

$$\bar{\nu}' = \nu\left(\frac{\nu}{\delta_1}\right)^3. \tag{B14}$$

Consequently, the corresponding effective attachment rate per monomer for $\sigma \leq s \leq \sigma + 3$ is given by

$$g_s = \tilde{r}_{s \to s+1} = 96\bar{\nu}'c_s m^4. \tag{B15}$$

We modify the effective theory Eq. (4) for the system with triangle-shaped monomers accordingly to account for the higher order of the initial attachment reaction.

5. Numerical solution of the kinetic rate equations

To numerically solve the resulting kinetic rate equations given by Eq. (4), we use Matlab's ode15s, a semi-implicit solver designed for solving stiff problems. Typically, semi-implicit ODE solvers allow one to provide the Jacobian matrix or a sparsity pattern of the Jacobian for the specific problem. The sparsity pattern indicates only the nonzero elements of the Jacobian without specifying them. For the given problem, Eq. (4), the sparsity pattern is easily found by observing that the temporal evolution of c_s depends only on itself, c_{s-1}, and the monomer concentration $m = c_1$. Hence, the sparsity pattern is given by a $S \times S$ matrix with ones only in its diagonal and its subdiagonal, as well as its first column and first row, while all other entries are zero. By providing the sparsity pattern for the Jacobian, the performance of the solver could be accelerated significantly, especially for large structure sizes S. In this way, Eq. (4) can be solved quickly despite the very small absolute and relative error tolerance of 10^{-12} that we use. The speed could be further increased by providing an analytic Jacobian, which is unnecessary in this case.

6. Growth rate for tubelike structures

In Sec. V F, we investigate the growth of two-dimensional tubes (cylinders) with a constant width (circumference) of six monomers and a variable height of $S/6$ monomers. Hence, in contrast to the case of symmetric target structures, the tubes have one dimension (the width) that is significantly shorter than the other one (the height). Thus, as soon as the structures close in the short dimension, they continue growing with a constant growth rate, since the structure's perimeter remains constant. To adopt the effective model accordingly, we determine the tubes' constant growth rate g_s. To this end, we note that, in the case of square and hexagonal building blocks, the number of possible binding sites p_s for the first monomer to form a new seed is $p_s = 2 \cdot 6$, since the monomer can attach to either the upper or the lower edge of the tube. In contrast, in the case of triangular monomers, $p_s = 2 \cdot 3$, since a row of width $l_s = 6$ monomers has only three binding sites. The rate of seed formation is then given by Eq. (B8) with the multiplicity factors q_s shown in Table II. Because of the domino effect following seed formation, $l_s = 6$ monomers in total are added as the result of a single seed-forming event in the case of square or triangular monomers ($l_s = 1$ for hexagonal monomers due to $\gamma = 1$). Hence, with Eq. (B10), we obtain the constant effective growth rate $g_s = a\bar{\nu}m^\gamma$, with $a = 144$, 108, and 12 for square, triangular, and hexagonal monomers, respectively. For simplicity, we assume that, up to size $s = 30$, clusters grow with the same (size-dependent) growth rate g_s as symmetric structures and for $s > 30$ with the constant growth rate derived above. The effective model modified this way accurately

matches the simulation data for tubelike structures, as shown in Fig. 7.

7. Interaction of oligomers

Equation (4) describes the basic reaction kinetics of reversible self-assembly. The basic theory can easily be extended by accounting for additional effects. For example, in Appendix E, we investigate how reactions between larger oligomers might affect the assembly time [see Fig. 15(b)]. To estimate this effect with the help of the effective theory, we additionally account for the reactions of any two clusters of size $i \geq \sigma$ and $j \geq \sigma$ with $i + j \leq S$ to a cluster of size $i + j$ at rate ν:

$$r_{(i,j) \to i+j} = \nu c_i c_j. \tag{B16}$$

To this end, we augment the right-hand side of Eq. (4) by the following terms:

$$\partial_t c_s = \cdots + \frac{1}{2}\nu \sum_{\substack{i,j \geq \sigma \\ i+j=s}} c_i c_j - \nu c_s \sum_{\substack{i \geq \sigma \\ i+s \leq S}} c_i, \tag{B17}$$

for all $\sigma \leq s \leq S$, where the factor $1/2$ in front of the first sum avoids double counting and serves as a stoichiometric factor in the case $i = j$.

8. Hierarchical self-assembly system

Furthermore, we use the effective theory to simulate the hierarchical self-assembly scenario in Sec. VI (cf. Fig. 9). To this end, we formally distinguish between weak and strong bonds of the monomers. This causes the monomers to first form either dimeric rhombuses or hexameric hexagons, as shown in Fig. 9.

In the first case, we simulate the irreversible dimerization of two (triangle-shaped) monomers via strong bonds into rhomboids at rate $\mu = \nu$. In the second case, we consider the formation of hexagons by explicitly simulating the nucleation process of triangle-shaped monomers depicted in Fig. 10 with detachment rate $\delta_{str} = 10^{-4} \cdot \delta_1$. The rhomboids and hexagons, once formed, are stable and react in the usual way as described by Eq. (4) mediated by the weak bonds. To simulate the assembly of these higher-order particles, we use Eq. (4) together with the morphological parameters for square- or hexagon-shaped particles, respectively, and the corresponding detachment rate for the weak bonds.

Interactions between (triangle-shaped) monomers and incomplete hexagons with larger oligomers are entirely neglected, which is justified by the time interval between the reactions mediated by strong and weak bonds. In the hexagonal case, the rate of detachment δ_{str} of the strong bonds must be neither too large nor too small relative to δ_1: In both cases, the formation of the hexagons would be slow, leading to an increase in the total buildup time. We find that

a ratio of $r_h := \delta_{\text{str}}/\delta_1 \approx 10^{-4}$ between the detachment rates of the strong and weak bonds minimizes about T_{90}.

APPENDIX C: SCALING THEORY FOR YIELD ONSET

Here, we analyze the solutions of the rate equations in the strong binding limit, Eq. (4), in a parameter regime where the assembly process has a low yield. The solutions are shown to exhibit scale invariance. Using this scale invariance, we can derive a scaling law for the parameter condition for yield onset, i.e., the condition to obtain a nonzero yield, in dependence on the target structure size. The same scaling law also describes how the optimal parameter value scales as a function of the target structure size, which we show rigorously in Appendix D.

1. Effective dynamics of monomers and clusters

Using the asymptotic form $f_s = as^\omega$ for the combinatorial factor, one can rewrite the sum on the right-hand side of the rate equations, Eq. (4a), governing the evolution of the concentration of monomers, as

$$\sum_{s=\sigma}^{S} s^\omega c_s = K \sum_{s=\sigma}^{S} s^\omega \frac{c_s}{K} = K\langle s^\omega \rangle. \quad (C1)$$

Here, $K := \sum_{s=\sigma}^{S} c_s$ denotes the total concentration of all complexes above the nucleation size, and $\langle s^\omega \rangle$ defines the ω moment of the distribution of cluster sizes. For $\omega = 1$, which holds for square- and triangle-shaped monomers, $\langle s^\omega \rangle$ reduces to the average cluster size $\langle s \rangle$. As long as the yield is zero, the average cluster size equals the total number of bound monomers divided by the total number of clusters: $\langle s \rangle = (C - m)/K$. For general ω, there is no such exact identity, but, to make progress, we approximate the ω moment by $\langle s^\omega \rangle \approx \langle s \rangle^\omega$. Since ω is usually close to 1 (e.g., for the hexagon-shaped monomers $\omega \approx 1/2$), we expect the approximation to be generally sufficiently accurate. Hence, as long as the yield is zero, the dynamics of the monomer concentration are governed by [cf. Eq. (4a)]

$$\partial_t m(t) = -a\bar{\nu} m^\gamma (C - m)^\omega K^{1-\omega} - \sigma\bar{\mu} m^\sigma. \quad (C2)$$

Moreover, the total concentration of complexes K increases by nucleation events and, thus, follows the equation

$$\partial_t K(t) = \bar{\mu} m^\sigma. \quad (C3)$$

Given initial conditions for both the concentrations of monomers m and complexes K, Eqs. (C2) and (C3) form a closed system of differential equations from which the number of particles that remain in the monomer pool and the number of complexes can be calculated.

2. Continuum limit of cluster assembly

However, to determine the condition under which nonzero yield is achieved in the self-assembly process, we need to know how the distribution of cluster sizes evolves: Nonzero yield is achieved if the outermost front of the cluster size distribution eventually reaches the target structure size S, i.e., $\lim_{t\to\infty} c_S(t) \neq 0$. To determine a condition for the onset of the yield, we rewrite Eq. (4c) in a continuum limit by considering $c(s) = c_s$ as a continuous function of the cluster size s. Since the cluster size s is typically large, we approximate $(s-1)^\omega \approx s^\omega$ and then use a second-order Taylor expansion $c_{s-1} \approx c(s) - \partial_s c(s) + \frac{1}{2}\partial_s^2 c(s)$ to arrive at

$$\partial_t c(s,t) = -a\bar{\nu} s^\omega m^\gamma \partial_s c(s,t) + \frac{1}{2} a\bar{\nu} s^\omega m^\gamma \partial_s^2 c(s,t). \quad (C4)$$

This constitutes an advection-diffusion equation for the time-dependent cluster size distribution $c(s,t)$ with coefficients depending on the cluster size s and the instantaneous monomer concentration $m(t)$. The drift velocity reads $v(s) = a\bar{\nu} s^\omega m^\gamma$, and the diffusion coefficient $D(s) = \frac{1}{2} a\bar{\nu} s^\omega m^\gamma$. A similar approach of approximating the system of rate equations by a continuous advection-diffusion equation has been used previously to describe virus capsid assembly [19,68,69].

Equation (C4) can be interpreted as a Fokker-Planck equation for $c(s,t)$. The corresponding jump moments for small time increments Δt are given by [note that $D(s) = \frac{1}{2}v(s)$]

$$\langle \Delta s \rangle = v(s)\Delta t,$$
$$\langle (\Delta s)^2 \rangle = v(s)\Delta t, \quad (C5)$$

where the average is taken over the cluster size distribution. Hence, the Fano factor (variance-to-mean ratio) $F = 1$; i.e., fluctuation effects are fairly large. Here, as we are interested in the asymptotic scaling for large target sizes S, we neglect the diffusive contribution against the advective contribution. However, as it turns out and is expected from the value of the Fano factor, the diffusive contribution is one of the main reasons why the yield and assembly time for small structure sizes deviate from their asymptotic scaling. Considering only the advective contribution, i.e., the average time evolution of the cluster distribution, and assuming that several complexes nucleate at $s = \sigma$ at time $t = 0$, the leading edge of the density profile evolves according to

$$\partial_t s(t) = v[s(t)] = a\bar{\nu} s(t)^\omega m(t)^\gamma, \quad (C6)$$

with initial condition $s(0) = \sigma$. In this advective limit, the onset condition for the existence of a finite yield reads $\lim_{t\to\infty} s(t) = S$; i.e., to obtain a finite yield, the leading

edge of the cluster size distribution must eventually reach the target structure size S.

In summary, Eqs. (C2), (C3), and (C6) constitute a closed set of equations that allows us to compute the onset of yield by the condition $S = s(\infty)$.

3. Analytic solution

We are able to find a closed analytic solution only for systems with square-shaped monomers ($\sigma = 4$, $\omega = 1$, and $\gamma = 2$) using elementary calculus. Specifically, the complex concentration K drops out of Eq. (C2) if $\omega = 1$, which can, therefore, be solved for the monomer concentration m. Using m in Eq. (C6) then gives the solution

$$s(\infty) = \sigma \frac{1 + \sqrt{1 - 4\tilde{\eta}}}{1 - \sqrt{1 - 4\tilde{\eta}}} \tag{C7}$$

with $\tilde{\eta} := \sigma \bar{\mu} C / (a \bar{\nu})$, which, for large S (and $\eta \ll 1$), implies

$$\frac{\delta_1^{\mathrm{on}}}{C\nu} \sim \frac{\mu}{\nu} S. \tag{C8}$$

4. Scaling analysis

To determine the asymptotic scaling exponents for general cluster morphologies, however, finding an explicit solution to the above equations is not necessary. It is sufficient to demonstrate that, in the limit of large target, structure sizes S, the equations obey a scale invariance, which can be exploited to determine the exponents without finding explicit solutions.

To reveal this scale invariance, we first note that, in the limit of large structure sizes, the nucleation term can be neglected in the dynamics of the monomer concentration Eq. (C2), since nucleation events are rare compared to the attachment of monomers to existing clusters. The rate equations can then be written in dimensionless form by measuring concentrations in units of C and time in units of the effective timescale $(a\bar{\nu}C^\gamma)^{-1}$. Specifically, with the variable transformations $m \to mC$, $K \to KC$, and $t \to t/(a\bar{\nu}C^\gamma)$, one obtains

$$\partial_t m = -m^\gamma (1 - m)^\omega K^{1-\omega}, \tag{C9a}$$

$$\partial_t K = \eta m^\sigma, \tag{C9b}$$

$$\partial_t s = s^\omega m^\gamma, \tag{C9c}$$

where we introduce the dimensionless parameter

$$\eta := \frac{\bar{\mu}}{a\bar{\nu}} C^{\sigma - \gamma - 1}. \tag{C10}$$

Using a scaling ansatz

$$m = \tilde{m}(\eta^z t), \tag{C11a}$$

$$K = \eta^x \tilde{K}(\eta^z t), \tag{C11b}$$

$$s = \eta^y \tilde{s}(\eta^z t), \tag{C11c}$$

with the dynamical exponent $z = (1 - \omega)/(2 - \omega)$ and the scaling exponents $x = (2 - \omega)^{-1}$ and $y = -(2 - \omega)^{-1}$ for the cluster density and the size s of the largest clusters, one can easily verify that the resulting set of equations for the scaling functions becomes independent of the parameter η:

$$\partial_{\tilde{t}} \tilde{m}(\tilde{t}) = -\tilde{m}^\gamma (1 - \tilde{m})^\omega \tilde{K}^{1-\omega}, \tag{C12a}$$

$$\partial_{\tilde{t}} \tilde{K}(\tilde{t}) = \tilde{m}^\sigma, \tag{C12b}$$

$$\partial_{\tilde{t}} \tilde{s}(\tilde{t}) = \tilde{s}^\omega \tilde{m}^\gamma, \tag{C12c}$$

where $\tilde{t} = \eta^z t$. Therefore, the scaling functions $\tilde{m}$, $\tilde{K}$, and $\tilde{s}$ still depend on the parameters σ, γ, and ω characterizing the morphology but are independent of η. The fact that the dynamics obey scale invariance can now be exploited to write the onset condition for a finite yield as

$$S = s(\infty) = \eta^y \tilde{s}(\infty) \sim \eta^y, \tag{C13}$$

since $\tilde{s}(\infty)$ is just a number that is independent of η. We conclude that there must be a scaling relation

$$\eta^{-1} \sim S^{2-\omega}, \tag{C14}$$

which, by using the definition of the dimensionless parameter η [Eq. (C10)] and the definitions for $\bar{\mu}$ and $\bar{\nu}$ from Eqs. (2) and (3), can be written as a scaling relation for the detachment rate:

$$\frac{\delta_1^{\mathrm{opt}}}{C\nu} \sim \frac{\delta_1^{\mathrm{on}}}{C\nu} \sim \left(\frac{\mu}{\nu}\right)^{1/(\sigma - \gamma - 1)} S^\phi, \tag{C15}$$

with the control parameter exponent

$$\phi = \frac{2 - \omega}{\sigma - \gamma - 1}. \tag{C16}$$

This relation is consistent with the result obtained by the closed analytical solution above for the system with square-shaped monomers.

Since nucleation is the time-limiting step, the assembly time T_{90} can be estimated as the time required for $0.9C/S$ nucleation events to happen. Denoting the effective nucleation rate resulting from $\delta_1 = \delta_1^{\mathrm{opt}}$ by $\bar{\mu}^{\mathrm{opt}}$, one, hence, obtains

$$T_{90}^{\min} \sim \frac{C/S}{\bar{\mu}^{\text{opt}}C^{\sigma}} \sim (C\nu)^{-1}\left(\frac{\mu}{\nu}\right)^{(\gamma-1)/(\sigma-\gamma-1)} S^{\theta}, \qquad \text{(C17)}$$

with the asymptotic time complexity exponent given by

$$\theta = \frac{(1-\omega)\sigma + \gamma + 2\omega - 3}{\sigma - \gamma - 1}. \qquad \text{(C18)}$$

Alternatively, as described in the main text, the time complexity exponent can also be read off directly from Eq. (C12). Specifically, by reintroducing the proper dimensions, the timescale of the dynamics described by Eq. (C12) is found to obey

$$\tau_{\text{assem}} \sim (a\bar{\nu}C^{\gamma}\eta^{z})^{-1} \sim \delta_1^{\alpha}, \qquad \text{(C19)}$$

with $\alpha = \gamma - 1 + z(\sigma - \gamma - 1)$. Using Eqs. (C15) and (C16), one, thus, obtains

$$\tau_{\text{assem}} \sim S^{\phi\alpha} := S^{\theta} \qquad \text{(C20)}$$

with θ given by Eq. (C18).

APPENDIX D: GENERAL SCALING SYMMETRY

A similar scaling symmetry that we found for Eq. (C9) can be shown to hold approximately even more generally, namely, for the entire dynamic system described by the rate equations, Eq. (4). Note that, in Eq. (C9), we assume that complexes can grow indefinitely, and, therefore, the derived scaling symmetry holds only as long as the yield of completed structures is zero. In contrast, a scaling symmetry for the entire system would have even more far-reaching consequences. It would prove, among other things, our basic assumption underlying the analysis in the previous section that the threshold rate for nonzero yield scales identically as the optimal rate to achieve 90% yield in the shortest time.

To see how a scaling symmetry also holds for the entire system, Eq. (4), we again use a continuous (hydrodynamic) approximation and write the dynamic system as a partial differential equation. To this end, we interpret the density $c(s)$ as a continuous function of a real variable s over the interval $[\sigma, S]$. Expanding the first term in Eq. (4) to first order in ∂_s, one obtains the advection equation

$$\partial_t c(s,t) = (s^{\omega}c(s) - \partial_s[s^{\omega}c(s)] - s^{\omega}c(s))a\bar{\nu}m(t)^{\gamma}$$
$$= -a\bar{\nu}m(t)^{\gamma}\partial_s[s^{\omega}c(s)], \qquad \text{(D1)}$$

where $m := c_1$ denotes the concentration of monomers. This can also be written as a continuity equation $\partial_t c(s,t) = -\partial_s J(s)$ with the flux (of cluster sizes) given by $J(s) = a\bar{\nu}m^{\gamma}s^{\omega}c(s)$. Note that this first-order expansion, where we neglect the diffusion term and higher-order contributions, becomes exact in the limit $S \to \infty$, where

advection becomes the dominant driving force for the transfer of cluster mass. The equation for the concentration of nuclei, Eq. (4b), translates into a boundary condition at $s = \sigma$, where the influx of complexes $J_{\text{in}} = \bar{\mu}m^{\sigma}$ must match the outflux toward larger complexes $J(\sigma) = a\bar{\nu}m^{\gamma}\sigma^{\omega}c(\sigma)$. At $s = S$, we set an absorbing boundary condition, which removes completed structures from the system. Finally, in the equation for the monomers, Eq. (4a), the sum is replaced by an integral, $\sum_{\sigma}^{S-1} s^{\omega}c(s) \to \int_{\sigma}^{S} s^{\omega}c(s)ds$, and we neglect the first term describing the loss of monomers by nucleation versus the second term describing the loss by attachment of monomers to clusters. The latter approximation is accurate whenever there is significant growth of the structures and S is sufficiently large so that, in total, there are many more attachment events than nucleation events.

Taken together, this gives the following set of hydrodynamic equations:

$$\partial_t m(t) = -a\bar{\nu}m(t)^{\gamma}\int_{\sigma}^{S} s^{\omega}c(s,t)ds, \qquad \text{(D2a)}$$

$$\partial_t c(s,t) = -a\bar{\nu}m(t)^{\gamma}\partial_s[s^{\omega}c(s,t)], \qquad \text{(D2b)}$$

with boundary conditions

$$J(\sigma,t) = a\bar{\nu}m(t)^{\gamma}\sigma^{\omega}c(\sigma,t) = \bar{\mu}m(t)^{\sigma}, \qquad \text{(D2c)}$$

$$c(S,t) = 0 \qquad \text{(D2d)}$$

and initial conditions

$$m(0) = C, \qquad \text{(D2e)}$$

$$c(s,0) = 0. \qquad \text{(D2f)}$$

Next, we use the variables transformations $m \to mC$, $c \to cC$, $t \to \tau/(\nu C)$, and $s \to xS$ to convert the equations into dimensionless form. In these dimensionless units, the effective attachment and nucleation rate are given by $\bar{\nu} = \tilde{\delta}_1^{1-\gamma}$ and $\bar{\mu} = \tilde{\mu}\tilde{\delta}_1^{2-\sigma}$, respectively, with $\tilde{\mu} := \mu/\nu$. Furthermore, we approximate the system by replacing the position of the lower boundary σ/S (in dimensionless units) with some small and constant ϵ. Hence, instead of a process in which clusters grow from a size σ to S, we consider one in which clusters grow only from size ϵS to S, thereby neglecting a fraction $(\epsilon S - \sigma)/S \approx \epsilon$ of all assembly steps. If $\omega > 1$, as we discuss in detail in Appendix D 2, this approximation can become inaccurate, because the growth of small clusters from size σ to ϵS defines the predominant timescale in the growth process of a cluster. In this case, we therefore use a different approximation; see Appendix D 2 for details. In dimensionless form, with the lower boundary of the integral approximated by the constant, the system becomes

$$\partial_\tau m(\tau) = -a\tilde{\delta}_1^{1-\gamma} m(\tau)^\gamma S^{1+\omega} \int_\epsilon^1 x^\omega c(x,\tau)dx, \quad \text{(D3a)}$$

$$\partial_\tau c(x,\tau) = -a\tilde{\delta}_1^{1-\gamma} m(\tau)^\gamma S^{\omega-1} \partial_x[x^\omega c(x,\tau)], \quad \text{(D3b)}$$

with the boundary conditions

$$J(\epsilon,\tau) = a\tilde{\delta}_1^{1-\gamma} S^\omega \epsilon^\omega m(\tau)^\gamma c(\epsilon,\tau) = \tilde{\mu}\tilde{\delta}_1^{2-\sigma} m^\sigma, \quad \text{(D3c)}$$

$$c(1,\tau) = 0 \quad \text{(D3d)}$$

and initial conditions

$$m(0) = 1, \quad \text{(D3e)}$$

$$c(x,0) = 0. \quad \text{(D3f)}$$

Next, we perform the following general scaling ansatz: $\tilde{\delta}_1 \to S^\phi \tilde{\delta}_1$; $\tau \to S^\theta \hat{\tau}$; $m \to S^\chi \hat{m}$ and $c \to S^\xi \hat{c}$. To show that the system exhibits scale invariance, the exponents ϕ, θ, ϕ, and ξ must be determined in a way so that the system of the new scaling variables (those with the hat) is independent of S. By equating powers of S on both sides of the equations after the scale transformation, this results in a set of linear equations for the exponents which must be simultaneously fulfilled. Since Eqs. (D3d) and (D3f) equate to zero, they pose no additional constraint on the exponents, and the initial condition Eq. (D3e) immediately implies that $\chi = 0$. The remaining three linear equations then become

$$-\theta = \phi(1-\gamma) + 1 + \omega + \xi, \quad \text{(D4a)}$$

$$\xi - \theta = \phi(1-\gamma) + \omega - 1 + \xi, \quad \text{(D4b)}$$

$$\phi(2-\sigma) = \phi(1-\gamma) + \omega + \xi. \quad \text{(D4c)}$$

Subtracting the second from the first equation yields $\xi = -2$, and solving the third equation then gives

$$\phi = \frac{2-\omega}{\sigma-\gamma-1}. \quad \text{(D5)}$$

The first or second equation can then be solved for θ, which gives the unique solution

$$\theta = \frac{(1-\omega)\sigma + \gamma + 2\omega - 3}{\sigma-\gamma-1}, \quad \text{(D6)}$$

consistent with the exponents derived in the previous section [cf. Eqs. (C16) and (C18)]. This scaling symmetry implies that it is sufficient to understand the behavior of the scale-free system (for the variables with the hat), whereupon the functions $m(t)$, $c(s,t)$, etc., for arbitrary (but sufficiently large) structure size S are obtained by appropriate rescaling. Specifically, the monomer concentration

and complex size distribution obey the scaling forms

$$m(t,\tilde{\delta}_1,S) = C\hat{m}(S^{-\theta}\tau, S^{-\phi}\tilde{\delta}_1), \quad \text{(D7a)}$$

$$c(s,t,\tilde{\delta}_1,S) = CS^{-2}\hat{c}(x,S^{-\theta}\tau, S^{-\phi}\tilde{\delta}_1). \quad \text{(D7b)}$$

To relate the exponent θ with the time complexity exponent, however, it is necessary that the condition, to realize a yield Y, translates into a scaleless condition depending only on the variables $\hat{m}$ and $\hat{c}$. This can be seen by noting that the yield being Y is equivalent to the amount of resources that remain in the system (i.e., that have not been absorbed by the absorbing boundary) being equal to $1 - Y$: $m + \int_\sigma^S sc(s)ds = 1 - Y$. Hence, performing the same variable and scale transformations, this condition translates into the scale-free form $\hat{m} + \int_\epsilon^1 x\hat{c}dx = 1 - Y$. Because of the scale invariance of the governing equation [Eq. (D3)], as well as the yield condition, the exponents ϕ and θ can be identified with the parameter and time complexity exponents measured in the simulations.

The general scaling symmetry implies that the same scaling behavior that is found to hold for the optimal value of the control parameter and the minimal assembly time, in fact, applies identically for the entire functions $T_Y(\tilde{\delta}_1)$, denoting the time required to achieve a yield $Y \in (0,1)$ as a function of $\tilde{\delta}_1$, whenever this has a finite value:

$$T_Y(\tilde{\delta}_1) = S^\theta \hat{T}_Y(S^{-\phi}\tilde{\delta}_1). \quad \text{(D8)}$$

In particular, this shows that, for the scaling behavior of T_Y, it is irrelevant which value we choose for the yield threshold Y (in the main text, we exclusively use $Y = 0.9$). Finally, the scaling symmetry also explains the collapse of the yield curves in Fig. 3, and, in particular, it shows that the threshold parameter values $\tilde{\delta}_1^{\text{on},Y}$ scale identically as the optimal parameter values $\tilde{\delta}_1^{\text{opt},Z}$, independently of the demanded values Y and Z for the yield.

1. Scaling theory for the dimerization barrier $\tilde{\mu}$

With the scaling ansatz in the last section, we transform the system into one that is independent of S, giving us the scaling forms Eqs. (D7a) and (D7b). However, through the boundary condition Eq. (D3c), the solution still depends on the ratio $\tilde{\mu} = \mu/\nu$ between the dimerization and the attachment rate. For the simulations in the main text, we set $\mu = \nu$, corresponding to $\tilde{\mu} = 1$. A ratio $\tilde{\mu} < 1$ corresponds to the case where the dimerization rate (or any other subnucleation binding reaction rate) is smaller than the typical attachment rate ν, e.g., as a result of cooperative binding effects. With a similar scaling ansatz as before, we can also eliminate the parameter $\tilde{\mu}$ from the system. Specifically, transforming the control parameter as $\tilde{\delta}_1 \to \tilde{\mu}^{\phi'}\tilde{\delta}_1$ and time as $\tau \to \tilde{\mu}^{\theta'}\tau$, the system can be shown to become independent of $\tilde{\mu}$ if

$$\phi' = \frac{1}{\sigma - \gamma - 1} \quad \text{and} \quad \theta' = \frac{\gamma - 1}{\sigma - \gamma - 1}. \qquad (D9)$$

We call ϕ' and θ' the complementary control parameter and time complexity exponent, respectively. The corresponding general scaling forms for m and c thereby read

$$m(t, \tilde{\delta}_1, S, \tilde{\mu}) = C\hat{m}(S^{-\theta}\tilde{\mu}^{-\theta'}\tau, S^{-\phi}\tilde{\mu}^{-\phi'}\tilde{\delta}_1), \qquad (D10a)$$

$$c(s, t, \tilde{\delta}_1, S, \tilde{\mu}) = CS^{-2}\hat{c}(x, S^{-\theta}\tilde{\mu}^{-\theta'}\tau, S^{-\phi}\tilde{\mu}^{-\phi'}\tilde{\delta}_1), \quad (D10b)$$

and, for the assembly time, we obtain

$$T_Y(\tilde{\delta}_1) = S^\theta \tilde{\mu}^{\theta'} \hat{T}_Y(S^{-\phi}\tilde{\mu}^{-\phi'}\tilde{\delta}_1). \qquad (D11)$$

This implies that whether or not the minimal assembly time (as a function of $\tilde{\delta}_1$) can be reduced by decreasing $\tilde{\mu}$ depends on the morphology of the building blocks: For hexagonal building blocks, we have $\theta' = 0$, so $T_{90}^{\min}$ cannot further be reduced. In contrast, for square- and triangle-shaped monomers, we have $\theta' = 1$, and so the minimal assembly time decreases proportionally with $\tilde{\mu}$ within a certain range of $\tilde{\mu}$ for which our effective theory gives an accurate description.

Note that, if the complementary exponents are nonzero, the efficiency of the simulation can be increased significantly by reducing $\tilde{\mu}$: First, reducing $\tilde{\mu}$ (for $\theta' > 0$) diminishes the assembly time and, thus, the simulation time. In addition, if $\phi' > 0$, the optimal parameter value decreases as well. Hence, fewer simulation steps are needed to simulate one time unit if δ_1 is close to the respective optimal value. This offers the possibility to speed up simulations that would normally be extremely inefficient (if the goal is to determine the minimal assembly time or its scaling properties). For example, we use this trick for systems in Appendix J that would otherwise take extremely long to simulate. With a reduced value of $\tilde{\mu}$, however, we are able to determine their scaling properties with reasonable simulation effort. Plotting the assembly time in units of $(C\nu)^{-1}\tilde{\mu}^{\theta'}$ and the control parameter in units of $(C\nu)\tilde{\mu}^{\phi'}$ allows one to compare simulation data obtained with different values of $\tilde{\mu}$.

Reducing $\tilde{\mu}$ decreases the nucleation rate relative to the growth rate of clusters. Thus, it has the same effect on the assembly dynamics as increasing the detachment rate $\tilde{\delta}_1$. Hence, we can also ask how the assembly time scales with the structure size if instead of $\tilde{\delta}_1$ we change only $\tilde{\mu}$ to control the assembly process. In terms of our scaling approach, this corresponds to switching the roles of $\tilde{\delta}_1$ and $\tilde{\mu}$. Specifically, we now use the scaling ansatz $\tilde{\mu} \to S^{\phi_\mu}\tilde{\delta}_1^{\phi'_\mu}\hat{\mu}; \ \tau \to S^{\theta_\mu}\tilde{\delta}_1^{\theta'_\mu}\hat{\tau}; \ c \to S^{\xi_\mu}\hat{c}$ on the system Eq. (D3). This yields two sets of linear equations for the exponents of S and $\tilde{\delta}_1$ on both sides of the equations. Solving the linear systems analogously, as shown above,

we obtain the exponents

$$\phi_\mu = \omega - 2, \qquad \theta_\mu = 1 - \omega, \qquad \xi_\mu = -2 \qquad (D12)$$

and complementary exponents

$$\phi'_\mu = \sigma - \gamma - 1, \qquad \theta'_\mu = \gamma - 1. \qquad (D13)$$

For the scaling of the assembly time, this implies

$$T_Y(\tilde{\mu}) = S^{\theta_\mu}\tilde{\delta}_1^{\theta'_\mu}\hat{T}_Y(S^{-\phi_\mu}\tilde{\delta}_1^{-\phi'_\mu}\tilde{\mu}). \qquad (D14)$$

By tuning the dimerization barrier, we find that one achieves the same minimal time complexity exponent $1 - \omega$ as by controlling δ_1 for the case $\gamma = 1$. Indeed, both these scenarios are essentially the same, since either of the two scenarios (decreasing $\tilde{\mu}$ or increasing $\tilde{\delta}_1$ in the case $\gamma = 1$) reduces the speed of nucleation while leaving the speed of cluster growth unaffected. Note, however, that the time complexity exponent in the $\tilde{\delta}_1$ scenario exists only if $\sigma > \gamma + 1$, whereas, for the $\tilde{\mu}$ scenario, θ_μ is independent of σ and γ. This implies that controlling $\tilde{\mu}$ allows one to achieve high resource and time efficiency in the case where $\tilde{\delta}_1$ is very small so that single bonds can already be considered stable on the relevant timescales. The corresponding nucleation size in this case is $\sigma = 2$, and the attachment order $\gamma = 1$. Assuming a growth exponent $\omega = 1/2$ for two-dimensional structures would, thus, yield a time complexity exponent $\theta_\mu = 1/2$. However, controlling the parameter $\tilde{\mu}$ in an experiment is presumably tricky, since it relies on allosteric or cooperative binding effects. In contrast, the parameter $\tilde{\delta}_1$ can be controlled more effectively, e.g., by changing the temperature or the monomer concentration, among other possibilities (cf. Sec. II).

2. Large growth exponents: 3D structures

This section investigates the self-assembly of structures characterized by growth exponents $\omega > 1$. In particular, three-dimensional structures can have growth exponents $\omega > 1$. To see this, we consider as an example the self-assembly of cube-shaped monomers into a three-dimensional cube-shaped structure of edge length $L = S^{1/3}$. This example corresponds to the three-dimensional generalization of the square-shaped monomer morphology discussed in the main text: As in the two-dimensional case, $\sigma = 4$ cube-shaped monomers form a stable nucleus. Subsequently, clusters grow by adding one face at a time in a domino effect that starts after a leading-order attachment event involving $\gamma = 2$ monomers has occurred. The face of a cube-shaped cluster of size s comprises $l_s = s^{2/3}$ monomers, and the total perimeter is, thus, given by $p_s = 6s^{2/3}$. Therefore, Eq. (B10) and after implies that the growth exponent is $\omega = 4/3$ in this case.
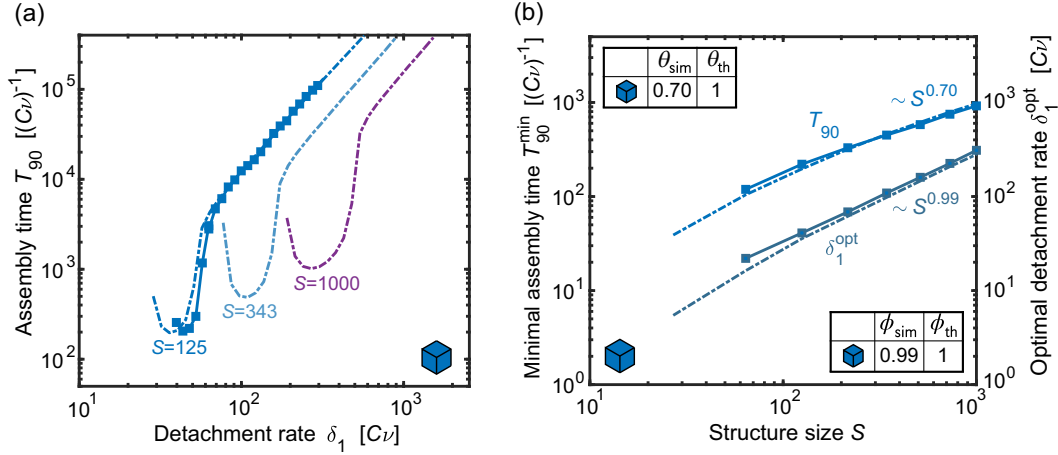
FIG. 14. Assembly time and optimal detachment parameter for three-dimensional structures. (a) The assembly time T_{90} for the self-assembly of cube-shaped monomers into three-dimensional cubic structures is plotted against the detachment rate δ_1. Markers show the results of stochastic simulations, while dash-dotted lines show the prediction of the effective model for three different sizes of the target structure. Stochastic simulations are performed with particle numbers $N = 500S$ and averaged over ten independent runs. The effective model is integrated with parameters $\sigma = 4$, $\gamma = 2$, $\omega = 4/3$, $a = 15$, and $\mu = 2\nu$; see Appendix D 2. (b) The minimal assembly time T_{90}^{min} and the optimal detachment rate $\tilde{\delta}_1^{opt}$ inferred from the stochastic simulations (markers and solid lines) and effective theory (dash-dotted lines) are plotted against the size S of the target structure. The tables show their respective scaling exponents inferred from the last three data points of the stochastic simulation in comparison with their theoretical values according to Eq. (D17).

For $\omega > 1$, the approximation of the lower boundary of the integral in Eq. (D3a), which we use to derive the scale invariance, is no longer fully accurate, as we elaborate in the following. Consequently, our previous results for the time complexity and control parameter exponent [Eqs. (D6) and (D5)] no longer accurately describe the scaling behavior for growth exponents larger than one. We recall that, to derive the scale invariance, we approximated the lower boundary σ of the integral in the equation for the monomer concentration, Eq. (D3a), by a small target-size-dependent value ϵS. In this way, a small fraction ϵ of all assembly steps is neglected. This approximation is accurate as long as $\omega \leq 1$ but becomes problematic when $\omega > 1$. The reason is that, despite neglecting only a small fraction of the assembly steps, the timescale for the growth of a cluster from size σ to ϵS,

$$T_{\sigma \to \epsilon S} = \int_\sigma^{\epsilon S} \frac{1}{m^\gamma \bar{\nu} S^\omega} ds \sim s^{1-\omega}|_{\epsilon S}^\sigma \approx \sigma^{1-\omega},$$

still exceeds the timescale for the growth from size ϵS to the final size S, $T_{\epsilon S \to S} \sim (\epsilon S)^{1-\omega}$. Hence, neglecting the slow initial assembly steps significantly underestimates the overall timescale of the assembly process.

A better approximation in the case where $\omega > 1$ is obtained by approximating the upper boundary of the integral instead. Specifically, since large clusters grow very fast on average, we truncate the growth process at a fixed cluster size K and assume that all subsequent growth steps are infinitely fast. On average, therefore, each assembly step from size σ to K consumes $(S - \sigma)/(K - \sigma) \approx S/K$

monomers. This stoichiometric factor needs to be included as a prefactor to the integral to guarantee that the total number of monomers is conserved. The equation for the monomer concentration replacing Eq. (D2a) is, thus, given by

$$\partial_t m(t) = -a\bar{\nu} m(t)^\gamma \frac{S}{K} \int_\sigma^K s^\omega c(s,t) ds. \quad \text{(D15)}$$

Using the same scaling ansatz as in Appendix D results in the linear system

$$-\theta = \phi(1-\gamma) + 1 + \xi, \quad \text{(D16a)}$$

$$\xi - \theta = \phi(1-\gamma) + \xi, \quad \text{(D16b)}$$

$$\phi(2-\sigma) = \phi(1-\gamma) + \xi, \quad \text{(D16c)}$$

which is readily solved to give

$$\phi = \frac{1}{\sigma - \gamma - 1}, \qquad \theta = \frac{\gamma - 1}{\sigma - \gamma - 1}. \quad \text{(D17)}$$

This suggests that, if ω exceeds 1, the time complexity and control parameter exponent remain constant as functions of ω. Note that, for $\omega = 1$, Eqs. (D5) and (D6) yield the same values for the exponents as Eq. (D17). Comparing with the results from stochastic simulations (cf. Fig. 14), we find that the control parameter exponent accurately matches the theoretical value $\phi_{th} = 1$ predicted by Eq. (D17), while the time complexity exponent is slightly overestimated by the theoretical value $\theta_{th} = 1$. The

estimates are, however, significantly better than those obtained with Eqs. (D5) and (D6), which would yield $\phi_{\text{th}} = 2/3$ and $\theta_{\text{th}} = 1/3$. The overestimation of the time complexity exponent results from the fact that we average the rate of monomer consumption over the entire growth process of a cluster [accounting for the constant prefactor S/K in Eq. (D15)]. This approximation particularly affects the kinetics in the regime of the ∪ shape, where clusters grow simultaneously, and, thus, the precise temporal dynamics of the monomer concentration are important. In contrast, in the regime characterized by the power-law scaling, clusters tend to nucleate and grow consecutively, and the approximation is, therefore, quite accurate. Indeed, we find that the assembly time simulated with the effective model at a constant, large value of $\tilde{\delta}_1$ (e.g., $\tilde{\delta}_1 = 10^4$) scales with large target structure sizes as $\sim S^{-1.00}$—fully in accordance with the theoretical exponents $\phi_{\text{th}} = \theta_{\text{th}} = 1$ and the power-law scaling $T_{90} \sim S^{\sigma-2}$. We note from Fig. 14(a) that the form of the ∪ shape broadens when S increases. Therefore, there can be only an approximate scaling relation but no exact one for the entire T_{90} curve as in the case where $\omega \leq 1$.

Three-dimensional structures with higher coordination numbers that self-assemble with attachment order $\gamma = 1$ most likely have a growth exponent $\omega \approx 2/3$ (since there cannot be a domino effect if $\gamma = 1$). In this case, Eq. (D6) suggests a time complexity exponent of $\theta_{\text{th}} = 1/3$. This value is smaller than the value obtained for $\gamma \geq 2$ according to Eq. (D17) for any realistic nucleation size σ. Hence, also in the case of three-dimensional structures, self-assembly is fastest if the attachment order is small. Structures with higher coordination numbers, therefore, tend to assemble faster. This is consistent with the results in Ref. [52], showing that high coordination structures typically have steeper free energy gradients.

Effective model for 3D structures. The prefactors for the effective model used for the plots in Fig. 14 are derived as follows. Once the first monomer has attached to the face of a three-dimensional structure, the second monomer forming the seed can attach in $q_s = 4$ equivalent ways (q_s binding sites) in the case of a large cluster and $q_s = 2$ ways in the case of a small cluster. Hence, with the total perimeter surface given by $p_s = 6s^{2/3}$, Eq. (B10) predicts a value for the multiplicity factor a between 12 and 24. Indeed, a good fit with the stochastic simulation is obtained with $a = 15$ in Fig. 14. Furthermore, we note that, in the nucleation process, the multiplicity for a dimer to grow into a trimer is 8, while the detachment rate of both the trimer and the dimer is $2\delta_1$. Hence, we obtain an overall prefactor of the effective nucleation rate of 2 (for 2D structures, the prefactors for the nucleation processes cancel to 1 for all three monomer morphologies; cf. Fig. 10). We account for the additional prefactor in the nucleation process by setting $\mu = 2\nu$ [cf. Eq. (2)]. To further improve the fit, we explicitly set the growth rates for the first four

postnucleation assembly steps to $g_s = 64\bar{\nu}m^\gamma$. This improves the fit, since the nucleus does not yet have a cube shape, as assumed for the growth of larger clusters.

APPENDIX E: THE ROLE OF CLUSTER-CLUSTER INTERACTIONS

In the analysis in the main text, we work under the simplifying assumption of an *ideal* assembly process, meaning that interactions among clusters are negligible. This assumption is frequently used in the literature on self-assembly. It is motivated by the observation from experiments and molecular dynamics simulations that intermediate assembly products are typically present only in very low concentrations [15,16,84]. Here, we investigate and quantify to which extent cluster-cluster interactions can influence the time efficiency of self-assembly and whether the morphology of the building blocks is relevant in this regard. To this end, we first introduce a simple measure that informs whether cluster interactions can *potentially* influence the assembly process. Afterward, we extend our effective model to include cluster interactions and estimate the effect that those might have on the assembly time.

To estimate the extent to which interactions among oligomers can potentially influence the dynamics, we define the *assembly ideality index* that is calculated under the assumption of an ideal assembly process:

$$I := \frac{S}{C} \int_0^{T_{90}} \nu K^2(t) \mathrm{d}t, \tag{E1}$$

where, as above, $K(t) := \sum_{s=\sigma}^{S-1} c_s(t)$ denotes the total concentration of incomplete complexes in the system at time t and we integrate over the time window $[0, T_{90}]$ to estimate the total number of cluster-cluster interactions until time T_{90}. The index, therefore, estimates the expected number of cluster interactions per completed structure until time T_{90} (note that the concentration of completed structures at 90% yield is $0.9C/S \approx C/S$), assuming the same reaction rate ν between two clusters as between a cluster and a monomer. Small values $I \ll 1$ indicate that cluster interactions occur only with a small probability I during the growth of a structure and can, thus, be neglected. Hence, the index provides a simple consistency check for the ideality assumption of a simulation. Note that assuming the same reaction rate ν between clusters as between clusters and monomers most likely overestimates the reaction rate between clusters, since larger clusters typically diffuse and, therefore, react more slowly than monomers, and only a fraction of their interactions would lead to stable configurations. Hence, the ideality index must be interpreted as an upper limit for the expected number of cluster interactions. Small values $I \ll 1$ are considered a sufficient condition for the ideality assumption. In contrast, large values $I > 1$
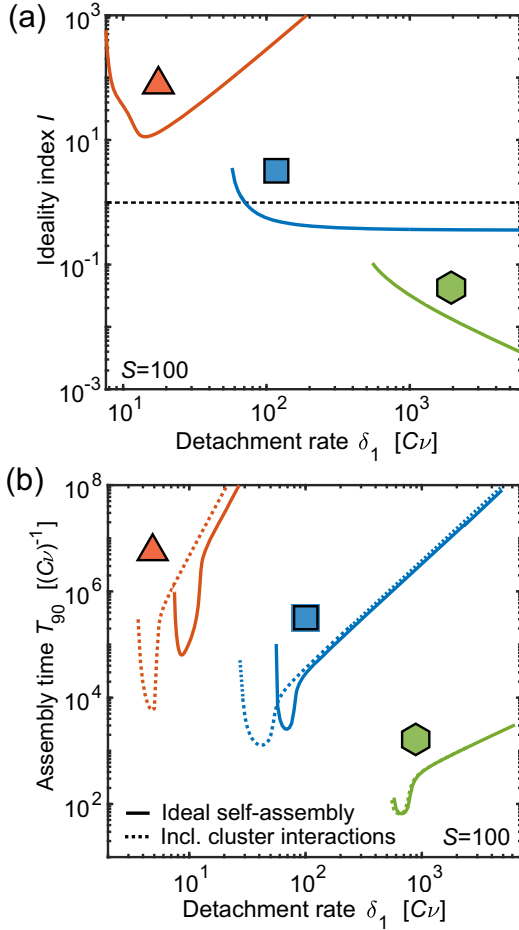
FIG. 15. Beyond ideal self-assembly: the role of cluster-cluster interactions. (a) The ideality index I, which estimates the number of interactions of an oligomer with other oligomers until time T_{90} [see definition Eq. (E1)], is plotted against the detachment rate $\tilde{\delta}_1$ for the assembly processes with three different particle morphologies ($S = 100$). The index is calculated numerically by integrating the effective model, Eq. (4). Values of I larger than 1 (dashed line) indicate that cluster interactions may be important for the assembly dynamics. Thus, a strong dependence of the potential role of cluster interactions on particle morphology is revealed. (b) Assembly time T_{90} with (drawn line) and without (dotted line) cluster interactions as a function of the detachment rate $\tilde{\delta}_1$. Assembly times are obtained by integrating the effective theory that is extended by additional terms accounting for all possible cluster reactions $c_s + c_p \overset{\nu}{\longrightarrow} c_{s+p}$ subject to the constraint $s + p \leq S$ (see Appendix B for details). Since, in an actual system, the rate of reactions between clusters would likely be smaller, this must be considered an upper estimate of the effect expected by cluster interactions. The minimal assembly time is only slightly reduced as a consequence of cluster interactions, and—as in the ideal case—the assembly time increases like $T_{90} \sim \tilde{\delta}_1^{\sigma-2}$ for large $\tilde{\delta}_1$.

indicate that cluster interactions might *potentially* influence the dynamics, depending on the specifics of the system.

Figure 15(a) shows the numerically determined ideality index as a function of the detachment rate δ_1 for the

assembly processes with the three different particle morphologies. The index reveals a strong dependence of the potential role of cluster-cluster interactions on the morphology of the constituents: While cluster-cluster interactions can be safely neglected for hexagonal-shaped monomers ($I \ll 1$), the magnitude of the index observed for the triangle-shaped monomers suggests that these interactions play a significant role in the assembly process. Furthermore, the index behaves differently as a function of the detachment rate $\tilde{\delta}_1$: While in the system with hexagonal monomers, cluster-cluster interactions become less critical when $\tilde{\delta}_1$ is increased, their potential influence further increases in the system with triangle-shaped monomers as $\tilde{\delta}_1$ grows. Hence, the ideality index seemingly correlates with the time efficiency of a self-assembly system: Particles with a morphology that assembles time efficiently tend to have a smaller ideality index, in general, than particles with a morphology that assembles less efficiently.

In order to estimate the extent to which cluster interactions affect the assembly time, we extend the effective model by additional terms accounting for reactions of any two clusters of sizes $i, j \geq \sigma$ with $i + j \leq S$ to a cluster of size $s = i + j$ at rate ν:

$$\partial_t c_s = \cdots + \frac{1}{2}\nu \sum_{\substack{i,j \geq \sigma \\ i+j=s}} c_i c_j - \nu c_s \sum_{\substack{i \geq \sigma \\ i+s \leq S}} c_i, \qquad (E2)$$

for all $\sigma \leq s \leq S$. The factor of $\frac{1}{2}$ in front of the first sum avoids double counting and serves as a stoichiometric factor in the case $i = j$. As before, assuming that clusters react at the same rate ν as monomers strongly overestimates the effect that cluster interactions have on the assembly time, and, thus, the model describes an upper limit for the impact that is to be expected by these additional interactions.

Figure 15(b) compares the assembly time of the ideal system with that obtained from integrating the extended model accounting for cluster interactions. The comparison suggests that cluster-cluster interactions can indeed reduce the assembly time in systems with triangles— and to a much lesser extent also in systems with square-shaped monomers. However, the reduction of the assembly time is relatively small, in general. The reason is that, as the size of a cluster increases, the number of its possible binding partners decreases: According to Eq. (E2), a cluster of size i can react only with clusters of size $j \leq S - i$. Therefore, cluster interactions are not effective in avoiding kinetic traps.

If we also take into account that the model probably overestimates the reaction rate between clusters, the reduction in assembly time due to cluster-cluster interactions will presumably be even smaller in any actual system that uses the respective types of particles. On the other hand, cluster-cluster interactions might enhance the occurrence of assembly errors and defects. For example, occasionally, clusters might
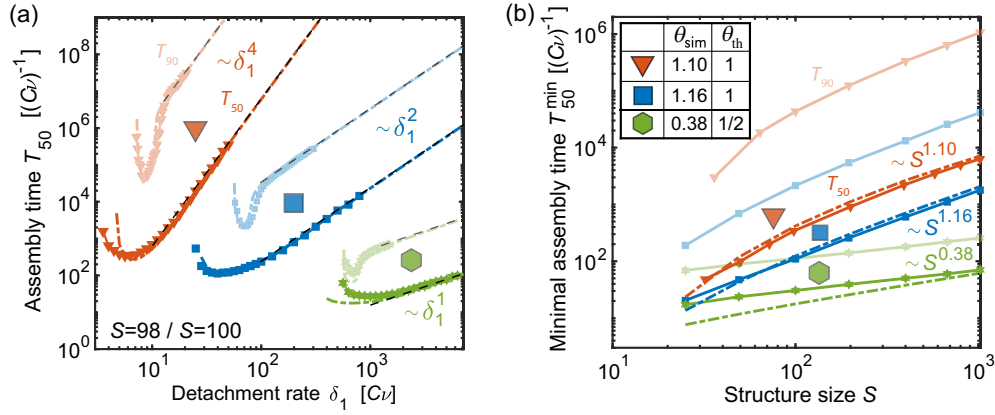
FIG. 16.   Assembly time T_{50} (solid) in comparison with assembly time T_{90} (faded). (a) The assembly times T_{50} (solid) and T_{90} (faded) required to achieve 50% or 90% yield, respectively, in the strong binding limit are plotted against the detachment rate δ_1 (in units of $C\nu$) for systems with triangle- (red), square- (blue), and hexagon-shaped (green) building blocks for a fixed target structure size $S = 98$ (triangle-shaped monomers) or $S = 100$ (square and hexagonal monomers). Markers represent averages of ten independent runs of the stochastic simulation [performed with $N = 500S$ and $N = 300S$ (triangle-shaped monomers)]. The dash-dotted colored lines represent the prediction of the effective theory obtained by numerically integrating Eq. (4). Black dashed lines show the analytic result for the assembly times given by Eqs. (31) and (32) in the slow nucleation regime (ranges for $\tilde{\delta}_1$ are chosen manually). (b) The minimal assembly time $T_{50}^{\min}$ (solid) and $T_{90}^{\min}$ (faded) inferred from the stochastic simulation (markers and solid lines) and effective theory (dash-dotted lines) are plotted against the size S of the target structure. Although the assembly times T_{90} and T_{50} may differ significantly, they exhibit approximately the same scaling in dependence on the structure size S. The scaling exponents θ_{sim} inferred from the last three data points of the stochastic simulation for T_{50} [and their theoretical asymptotic values θ_{th}; cf. Eq. (17)] are summarized in the table [compare with the analogous exponents θ_{sim} for T_{90} exhibited in the table in Fig. 4(b)].

react that do not fit together perfectly, thus leaving defects and holes or distorting the structure. Operating at a small ideality index might, thus, be preferable for general self-assembly systems, as it might enhance the robustness of the process against errors. In conclusion, the morphology of the constituents could be an essential determinant not only of the time efficiency, but also of the reliability of the assembly process.

APPENDIX F: INVARIANCE OF THE SCALING RESULTS TO THE YIELD THRESHOLD

In the main text and the appendix so far, we have studied the behavior of the assembly time T_{90} required to achieve 90% yield. Alternatively, one could also choose a different threshold for the yield, e.g., 50%, and analyze the behavior of the corresponding assembly time T_{50}. For our scaling results to be informative and valuable, it is important that these results are invariant to the yield threshold and that the qualitative behavior of the assembly time T_Y is the same, independent of the yield threshold Y. We verify this in the following by repeating the same plots shown in the main text for T_{90} now with T_{50}.

Figure 16(a) shows the assembly time T_{50} as a function of the detachment rate δ_1 in comparison with T_{90} for the different particle morphologies in the strong binding limit. We find that T_{50} and T_{90} behave qualitatively similar, and, in particular, both exhibit the same scaling $\sim \tilde{\delta}_1^{\sigma-2}$ for large enough δ_1. However, the values of the assembly times T_{50} and T_{90} can differ quite dramatically depending on the

particle morphology. In particular, differences in the minimal assembly times between the different particle morphologies tend to become less pronounced for a lower yield threshold. This is consistent with our analysis of the regime in which nucleation is very slow (cf. Sec. V D). In particular, Eq. (31) and (32) [visualized as black dashed lines in Fig. 16(a)] predict that the assembly time in this regime depends on the yield threshold Y as $T_Y \sim (1 - Y)^{1-\sigma}$. Thus, the assembly time in this regime depends more strongly on the demanded resource efficiency if the nucleation size σ is larger. However, the scaling of the minimal assembly time with the target structure size (i.e., the time complexity exponent θ) remains independent of the yield threshold as we verify in Fig. 16(b): Comparing the measured time complexity exponents for T_{50} with those measured for T_{90} [see the table in Fig. 4(b)] shows that they are almost identical. Thus, despite the assembly time as such might depend strongly on the yield threshold, all scaling results are invariant to the yield threshold. Differences in the assembly times between the different particle morphologies studied in this work tend to become less pronounced if the requirement for resource efficiency is relaxed.

APPENDIX G: SEEDING

Seeding refers to the experimental method of initially putting a certain number of stable nuclei (seeds) into the system. In systems that enable the formation of different
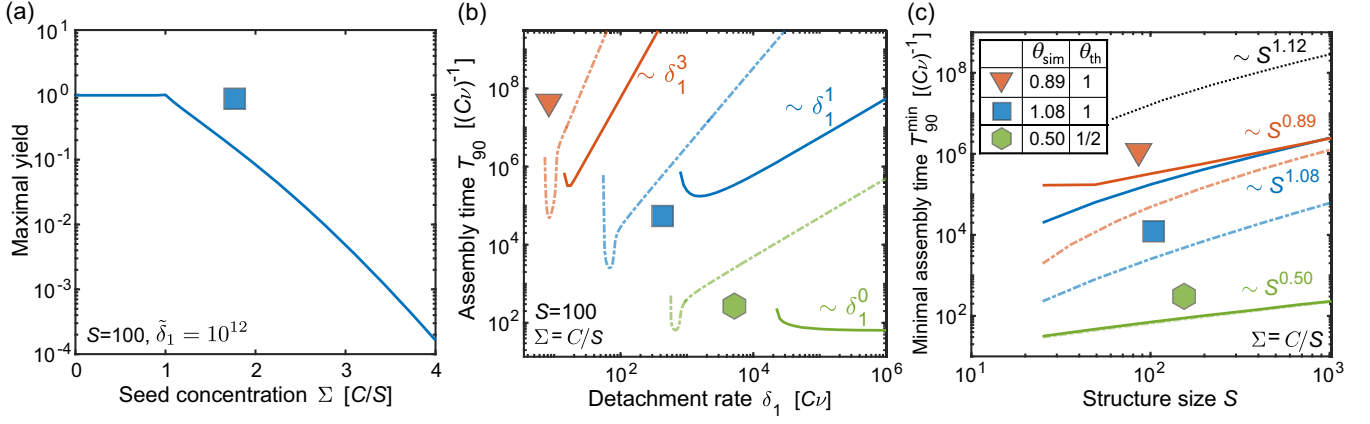
FIG. 17. Yield and assembly time in seeded systems (according to the effective model). (a) The maximal yield of target structures of size $S = 100$ in the effective model in the limit of large $\tilde{\delta}_1$ (here, $\tilde{\delta}_1 = 10^{12}$) is plotted against the initial concentration Σ of nuclei (seeds). Parameters of the effective model are chosen to simulate the self-assembly of square-shaped monomers ($\sigma = 4$, $\gamma = 2$, $\omega = 1$, and $a = 5.3$), and the initial monomer concentration is $m(0) = C - \sigma\Sigma$. (b) Assembly time T_{90} as a function of $\tilde{\delta}_1$ for seeded systems with initial seed concentration $\Sigma = C/S$ (drawn lines) and corresponding unseeded systems (dash-dotted lines, faded). (c) Minimal assembly time $T_{90}^{\min}$ as a function of S for seeded systems with initial seed concentration $\Sigma = C/S$ (drawn lines) and corresponding unseeded systems (dash-dotted lines, faded). The black dashed line shows the assembly time for a putative system of triangular monomers in which the first assembly step after nucleation has attachment order $\gamma' = 3$ instead of $\gamma' = 4$. The table shows the time complexity exponents θ_{sim} measured in the range of target structure sizes $400 \leq S \leq 1000$ and their corresponding theoretical values θ_{th}.

structures, like, e.g., algorithmic self-assembly, seeding is an integral part, since the composition of the seed determines the structure or the pattern that grows (i.e., the algorithm that is performed via algorithmic self-assembly) [35]. From a kinetic point of view, we expect that seeding should reduce the assembly time by eliminating the requirement of the rate-limiting nucleation reactions, as the seeds can start growing right from the beginning. First, in order to determine the optimal number (concentration) of seeds, we integrate the effective model with an initial concentration Σ of stable nuclei $[c_\sigma(0) = \Sigma]$ and corresponding monomer concentration $m(0) = C - \sigma\Sigma$ (to guarantee a fixed total number of resources). We find that the maximum yield (obtained in the limit of large $\tilde{\delta}_1$) is 1 if the seed concentration is smaller or equal C/S and decreases rapidly if this concentration is exceeded [Fig. 17(a)]. This is intuitive, since an initial concentration $\Sigma = C/S$ of nuclei exactly equals the concentration of completed structures when the yield is 100%.

Next, we study the effect of seeding on the assembly time by integrating the effective model with initial concentration $\Sigma = C/S$ of nuclei (corresponding to the number N/S of seeds required for 100% yield) and corresponding monomer concentration $m(0) = C - \sigma\Sigma$ [cf. Figs. 17(a) and 17(c)]. Surprisingly, contrary to our expectation, we find that the assembly time increases due to seeding if $\gamma > 1$ (square and triangular monomers). This is because if nuclei are already present in the system, the control parameter $\tilde{\delta}_1$ must be chosen even larger to suppress

formation of additional nuclei, which would lead to an excess of clusters. Increasing $\tilde{\delta}_1$ simultaneously decreases the growth rate if $\gamma > 1$, making the overall process less time efficient. In contrast, for hexagon-shaped monomers ($\gamma = 1$), the growth rate is not affected by $\tilde{\delta}_1$ and the minimal assembly time remains the same. With structure growth (rather than nucleation) being rate limiting in seeded systems, increasing $\tilde{\delta}_1$ beyond the optimal value makes the assembly time increase only proportional to $\tilde{\delta}_1^{\gamma-1}$ (rather than $\sim\tilde{\delta}_1^{\sigma-2}$ in the case of unseeded systems [compare Fig. 4(a)]). In particular, in the case of hexagonal building blocks ($\gamma = 1$), the assembly time is, thus, constant as a function of $\tilde{\delta}_1 \geq \tilde{\delta}_1^{\text{opt}}$.

In the framework of our scaling analysis, the initial condition for seeded systems, $c_\sigma(t = 0) \sim S^{-1}$, is consistent with the scale invariance described by the scaling forms for the total concentration of clusters, $K(t, S) \sim S^{-1}\hat{K}(S^{-\theta}\tilde{t})$ (cf. Appendix C), and for the cluster size density, $c(s, t, S) \sim S^{-2}\hat{c}(S^{-1}s, S^{-\theta}\tilde{t})$ [cf. Eq. (24)] (note that, by rescaling the cluster size $s \to S^{-1}s$, we get an additional factor of S^{-1} for the cluster size density). Furthermore, the initial condition $m(0) = C - C/S$ is approximately consistent with the scaling form for the monomer concentration $m(t, S) \sim \hat{m}(S^{-\theta}\tilde{t})$ [cf. Eq. (23)] in the limit of large S where C/S is negligible and $m(0) \approx C$. Therefore, the scaling analysis in Appendix D can be performed identically for seeded systems with the modified initial conditions [cf. Eqs. (D3e) and (D3f)]. Consequently, the scaling behavior should be the same for seeded systems with initial concentration of seeds

$\Sigma \sim S^{-1}$. This is confirmed by Fig. 17(c) for square and hexagonal monomers. In the case of triangular monomers, we observe a slightly smaller time complexity exponent compared to the unseeded system (but still close to its theoretical value). This anomalous behavior is likely caused by the higher effective order $\gamma' = 4$ of the first assembly step following nucleation (cf. Appendix B) (which we do not explicitly account for in the scaling analysis). To see this, we simulate a putative system of triangular monomers in which the attachment order of the first step likewise equals $\gamma = 3$ (black dashed line). The assembly time in this case is significantly larger than for the true system, and the scaling exponent is very close to that of the unseeded system, confirming that the postnucleation assembly step is responsible for the drop in the exponent.

Hence, in general, this shows that our scaling results are very robust and apply likewise to seeded systems. In line with our general conclusions, using hexagonal building blocks (or generally monomers with $\gamma = 1$) appears to be highly favorable, as the minimal assembly time does not increase as a consequence of seeding and the assembly time even becomes a constant function of $\tilde{\delta}_1$.

APPENDIX H: ANNEALING PROTOCOLS

The assembly dynamics are controlled by the ratio between the frequency of detachment events δ_1 and the frequency of attachment events νm. Initially, the monomer concentration is $m(0) = C$, but, as more and more monomers get consumed during the assembly process, their concentration gradually decreases. Consequently, the ratio between the detachment and attachment rate increases during the assembly process. In order to counteract this effect, a frequently used experimental approach consists in "annealing" the system by decreasing the temperature [35]. Typically, one starts at a high temperature and gradually cools the system to room temperature. Since the detachment rate decreases with decreasing temperature, $\delta_1 \sim e^{-E_B/(k_B T)}$, if applied optimally, annealing allows one to keep the ratio between the detachment rate and attachment rate constant during the assembly process. Here, we show that our scaling analysis also applies to the case of an annealing protocol in which the temperature adapts instantaneously to the momentary concentration of monomers such that the ratio between detachment and attachment rate remains constant. To describe such an annealing protocol with the effective theory presented in Sec. V B and Appendix D, we need only to replace the control parameter $\tilde{\delta}_1 = \delta_1/(C\nu)$ with a time-dependent control parameter $\tilde{\delta}_1(t) = \tilde{\delta}_1(t = 0)[m(t)/C]$ that gradually decreases proportional to the monomer concentration. This replacement introduces additional factors of m only in the dynamic equations, but the monomer concentration does not scale with S since $m(t) = \hat{m}(S^{-\theta}\tilde{t})$ [cf. Eq. (23)]. Hence, the scaling analysis can be performed in exactly the same way.

This implies that all our scaling results apply identically under annealing. Note that we could also use any other annealing protocol of the form $\tilde{\delta}_1(t) = \tilde{\delta}_1(0)f(S^{-\theta}\tilde{t})$ to maintain the scale invariance.

Figure 20 shows the average minimal assembly time simulated stochastically with the above annealing protocol plotted against the target structure size (for square-shaped monomers). The minimal assembly time is determined by tuning the parameter $\tilde{\delta}_1(t = 0)$. We find that the self-assembly efficiency is indeed significantly increased as a result of annealing. However, the plot confirms that the scaling of the assembly time remains invariant under annealing.

APPENDIX I: SYSTEMS WITH UNLIMITED CLUSTER GROWTH

So far, we have considered systems with self-limiting cluster growth by assuming periodic boundary conditions of the target structures. However, our scaling analysis can straightforwardly be generalized to systems with unlimited growth of the structures. The only requirement is that the growth rate of large clusters consistently scales with the cluster size as $\sim s^\omega$, with some arbitrary growth exponent ω.

To describe unlimited cluster growth with the effective theory presented in Sec. V B and Appendix D, we have only to drop the absorbing boundary at $s = S$, which removes finished structures from the system, and replace the upper bound in the integral in the equation for the monomer concentration, Eq. (22) [respectively, Eq. (D3a)], by infinity, to integrate clusters of arbitrary size. Clearly, these changes do not affect the scaling analysis. The natural generalization of the yield Y_S represents the fraction of resources that are bound in clusters of size larger than or equal to S. The yield can thereby be written in the same form as Eq. (25) by identifying $1 - Y_S$ with the number of resources that are either available as monomers or bound in clusters of size smaller than S. In the form of Eq. (25), the scale invariance of the yield [cf. Eq. (26)] and of the corresponding assembly time T_{90} (likewise defined as the time required to achieve 90% yield) [cf. Eq. (27)] becomes evident. Hence, with the yield and assembly time defined in this way, all scaling results derived for the case of self-limiting cluster growth can directly be transferred to the case of unlimited cluster growth.

In principle, we could also define the yield and assembly time in a different way as long as it is consistently scale invariant. For example, the assembly time could alternatively be defined as the time when the average size of a cluster exceeds S. The scale invariance of this condition becomes evident by using the scaling form for the complex concentration, Eq. (24) [respectively, Eq. (D7)]:

$$S < \frac{\int_\sigma^\infty s c(s,t)ds}{\int_\sigma^\infty c(s,t)ds} = \frac{\int_\epsilon^\infty Sx S^{-2}\hat{c}(x, S^{-\theta}\tilde{t})S dx}{\int_\epsilon^\infty S^{-2}\hat{c}(x, S^{-\theta}\tilde{t})S dx}, \quad \text{(I1)}$$
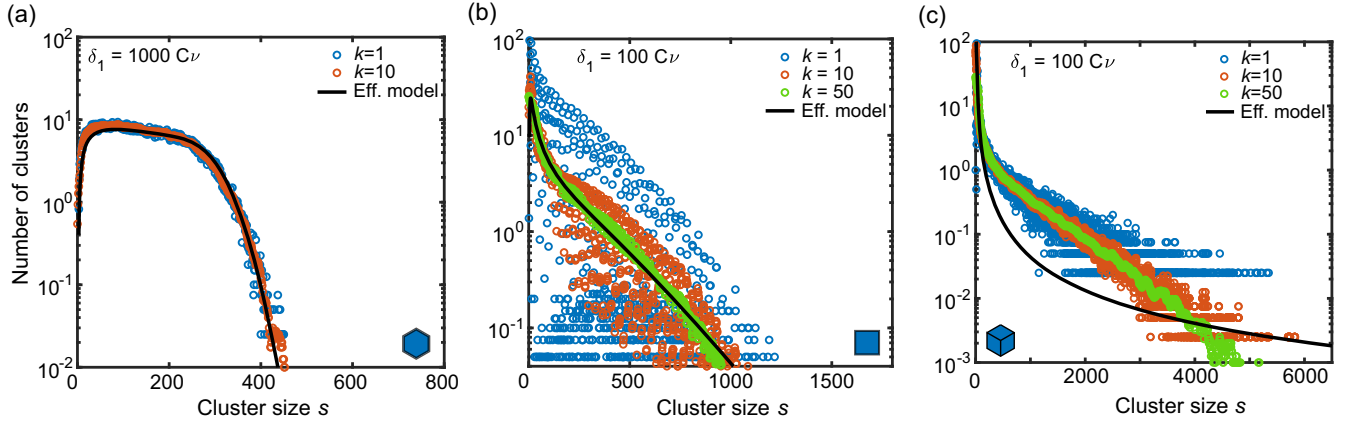
FIG. 18.   Final cluster size distribution. The final cluster size distributions for (a) hexagon-shaped ($\omega \approx 1/2 > 1$), (b) square-shaped ($\omega = 1$), and (c) 3D cube-shaped ($\omega = 4/3$) monomers are shown for the respective values of the detachment rate δ_1 indicated in each plot. The target structure size in the simulation is chosen very large [$S = 3025$ (a),(b) and $S = 8000$ (c)] so that the yield is zero in each case, thereby effectively simulating unlimited cluster growth. The initial number of monomers is $N = 3 \times 10^5$, and the distribution is averaged over 40 independent simulation runs (blue circles). Subsequently, moving averages are taken over windows of $k = 10$ (red circles) and $k = 50$ (green circles) cluster sizes. The averaged cluster size distribution decreases exponentially for large cluster sizes for square- and cube-shaped monomers (i.e., for $\omega \geq 1$). Black lines show the final cluster size distribution obtained by integrating the effective model.

since all factors of S on the right and on the left-hand side cancel. Note that it is not possible to include the monomers as clusters of size 1 in the definition of the mean cluster size, since the contribution of the monomers and the complexes in the denominator, $\hat{m} + S^{-1} \int_\epsilon^\infty \hat{c}(x)dx$, would scale differently with S and would, thus, violate the scale invariance. However, one can additionally demand that the monomer concentration drop below a certain value X, which gives the additional scale-free condition $\hat{m}(S^{-\theta}\tilde{t}) \leq X$ for the assembly time. This shows that our scaling results are very general and likewise apply to systems with unlimited cluster growth.

Cluster size distribution. It is interesting to characterize the final cluster size distribution in systems with unlimited cluster growth. To this end, we choose a target structure size in the simulation that is very large and ensures that the final yield is zero. Therefore, the size limit does not affect the system's evolution, effectively mimicking a system with unlimited cluster growth. Figure 18 shows the final cluster size distributions for hexagon-, square-, and 3D cube-shaped monomers, which are characterized by different values of the growth exponent ($\omega < 1$, $\omega = 1$ and $\omega > 1$, respectively). Since the growth rates of leading and subleading attachment reactions differ strongly, the distributions are scattered over broad ranges (in number of clusters, respectively cluster sizes) for square- and cube-shaped monomers (also compare with Fig. 2). The distributions can be narrowed down by calculating moving averages in the number of clusters over increasingly larger windows of k subsequent cluster sizes (Fig. 18 shows the distributions calculated for $k = 1$, $k = 10$, and $k = 50$).

We find that these averaged distributions decrease exponentially with large cluster sizes for square- and

cube-shaped monomers. In contrast, for hexagonal monomers, due to the smaller growth exponent, the distribution is more compact and has a significantly shorter tail. In the cases of hexagon- and square-shaped monomers, the averaged final distributions coincide well with the distribution obtained by integrating the effective model, Eq. (4) (black curves). However, in the case of cube-shaped monomers, the effective model predicts a distribution that decreases as a power law rather than exponentially, thus implying a significantly longer tail than the stochastically simulated distribution. The reason for this discrepancy is that if single clusters quickly grow very large in the three-dimensional system, the timescale of the subleading processes (domino effect, which scales as $\sim s^{2/3}$) can no longer be neglected against that of the leading-order processes [cf. Eq. (B9)]. Therefore, the effective growth rate of large clusters is actually smaller than the growth rate predicted by Eq. (B10). To accurately describe the distribution for three-dimensional structures, the timescale of the domino effect would have to be considered explicitly in Eq. (B9). On the other hand, self-limitation of cluster growth, as assumed in our original model, prevents single clusters from growing extremely large. Thus, the effective model accurately describes the assembly time for limited cluster growth, as seen in Fig. 14.

APPENDIX J: HETEROGENEOUS SYSTEMS

Next, we study how our results generalize to heterogeneous systems with different species of monomers, each species binding specifically only with certain other species. Specifically, we consider a fully heterogeneous system of square-shaped monomers such that each species (labeled

FIG. 19. Heterogeneous systems. Simulating the self-assembly of heterogeneous structures (information-rich structures), we assume that each species (labeled by an index $1\ldots S$) is unique and binds only with the respective neighboring species, as shown. Thus, each species appears only once in a finished structure and at a specified position. The underlying kinetics for self-assembly are the same as in the homogeneous case, but the different species can have distinct binding and detachment rates or varying concentrations. In Appendix J, we investigate how self-assembly behaves for various heterogeneous systems.

$1\ldots S$) can bind only with four other species ($i+1$, $i-1$, $i-L$, and $i+L$), and each site within a cluster can be occupied only by a specified species; see Fig. 19. We investigate five different relevant cases: (i) all species have the same reaction rates ν and δ_1 and are available in the same copy numbers N; (ii) the reaction rates ν_i for the different species are chosen randomly; (iii) the detachment rates of four specified species that are able to form a nucleus are set to zero; (iv) four specified species that are able to form a nucleus are provided in much larger concentrations; and (v) investigating the role of the boundary conditions, we assume that species at the boundary limit the growth of the clusters such that structures with explicit boundaries self-assemble.

1. Identical reaction rates

First, we consider the case in which all species have the same attachment and detachment rates and are available in identical concentrations. In this ideal case, one can show mathematically that the system behaves equivalent to a homogeneous system in the limit of large particle numbers. To establish this equivalence, the key insight is that, assuming periodic boundary conditions of the structures, the concentrations of all species are identical, since all species are subject to the same dynamics. We show this rigorously for the evolution of the monomer concentration. To this end, we first have to introduce some notation: We denote the species i of a complex as the species with the lowest label of the constituent monomers of the complex. A complex in the heterogeneous system is then uniquely characterized by its species index and its morphology, which specifies the size of a cluster and its shape. We denote the morphology by another index ζ that takes values in some countable set of numbers characterizing the morphology via some mapping, which we do not further specify. By $c_{j,\zeta}(t)$, we denote the concentration of complexes of species j and morphology ζ at

time t. We use $nn(i)$ as a shorthand notation of the set of indices $\{i+1, i-1, i+L, i-L\}$ with which species i is able to bind. Furthermore, $nn(i,j,\zeta)$ denotes the set of indices of species that are direct neighbors of i and constituents of the complex (j,ζ). The temporal change of the monomer concentration of species i is then given by

$$\partial_t m_i = -\mu_{\text{het}} m_i \sum_{j \in nn(i)} m_j - \sum_{\substack{(j,\zeta) \\ i \notin (j,\zeta) \\ |nn(i,j,\zeta)| \geq 1}} \nu m_i c_{j,\zeta} + \sum_{\substack{(j,\zeta) \\ i \in (j,\zeta) \\ |nn(i,j,\zeta)| = 1}} \delta_1 c_{j,\zeta}.$$

(J1)

The three terms on the right account for the loss and gain of monomers due to nucleation (with rate μ_{het}), attachment to complexes, and detachment from complexes, respectively. For the attachment reactions with monomers of species i, we consider all complexes that contain at least one neighbor of species i but not species i itself. Similarly, the complexes from which a monomer of species i can detach are those that contain species i and exactly one neighbor of i. Next, we use that the average concentrations of different monomer and complex species must be the same ($m_i = m_j$ and $c_{i,\zeta} = c_{j,\zeta}$), since they are subject to equivalent dynamics. Hence, the dimerization term simply becomes $-4\mu_{\text{het}} m_i^2$. We furthermore note that, for a given cluster morphology ζ, there is exactly one species index for each binding site of the cluster fulfilling the conditions in the sum of the attachment term. Thus, the second term can be written as

$$-\sum_{\zeta} m_i c_{i,\zeta} \tilde{b}(\zeta) \nu,$$

(J2)

where $\tilde{b}(\zeta)$ denotes the number of binding sites of the cluster with morphology ζ and the sum goes over all possible cluster morphologies. Note that the concentration of a specified cluster species i equals the total concentration of complexes with the same morphology divided by S: $c_{i,\zeta} = c_{\zeta}/S$. Finally, for the sum in the last term, there is exactly one species index fulfilling the conditions for each site in the cluster that has exactly one neighbor. Denoting the number of sites with one neighbor for a given cluster morphology by $b_1(\zeta)$, the last term thus becomes

$$\sum_{\zeta} c_{i,\zeta} b_1(\zeta) \delta_1.$$

(J3)

Hence, by setting $\mu = 4\mu_{\text{het}}$, this gives the same dynamic equation for the monomer concentration as one would have for a homogeneous system:

$$\partial_t m = \mu m^2 - \sum_{\zeta} m c_{\zeta} \tilde{b}(\zeta) \nu + \sum_{\zeta} c_{\zeta} b_1(\zeta) \delta_1. \quad (J4)$$

Setting $\mu = 4\mu_{\mathrm{het}}$ makes sense intuitively, since, in the homogeneous case, two monomers can dimerize in four possible constellations (four possible binding sites). In contrast, in the heterogeneous case, they can dimerize in only one constellation.

Similarly, the evolution of the cluster concentrations can be written independently of the species indices in a form equivalent to that of a homogeneous system. Thereby, we use the notation $\zeta = \zeta' + 1$ to indicate that the morphology ζ equals ζ' once we add a single particle to the boundary of ζ'. Similarly, $\zeta' +_1 1$ indicates that we add one particle to ζ' so that it has exactly one neighbor. The evolution of cluster concentrations then reads

$$\partial_t c(\zeta) = \sum_{\zeta = \zeta' + 1}^{\zeta'} \nu m c(\zeta') - \nu m \tilde{b}(\zeta) c(\zeta)$$
$$+ \sum_{\zeta' = \zeta +_1 1}^{\zeta'} c(\zeta') \delta_1 - \delta_1 b_1(\zeta) c(\zeta). \quad (J5)$$

Together, this shows that the heterogeneous process decouples into S homogeneous processes, one for each species. In other words, this means that a heterogeneous process with total particle number $N \cdot S$ behaves identically on average as a homogeneous process with particle number N. Because of this mathematical equivalence, all scaling laws derived for homogeneous systems directly transfer to heterogeneous systems in the ideal case. Next, we check how robust these scaling laws are if we deviate from the ideal case, e.g., by varying the rate constants or the particle numbers of the different species or by assembling structures with explicit boundaries rather than periodic boundary conditions.

2. Random heterogeneous binding rates

First, we consider a system with random heterogeneous binding rates. Clearly, for distinct species, the assumption of identical binding rates for all species is an idealization. More realistically, the rates vary to some extent. We, therefore, simulate a system with random heterogeneous nucleation and attachment rates for the different species, drawn independently from a (truncated) normal distribution with a coefficient of variation of 50%. The normal distribution is truncated for values below 20% of the mean to ensure that individual rates do not become negative or extremely small. We measure the assembly time accordingly in units of $(C\nu_{\mathrm{avg}})^{-1}$, where ν_{avg} is the average attachment rate of all species. For each run, the binding rates are chosen independently, and the assembly times are averaged over 10–20 independent runs. The measured time complexity exponent is identical to the one of the system with homogeneous rates, demonstrating the robustness of the scaling behavior. However, the time efficiency, in general, is reduced as a consequence of heterogeneous binding rates. We attribute this to the fact that the ratio between the average effective nucleation and effective
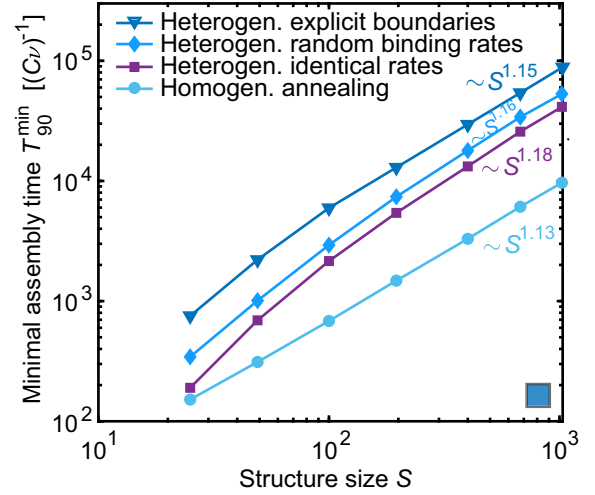

FIG. 20. Robustness of the scaling behavior. The minimal time required to achieve 90% yield ($T_{90}^{\min}$) is plotted in dependence of the target structure size S for different variants of the model. Diamond and triangle markers show $T_{90}^{\min}$ for heterogeneous systems with random heterogeneous binding rates drawn from a (truncated) Gaussian with coefficient of variation 50% (cf. Appendix J 2) or structures with explicit boundaries (cf. Appendix J 3), respectively. Squares and circles show $T_{90}^{\min}$ for the heterogeneous system with identical rates and periodic boundaries [equivalent to the homogeneous system; cf. Fig. 4(b)] as well as for a homogeneous system under annealing [i.e., with a time-dependent detachment rate $\delta_1(t) \sim m(t)$] (cf. Appendix H). Data points are obtained by averaging 10–20 independent runs of the stochastic simulation performed with $N = 500$ particles per species (heterogeneous case) or $N = 500S$ particles (homogeneous case). While the assembly time varies for the different model variants, the measured time complexity exponents are largely invariant.

growth rate increases with the coefficient of variation of the distribution of the rate constants (since the average nucleation and growth rate correspond to moments of different order of the distribution). Thus, increasing the coefficient of variation of the rate constants is equivalent to increasing the dimerization barrier $\tilde{\mu} = \mu/\nu$, which, according to Eq. (D11), increases $T_{90}^{\min}$.

3. Structures with explicit boundaries

Heterogeneous systems furthermore offer an opportunity to test the impact of the boundary conditions of the clusters. In the homogeneous case, we considered the self-assembly of closed structures, implemented by periodic boundary conditions of the clusters. The geometry of the clusters thereby limits their further growth. In a heterogeneous system, structure growth is automatically limited if the species required for further growth are missing. Therefore, we think that heterogeneous systems realized, e.g., with DNA bricks as used in Refs. [40,50] provide a valuable testing ground for our scaling results. However, the

emerging structures differ in their boundary conditions from the systems we considered in the main text (periodic versus explicit boundaries) [85]. It is, therefore, important to understand whether and how the boundary conditions influence the assembly time. Simulating clusters with explicit boundaries, we find that the assembly time slightly increases, but its scaling behavior as a function of the target structure size remains invariant (cf. Fig. 20). Statistically, i.e., when averaging over many simultaneously growing clusters, explicit boundaries merely reduce the effective average growth rate of clusters, since clusters stop growing once they reach the boundary. Hence, changing the boundary conditions is effectively equivalent to changing the binding rate ν, which affects the assembly time but not its scaling behavior. This suggests that our scaling results are generally robust with respect to the boundary conditions limiting structure growth.

4. Heterogeneous detachment rates

The question arises whether, by choosing the reaction rates for the heterogeneous system in a concerted way, the assembly time can be reduced. Since nucleation is the rate-limiting process, a simple idea to achieve this is to increase the effective rate of seed formation by making the bonds between four specified neighboring species very strong. Specifically, to test this scenario, we set the detachment rates for the four bonds between species 1, 2, $L + 1$, and $L + 2$ to zero in the simulation. The intention is that the specified species will quickly form the required seeds that can subsequently grow into complete structures. However, the simulation reveals that the minimal assembly time increases significantly as a result of this modification of the reaction rates. In fact, to achieve a yield of 90%, the uniform detachment rate δ_1 of the remaining bonds has to be chosen significantly larger than in the homogeneous case, making the simulation extremely inefficient. We, therefore, had to simulate the scenario with a reduced nucleation rate $\mu/\nu = 0.1$ (see Appendix D 1) and correspondingly plotted the assembly time in units of $(C\nu)^{-1}\mu/\nu$ [assuming a complementary time complexity exponent of 1 as in the homogeneous case (cf. Appendix D 1)]. Figure 21 indeed shows a strong increase of the minimal assembly time as compared to the system with identical rates (purple line in Fig. 21). The reason for this increase is the same as in the scenario of seeding discussed in Appendix G: By enhancing the nucleation rate of four selected species, δ_1 has to be chosen much larger for all remaining species to avoid spurious nucleation. This drastically reduces the effective growth rate and, hence, the overall time efficiency. In fact, if the timescale of nucleation of the four species becomes negligible against the overall assembly time, there is effectively no difference from putting a corresponding number of seeds into the system. Therefore, in the limit of large S, the assembly time is approximately the same as for the homogeneous seeded system; compare Fig. 17(c) and
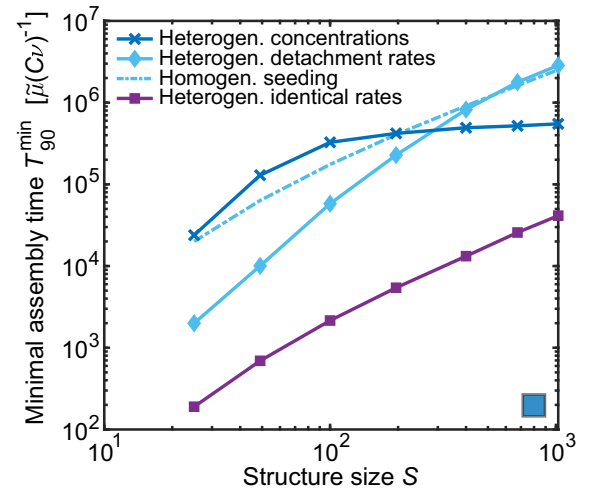

FIG. 21. Heterogeneous concentrations and heterogeneous detachment rates. The dark blue line with crosses shows the minimal assembly time of a heterogeneous system in which species 1, 2, $L + 1$, and $L + 2$ (cf. Fig. 19) are provided in sixfold higher concentration compared to the other species. The bright blue line with diamond markers shows the minimal assembly time when the detachment rates for bonds between species 1, 2, $L + 1$, and $L + 2$ are set to zero. In both cases, data points are obtained by averaging ten independent runs of the stochastic simulation performed with $N = 500$ particles per species and a dimerization barrier $\tilde{\mu} := \mu/\nu = 0.1$ to increase the efficiency of the simulations (cf. Appendix D 1) (note that the y axis is scaled in units of $\tilde{\mu}$ accordingly). For comparison, the blue dash-dotted line shows $T_{90}^{\min}$ for the homogeneous system under seeding determined with the effective model (cf. Appendix G), and the purple line with square markers shows $T_{90}^{\min}$ for the heterogeneous system with identical rates and concentrations [equivalent to the homogeneous system; cf. Fig. 4(b)].

the dash-dotted line in Fig. 21. Note that the minimal assembly time will not increase if hexagonal instead of square-shaped monomers are used, because the growth rate of hexagonal monomers is independent of $\tilde{\delta}_1$.

5. Heterogeneous concentrations

In Ref. [25], the authors describe how the yield in heterogeneous systems can be increased by providing a few neighboring species in significantly larger concentrations compared to the other species. In this way, the authors argued that specific assembly pathways are favored, leading to higher yields as resources are used more effectively. Here, we want to analyze the time efficiency of this strategy. To this end, we set the initial particle number of species 1, 2, $L + 1$, and $L + 2$ to $6N$, while all other species are provided in N copies per species. In order not to restrict the achievable yield, we do not count the additional $20N$ monomers when calculating the yield. Nevertheless, we find that, to achieve an elevated yield of 90%, the detachment rate must be chosen very large,

causing the simulation to become extremely inefficient. We, therefore, again simulate this scenario with a reduced dimerization rate $\mu/\nu = 0.1$ [and plot the assembly time in units of $(C\nu)^{-1}\mu/\nu$] as in the last paragraph. As for heterogeneous detachment rates, providing four species in higher concentrations increases the effective nucleation rate. Consequently, the assembly time increases drastically, since $\tilde{\delta}_1$ must be chosen large to suppress spurious nucleation (cf. Fig. 21). Specifically, the rate of nucleation of species 1, 2, $L + 1$, and $L + 2$ is $\bar{\mu}(6N)^4$, while the nucleation rate of each of the $S - 4$ remaining nucleus species is $\bar{\mu}N^4$. Hence, providing species in heterogeneous concentrations has a similar effect as scaling the (homogeneous) dimerization rate parameter $\tilde{\mu} := \mu/\nu$ as $\tilde{\mu} \to \tilde{\mu}(1 + 6^4/S)$. With Eq. (D11), this suggests that the assembly time scales roughly as $T_{90}^{\min} \sim 6^4 + S$. Hence, as long as S is small compared to 6^4, we expect the minimal assembly time to increase only very slowly with S. Indeed, Fig. 21 shows that the assembly time transiently becomes almost constant as a function of S. In the limit of large S, we expect that the effect of heterogeneous concentrations declines more and more (provided the concentrations remain constant) and the assembly time ultimately behaves similarly as in the homogeneous case (purple line in Fig. 21). Hence, we find that neither the yield nor the time efficiency can be enhanced through heterogeneous concentrations, since heterogeneous concentrations increase the effective nucleation rate, which reduces the efficiency according to Eq. (D11). In their work [25], Murugan, Zou, and Brenner consider a system of one-dimensional rings that self-assemble irreversibly. Furthermore, they work in a regime where the yield is very low. We expect that their finding of a yield increase due to heterogeneous concentrations is associated with these system specificities and has a lot to do with avoiding stochastic effects that arise in irreversible systems [53].

6. Hierarchical heterogeneous self-assembly

We realize that any measures leading to an increase in the effective nucleation rate are counterproductive in enhancing self-assembly efficiency. The key to increasing time efficiency instead lies in enhancing the speed of cluster growth by reducing the attachment order γ. Hierarchical self-assembly, as discussed in Sec. VI, offers a great opportunity in this regard, particularly for heterogeneous systems. For example, we show that rectangular building blocks with six binding sites that self-assemble on gap (cf. Fig. 8 and Sec. VI) are topologically equivalent to hexagons and, thus, self-assemble with the same efficiency as hexagonal monomers. In our heterogeneous system with square-shaped building blocks, this could be realized by making the bonds between every second pair of species [i.e., 1–2; 3–4; etc., and then $(L + 2)$—$(L + 3)$; $(L + 4)$—$(L + 5)$; etc.] significantly stronger than the other bonds.
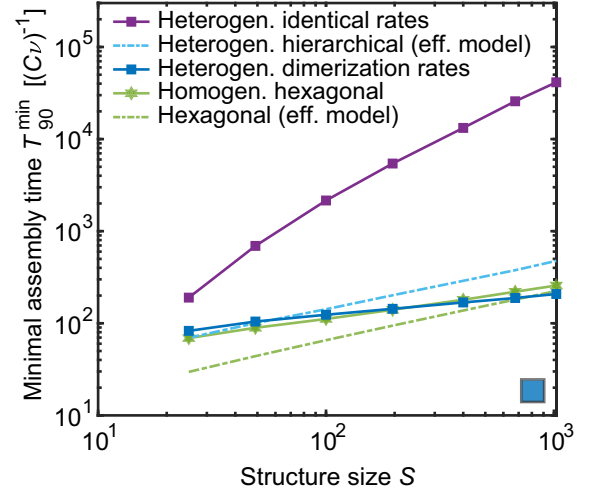

FIG. 22. Improving heterogeneous self-assembly: hierarchical self-assembly and heterogeneous dimerization rates. The blue dash-dotted line shows the minimal assembly time obtained with the effective model for the hierarchical self-assembly scenario in which square-shaped monomers first form rectangle-shaped dimers ($\delta_1^{\text{strong}} = 0$), which then assemble on gap as shown in Fig. 8. Notice that this hierarchical scenario can be realized only with heterogeneous building blocks. The dark blue line with square markers shows the minimal assembly time of a heterogeneous system in which only species 1 and 2 can dimerize with rate $\mu_{12} = \nu$ and the dimerization rates of all other pairs of species as well as the detachment rates are zero. Simulations are performed with $N = 1000$ particles per species and averaged over 10 independent runs. The purple line with square markers shows the assembly time in a heterogeneous system with identical rates for comparison. Furthermore, the green line with star markers and the green dash-dotted line show the assembly time of a homogeneous system with hexagonal monomers (stochastic simulation and effective model, respectively) copied from Fig. 4(b) for comparison. With hierarchical self-assembly or heterogeneous detachment rates, the required assembly time can be reduced drastically.

This will induce the monomers to form rectangle-shaped dimers with six binding sites, which then assemble into the final structures with a minimal time complexity exponent $\theta \approx 1/2$. We simulate this scenario with the effective model (cf. Appendix B) analogous to the hierarchical assembly scenarios in Fig. 9 in the main text. Thereby, we simulate the dimerization of monomers with rate $\mu = \nu$ and detachment rate $\delta_1^{\text{strong}} = 0$ and the subsequent self-assembly of the dimers with the same parameters $\sigma = 3$, $\gamma = 1$ and $\omega = 1/2$ as in the self-assembly of hexagonal monomers. By tuning the detachment rate $\tilde{\delta}_1^{\text{weak}}$ for the bonds between the dimers, we obtain the minimal assembly time plotted in Fig. 22 against the target structure size. Comparing with Fig. 4(b) (or the green dash-dotted line in Fig. 22), we find that the minimal assembly time is roughly a factor of 2 larger than in the case of hexagonal monomers (since the dimers first have to form in the first assembly step).

Hence, the ability to control each bond individually in a heterogeneous system offers plentiful additional possibilities to reduce the attachment order via hierarchical self-assembly and increase the time efficiency relative to a homogeneous system.

7. Heterogeneous dimerization rates

Another possibility to increase self-assembly efficiency is by reducing the nucleation rate. In principle, this can be achieved with heterogeneous dimerization rates. Specifically, let us assume that only species 1 and 2 can form a dimer (with rate $\mu_{12} = \nu$), whereas all other pairs of species are unable to form dimers ($\mu_{ij} = 0$ for $i, j \neq 1, 2$), but the monomers of those species can still attach to existing clusters with rate ν. In nature, such scenarios are quite common. For example, flagellin proteins, the building blocks of flagellae in *Escherichia coli* do not react freely in the cytosol [1]. Instead, the molecules only attach irreversibly to the end of an existing flagellum when they are passed through the flagellum [86].

The fact that only species 1 and 2 can dimerize ensures that just the right number of nuclei will form, and, thus, the detachment rates of all species can be chosen arbitrarily small. Thereby, self-assembly becomes very efficient. Indeed, simulating this scenario with square monomers and detachment rate $\delta_1 = 0$ (cf. Fig. 22), we find that roughly the same time efficiency is achieved as with hexagonal monomers where the detachment rate is chosen optimally; compare with Fig. 4(b). Roughly the same minimal assembly time is also achieved by optimizing the dimerization rate in a homogeneous system (cf. Appendix D 1).

Controlling the dimerization rate is very effective, and, in the heterogeneous case, it does not even require fine-tuning of the rate constants. However, implementing such a scenario with artificial components would be very challenging, since it requires a particular molecular design that causes strong allosteric (cooperative) binding effects or the use of specialized enzymes. In contrast, controlling the ratio $\tilde{\delta}_1$ between detachment and attachment rates is simpler, as it can be tuned by several experimentally accessible parameters (cf. the discussion in Sec. II).

In conclusion, due to the exact equivalence between homogeneous and heterogeneous systems in the ideal case, both types of systems behave very similarly and can be typically related to each other. We found that the scaling properties of the assembly time in heterogeneous systems are robust to noise (e.g., random variations of the rate parameters) and to modifications regarding the boundary conditions. However, the (transient) scaling behavior is affected by concerted modifications, e.g., of the detachment rates or by using heterogeneous concentrations. The resulting effects on the assembly time can qualitatively be well understood on the basis of the results on homogeneous systems.

[1] D. Kushner, *Self-assembly of biological structures*, Bacteriological reviews **33**, 302 (1969).

[2] G. M. Whitesides and B. Grzybowski, *Self-assembly at all scales*, Science **295**, 2418 (2002).

[3] G. M. Whitesides, J. P. Mathias, and C. T. Seto, *Molecular self-assembly and nanochemistry: A chemical strategy for the synthesis of nanostructures*, Science **254**, 1312 (1991).

[4] S. Zhang, *Fabrication of novel biomaterials through molecular self-assembly*, Nat. Biotechnol. **21**, 1171 (2003).

[5] A. C. Mendes, E. T. Baran, R. L. Reis, and H. S. Azevedo, *Self-assembly in nature: using the principles of nature to create complex nanobiomaterials*, Wiley Interdiscip. Rev. **5**, 582 (2013).

[6] C. Sigl, E. M. Willner, W. Engelen, J. A. Kretzmann, K. Sachenbacher, A. Liedl, F. Kolbe, F. Wilsch, S. A. Aghvami, U. Protzer, M. F. Hagan, S. Fraden, and H. Dietz, *Programmable icosahedral shell system for virus trapping*, Nat. Mater. **20**, 1281 (2021).

[7] M. G. Mateu, *Assembly, stability and dynamics of virus capsids*, Arch. Biochem. Biophys. **531**, 65 (2013).

[8] I. G. Johnston, A. A. Louis, and J. P. Doye, *Modelling the self-assembly of virus capsids*, J. Phys. Condens. Matter **22**, 104101 (2010).

[9] F. M. Menger, *The structure of micelles*, Acc. Chem. Res. **12**, 111 (1979).

[10] G. M. Whitesides and M. Boncheva, *Beyond molecules: Self-assembly of mesoscopic and macroscopic components*, Proc. Natl. Acad. Sci. U.S.A. **99**, 4769 (2002).

[11] R. Freeman, M. Han, Z. Álvarez, J. A. Lewis, J. R. Wester, N. Stephanopoulos, M. T. McClendon, C. Lynsky, J. M. Godbe, H. Sangji, E. Luijten, and S. I. Stupp, *Reversible self-assembly of superstructured networks*, Science **362**, 808 (2018).

[12] P. Ceres and A. Zlotnick, *Weak protein-protein interactions are sufficient to drive assembly of hepatitis B virus capsids*, Biochemistry **41**, 11525 (2002).

[13] J. Grant, R. L. Jack, and S. Whitelam, *Analyzing mechanisms and microscopic reversibility of self-assembly*, J. Chem. Phys. **135**, 214505 (2011).

[14] A. Zlotnick, *Are weak protein-protein interactions the general rule in capsid assembly?*, Virology **315**, 269 (2003).

[15] D. C. Rapaport, *Role of reversibility in viral capsid growth: A paradigm for self-assembly*, Phys. Rev. Lett. **101**, 186101 (2008).

[16] M. F. Hagan, *Modeling viral capsid assembly*, Adv. Chem. Phys. **155**, 1 (2014).

[17] J. D. Perlmutter and M. F. Hagan, *Mechanisms of virus assembly*, Annu. Rev. Phys. Chem. **66**, 217 (2015).

[18] W. M. Gelbart, A. Ben-Shaul, and D. Roux, *Micelles, Membranes, Microemulsions, and Monolayers* (Springer Science & Business Media, New York, 2012).

[19] D. Endres and A. Zlotnick, *Model-based analysis of assembly kinetics for virus capsids or other spherical polymers*, Biophys. J. **83**, 1217 (2002).

[20] M. F. Hagan, O. M. Elrad, and R. L. Jack, *Mechanisms of kinetic trapping in self-assembly and phase transformation*, J. Chem. Phys. **135**, 104115 (2011).

[21] E. J. Deeds, J. A. Bachman, and W. Fontana, *Optimizing ring assembly reveals the strength of weak interactions*, Proc. Natl. Acad. Sci. U.S.A. **109,** 2348 (2012).

[22] S. Whitelam and R. L. Jack, *The statistical mechanics of dynamic pathways to self-assembly*, Annu. Rev. Phys. Chem. **66,** 143 (2015).

[23] W. M. Jacobs and D. Frenkel, *Self-assembly of structures with addressable complexity*, J. Am. Chem. Soc. **138,** 2457 (2016).

[24] K. Baek, I. Hwang, I. Roy, D. Shetty, and K. Kim, *Self-assembly of nanostructured materials through irreversible covalent bond formation*, Acc. Chem. Res. **48,** 2221 (2015).

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

[26] F. M. Gartner, I. R. Graf, and E. Frey, *The time complexity of self-assembly*, Proc. Natl. Acad. Sci. U.S.A. **119,** e2116373119 (2022).

[27] K. F. Wagenbauer, C. Sigl, and H. Dietz, *Gigadalton-scale shape-programmable DNA assemblies*, Nature (London) **552,** 78 (2017).

[28] S. Katen and A. Zlotnick, *Methods in Enzymology*, 1st ed. (Elsevier, New York, 2009), Vol. 455, pp. 395–417.

[29] B. A. Grzybowski, C. E. Wilmer, J. Kim, K. P. Browne, and K. J. Bishop, *Self-assembly: From crystals to cells*, Soft Matter **5,** 1110 (2009).

[30] A. Zlotnick, *To build a virus capsid: An equilibrium model of the self assembly of polyhedral protein complexes*, J. Mol. Biol. **241,** 59 (1994).

[31] L. O. Hedges, R. V. Mannige, and S. Whitelam, *Growth of equilibrium structures built from a large number of distinct component types*, Soft Matter **10,** 6404 (2014).

[32] M. F. Hagan and G. M. Grason, *Equilibrium mechanisms of self-limiting assembly*, Rev. Mod. Phys. **93,** 025008 (2021).

[33] J. N. Israelachvili, D. J. Mitchell, and B. W. Ninham, *Theory of self-assembly of hydrocarbon amphiphiles into micelles and bilayers*, J. Chem. Soc., Faraday Trans. 2 **72,** 1525 (1976).

[34] P. C. Hiemenz and T. P. Lodge, *Polymer Chemistry* (CRC Press, Boca Raton, 2007).

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

[36] M. Lenz and T. A. Witten, *Geometrical frustration yields fibre formation in self-assembly*, Nat. Phys. **13,** 1100 (2017).

[37] P. Masson, C. Hulo, E. De Castro, H. Bitter, L. Gruenbaum, L. Essioux, L. Bougueleret, I. Xenarios, and P. Le Mercier, *ViralZone: Recent updates to the virus knowledge resource*, Nucl. Acids Res. **41,** D579 (2012).

[38] C. A. Kerfeld and M. R. Melnicki, *Assembly, function and evolution of cyanobacterial carboxysomes*, Curr. Opin. Plant Biol. **31,** 66 (2016).

[39] M. F. Hagan, *Modeling viral capsid assembly*, Adv. Chem. Phys. **155,** 1 (2014).

[40] Y. Ke, L. L. Ong, W. M. Shih, and P. Yin, *Three-dimensional structures self-assembled from DNA bricks*, Science **338,** 1177 (2012).

[41] B. Alberts, *Molecular Biology of the Cell* (Garland Science, New York, 2017).

[42] A. Ianiro, H. Wu, M. M. van Rijt, M. P. Vena, A. D. Keizer, A. C. C. Esteves, R. Tuinier, H. Friedrich, N. A. Sommerdijk, and J. P. Patterson, *Liquid-liquid phase separation during amphiphilic self-assembly*, Nat. Chem. **11,** 320 (2019).

[43] S. Li, P. Roy, A. Travesset, and R. Zandi, *Why large icosahedral viruses need scaffolding proteins*, Proc. Natl. Acad. Sci. U.S.A. **115,** 10971 (2018).

[44] X. Ye, J. Chen, M. Engel, J. A. Millan, W. Li, L. Qi, G. Xing, J. E. Collins, C. R. Kagan, J. Li *et al.*, *Competition of shape and interaction patchiness for self-assembling nanoplates*, Nat. Chem. **5,** 466 (2013).

[45] H. Fang, M. F. Hagan, and W. B. Rogers, *Two-step crystallization and solid-solid transitions in binary colloidal mixtures*, Proc. Natl. Acad. Sci. U.S.A. **117,** 27927 (2020).

[46] S. Auer and D. Frenkel, *Prediction of absolute crystal-nucleation rate in hard-sphere colloids*, Nature (London) **409,** 1020 (2001).

[47] D. Hayakawa, T. E. Videbaek, D. M. Hall, H. Fang, C. Sigl, E. Feigl, H. Dietz, S. Fraden, M. F. Hagan, G. M. Grason *et al.*, *Geometrically programmed self-limited assembly of tubules using DNA origami colloids*, Proc. Natl. Acad. Sci. U.S.A. **119,** e2207902119 (2022).

[48] M. F. Hagan and D. Chandler, *Dynamic pathways for viral capsid assembly*, Biophys. J. **91,** 42 (2006).

[49] C. G. Evans, J. O'Brien, E. Winfree, and A. Murugan, *Pattern recognition in the nucleation kinetics of nonequilibrium self-assembly*, Nature (London) **625,** 500 (2024).

[50] B. Wei, M. Dai, and P. Yin, *Complex shapes self-assembled from single-stranded DNA tiles*, Nature (London) **485,** 623 (2012).

[51] W. M. Jacobs, A. Reinhardt, and D. Frenkel, *Rational design of self-assembly pathways for complex multicomponent structures*, Proc. Natl. Acad. Sci. U.S.A. **112,** 6313 (2015).

[52] A. Reinhardt, C. P. Ho, and D. Frenkel, *Effects of coordination number on the nucleation behaviour in many-component self-assembly*, Faraday Discuss. **186,** 215 (2016).

[53] F. M. Gartner, I. R. Graf, P. Wilke, P. M. Geiger, and E. Frey, *Stochastic yield catastrophes and robustness in self-assembly*, eLife **9,** e51020 (2020).

[54] C. Chen, C. C. Kao, and B. Dragnea, *Self-assembly of brome mosaic virus capsids: Insights from shorter timescale experiments*, J. Phys. Chem. A **112,** 9405 (2008).

[55] A. Ben-Shaul and W. M. Gelbart, *Statistical thermodynamics of amphiphile self-assembly: Structure and phase transitions in micellar solutions*, in *Micelles, Membranes, Microemulsions, and Monolayers* (Springer, New York, 1994), pp. 1–104.

[56] S. Arrhenius, *Über die reaktionsgeschwindigkeit bei der inversion von rohrzucker durch säuren*, Z. Phys. Chem. **4U,** 226 (1889).

[57] K. J. Laidler, *The development of the Arrhenius equation*, J. Chem. Ed. **61,** 494 (1984).

[58] A. Ben-Ari, L. Ben-Ari, and G. Bisker, *Nonequilibrium self-assembly of multiple stored targets in a dimer-based system*, J. Chem. Phys. **155,** 234113 (2021).

[59] M. Faran and G. Bisker, *Nonequilibrium self-assembly time forecasting by the stochastic landscape method*, J. Phys. Chem. B **127**, 6113 (2023).

[60] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, England, 2010).

[61] D. T. Gillespie, *Stochastic simulation of chemical kinetics*, Annu. Rev. Phys. Chem. **58**, 35 (2007).

[62] Note that, in the system with triangle-shaped monomers, the attachment order for the first assembly step following nucleation is larger, $\gamma = 4$, compared to all subsequent assembly steps, where $\gamma = 3$.

[63] Since cluster growth is by the addition of monomers, stochastic effects result in some interface roughness (akin to Eden growth [87]) leading to a larger surface-to-volume ratio than given by $p_s = a\sqrt{s}$ and, hence, to an effective growth exponent $\omega > 1/2$. Since this affects only the system with hexagon-shaped monomers, we do not discuss this additional effect here further and simply approximate ω by $1/2$ for hexagon-shaped monomers.

[64] Unlike for hexagon-shaped building blocks, in the case of square and triangular building blocks, the prefactor a additionally includes the multiplicity q_s of the seed-forming reaction pathways. For example, the multiplicity considers that the seed can typically grow in two directions unless it forms directly at the corner of an edge.

[65] The combinatorial factors f_s for clusters of small size s typically deviate from the asymptotic form as^ω. Note, however, that, for increasing target size S, the approximation becomes ever more accurate, as the asymptotic form of f_s applies to an increasingly large fraction of all assembly steps.

[66] In general, $\beta > \alpha$ holds if $\sigma > \gamma + 1$ and $\omega < 3/2$, meaning that this must be true, in general.

[67] The latter is particularly pronounced in the system with hexagonal monomers, where, due to the absence of a domino effect that fills up the edges of a cluster after an initial seed has formed, the typical morphologies of the cluster exhibit some interface roughness during growth. Note that the assembly time obtained from integrating the effective model in which we use $\omega = 1/2$ coincides with the analytic estimate $\theta_{\text{th}} = 1/2$ for the system with hexagonal building blocks. Furthermore, using a value of ω slightly larger than $1/2$ in Eq. (17) (which would be plausible due to the roughening effect) consistently results in a value of θ smaller than $1/2$ as observed from the simulation data. For example, the measured value $\theta \approx 0.4$ is obtained with $\omega = 0.6$ according to Eq. (17).

[68] A. Zlotnick, J. M. Johnson, P. W. Wingfield, S. J. Stahl, and D. Endres, *A theoretical model successfully identifies features of hepatitis B virus capsid assembly*, Biochemistry **38**, 14644 (1999).

[69] A. Y. Morozov, R. F. Bruinsma, and J. Rudnick, *Assembly of viruses and the pseudo-law of mass action*, J. Chem. Phys. **131**, 155101 (2009).

[70] Note that some of the approximations used here actually become exact in the limit of large S. For example, neglecting the diffusion term and higher-order terms in the advection equation is irrelevant in the limit $S \to \infty$, as their contribution becomes negligible against the leading-order advection term [see Eqs. (C4) and (C5)], showing that the advective contribution to cluster growth equals the square of the diffusive contribution. Similarly, neglecting the nucleation term in Eq. (22) becomes increasingly irrelevant as S increases.

[71] A. Zlotnick and S. Mukhopadhyay, *Virus assembly, allostery and antivirals*, Trends Microbiol. **19**, 14 (2011).

[72] C. Packianathan, S. P. Katen, C. E. Dann III, and A. Zlotnick, *Conformational changes in the hepatitis B virus core protein are consistent with a role for allostery in virus assembly*, J. Virol. **84**, 1607 (2010).

[73] E. Bekyarova, Y. Ni, E. B. Malarkey, V. Montana, J. L. McWilliams, R. C. Haddon, and V. Parpura, *Applications of carbon nanotubes in biotechnology and biomedicine*, J. Biomed. Nanotechnol. **1**, 3 (2005).

[74] M. F. De Volder, S. H. Tawfick, R. H. Baughman, and A. J. Hart, *Carbon nanotubes: Present and future commercial applications*, Science **339**, 535 (2013).

[75] S. K. Kulkarni and S. K. Kulkarni, *Nanotechnology: Principles and Practices* (Springer, New York, 2015).

[76] We remark that the precondition for the applicability of our theory is that the growth rate of large clusters consistently scales as $\tilde{r}_{s \to s+1} \sim s^\omega$ with some arbitrary growth exponent ω.

[77] J. E. Baschek, H. C. R Klein, and U. S. Schwarz, *Stochastic dynamics of virus capsid formation: Direct versus hierarchical self-assembly*, BMC Biophys. **5**, 1 (2012).

[78] C. Li, J. C.-Y. Wang, M. W. Taylor, and A. Zlotnick, *In vitro assembly of an empty picornavirus capsid follows a dodecahedral path*, J. Virol. **86**, 13062 (2012).

[79] S. Flasinski, A. Dzianott, J. A. Speir, J. E. Johnson, and J. J. Bujarski, *Structure-based rationale for the rescue of systemic movement of brome mosaic virus by spontaneous second-site mutations in the coat protein gene*, J. Virol. **71**, 2500 (1997).

[80] S. J. Hanslip, N. R. Zaccai, A. P. Middelberg, and R. J. Falconer, *Assembly of human papillomavirus type-16 virus-like particles: Multifactorial study of assembly and competing aggregation*, Biotechnol. Prog. **22**, 554 (2006).

[81] D. Willits, X. Zhao, N. Olson, T. Baker, A. Zlotnick, J. Johnson, T. Douglas, and M. Young, *Effects of the cowpea chlorotic mottle bromovirus β-hexamer structure on virion assembly*, Virology **306**, 280 (2003).

[82] M. A. Boettcher, H. C. Klein, and U. S. Schwarz, *Role of dynamic capsomere supply for viral capsid self-assembly*, Phys. Biol. **12**, 016014 (2015).

[83] There are either no or only a small number of successive lower-order processes for the assembly steps immediately following nucleation. Hence, the growth rates for small clusters behave a bit differently, as we discuss below, specifically for triangle-shaped building blocks. In the other cases, however, good agreement with the simulation data can be obtained even if we do not treat the initial assembly steps separately and assume the same growth rate for them as for larger clusters.

[84] P. E. Prevelige, Jr., D. Thomas, and J. King, *Nucleation and growth phases in the polymerization of coat and scaffolding subunits into icosahedral procapsid shells*, Biophys. J. **64,** 824 (1993).

[85] Note also that explicit boundaries violate the equivalence between homogeneous and heterogeneous systems (cf. Appendix J 1) due to the distinct properties of the boundary species.

[86] C. J. Jones and S.-I. Aizawa, *The bacterial flagellum and flagellar motor: Structure, assembly and function*, Advances in Microbial Physiology **32,** 109 (1991).

[87] E. Murray, *A two-dimensional growth process*, in *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability* (University of California Press, Berkeley, 1960), Vol. 4, p. 223.