Letter

Entropic size control of self-assembled filaments

Maximilian C. Hübl* and Carl P. Goodrich†
Institute of Science and Technology Austria (ISTA), Am Campus 1, 3400 Klosterneuburg, Austria

(Received 7 July 2025; accepted 6 January 2026; published 5 March 2026)

Controlling the size and shape of assembled structures is a fundamental challenge in self-assembly and is highly relevant in material design and biology. Here, we show that specific but promiscuous short-range binding interactions make it possible to economically assemble linear filaments of user-defined length. Our approach leads to independent control over the mean and width of the filament size distribution and allows us to smoothly explore design trade-offs between assembly quality (spread in size) and cost (number of particle species). We employ a simple hierarchical assembly protocol to minimize assembly times and show that multiple stages of hierarchy make it possible to extend our approach to the assembly of higher-dimensional structures. Our work provides a conceptually simple solution to size control that is applicable to a broad range of systems, from DNA nanoparticles to supramolecular polymers and beyond.

DOI: 10.1103/68rs-3qgn

I. INTRODUCTION

One of the hardest qualities to control in the design of self-assembling structures is also among the simplest: size. Nevertheless, achieving precise control over a structure's size—the number of subunits in the structure—is extremely important in many areas, from biology [1–4] and biomedical applications [5–7] to photonic materials [8–12] and nanofabrication [13–16], because size is often directly tied to the function of the assembled object. To control the function, one must first control the size.

The challenge is relating the global property of structure size to the local properties of the assembling particles: How should the particles "know" when to stop aggregating, and how can that knowledge be communicated via the particles' interactions? In some approaches to this problem, information is encoded in the particle geometry, which can lead to structures that self-close on controllable length scales. These strategies have been successfully demonstrated in the assembly of sheets, tubules, shells, or even more complex manifolds [7,17–20]. Other approaches construct interactions such that the energy of a structure contains terms that scale differently with size, leading to a preferred size determined by the competition of these terms. Such cooperative effects can be mediated through long-range repulsion [21,22], geometric frustration [23–25], or even frustration between crystalline domains [26].

However, such approaches are particularly challenging for one-dimensional (1D) structures such as linear filaments, as they allow only limited cooperativity between particles, constraining the potential mechanisms through which information can be transmitted from one end of the filament to the other. Le Roy *et al*. [25] demonstrate a powerful, yet experimentally demanding approach to 1D size control through geometric frustration with deformable particles, while other approaches exploit complex, nonequilibrium mechanisms that play an important role in biology [27–31]. Another potential approach is through multiple particle species with specific interactions. Whereas generic single-species assembly leads to a wide distribution of filament lengths [Fig. 1(a)], Fig. 1(b) shows a scheme for designing particles to assemble in a precisely ordered sequence. However, while this can lead to size control [32], it requires as many particle species as the desired filament is long, which makes experimental realizations highly costly and impractical, even for only moderately long filaments.

Here, we show how to achieve the size-control benefits of such addressable assembly, but with a fixed and reasonable number of species by exploiting configurational entropy. Using multiple particle species with programmable interactions, we assemble filaments whose length can be freely tuned by adjusting particle concentrations or binding energies—independently of the number of particle species. This tunability allows us to smoothly explore design trade-offs between assembly quality (spread in size) and cost (number of particle species), where nonaddressable assembly (low cost, low quality) and fully addressable assembly (high cost, high quality) are recovered as limiting cases. We find that high-quality assembly of long filaments requires long assembly times, and we propose a simple hierarchical assembly protocol that can speed up the assembly process by over 5 orders of magnitude, thereby making high-quality assembly feasible on experimentally accessible timescales. Since our approach does not rely on any secondary interactions beyond specific bond formation, it is directly applicable to many programmable self-assembly platforms, such as

*Contact author: maximilian.huebl@ist.ac.at
†Contact author: carl.goodrich@ist.ac.at

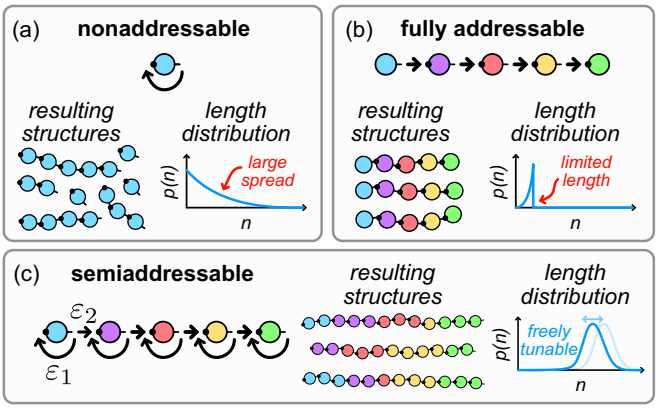

FIG. 1. Different filament assembly strategies. (a) Nonaddressable assembly with short-ranged interactions leads to a very broad length distribution. (b) Fully addressable assembly makes it possible to target a specific length, but that length is limited by the number of particle species, making this strategy not scalable. (c) Semiaddressable assembly, where particles bind promiscuously, makes it possible to engineer a narrow and tunable distribution of structure sizes. In all sketches, black arrows indicate the binding rules, i.e., the specific bonds that are allowed to form.

DNA-based systems [11,18,20,33–39], systems with shape-complementary interactions [40,41], and certain *de novo* proteins [42,43] or other supramolecular polymers [44–46].

II. A TUNABLE EQUILIBRIUM SIZE DISTRIBUTION

The fundamental quantity of interest in this Letter is the equilibrium length distribution $p(n)$ of the filaments, and it is our goal to design the mean $\langle n\rangle$ and width σ of this distribution. The challenge of size control becomes apparent if we look at nonaddressable (single species) filaments, as shown in Fig. 1(a). The equilibrium length distribution can be computed from the structure partition functions [32,47–52] and is given by

$$p_{\mathrm{na}}(n) = \frac{e^{-\lambda n}}{C_{\mathrm{na}}}, \tag{1}$$

where $\lambda = -\beta(\varepsilon + \mu)$ depends on the binding energy ε, chemical potential μ, and inverse temperature $\beta = 1/kT$, and where the normalization constant C_{na} is proportional to the partition function of the system (see Appendix A).

The important point is that the equilibrium assembly of a single-particle species is characterized by a single control parameter $\lambda > 0$ (nonpositive values are prohibited, as these would lead to diverging particle concentrations [4]). Adjusting λ allows one to tune the characteristic length of the aggregates, but since the length distribution is exponential [53], the width of the distribution is always comparable to the mean $\sigma \sim \langle n\rangle$. Achieving size control, i.e., lowering σ below this baseline, is not possible without some additional design strategy [4,54]. As discussed above, one possibility is to introduce multiple particle species with specific interactions, which allows one to precisely select the length of the filaments [32,54] [Fig. 1(b)]. However, a direct, fully addressable design requires the number of species and interactions to grow linearly with the desired target length, making this approach costly and impractical in experiments [55].

To avoid both of these limitations, we now introduce addressability in a more careful way. The resulting "semiaddressable" design [51] will allow us to freely tune the mean and width of the length distribution. Our design consists of m particle species that each can bind to themselves to form single-species filaments, just as the nonaddressable system discussed above. However, there are additional interactions that allow the "right side" of particle species i to bind to the "left side" of species $i+1$, so that the single-species filaments can be joined sequentially, as shown in Fig. 1(c).

The idea behind this design becomes clear if we imagine for the moment that the assembly proceeds hierarchically, so that the m species initially bind only to themselves and exclusively form single-species filaments. Once the single-species filaments have reached equilibrium, we freeze them and turn on the cross-species interactions, letting the single-species filaments combine in groups of up to m to form multispecies filaments. The length of a multispecies filament is the sum of m single-species filament lengths, which are uncorrelated and exponentially distributed random variables, as discussed above. Appealing to the central limit theorem, we may therefore expect that the length distribution of the multispecies filaments approaches a normal distribution as the number of species m is increased. If this is the case, the mean and variance of the filament length can be precisely and independently controlled.

We now make these statements rigorous and show that this design works as intended, even if the assembly does not proceed hierarchically. We make the simplifying assumption that all m particle species are supplied at the same chemical potential μ, and we denote the same-species binding energy by ε_1 and the cross-species binding energy by ε_2. We can immediately write down the length distribution by noting that the probability of observing a length-n filament is proportional to the number of ways it can be decomposed into k single-species filaments. Summing over all the ways k sequential species can be picked out of m species, and keeping track of the binding energies and chemical potentials, leads us to the semiaddressable length distribution:

$$p_{\mathrm{sa}}(n) = \frac{e^{-\lambda n}}{C_{\mathrm{sa}}} \sum_{k=1}^{m} (m-k+1) e^{k\beta\Delta\varepsilon} \binom{n-1}{k-1}, \tag{2}$$

where $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$ and $\lambda = -\beta(\varepsilon_1 + \mu)$.

Analytic expressions for the normalization constant C_{sa}, mean $\langle n\rangle$, and standard deviation σ of $p_{\mathrm{sa}}(n)$ can be found in Appendix A. We can gain valuable intuition by considering the limit where the mean $\langle n\rangle$ of the distribution is much larger than the number of species m. In this limit, we can neglect partially assembled filaments (i.e., filaments that do not contain all m species), and the general distribution can be simplified considerably. Specifically, if

$$\langle n\rangle \gg 2me^{-\beta\Delta\varepsilon}, \tag{3}$$

the assembly is dominated by "complete" filaments, and we can neglect all but the $k = m$ term in the sum of Eq. (2). To simplify further, we also approximate the binomial coefficient as $\binom{n-1}{k-1} \approx (n - k/2)^{k-1}/(k-1)!$, after which the length

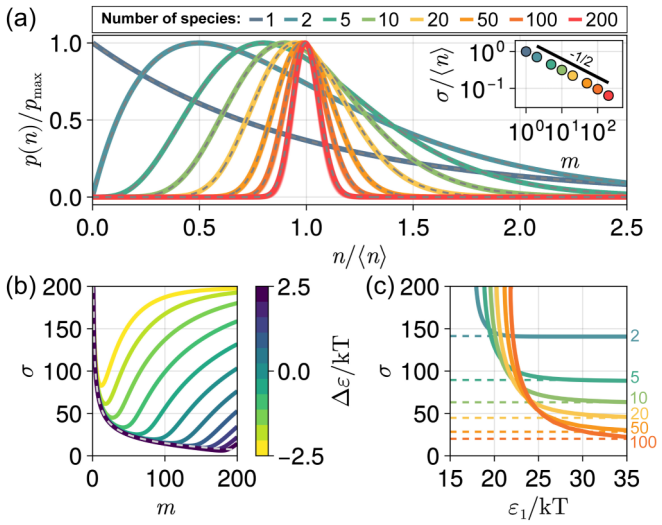

FIG. 2. Equilibrium properties of semiaddressable filaments. (a) Scaled equilibrium length distributions for different numbers of particle species m, with target lengths range from $\langle n\rangle = 500$–1000. The scaled distributions corresponding to different target lengths lie on top of one another and are excellently approximated by the limiting distribution given by Eq. (4) (dashed gray lines). For ease of viewing, all distributions are scaled by their maximum value. The inset shows how the relative peak width scales with m (colored points); the black line shows theoretical $m^{-1/2}$ scaling. The total particle concentration is $\rho_{\text{tot}} = 0.1\phi_0$ and $\Delta\varepsilon = 0$. (b) Peak width as a function of the number of species m, for different $\Delta\varepsilon$ at fixed $\langle n\rangle = 200$. The dashed gray line shows the limiting expression for σ given in Eq. (5). (c) Peak width as a function of the same-species binding energy ε_1 at fixed $\langle n\rangle = 200$, for different species m (indicated by color and the numbers on the right). Dashed lines show the approximate limiting peak widths $\bar{\sigma} = \langle n\rangle/\sqrt{m}$.

distribution assumes the simpler form

$$p_{\text{sa}}(n) \approx \begin{cases} 0, & \text{if } n < \frac{m}{2}, \\ \frac{\lambda^m}{(m-1)!}\left(n - \frac{m}{2}\right)^{m-1} e^{-\lambda\left(n-\frac{m}{2}\right)}, & \text{else.} \end{cases} \tag{4}$$

This limiting distribution is an *Erlang* distribution, shifted by $m/2$. The *Erlang* distribution describes sums of exponentially distributed variables, so it is not surprising that our general distribution reduces to it in the limit where all filaments are complete. Note that the limiting distribution is independent of the cross-species binding energy ε_2 because we are in a regime where filaments always contain $m-1$ cross-species bonds. The mean and standard deviation of the limiting distribution have the convenient analytical expressions:

$$\langle n\rangle = m\left[\frac{1}{2} + \frac{1}{\lambda}\right], \quad \sigma = \frac{\sqrt{m}}{\lambda}. \tag{5}$$

Figure 2(a) shows that the semiaddressable length distribution has a pronounced peak, whose width decreases as the number of species increases, confirming the effectiveness of our design. Moreover, in the large-$\langle n\rangle$ limit, the distributions for different target lengths can be collapsed onto a single curve; Fig. 2(a) shows the scaled length distribution for different numbers of particle species. We view the ratio of standard deviation to mean, $\sigma/\langle n\rangle$, which we call the "relative peak width," as a measure of the quality of the assembly—the smaller the relative peak width, the higher the quality. The inset in Fig. 2(a) confirms that the relative peak width scales as $m^{-1/2}$, which is expected from the central limit theorem and also follows directly from Eq. (5) if $\langle n\rangle \gg m$.

Figure 2(a) makes it clear that our approach allows us to freely tune the average filament length independently of the number of species. This is in contrast to fully addressable assembly, where the number of species imposes a hard limit on the achievable length. Here, we can vary particle concentrations or binding energies to adjust the mean of the distribution, while adding more species to the system increases the quality of the assembly. To put this in perspective: in DNA origami-based systems, for example, it is possible to create over 20 distinct particle species [17,18], which would lead to a size dispersity of roughly 20%.

Equation (3) implies that the value of $\Delta\varepsilon$ significantly affects whether the large-$\langle n\rangle$ limit can be reached. The consequences of this are shown in Fig. 2(b), where the peak width is shown as a function of the number of species for different values of $\Delta\varepsilon$ and at constant $\langle n\rangle = 200$. For small $\Delta\varepsilon$, the assembly quality deteriorates if too many species are used, since $\Delta\varepsilon$ controls how favorable it is for different single-species filaments to join together. If $\Delta\varepsilon$ is too low, not all single-species filaments can aggregate, and the assembly is dominated by incomplete filaments that contain fewer than m segments, negatively impacting assembly quality. On the other hand, if $\Delta\varepsilon$ is too high, the time it takes for filaments to break apart becomes excessively long, and the filaments can become kinetically arrested.

Another way to look at this effect is shown in Fig. 2(c), where the peak width is shown as a function of ε_1, for $\langle n\rangle = 200$ and various m. As ε_1 is increased, the peak width decreases to a minimal value that depends on the number of species, roughly given by $\bar{\sigma} = \langle n\rangle/\sqrt{m}$ (as long as $\langle n\rangle \gg m$). However, if energies cannot be raised high enough to achieve optimal σ, it is more favorable to employ a lower number of particle species. For example, as shown in Fig. 2(c), using 100 instead of 50 species results in a lower peak width only if energies are above $\sim$25 kT. Since high binding energies are often associated with long equilibration times, this need for high energies suggests a trade-off between assembly quality and assembly time, which we will explore in the "Assembly kinetics and the resulting design trade-offs" section.

III. SIZE CONTROL IN PRACTICE

To better understand how our theoretical results can be applied in practice, we now walk through an explicit example, with the goal of assembling size-controlled filaments of length $\langle n\rangle = 100$ using $m = 10$ particle species, which is experimentally feasible with current capabilities in DNA origami nanoparticles [18], for example.

In our grand-canonical description, the equilibrium length distribution is controlled by four quantities: the number of species m, the two binding energies ε_1 and ε_2, and the chemical potential μ. Given these four quantities, we can determine the mean filament length $\langle n\rangle$, the peak width σ, and the total particle concentration ρ_{tot}, which are given by rather

complicated nonlinear equations, Eqs. (A11)–(A13) in Appendix A. (The equations for $\langle n\rangle$ and σ reduce to the simple limiting expression Eq. (5), for high $\langle n\rangle$ and/or $\Delta\varepsilon$.) To find the parameters required to achieve a desired assembly outcome, we need to numerically invert Eqs. (A11)–(A13). Since we have three equations and seven quantities in total, we need to fix four of the quantities in order to solve for the remaining three.

In our example, we want to find the binding energies required to assemble filaments of length $\langle n\rangle = 100$ with $m = 10$ particle species. Thus, we need to fix two more quantities to close the system of equations. Since experiments and simulations are generally performed with fixed particle concentration (canonical ensemble), we pick a total particle concentration $\rho_{\mathrm{tot}} = 0.1\phi_0$, where ϕ_0 is the reference concentration described in Appendix A. Finally, we also fix the desired peak width σ, which we choose to be equal to $\bar{\sigma} = \langle n\rangle/\sqrt{m} \approx 31.6$, close to the maximum quality achievable with $m = 10$ particle species. We can now use Eqs. (A11)–(A13) to solve for the required binding energies and chemical potential, leading to $\varepsilon_1 \approx 29.6\,\mathrm{kT}$, $\varepsilon_2 \approx 29.1\,\mathrm{kT}$, and $\mu \approx -29.7\,\mathrm{kT}$.

Whether these energies can be realized in practice depends on the experimental system at hand. While achieving binding energies around and above 30 kT is feasible using DNA nanotechnology [56], other experimental platforms may not be able to reach energies as high. These systems will be limited to assemblies with shorter filaments or a broader length distribution.

Finally, to realize a given design in experiments or simulations, we need to compute the particle concentrations of every species. The concentrations are fully determined by our assumption that all species are supplied at the same chemical potential μ. However, even if the chemical potentials are uniform, the concentrations are generally not. We calculate the concentrations analytically via Eq. (C3) in Appendix C. Once the binding energies and particle concentrations are computed, the filament design is ready to be realized in practice.

IV. ASSEMBLY KINETICS AND THE RESULTING DESIGN TRADE-OFFS

The preceding discussion and the data shown in Fig. 2(c) show that assembling long filaments with high quality can require binding energies greater than 30 kT. This suggests important practical limitations of our approach. Since the binding energies determine the timescale on which bonds can break and the system can equilibrate, we expect to find an unavoidable trade-off between filament length, quality, and assembly time.

To explore this trade-off, we investigate the assembly kinetics of our $\langle n\rangle = 100$ example from the previous section by performing Markov-based kinetic simulations, as described in Appendix B. The results of these simulations are shown in Fig. 3(a), which shows the time-dependent average ensemble length $\langle n\rangle(t)$. We perform multiple simulations corresponding to different designs, each targeting a different peak width. For each target peak width, we solve for the required binding energies and particle concentrations, as described in the previous section. We use the equilibration time of the nonaddressable system, τ_{na}, as a reference timescale (see Appendix B and the Supplemental Material [57] for details).

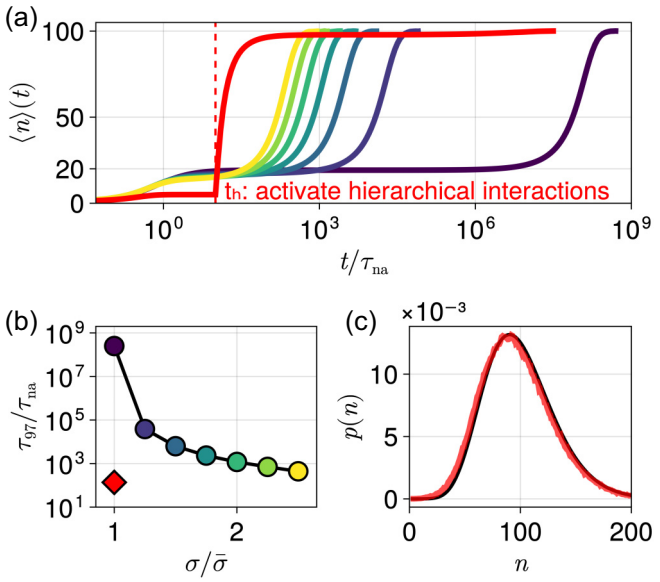

FIG. 3. Kinetics of semiaddressable filaments. (a) Time-dependent average ensemble length for $m = 10$ species and a target length of $\langle n\rangle(t = \infty) = 100$, for target peak widths $\sigma/\bar{\sigma}$ ranging from 1 (dark purple) to 2.5 (yellow), which are achieved by varying the binding energies. We use $\bar{\sigma} = \langle n\rangle/\sqrt{m}$, which is approximately equal to the peak width of the limiting distribution, Eq. (5), as a reference. The red curve shows hierarchical assembly, targeting $\sigma/\bar{\sigma} = 1$. Time is measured in units of the nonaddressable equilibration time τ_{na}. (b) Assembly time τ_{97} as a function of the target peak width for direct assembly (circles) and hierarchical assembly (diamond), showing the trade-off between assembly quality and time. For non-hierarchical systems, colors indicate peak width and are identical to the ones used in the main panel. (c) Filament length distribution for hierarchical assembly at τ_{97} (red) compared to the equilibrium length distribution (black).

Figure 3(a) shows that the assembly time of an addressable multispecies system can be orders of magnitude higher compared to a nonaddressable system, and it confirms that the target peak width σ has a strong impact, since smaller σ requires higher binding energies (see Table S1 in the Supplemental Material [57]).

Illustrating this further, we can obtain a rough estimate of the equilibration times for a typical experimental system. For nanoparticles with $a = 20\,\mathrm{nm}$ radius, their Smoluchowski aggregation rate is given by $\alpha = 16\pi\, Da$, where $D \approx 10^{-11}\,\mathrm{m^2/s}$ is the particle diffusion constant. Using $\tau_{\mathrm{na}} = \langle n\rangle/(\alpha\rho_{\mathrm{tot}})$ (see Appendix B) at a total concentration of $\rho_{\mathrm{tot}} = 10\,\mathrm{nM}$, this results in $\tau_{\mathrm{na}} \approx 1\,\mathrm{s}$, implying that the assembly of the higher quality assemblies shown in Fig. 3(a) becomes difficult on experimentally accessible timescales. Moreover, this estimate of the equilibration time was made under the assumption of isotropic particles and equal aggregation rates among all filaments, irrespective of their lengths. The aggregation rate in real systems generally decreases with increasing structure size; actual equilibration times are likely significantly greater than this basic estimate.

To promote the assembly of high-quality distributions on experimentally feasible timescales, we now consider a hierarchical strategy similar to the one we used to motivate our design in the first place: We first form the single-species filaments and only later turn on the cross-species interactions that allow filaments to join sequentially. In this hierarchical approach, we initially choose ε_1 such that all m species assemble single-species filaments of average length $\langle n_{\mathrm{na}}\rangle = \langle n\rangle/(2m)$ [58], while disallowing any cross-species bonds from forming. Here, $\langle n\rangle = 100$ is the desired eventual average length of the fully assembled filaments. After a time $t_{\mathrm{h}} \geqslant \tau_{\mathrm{na}}$, we turn on the cross-species interactions and set both ε_1 and ε_2 to the values required for equilibrium assembly with the desired $\langle n\rangle$ and σ. Experimentally, such a rapid change in interactions could, for example, be achieved through a change of temperature [59,60] or through the addition of DNA linkers [61,62]. While we assume here that the switching of interactions is instantaneous, in the Supplemental Material [57], we investigate the effects of a finite switching duration and find that the instantaneous approximation is valid as long as the switching timescale is small compared to τ_{na}.

The hierarchical assembly kinetics are shown as the red curve in Fig. 3(a), demonstrating that hierarchical assembly leads to a drastic speed-up compared to all-at-once assembly. Figure 3(b) shows the measured assembly time, τ_{97}, defined as the first time at which $\langle n\rangle(\tau_{97}) = 0.97\langle n\rangle$, as a function of the target peak width, showing a stronger-than-exponential dependence for nonhierarchical assembly. Compared to this, hierarchical assembly times are more than 5 orders of magnitude lower.

Note though that the hierarchical system is not completely equilibrated at τ_{97}. Once the hierarchical step at t_{h} is completed, the hierarchical system assembles under the same conditions as the highest quality nonhierarchical system [dark purple curve in Fig. 3(a)], and the final relaxation to $\langle n\rangle = 100$ is thus governed by a similar long timescale. However, hierarchical assembly gets us very close to the equilibrium state very quickly—for most practical purposes, the hierarchical length distribution at τ_{97} will be indistinguishable from the equilibrium distribution, Eq. (2), as shown in Fig. 3(c). In the Supplemental Material [57], we also consider a stronger form of hierarchy, where we freeze the single-species filaments after the initial equilibration period, which yields similar results.

V. DISCUSSION

Our results show that size control of linear filaments can be achieved by carefully designing specific, short-range interactions between an experimentally reasonable number of species, following the semiaddressable scheme presented in Fig. 1(c). This leads to an equilibrium distribution of filament lengths whose mean and width can be freely tuned. We have identified two fundamental trade-offs that govern optimal design. The first trade-off is between quality and cost: As shown by Fig. 2(a), higher cost (higher number of species m) leads to higher quality (lower $\sigma/\langle n\rangle$). The second trade-off is between quality and equilibration time: For fixed cost (fixed m), both quality and equilibration times depend on binding energies, with higher quality requiring higher equilibration times (Fig. 3). This second trade-off can be mitigated through a simple hierarchical protocol, as discussed.

In Appendix D, we perform a sensitivity analysis of the example shown in Fig. 3, where we investigate the effects of noisy binding energies and particle concentrations. We find that the general shape of the length distribution is robust against perturbations. However, setting ε_1 and ε_2 precisely are essential to achieving the desired mean of the length distribution. Moderate cross-talk interactions and variations in particle concentrations do not significantly affect the results.

In our approach, size control is achieved through a competition between bulk free energy and configurational entropy. This entropy arises from the possible arrangements of the domain walls separating different particle species: Longer filaments allow for a larger number of domain wall configurations [which are counted by the binomial coefficient in Eq. (2)], and are therefore entropically favored. On the other hand, adding particles to a filament comes with a free energy penalty quantified by λ. The balance between these two effects, which scale differently with structure size, leads to a preferred filament length that is continuously tunable by varying binding energies or particle concentrations.

We have focused our analysis on the assembly of one-dimensional filaments, which pose a particular challenge for many other strategies for size control. However, our approach could be extended to form higher-dimensional structures with controlled size and shape. In Appendix E, we show how pairing our approach with additional layers of hierarchical assembly makes it possible to control the size and shape of two-dimensional sheets. In short, we introduce "vertical" interactions so that size-controlled filaments stack on top of each other, which allows us to iterate our approach and also achieve size control in the second dimension. By introducing multiple bond types for these vertical interactions, the height h and width w of the 2D sheet can be controlled in a way that is closely analogous to the 1D case. We demonstrate this in Fig. 4, which shows an example where $\langle w\rangle \approx \langle h\rangle \approx 100$, and with $\sigma_w/\langle w\rangle \approx 0.05$ and $\sigma_h/\langle h\rangle \approx 0.38$. This is only one of many potential ways our approach could be generalized to higher dimensions, and we expect that exploiting the trade-off between binding energy and configurational entropy will be a fruitful approach to steering self-assembly outcomes in a variety of contexts.

ACKNOWLEDGMENTS

We thank Maitane Muñoz-Basagoiti for helpful discussions. The research was supported by the Gesellschaft für Forschungsförderung Niederösterreich under Project No. FTI23-G-011.

DATA AVAILABILITY

The data that support the findings of this article are openly available [63].

APPENDIX A: THE EQUILIBRIUM DISTRIBUTION

Here, we provide a detailed derivation of the semiaddressable length distribution [Eq. (2)]. As in the main text, we

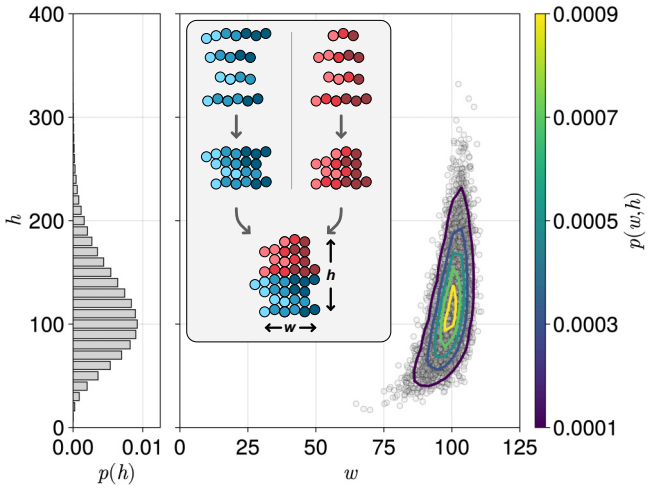

FIG. 4. Size distribution of hierarchically assembled two-dimensional sheets. The main panel shows the probability $p(w, h)$ of observing a sheet with width w and height h, where width is defined as the mean length of the constituent filaments and height is given by the number of filaments in a sheet. Data points correspond to individual sheets and contour lines show a two-dimensional histogram. The panel on the left shows the marginal distribution of sheet heights $p(h)$. The inset shows a cartoon sketch of sheet assembly, starting from two filament species (shown in blue and red, different shades indicate the different particle species needed for the initial filament assembly), each with the same length distribution but distinct vertical binding sites. These filaments first come together to assemble single-species sheets, which then, in turn, combine into multispecies sheets.

assume that all particle species are supplied at equal chemical potential μ and that all bonds between the same species have the same binding energy ε_1, and that bonds between different species have binding energy ε_2. Particles of species i may only bind to other particles of species i, or to particles of species $i+1$. Under these conditions, the partition function of any filament of length n that consists of k different single-species segments is given by

$$Z(n,k) = V\,\Omega(n,k)\phi_0^n e^{\beta[n\mu+(n-k)\varepsilon_1+(k-1)\varepsilon_2]}, \tag{A1}$$

where V is the system volume, ϕ_0 is an arbitrary reference concentration, and $\Omega(n,k)$ is the rotational and vibrational partition function of the filament, which depends on the binding interactions between the particles. If we neglect interactions between non-neighboring particles, then a filament is one-particle reducible [64], so that the entropy of a filament can be written as

$$\Omega(n,k) = 8\pi^2\,\omega^{n-1}, \tag{A2}$$

where ω is the per-bond entropy (assumed, for simplicity, to be the same for both bond types), and the factor of $8\pi^2$ is due to center-of-mass rotations of the entire chain (in three dimensions). The per-bond entropy ω, which depends on the microscopic interactions between the particles, can then simply be absorbed into the binding energies,

$$\varepsilon_{1/2} \to \varepsilon_{1/2} + \frac{1}{\beta}\log(\omega\phi_0), \tag{A3}$$

making our description agnostic to the system-dependent microscopic details.

To obtain the partition function of all structures of length n, we need to count the ways in which a length-n filament can be subdivided into k segments, which is given by $\binom{n-1}{k-1}$. Moreover, we need to keep in mind that the segments need not start at species 1 and end at species m, so we need to sum over all the $m-k+1$ ways we can select k contiguous species. Combining these considerations with Eq. (A1), and defining $\Delta\varepsilon = \varepsilon_2 - \varepsilon_1$ and $\lambda = -\beta(\varepsilon_1+\mu)$, gives the partition function for an arbitrary length-n filament

$$\frac{Z(n)}{8\pi^2 V\phi_0} = e^{-\lambda n-\beta\varepsilon_2}\sum_{k=1}^{m}(m-k+1)\binom{n-1}{k-1}e^{k\beta\Delta\varepsilon}. \tag{A4}$$

The probability of observing a filament of length n is proportional to $Z(n)$, so to obtain the length distribution, we need to compute the normalization constant

$$Z = \sum_{n=1}^{\infty} Z(n). \tag{A5}$$

To this end, we first reorder the sums so that we can perform the sum over n first. Looking at only the factors that depend on n in Eq. (A4), we have

$$z(k) = \sum_{n=k}^{\infty}\binom{n-1}{k-1}e^{-\lambda n}, \tag{A6}$$

where the lower summation bound starts at k, since a filament containing k segments cannot have fewer than k particles. This sum can be performed with help from the generating function of the binomial coefficients, $\sum_{n=k}^{\infty}\binom{n}{k}e^{-\lambda n} = e^{-\lambda(k+1)}/(1-e^{-\lambda})^k$. Using this identity for our purposes, we obtain

$$z(k) = \left[\frac{1}{e^{\lambda}-1}\right]^k. \tag{A7}$$

Inserting this expression back into Eq. (A4), we have

$$\frac{Z}{8\pi^2 V\phi_0} = e^{-\beta\varepsilon_2}\sum_{k=1}^{m}(m-k+1)q^k, \tag{A8}$$

where we defined $q = e^{\beta\Delta\varepsilon}/(e^{\lambda}-1)$ for convenience. We can now easily perform the sum over k and finally find

$$\frac{Z}{8\pi^2 V\phi_0} = \frac{e^{\beta\mu}}{1-e^{-\lambda}}\frac{q^{m+1}-(m+1)q+m}{(1-q)^2}. \tag{A9}$$

Canceling the common factors contained in both $Z(n)$ and Z from the above expression leads to the normalization constants C_{na} (if $m=1$) and C_{sa} (if $m \geqslant 2$) of the main text.

From Eq. (A9), we can compute the average length of the structures in the ensemble

$$\langle n\rangle = \frac{1}{Z}\sum_{n=1}^{\infty} Z(n)n. \tag{A10}$$

After a somewhat lengthy calculation, we find

$$\langle n\rangle = \frac{mq^{m+2}+(m+2)(1-q^m)q-m}{(1-e^{-\lambda})(q^{m+1}-(m+1)q+m)(q-1)}. \tag{A11}$$

Similarly, for the second moment,

$$\langle n^2\rangle = \frac{m(rq^{m+3} - rq^{m+2} + rq^2 - (r+1)q - 4q^{m+1} - 1)(q-1) - 2(rq^2 - (r+1)q - 2)(q^m - 1)q + m^2q^{m+1}(q-1)^2}{(1-e^{-\lambda})^2(q^{m+1} - (m+1)q + m)(1-q)^2}, \tag{A12}$$

where we also defined $r = e^{-\beta\Delta\varepsilon}(1-e^{-\lambda})$. From the moments, we can then compute the width $\sigma = \sqrt{\langle n^2\rangle - \langle n\rangle^2}$ of the distribution. The total particle concentration is given by $\rho_{\text{tot}} = \langle n\rangle Z/V$, which gives

$$\frac{\rho_{\text{tot}}}{8\pi\phi_0} = e^{\beta\mu}\frac{mq^{m+2} + (m+2)(1-q^m)q - m}{(1-e^{-\lambda})^2(q-1)^3}. \tag{A13}$$

We can rederive the limiting expressions, Eq. (5), directly from these general relations. Taking the limit $\Delta\varepsilon \to \infty$ and keeping only the leading terms, we have

$$\langle n\rangle \approx \frac{m}{1-e^{-\lambda}} = m\left[\frac{1}{2} + \frac{1}{\lambda}\right] + O(\lambda), \tag{A14}$$

$$\sigma \approx \sqrt{\frac{me^{\lambda}}{(e^{\lambda}-1)^2}} = \frac{\sqrt{m}}{\lambda} + O(\lambda), \tag{A15}$$

where we have also expanded in a power series around $\lambda = 0$, corresponding to large $\langle n\rangle$. The total particle concentration also simplifies to

$$\frac{\rho_{\text{tot}}}{8\pi^2\phi_0} \approx \frac{me^{\beta\mu}}{(1-e^{-\lambda})^2}\left[\frac{e^{\beta\Delta\varepsilon}}{e^{\lambda}-1}\right]^{m-1}, \tag{A16}$$

in the $\Delta\varepsilon \to \infty$ limit.

APPENDIX B: ASSEMBLY KINETICS AND SIMULATION DETAILS

In simulations, we treat the assembly process as a Markov chain, consisting of $N = 5\times10^7$ particles in a volume V, and simulate it using the Gillespie algorithm [65]. We assume that two particles (more precisely: two binding sites) encounter each other at a constant rate α/V, and that every encounter between compatible species leads to the formation of a bond. The bonds between particles of the same species break at a constant rate δ_1, and the bonds between particles of different species break at a rate δ_2. These rate constants are related to the binding energies via $\alpha/\delta_1 = e^{\beta\varepsilon_1}/(8\pi^2\phi_0)$ and $\alpha/\delta_2 = e^{\beta\varepsilon_2}/(8\pi^2\phi_0)$ [66], where ϕ_0 is the reference concentration from Appendix A. For single-species assembly, this model describes the stochastic kinetics of Flory-Stockmayer polymerization [67–69]. It is important to note that the assumption of a constant aggregation rate for all filaments is a significant approximation and generally leads to an underestimation of equilibration times. Actual aggregation rates are highly system specific, and a more detailed description involving orientational, steric, or even long-range (electrostatic) interactions will be important to establish in the future on a case-by-case basis.

As mentioned in the main text, we use the single-species (nonaddressable) equilibration time $\tau_{\text{na}} = \langle n\rangle/(\alpha\rho_{\text{tot}})$ (see the Supplemental Material [57] for a derivation) as a reference time scale, where $\rho_{\text{tot}} = N/V$ is the total particle concentration.

We can gain a more intuitive understanding of the long equilibration times reported in Fig. 3 if we consider the assembly kinetics more closely. At early times, a system consists predominantly of free monomers, and bond-breaking events are rare compared to binding events, which means that bonds can be viewed as unbreakable early on. Depending on particle concentrations, any particle has a roughly 50% chance of encountering either another particle of the same species or one of the subsequent species, which means that the expected filament length at the end of this initial growing phase is about $2\,m$. Note that this is not exact, since we use nonuniform particle concentrations to maintain constant μ, as described in the main text and Appendix C.

After this initial growth, filaments become kinetically arrested. While they can still grow from the ends, to grow one of the segments in the bulk, the filament has to break open first. Equilibrating the lengths of all filament segments, therefore, requires a large number of bond-breaking events, which means that the approach to equilibrium is governed by the bond-breaking timescales δ_1 and δ_2. These timescales are determined by the binding energies required to assemble filaments with a given length $\langle n\rangle$ and peak width σ. Due to the exponential relationship between energies and rates, even a small change in binding energies can have a large effect on the equilibration time.

APPENDIX C: CONVERTING BETWEEN GRAND CANONICAL AND CANONICAL ENSEMBLE

As discussed in the main text and Appendix A, we carry out our analytical calculations at uniform chemical potentials for all particle species. In order to compare our equilibrium calculations to the kinetic simulations, which are performed at fixed particle concentrations, we first need to relate the chemical potential μ to the concentrations. To this end, we now analytically compute the particle concentrations as functions of μ.

The concentration ρ_i of particle species $i \in (1, ..., m)$ is given by

$$\rho_i = \frac{1}{V}\sum_{s\in\mathcal{S}} n_i^s Z(s), \tag{C1}$$

where s is a specific structure, $\mathcal{S}$ is the set of all filaments, and n_i^s is the number of particles of species i in s. Because n_i^s enters directly into this sum, we cannot immediately group filaments by their total length, as we did above, and we need to be mindful of the filaments' compositions.

The trick to computing this sum is to split it into three parts: the section of the filament to the left of the species i whose concentration we want to compute, the section consisting only of species i, and the section to the right of the species i. We denote the length of the section to the left by ℓ_-, and the number of species in that section by k_-. Similarly, the

length and number of species on the right are denoted by ℓ_+ and k_+. We also denote the number of particles of species i simply by n. With this notation, the expression for ρ_i takes the form

$$\frac{\rho_i}{8\pi^2\phi_0} = \left[e^{-\beta\varepsilon_1}\sum_{n=0}^{\infty} n e^{-\lambda n}\right] \times \left[\sum_{\ell_-=0}^{\infty} e^{-\lambda\ell_-}\sum_{k_-=0}^{i-1}\binom{\ell_- - 1}{k_- - 1}e^{\beta\Delta\varepsilon k_-}\right] \times \left[\sum_{\ell_+=0}^{\infty} e^{-\lambda\ell_+}\sum_{k_+=0}^{m-i}\binom{\ell_+ - 1}{k_+ - 1}e^{\beta\Delta\varepsilon k_+}\right], \tag{C2}$$

where independent sums have been separated by brackets.

The sums over $\ell_\pm$ and $k_\pm$ are almost identical to the sum we already computed in Appendix A. After a straightforward calculation, we find

$$\frac{\rho_i}{8\pi^2\phi_0} = \frac{e^{\beta\mu}}{(1-e^{-\lambda})^2} \times \frac{1-q^i}{1-q} \times \frac{1-q^{m-i+1}}{1-q}, \tag{C3}$$

where q is defined as above and also depends on μ. Note that this expression is only valid for $1 < i < m$. If $i = 1$, the second factor on the right-hand side needs to be replaced by 1, as there cannot be a segment to the left of the first species. Similarly, if $i = m$, the last factor needs to be replaced by 1.

This calculation shows that the concentrations of the different particle species are not necessarily uniform, even if the chemical potential is uniform. The particle numbers $N_i = V\rho_i$ and binding energies used in simulations are listed in Table S1 in the Supplemental Material [57], which shows significant variations between species.

APPENDIX D: SENSITIVITY ANALYSIS VIA A TRANSFER MATRIX METHOD

In our analysis of the equilibrium length distribution so far, we have assumed that all particle species are supplied at equal chemical potential, that all binding energies are equal to ε_1 or ε_2 (depending on the contact), and that there are no nonspecific crosstalk interactions. In this section, we investigate what happens if these assumptions are relaxed. In other words, here we investigate the length distribution as a function of $m(m+1)$ parameters: m^2 binding energies and m particle concentrations.

Allowing all energies and concentrations to vary makes the problem much more complicated and prevents us from writing down an explicit expression for the length distribution. To make progress, we adopt and extend the transfer matrix ("Feynman diagram") method of Ref. [32], which allows us to quickly compute all relevant quantities through matrix manipulations.

The basic idea is that the partition function of any one-dimensional aggregate can be expressed using a transfer matrix

$$T_{ij} = E_{ij}y_j, \tag{D1}$$

where the (nonsymmetric) matrix $E_{ij} = e^{\beta\varepsilon_{ij}}$ is given by the binding energies ε_{ij} controlling the contact of the left of particle species j to the right of particle species i, and $y_i = e^{\beta\mu_i}$ is a vector of exponentiated chemical potentials. Our analytical calculations from before correspond to a special case of this, where E_{ij} is a banded matrix with equal elements along the diagonals, and all elements of y_i are equal.

Using the transfer matrix, the partition function of a length-n filament can be written as

$$\frac{Z(n)}{8\pi^2 V\phi_0} = \sum_{ij} y_i[T^{n-1}]_{ij}, \tag{D2}$$

and the full partition function $Z = \sum_n Z(n)$ is given by

$$\frac{Z}{8\pi^2 V\phi_0} = \sum_{ij} y_i[\mathbb{I} - T]^{-1}_{ij}. \tag{D3}$$

These two expressions make it possible to compute the full-length distribution, $p(n) = Z(n)/Z$, as a function of all m^2 energies and m particle concentrations.

We can also apply the transfer-matrix approach to directly compute the mean filament length and peak width. Straightforward calculations lead to the following expressions for the first two moments of the length distribution:

$$\frac{\langle n\rangle}{8\pi^2 V\phi_0} = \frac{1}{Z}\sum_{ij} y_i[\mathbb{I} - T]^{-2}_{ij}, \tag{D4}$$

$$\frac{\langle n^2\rangle}{8\pi^2 V\phi_0} = \frac{1}{Z}\sum_{ijk} y_i[\mathbb{I}_{ik} + T_{ik}][\mathbb{I} - T]^{-3}_{kj}, \tag{D5}$$

from which we can compute the width $\sigma = \sqrt{\langle n^2\rangle - \langle n\rangle^2}$ of the distribution.

As before, we need to relate particle concentrations to chemical potentials if we wish to carry out calculations in the canonical ensemble. Using the transfer-matrix formalism, the concentrations $\rho_i = (\beta V)^{-1}\partial Z/\partial\mu_i$ of each particle species are given by

$$\frac{\rho_i}{8\pi^2\phi_0} = \left(\sum_j y_j[\mathbb{I} - T]^{-1}_{ji}\right)\left(\sum_k [\mathbb{I} - T]^{-1}_{ik}\right). \tag{D6}$$

By numerically inverting this relation to solve for μ_i as a function of ρ_i, we can compute the length distribution and all its properties as functions of the particle concentrations.

We are now in a position to evaluate the sensitivity of the length distribution to variations in the binding energies and particle concentrations, which we do here in the context of the example system discussed in the "Size-control in practice" section and shown in Fig. 3. From this ideal baseline, we then iteratively increase the noise level η and observe the changes in the mean and the relative peak width of the length distribution.

First, we investigate what happens when the binding energies ε_1 and ε_2 vary among different species. Specifically, we set the same-species and cross-species binding energies of particle species i to $\varepsilon^i_{1/2} = \varepsilon_{1/2} + \xi^i_{1/2}$, where $\xi^i_{1/2}$ are random variables drawn from a normal distribution with a mean of zero and a standard deviation of η. Figure 5(a) shows that the

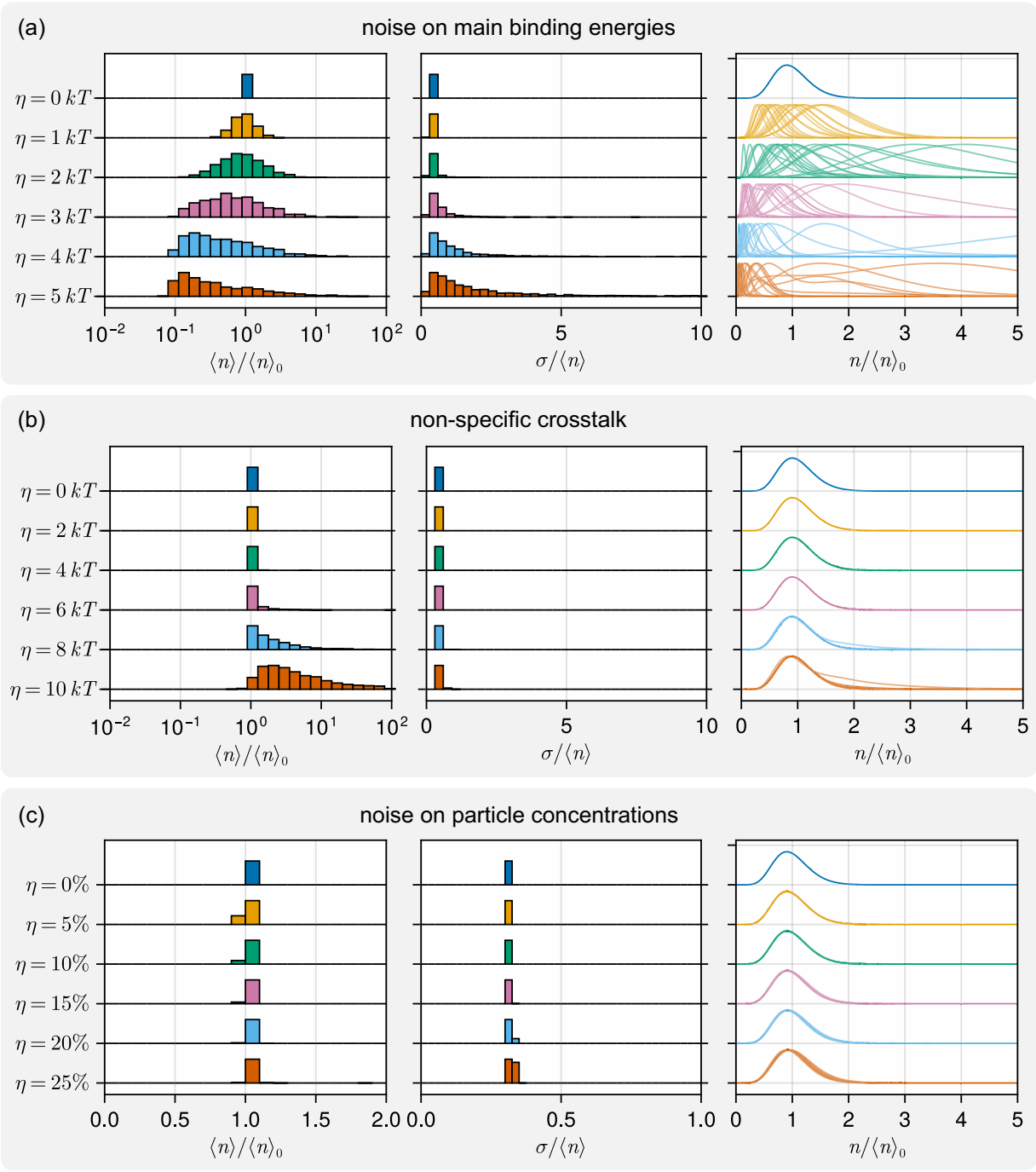

FIG. 5. Sensitivity of the length distribution for a system of $m = 10$ species targeting the assembly of a $\langle n \rangle = 100$ filament with peak width $\sigma = \langle n \rangle / \sqrt{m}$. All panels show histograms of the shift of the mean (*left*), the relative peak width (*center*), and a subset of the sampled length distributions (*right*) for 1000 realizations of random parameter noise. (a) Sensitivity with respect to binding energies ε_1 and ε_2: Both main binding energies were subjected to normal-distributed noise of standard deviation η. (b) Sensitivity with respect to crosstalk. Main binding energies remain unaffected, but there are random nonspecific (all-to-all) interactions sampled from a folded normal distribution with standard deviation η. (c) Sensitivity with respect to concentrations. All particle concentrations ρ_i were scaled by a factor $1 + \xi_i$, where ξ_i are normally distributed random variables with standard deviation η.

mean of the length distribution can shift by a factor of up to 10 in both directions, especially at $\eta = 5\,kT$ (about 18% of ε_1), and the distribution also tends to broaden at high noise. At lower noise levels, the effects are more manageable: while the mean of the distribution may shift by about a factor of 3 for $\eta \leqslant 2\,kT$, the relative peak width remains mostly unaffected. This indicates that it may be challenging to program a specific length if energies cannot be tightly controlled (less than about 10% variability).

Second, we investigate the effects of nonspecific crosstalk interactions. For this, we keep the main binding energies ε_1 and ε_2 noise-free but add random, nonspecific interactions

between all particle species. These crosstalk interactions are given by the absolute values of random variables sampled from a normal distribution with a mean of zero and a standard deviation of η (folded normal distribution). Interestingly, the equilibrium distribution is not very sensitive to this nonspecific crosstalk, as long as its magnitude is small compared to the main interactions ε_1 and ε_2.

Finally, we also investigate the effects of noise on particle concentrations. For this, we leave binding energies unaffected but scale each particle concentration ρ_i by a factor $1+\xi_i$, where ξ_i are again normally distributed with zero mean and a standard deviation of η. Figure 5(c) shows that the assembly is rather insensitive to variations in particle concentrations—even with variations as large as $\eta = 25\%$, the equilibrium length distribution remains almost completely unaffected.

APPENDIX E: ASSEMBLY OF TWO-DIMENSIONAL SHEETS

Figure 4 shows the size distribution of two-dimensional sheets, which are assembled from vertically "stacked" filaments. The assembly protocol employed here is the following: We first assemble filaments consisting of $m_1 = 10$ species and targeting a peak width of $\sigma = \langle n\rangle/\sqrt{m}$, which corresponds to the same conditions as in the hierarchical system shown in Fig. 3. After this initial stage of assembly is complete, we freeze the distribution of filaments and activate binding sites on the sides of the filaments, so that the filaments can aggregate vertically. We assume that filaments are rigid and that no branching occurs; i.e., every filament may only bind to a single other filament above and below. The total binding energy between two filaments then depends on their overlap; we assume that the vertical binding energy between two filaments of length n_i and n_j is given by

$$\varepsilon_{\mathrm{v}} = \varepsilon_{\mathrm{v},0}\min(n_i, n_j), \tag{E1}$$

where $\varepsilon_{\mathrm{v},0}$ is the per-particle vertical energy; we set $\varepsilon_{\mathrm{v},0} = 0.1\,\mathrm{kT}$. We neglect any entropic contributions resulting from the different possible binding offsets between two filaments, and assume that filaments are always overlapping without any "overshoot," resulting in $\min(n_i, n_j)$ particle contacts.

For only a single species of filament, the "height" distribution of the resulting sheets, $p(h)$, is monotonically decaying and is not well controllable, analogously to the single-species filament distribution, Eq. (1), as shown in Fig. S2 in the Supplemental Material [57]. To obtain more control over sheet heights, we introduce $m_2 = 10$ species of filaments, each consisting of $m_1 = 10$ distinct particle species, leading to a total species count of $m = 100$. (Even though $m = 100$ species is pushing the boundaries of current experimental feasibility, the crucial point is that this does not scale with the size of the system, motivating an experimental challenge for the coming decade.) In full analogy to the one-dimensional case shown in Fig. 1(c), each filament species can bind vertically to itself or to the subsequent filament species. For simplicity, we further assume that all interactions during the final assembly step of assembly are infinitely strong, meaning that bonds no longer break and that each assembled sheet contains all 10 filament species.

Under these assumptions, we can obtain the size distributions for the multispecies sheets by sampling different sheets obtained from a single-species simulation and concatenating them together. We start from a population of 16 698 single-species sheets (the size distribution of which is shown in Fig. S1 in the Supplemental Material [57]), and then sample 10^6 groups of 10 sheets each, which we concatenate vertically to simulate a multispecies sheet, leading to the size distribution as shown in Fig. 4.

[1] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 7th ed. (W. W. Norton & Company, New York, 2022).

[2] N. Rhind, Cell-size control, Curr. Biol. **31**, R1414 (2021).

[3] M. B. Ginzberg, R. Kafri, and M. Kirschner, On being the right (cell) size, Science **348**, 1245075 (2015).

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

[5] Y. Kim, P. Dalhaimer, D. A. Christian, and D. E. Discher, Polymeric worm micelles as nano-carriers for drug delivery, Nanotechnology **16**, S484 (2005).

[6] M. Rad-Malekshahi, L. Lempsink, M. Amidi, W. E. Hennink, and E. Mastrobattista, Biomedical applications of self-assembling peptides, Bioconjugate Chem. **27**, 3 (2016).

[7] 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).

[8] P. Vukusic and J. R. Sambles, Photonic structures in biology, Nature (London) **424**, 852 (2003).

[9] E. R. Dufresne, H. Noh, V. Saranathan, S. G. J. Mochrie, H. Cao, and R. O. Prum, Self-assembly of amorphous biophotonic nanostructures by phase separation, Soft Matter **5**, 1792 (2009).

[10] D. Satyabola, A. Prasad, H. Yan, and X. Zhou, Bioinspired photonic systems directed by designer DNA nanostructures, ACS Appl. Opt. Mater. **3**, 552 (2025).

[11] D. Hayakawa, T. E. Videbæk, G. M. Grason, and W. B. Rogers, Symmetry-guided inverse design of self-assembling multiscale DNA origami tilings, ACS Nano **18**, 19169 (2024).

[12] A. Hensley, T. E. Videbæk, H. Seyforth, W. M. Jacobs, and W. B. Rogers, Macroscopic photonic single crystals via seeded growth of DNA-coated colloids, Nat. Commun. **14**, 4237 (2023).

[13] F. A. Aldaye, A. L. Palmer, and H. F. Sleiman, Assembling materials with DNA as the guide, Science **321**, 1795 (2008).

[14] A. K. Pearce, T. R. Wilks, M. C. Arno, and R. K. O'Reilly, Synthesis and applications of anisotropic nanoparticles with precisely defined dimensions, Nat. Rev. Chem. **5**, 21 (2021).

[15] A. Michelson, A. Subramanian, K. Kisslinger, N. Tiwale, S. Xiang, E. Shen, J. S. Kahn, D. Nykypanchuk, H. Yan, C.-Y. Nam, and O. Gang, Three-dimensional nanoscale metal,

metal oxide, and semiconductor frameworks through DNA-programmable assembly and templating, Sci. Adv. **10**, eadl0604 (2024).

[16] A. Michelson, L. Shani, J. S. Kahn, D. C. Redeker, W.-I. Lee, K. R. DeOlivares, K. Kisslinger, N. Tiwale, H. Yan, A. Pattammattel, C.-Y. Nam, V. S. Pribiag, and O. Gang, Scalable fabrication of chip-integrated 3D-nanostructured electronic devices via DNA-programmable assembly, Sci. Adv. **11**, eadt5620 (2025).

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

[18] T. E. Videbæk, D. Hayakawa, G. M. Grason, M. F. Hagan, S. Fraden, and W. B. Rogers, Economical routes to size-specific assembly of self-closing structures, Sci. Adv. **10**, eado5979 (2024).

[19] C. M. Duque, D. M. Hall, B. Tyukodi, M. F. Hagan, C. D. Santangelo, and G. M. Grason, Limits of economy and fidelity for programmable assembly of size-controlled triply periodic polyhedra, Proc. Natl. Acad. Sci. USA **121**, e2315648121 (2024).

[20] R. Saha, D. Hayakawa, T. E. Videbaek, M. Price, W.-S. Wei, J. Pombo, D. Duke, G. Arya, G. M. Grason, W. B. Rogers, and S. Fraden, Modular programming of interaction and geometric specificity enables assembly of complex DNA origami nanostructures, Nat. Commun. **16**, 11392 (2025).

[21] F. Sciortino, S. Mossa, E. Zaccarelli, and P. Tartaglia, Equilibrium cluster phases and low-density arrested disordered states: The role of short-range attraction and long-range repulsion, Phys. Rev. Lett. **93**, 055701 (2004).

[22] T. D. Nguyen, B. A. Schultz, N. A. Kotov, and S. C. Glotzer, Generic, phenomenological, on-the-fly renormalized repulsion model for self-limited organization of terminal supraparticle assemblies, Proc. Natl. Acad. Sci. USA **112**, E3161 (2015).

[23] G. M. Grason, Perspective: Geometrically frustrated assemblies J. Chem. Phys. **145**, 110901 (2016).

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

[25] H. Le Roy, M. M. Terzi, and M. Lenz, Collective deformation modes promote fibrous self-assembly in deformable particles, Phys. Rev. X **15**, 011022 (2025).

[26] L. Koehler, M. Eder, C. Karfusehr, V. Ouazan-Reboul, P. Ronceray, F. C. Simmel, and M. Lenz, Topological defect engineering enables size and shape control in self-assembly, arXiv:2504.13073.

[27] D. Johann, C. Erlenkämper, and K. Kruse, Length regulation of active biopolymers by molecular motors, Phys. Rev. Lett. **108**, 258103 (2012).

[28] A. Melbinger, L. Reese, and E. Frey, Microtubule length regulation by molecular motors, Phys. Rev. Lett. **108**, 258104 (2012).

[29] M. Striebel, F. Brauns, and E. Frey, Length regulation drives self-organization in filament-motor mixtures, Phys. Rev. Lett. **129**, 238102 (2022).

[30] H.-S. Kuan and M. D. Betterton, Biophysics of filament length regulation by molecular motors, Phys. Biol. **10**, 036004 (2013).

[31] S. Datta, M. L. Saha, and P. J. Stang, Hierarchical assemblies of supramolecular coordination complexes, Acc. Chem. Res. **51**, 2047 (2018).

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

[33] M. He, J. P. Gales, E. Ducrot, Z. Gong, G.-R. Yi, S. Sacanna, and D. J. Pine, Colloidal diamond, Nature (London) **585**, 524 (2020).

[34] C. A. Mirkin, R. L. Letsinger, R. C. Mucic, and J. J. Storhoff, A DNA-based method for rationally assembling nanoparticles into macroscopic materials, Nature (London) **382**, 607 (1996).

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

[36] Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck, and D. J. Pine, Colloids with valence and specific directional bonding, Nature (London) **491**, 51 (2012).

[37] Y. Wang, Y. Wang, X. Zheng, E. Ducrot, J. S. Yodh, M. Weck, and D. J. Pine, Crystallization of DNA-coated colloids, Nat. Commun. **6**, 7253 (2015).

[38] M.-P. Valignat, O. Theodoly, J. C. Crocker, W. B. Russel, and P. M. Chaikin, Reversible self-assembly and directed assembly of DNA-linked micrometer-sized colloids, Proc. Natl. Acad. Sci. USA **102**, 4225 (2005).

[39] W. M. Jacobs and W. B. Rogers, Assembly of complex colloidal systems using DNA, Annu. Rev. Condens. Matter Phys. **16**, 443 (2025).

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

[41] S. Sacanna, M. Korpics, K. Rodriguez, L. Colón-Meléndez, S.-H. Kim, D. J. Pine, and G.-R. Yi, Shaping colloids for self-assembly, Nat. Commun. **4**, 1688 (2013).

[42] P.-S. Huang, S. E. Boyken, and D. Baker, The coming of age of de novo protein design, Nature (London) **537**, 320 (2016).

[43] N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates, and D. Baker, Computational design of self-assembling protein nanomaterials with atomic level accuracy, Science **336**, 1171 (2012).

[44] T. F. A. d. Greef and E. W. Meijer, Supramolecular polymers, Nature (London) **453**, 171 (2008).

[45] T. Aida, E. W. Meijer, and S. I. Stupp, Functional supramolecular polymers, Science **335**, 813 (2012).

[46] H. M. M. t. Eikelder, B. Adelizzi, A. R. A. Palmans, and A. J. Markvoort, Equilibrium model for supramolecular copolymerizations, J. Phys. Chem. B **123**, 6627 (2019).

[47] E. D. Klein, R. W. Perry, and V. N. Manoharan, Physical interpretation of the partition function for colloidal clusters, Phys. Rev. E **98**, 032608 (2018).

[48] A. I. Curatolo, O. Kimchi, C. P. Goodrich, R. K. Krueger, and M. P. Brenner, A computational toolbox for the assembly yield of complex and heterogeneous structures, Nat. Commun. **14**, 8328 (2023).

[49] M. Holmes-Cerfon, S. J. Gortler, and M. P. Brenner, A geometrical approach to computing free-energy landscapes from short-ranged potentials, Proc. Natl. Acad. Sci. USA **110**, E5 (2013).

[50] M. Holmes-Cerfon, Sticky-sphere clusters, Annu. Rev. Condens. Matter Phys. **8**, 77 (2017).

[51] M. C. Hübl and C. P. Goodrich, Accessing semiaddressable self-assembly with efficient structure enumeration, Phys. Rev. Lett. **134**, 058204 (2025).

[52] M. C. Hübl, T. E. Videbæk, D. Hayakawa, W. B. Rogers, and C. P. Goodrich, A polyhedral structure controls programmable self-assembly, Nat. Phys. (2026), doi: 10.1038/s41567-025-03120-3.

[53] Strictly speaking, the distribution is *geometric*, since n is an integer and $1 \leqslant n < \infty$.

[54] M. P. Brenner, *Mathematics and Materials Ideas About Self-Assembly*, IAS/Park City Mathematics Series (American Mathematical Society, Providence, 2017), Vol. 23.

[55] This remains true if the twofold symmetry of a filament is exploited, which cuts down the required particle species by half but leaves the prohibitive linear scaling unaffected.

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

[57] See Supplemental Material at https://link.aps.org/supplemental/10.1103/68rs-3qgn for additional details of the kinetic calculations and supplementary data, which includes Ref. [70].

[58] The length of $\langle n \rangle/(2m)$ is necessary since there is a roughly 50% chance that a single-species segment binds with another segment of the same species before encountering a filament of the subsequent species.

[59] O. G. Hayes, B. E. Partridge, and C. A. Mirkin, Encoding hierarchical assembly pathways of proteins with DNA, Proc. Natl. Acad. Sci. USA **118**, e2106808118 (2021).

[60] E. W. Gehrels, W. B. Rogers, Z. Zeravcic, and V. N. Manoharan, Programming directed motion with DNA-grafted particles, ACS Nano **16**, 9195 (2022).

[61] H. Xiong, D. van der Lelie, and O. Gang, Phase behavior of nanoparticles assembled by DNA linkers, Phys. Rev. Lett. **102**, 015504 (2009).

[62] J. Lowensohn, B. Oyarzún, G. N. Paliza, B. M. Mognetti, and W. B. Rogers, Linker-mediated phase behavior of DNA-coated colloids, Phys. Rev. X **9**, 041054 (2019).

[63] M. C. Hübl and C. P. Goodrich, Entropic size control of self-assembled filaments, Zenodo (2025), doi: 10.5281/zenodo.17898642.

[64] M. Kardar, *Statistical Physics of Particles* (Cambridge University Press, New York, 2007).

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

[66] N. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science Publishers, Amsterdam, 1992).

[67] P. J. Blatz and A. V. Tobolsky, Note on the kinetics of systems manifesting simultaneous polymerization-depolymerization phenomena, J. Phys. Chem. **49**, 77 (1945).

[68] W. H. Stockmayer, Theory of molecular size distribution and gel formation in branched-chain polymers, J. Chem. Phys. **11**, 45 (1943).

[69] P. G. J. v. Dongen and M. H. Ernst, Kinetics of reversible polymerization, J. Stat. Phys. **37**, 301 (1984).

[70] R. D. Vigil, On equilibrium solutions of aggregation–fragmentation problems, J. Colloid Interface Sci. **336**, 642 (2009).