One-dimensional modeling of the self-assembly of complex elastic particles

Pawat Akarapipattana
advisor: Martin Lenz

September 6, 2023

Abstract

Self-assembly is a spontaneous aggregation of small subunits into a larger structure. We focus on a specific kind of self-assembly phenomenon in cells where protein molecules aggregate into finite-size fibers. This phenomenon involves deformations and elastic properties of the protein molecules. However, this behavior has been shown to be generic beyond any specific types of proteins. Currently, the exact mathematical descriptions of the deformation are known only for a one-dimensional aggregate model consisting of simple subunits. Real-world protein molecules, on the other hand, pose more complexity – both in terms of their shapes and their elastic properties – and a more complex model is required.

In this work, we attempt to narrow down the difference between real-world systems and the existing mathematical models by introducing complexity via randomness of the positions of protein binding sites. Our model displays a novel oscillatory deformation mode that has not been previously observed. We offer a quantitative explanation of this kind of deformation mode. Moreover, our findings suggest a universality in aggregate size independent of the elastic properties of the subunits.

1 Introduction

The self-assembly process is broadly described as a spontaneous aggregation of small particles into a larger, and possibly more complex, structure. This building process is common in living organisms at the molecular level [1]. Intuitively, it is indeed easier for complex functioning structures in cells to be assembled from small similar building blocks rather than to be entirely constructed as a whole. First, it is less energetically costly to build small structures, and small molecules require fewer metabolism pathways to construct. Self-assembly is also appealing to the field of material science because of its bottom-up and cost-efficient nature [2, 3].

In biological systems, the self-assembly aggregation process spontaneously terminates at a certain size when it reaches the target structure. This kind of self-assembly is called self-limiting assembly (SLA) and will be the main focus of this work. SLA can be governed by the intrinsic structures of the building blocks themselves, or by external factors as shown in Fig.1 [4]. In the former case, known as self-closing assembly, the assembly process is governed by intrinsic anisotropy of the building blocks, which gives rise to overall curvature of the aggregate. This eventually allows the aggregate

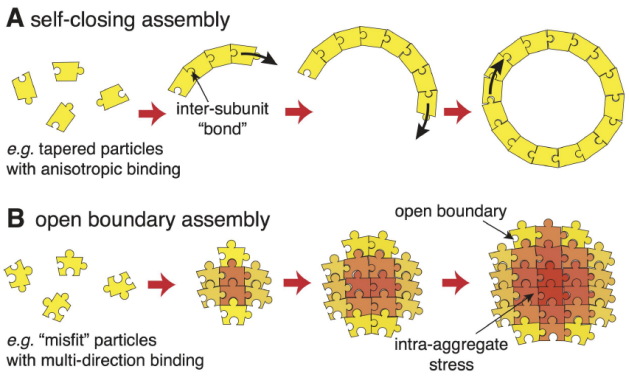


Fig. 1: (Taken from [4]) A schematic showing two mojor kind of SLA. **A**: self-closing assembly, where the size of the aggregate is controlled intrinsically by the geometry of subunits. Anisotropy of each subunit contributes to overall curvature which allows the aggregate to close onto itself and stop the process. **B**: open-boundary assembly, where the size of the aggregate is controlled by the competition between intra-aggregate frustration and inter-aggregate interaction with the environment.

to close onto itself and thus stop the aggregation process. On the other hand, the assembly process can also be controlled by external interactions such as surface tension. This happens in a system where the aggregates have open boundaries to interact with the environment, thus the name open boundary assembly. This interaction is generally proportional to the inverse of the boundary size of the aggregate so it favors the aggregate of infinite size – boundary growth becomes negligible compared to volume growth in the limit of infinitely many subunits. Thus, for this kind of aggregate to be self-limiting, another interaction that introduces energy penalty as some function of the aggregate size is necessary. For a small aggregate, the main energic cost comes from surface tension. As the aggregate grows in size, the energy penalty from its size becomes more expensive. The size of open boundary SLA is thus controlled by the competition of these two interactions.

Functional proteins in cells also follow the process of self-assembly. However, protein folding procedures are complicated and there are many points of failure. Errors in the correction mechanisms in cells can permit pathological structures known as amyloids. Amyloids are aggregates of imperfect proteins that, instead of their normal functioning form, assemble themselves into thread-like structures called fibrils. The presence of amyloids, both within and around cells, has been associated with many important diseases including Alzheimer's disease [5, 6], Pakinson's disease [7], and type II diabetes [8]. (The lists of many other amyloid-associated diseases can be found on [9].) For these reasons, the understanding of amyloid structures will lead to both the understanding of the fibril formations and the understanding of potential causes of many important diseases.

Amyloids are not the only pathological fiber structure that occurs in human body. Gene mutations can alter the structure of proteins causing them to assemble into pathological structures. One example is sickle cell disease, where a mutation causes hemoglobin protein to form fiber instead of a globular structure [10].

The formation of pathological fibers is not specific to the encodings of the amino sequences of the proteins, unlike the functioning structure counterparts, but is a generic physiochemical property of a universal property of peptide chain beyond the scope of biomolecular chemistry [9]. In order to explain this phenomenon, it is necessary to

develop a new framework where the focus is shifted from attempting to capture every detail from experiments to constructing a minimum model capable of capturing this phenomenon. This task aligns with the goals of physicists; to look for a set of general rules rather than to study a specific behavior in detail.

One of the well-known size-controlling mechanisms in self-assembly is deformation inside the aggregate. It has been shown that fiber formation is affected by internal distortions [11, 12]. However, less is known about how the length of fibers is controlled through this intra-fiber deformation.

Previously, the robustness of fiber structure is studied in [13]. There, the authors consider open-boundary SLA in which the source of energy penalty is elastic deformation inside the bulk of the aggregate. The size and the geometry of the aggregate are determined through the competition between deformation energy and surface energy. It has been observed that fibers can form consistently over a vast range of surface tension and elastic constants of the subunit. In the subsequent work, the focus is extended onto a more analytically tractable model of subunits made of hard and soft springs. It is observed that fiber-like structures also emerge consistently as well in this setting [14]. Moreover, the deformation modes of the aggregate have been shown to be a useful description of the origin of the fiber structures.

Although it has demonstrated how fibrous structure can easily arise just from the competition between frustration and surface energy, the previous observations were made on a model with simple subunit shapes. Protein molecules, on the other hand, are complex by nature. To make the previous model more realistic, it is important to incorporate this complexity into the model. The new model will tell us if this previously observed behavior persists when the complexity of the model is increased and if this increase in complexity will give rise to any new behaviors or general properties not previously observed.

Here, we increase the complexity of previous models by randomizing the shape of subunits and spring constants. This randomness will give us insights into the aggregate behaviors not specific to the shape and elastic property of the subunit – in analogy to how fibril formation is common and not specific to any specific amino sequence of the protein.

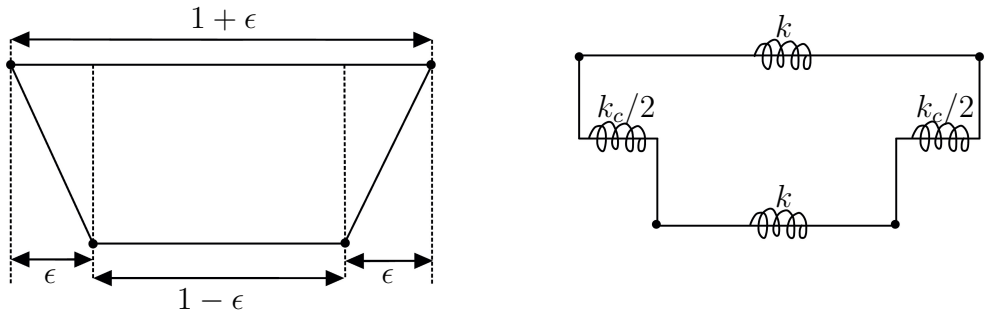
2 Simple model for 1D SLA

In this section, Although this toy model lives in 1-dimension where the only possible aggregate is fiber, far from any realistic protein self-assemblies, it will provide a relevant framework and intuitions necessary to understand more complex models.

Consider a one-dimensional aggregate with an open boundary sitting in a solution. The energy of the system comes from 2 parts: surface tension Γ (per contact point) and deformation energy in the bulk E_{bulk}. We can write the energy per subunit as a function of the number of particles N in an aggregate as

$$e(N) = \left(E_{\text{bulk}}(N) + 2\Gamma\right)/N. \tag{2.1}$$

E_{bulk} is the main focus of this phenomenon. Since Γ is assumed to be constant during the formation of aggregates, the size of the aggregate will be determined solely by the functional shape of E_{bulk}. We shall assume that the main source of the deformation energy is the accumulation of subunits deformation. For an aggregate made up of ill-fitting subunits, each subunit undergoes a deformation to fit into its neighbor. This deformation is accumulative and gives rise to frustration energy that grows faster than aggregate size N.



(a) (b)

Fig. 2: (a) geometric description of the simple trapezoid model for SLA. We choose the particle to have unit width by convention and control the slope of the non-parallel edges with ϵ. Note that even though the illustration is in 2 dimensions, only the position along the x-coordinate of each vertex point matters. (b) springs and their corresponding spring constants of the trapezoid subunit. The springs on the non-parallel edges are placed horizontally since only the horizontal position matters. The convention for the non-parallel edges' springs is chosen in the way that each interface has spring constant k_c after two particles are merged.

In [14], the authors consider a model for SLA composed of trapezoid elastic particles. Because trapezoids cannot be put together to form n aggregate without deforming their shape, constructing a 1-dimensional aggregate from this shape costs deformation energy – a phenomenon known as frustration. This is the simplest 1-dimensional model that can display the frustration behavior of the subunits in an aggregate. Each edge of a particle is modeled as a spring as shown in Fig.2. The trapezoids have a unit width by convention. Note that, however, this width can be arbitrarily and, as we shall see, does not affect the energy of the overall aggregate.

We parametrize how misfit a particle is by $\epsilon \in [0, 1]$ with $\epsilon = 0$ meaning that the particle is flat (or rectangle). Indeed, if the subunits are flat, they can perfectly assemble themselves into the minimum energy configuration without any deformation cost. However, for $\epsilon > 0$, the subunits need to deform themselves to form an aggregate.

We assign spring constant k to the two parallel edges of the trapezoid and $k_c/2$ to the springs on the other two edges. To form a 1-dimensional aggregate from the subunits, we connect two particles together; merging 2 vertices of the edge together but leaving the edges intact. In this way, the two springs on the edges behave effectively like a single spring with the coefficient k_c.

From this description, the deformation energy of an aggregate composed of N particles can be written as

$$\begin{aligned} E_{\text{bulk}}^{\text{trap}} = \sum_{i=-N/2}^{N/2-1} \Bigg[& \frac{k}{2}\left(x_{i+1}^{\uparrow} - x_i^{\uparrow} - (1+\epsilon)\right)^2 + \frac{k}{2}\left(x_{i+1}^{\downarrow} - x_i^{\downarrow} - (1-\epsilon)\right)^2 \\ & + \frac{k_c}{4}\left(x_{i+1}^{\uparrow} - x_{i+1}^{\downarrow} - \epsilon\right)^2 + \frac{k_c}{4}\left(x_i^{\downarrow} - x_i^{\uparrow} - \epsilon\right)^2 \Bigg], \end{aligned} \tag{2.2}$$

where $x_i^{\uparrow}$ and $x_i^{\downarrow}$ are respectively the position of the upper and lower vertex of the i^{th} column of the aggregate. The summation convention is chosen so that the 0^{th} particle is the middle of the aggregate. To look for an equilibrium configuration, we minimize the deformation energy with respect to $x_i^{\uparrow}$ and $x_i^{\downarrow}$. Combining this and force balance

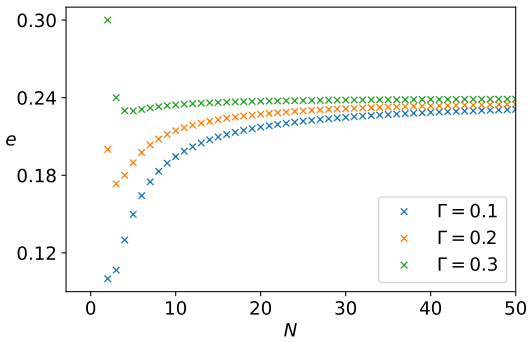


Fig. 3: Total aggregate energy per subunit as defined in (2.1) as a function of N at different value of surface tension Γ. We can see a slight shift of the minimum of this function as Γ increases.

in the bulk gives a 2-step recursion equation

$$\delta_{i+1} - 2\delta_i + \delta_{i-1} = 2\frac{k_c}{k}\delta_i, \tag{2.3}$$

where $\delta_i = x_i^{\uparrow} - x_i^{\downarrow}$. δ_i can be interpreted as a relative deformation between two layers of the aggregate. This equation can be solved with a simple ansatz $\delta_i = C_1 x_+^i + C_2 x_-^i$. The vertex positions $x_\pm$ can be obtained directly from (2.3) while C_1, C_2 can be obtained from force boundary equations at the boundaries at both sides of the aggregate. In the end, we have

$$\delta_i = \frac{\alpha\epsilon}{\sqrt{4\alpha+\alpha^2}} \frac{\sinh\left(\frac{i-(N-1)/2}{l_0}\right)}{\cosh\left(\frac{N-1}{2l_0}\right)}, \tag{2.4}$$

where $\alpha = 2\frac{k_c}{k}$ and $l_0 = -\log\left(1 + \frac{\alpha}{2} + \sqrt{\left(\frac{\alpha}{2}\right)^2 + \alpha}\right)$. This solution is antisymmetric, which is consistent with our setting; since each particle is antisymmetric, the whole aggregate should have mirror symmetry. Note that the deformation depends only on the ratio of k_c and k not each of them individually. Moreover, ϵ only contributes as a scaling factor in the deformation. Simply put, the particles near the boundary of the aggregate relax exponentially to their rest shapes as the sine hyperbolic function grows (decays) exponentially for a large positive-valued (negative-valued) argument.

The understanding of this simple model will be a foundation to understand more complex models coming up the in later sections. Now we know that the relative displacement of vertex positions is a useful quantity to describe the deformation with. We shall consider a similar quantity in the generalized model. The exponential decay behavior at the boundary of this model will also guide us on what to expect in the more complex models.

Lastly, this model displays size-limiting behavior when surface tension is introduced. In Fig.3, we can observe a minimum of the aggregate total energy per subunit (2.1) that appears when Γ is increased.

3 Complex 1D SLA

While the simple 1D SLA model is able to show size-limiting behavior, it is too simple to take into account the complexity of subunits that exists in biological self-assembly

systems. Lastly, since it is not possible to obtain a closed-form analytical solution of this model, we shall outline the numerical method we have used in this work in 3.3.

3.1 Generalized 1D SLA model

From a simple trapezoid particle described in section 2, we can make three main generalizations. First, we can vary the number of vertices; instead of having two vertices per side-edge, we can have l vertices per side-edge. The first part of randomness is introduced through the resting position of particle vertices. This randomness makes the particle misfit. We expect the degree of frustration to depend on the variance of the vertex positions. Lastly, we introduce the second part of randomness via spring constants. This part of randomness will mimic a generic response of each particle to the frustration when they form an aggregate.

The energy function that captures all the features previously mentioned is

$$\begin{aligned} E_{\text{bulk}} = \sum_{i=-N/2}^{N/2-1} \Bigg[& \sum_{ab}^{l} J_{ab}(Y_{i,b} - Y_{i,a} - \epsilon_{ab}^{L})^2 + \sum_{ab}^{l} K_{ab} \left(Y_{i+1,b} - Y_{i+1,a} - \epsilon_{ab}^{R}\right)^2 \\ & + \sum_{ab}^{l} L_{ab} \left(Y_{i,b} - Y_{i+1,a} - d - \epsilon_{ab}^{C}\right)^2 \Bigg]. \end{aligned} \tag{3.1}$$

Here, the index i represents the particles, as in the simple trapezoid particle, but the index a (and b) represents the vertices on the side edge of a particle. We have $Y_{i,a}$ as an x-axis position of a vertex so that $\mathbf{Y}_i$ is an l-vector corresponding to the i$^{\text{th}}$ column of an aggregate. We define the rest position of each vertex of a particle as $Y_a^{(0),L}$ and $Y_a^{(0),R}$ for the left and the right edge respectively. These lengths will be passed into the energy function by the following definitions of spring neutral lengths

$$\begin{aligned} \epsilon_{ab}^{L} &= Y_b^{(0),L} - Y_a^{(0),L} \\ \epsilon_{ab}^{R} &= Y_b^{(0),R} - Y_a^{(0),R} \\ \epsilon_{ab}^{C} &= Y_b^{(0),L} - Y_a^{(0),R}. \end{aligned} \tag{3.2}$$

$\mathbf{J}$, $\mathbf{K}$, and $\mathbf{L}$ are spring constant matrices containing the spring constants for all possible connections between vertices. $\mathbf{J}$ is for the intra-connections within the left side, $\mathbf{J}$ within the right side, and $\mathbf{K}$ is for the interconnections between two sides of the particle. We shall restrict $\mathbf{J}$, $\mathbf{K}$, and $\mathbf{L}$ to be positive definite, so that the deformation energy has a unique minimum. Finally, d is a linear shift between two adjacent interfaces of the particles in the aggregate resulting solely from the width of a subunit. We explicitly write this shift so that $\mathbf{Y}_i$ only contains the information on the deformation. This point will be made clear later in the calculations. In Fig.4, we show a schematic of a particle with $l = 4$ as well as an example of a frustrated aggregate in this generalized model to give a better idea of the notations.

3.2 Analytic description of the deformation

In this section, we shall provide mathematical descriptions of the generalized SLA model.

Before we begin, it is useful to introduce a more compact notation to make the calculation easier. First, we introduce a $2l$-vector

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{Y}_i \\ \mathbf{Z}_i \end{pmatrix}, \tag{3.3}$$

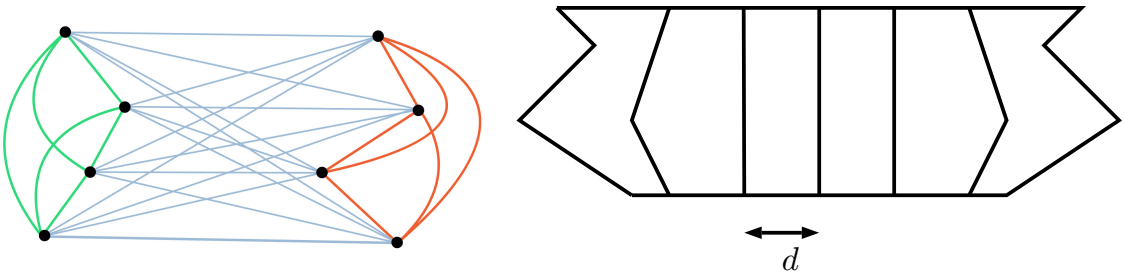


Fig. 4: (left) an example of a fully random subunit with $l = 4$. Spring constants on the green edges are described by the matrix $\mathbf{J}$, on the orange edges by $\mathbf{K}$, and on the blue edges by $\mathbf{L}$. Vertices on the left have the rest positions given by $\mathbf{Y}^L$, and on the right by $\mathbf{Y}^R$. (right) a schematic example of a 1D aggregate made up of six antisymmetrical subunits. Two ends of the aggregate take the rest shape of the particle. As we go deeper into the center of the aggregate, the edges of subunits are more and more deformed from their original shapes. d is the width between two interfaces "deep inside the aggregate" where each subunit is fully deformed.

with $\mathbf{Z}_i = \mathbf{Y}_{i+1}$. This will be useful later when we need to write a recursion relation for vertex positions. Then, we introduce a $2l$-vector containing the information on the boundary

$$\mathbf{B}^\intercal = \begin{pmatrix} \mathbf{B}^L \\ \mathbf{B}^R \end{pmatrix} \text{ with } \begin{array}{l} B^L_a = -2\sum_b J_{ab}\epsilon^L_{ba} - \sum_b L_{ba}\left(\epsilon^C_{ba}\right), \\ B^R_a = -2\sum_b K_{ab}\epsilon^R_{ba} + \sum_b L_{ab}\left(\epsilon^C_{ab}\right) \end{array}, \tag{3.4}$$

and the shift

$$\mathbf{D} = d\begin{pmatrix} -\mathbf{L}^\intercal\mathbf{1} \\ \mathbf{L}\mathbf{1} \end{pmatrix}. \tag{3.5}$$

Lastly, we introduce a $2l \times 2l$ matrix that contains the information about all spring constants in a particle

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} \\ \mathbf{M}_{12}^\intercal & \mathbf{M}_{22} \end{pmatrix}, \tag{3.6}$$

with

$$\begin{aligned} (\mathbf{M}_{11})_{ab} &= -2J_{ab} + \left(2\sum_c J_{ac} + \sum_c L_{ca}\right)\delta_{ab} \\ (\mathbf{M}_{22})_{ab} &= -2K_{ab} + \left(2\sum_c K_{ac} + \sum_c L_{ac}\right)\delta_{ab} \\ (\mathbf{M}_{12})_{ab} &= -L_{ab}. \end{aligned} \tag{3.7}$$

With this notation, we can concisely write the energy function as

$$E_{\text{bulk}} = \sum_i \left[\mathbf{X}_i^\intercal \mathbf{M}\mathbf{X}_i + 2(\mathbf{B}^\intercal + \mathbf{D}^\intercal)\mathbf{X}_i\right] + \text{const}, \tag{3.8}$$

The constant term is given by $\text{const} = \sum_{ab}\left(J_{ab}\left(\epsilon^L_{ab}\right)^2 + K_{ab}\left(\epsilon^R_{ab}\right)^2 + L_{ab}\left(d + \epsilon^C_{ab}\right)^2\right)$. However, its only importance is when we want to numerically compute the energy function. We would neglect this term from now on. Note that each row/column of $\mathbf{M}$ sums to zero. This is the consequence of translational invariance of the energy function i.e. E_{bulk} stays invariant under addition of constant to all $\mathbf{X}_i$.

To study the deformed configuration at the equilibrium, we minimize the energy with respect to $Y_{k,c}$ and d. This minimization gives the equations with the same

algebraic structure for all index c so we can compactly write

$$\frac{\partial E_{\text{bulk}}}{\partial \mathbf{Y}_k} = 0 = 2\mathbf{M}_{11}\mathbf{Y}_k + 2\mathbf{M}_{12}\mathbf{Y}_{k+1} + 2\mathbf{M}_{12}\mathbf{Y}_{k-1} + 2\mathbf{M}_{22}\mathbf{Y}_k + 2\left(\mathbf{B}^L + \mathbf{B}^R - d\left(\mathbf{M}_{12}^\intercal - \mathbf{M}_{12}\right)\mathbf{1}\right), \tag{3.9}$$

for $k = -N/2 + 1 \dots N/2 - 1$, that is, for edges that are not exposed to the solvent. This equation tells us how the forces exerted on each interface are balanced at the equilibrium. We can write this in the notation of $\mathbf{X}_i$ as follow.

$$\mathbf{X}_{k+1} = \boldsymbol{\mathcal{M}}\mathbf{X}_k - \mathbf{M}_{12}^{-1}\left(\mathbf{B}^L + \mathbf{B}^R\right) \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}, \tag{3.10}$$

where

$$\boldsymbol{\mathcal{M}} = \begin{pmatrix} \mathbf{0} & \mathbb{I} \\ -\mathbb{I} & -\mathbf{M}_{12}^{-1}\left(\mathbf{M}_{11} + \mathbf{M}_{22}\right) \end{pmatrix}. \tag{3.11}$$

Instead, for $k = -N/2$ and $k = N/2$, we obtained two boundaries condition of the aggregate; for the left and for the right respectively

$$\mathbf{Y}_{-N/2+1} = -\mathbf{M}_{12}^{-1}\mathbf{M}_{11}\mathbf{Y}_{-N/2} - \mathbf{M}_{12}^{-1}(\mathbf{B}^L + d\mathbf{M}_{12}^\intercal\mathbf{1}), \tag{3.12}$$

$$\mathbf{Y}_{N/2-1} = -\mathbf{M}_{12}^{-1}\mathbf{M}_{22}\mathbf{Y}_{N/2} - \mathbf{M}_{12}^{-1}(\mathbf{B}^R - d\mathbf{M}_{12}\mathbf{1}). \tag{3.13}$$

These left and right boundarie condition depend only on the detail corresponding to their side. Lastly, minimizing deformation energy with respect to d yields

$$\frac{\partial E_i}{\partial d} = 0 = -2\sum_{ab}^{l} L_{ab}\left(Y_{i,b} - Y_{i+1,a} - d - \epsilon_{ab}^C\right). \tag{3.14}$$

as a function of central spring constrant matrix $\mathbf{M}_{12}$, vertex positions $\mathbf{Y}_i$, and an implicit function of rest positions $\mathbf{Y}_0^{R,L}$. Although this equation is true throughout the aggregate, we have to solve it deep in the aggregate where the shift is purely the result of subunit width to make it consistent with the original definition of d. We will clarify this notion of "deep in the aggregate" shortly when we separate two scales of $\mathbf{X}_i$.

In this notation, we have instead a 1-step vectorial recursion equation which is easier to work with. We can get rid of the extra term by simply introducing $\mathbf{X}_k = \mathbf{X}_\infty + \delta\mathbf{X}_k$ where $\mathbf{X}_\infty$ is constant in k, that is, it satisfies

$$\mathbf{X}_\infty = \boldsymbol{\mathcal{M}}\mathbf{X}_\infty - \mathbf{M}_{12}^{-1}\left(\mathbf{B}^L + \mathbf{B}^R\right) \otimes \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \tag{3.15}$$

Solving this explicitly gives

$$\mathbf{X}_\infty = \begin{pmatrix} -(\mathbf{M}_{11} + 2\mathbf{M}_{12} + \mathbf{M}_{22})^\dagger(\mathbf{B}^L + \mathbf{B}^R) \\ -(\mathbf{M}_{11} + 2\mathbf{M}_{12} + \mathbf{M}_{22})^\dagger(\mathbf{B}^L + \mathbf{B}^R) \end{pmatrix} \equiv \begin{pmatrix} \mathbf{Y}_\infty \\ \mathbf{Y}_\infty \end{pmatrix}. \tag{3.16}$$

where † denotes Moore-Penrose psuedoinverse. $\mathbf{X}_\infty$ represents the shape of the interfaces of the particles "deep inside the aggregate", where the frustration is fully realized i.e. the frustration no longer deforms the shapes of the interfaces from one interface to the next. In the case of oscillating deformation, as we will see later on, $\mathbf{X}_\infty$ can be interpreted as the shape about which the oscillations take place. Note also that $\mathbf{X}_\infty$ is fully determined by the shape and spring constants of the subunit. On the other hand, δX_i denotes how much each vertex displaces from this shape. This quantity is an analogy to δ_i in (2.4). (3.16) can also be equivalently obtained by solving (3.9) with all $\mathbf{Y}_k$ replaced by $\mathbf{Y}_\infty$.

It is worth commenting that in the case of an aggregate made of antisymmetric subunits, namely $\mathbf{J} = \mathbf{K}$, $\mathbf{L}$ is symmetric, and $\mathbf{Y}^{(0),L} = const - \mathbf{Y}^{(0),R}$, $\mathbf{B}^L = -\mathbf{B}^R$ and all element of $\mathbf{Y}_\infty$ vanishes. This means that antisymmetric subunits will always be deformed to flat particles deep inside the aggregate.

Having the notion of "deep in the aggregate" well constructed, we return to the linear shift d. As originally defined, d is the width between two interfaces that is solely the result of particle size. Therefore, we have to compute d from $\mathbf{Y}_\infty$ in (3.14) – indeed, $\mathbf{Y}_\infty$ stays the same from interface to interface making it a good anchor point to compute subunit width. This gives

$$d = \frac{\sum_{ab} L_{ab} \left(\left(Y_{\infty,b} - Y_b^{(0),L} \right) - \left(Y_{\infty,a} - Y_a^{(0),R} \right) \right)}{\sum_{ab} L_{ab}}. \tag{3.17}$$

In essence, the Moore-Penrose inverse acts in the orthogonal complement subspace of a kernel of a matrix (see appendix A). As a consequence, Moore-Penrose inverse automatically fixes the shift of X_∞ solutions to 0. However, there are infinitely many solutions of X_∞ that satisfy (3.15) with different values of shift. Here, the physics tells us that if a particle has a finite width, squeezing them together will cause both deformations and a linear translational shift on each interface. This linear shift of the interfaces can cover up the actual deformation. Hence, this is the reason that we have to introduce the shift d explicitly in the energy function (3.1) to isolate this effect from X_i making the usage of Moore-Penrose inverse correct.

With this, we can isolate the deformations from the boundary effect and we are left with

$$\delta \mathbf{X}_{k+1} = \boldsymbol{\mathcal{M}} \delta \mathbf{X}_k. \tag{3.18}$$

This is one step recursion equation whose solutions are simply governed by the eigenvalue and eigenvector of $\boldsymbol{\mathcal{M}}$[1]. Indeed, the general solution is

$$\delta \mathbf{X}_k = \sum_{a=1}^{2l} c_i \lambda_a^k \mathbf{v}_a, \tag{3.19}$$

where λ_a and $\mathbf{v}_a$ are the a^{th} eigenvalue and its corresponding eigenvector of $\boldsymbol{\mathcal{M}}$ respectively. Since the eigenvalues and eigenvectors of $\boldsymbol{\mathcal{M}}$ come in pairs (see appendix B), we obtain directly the deformation of each interface

$$\delta \mathbf{Y}_k = \sum_a \left(c_{+,a} \lambda_{+,a}^k + c_{-,a} \lambda_{-,a}^k \right) \mathbf{V}_a, \tag{3.20}$$

where $\mathbf{V_a}$ is the ath eigenvector of $-\mathbf{M}_{12}^{-1}(\mathbf{M}_{11} + \mathbf{M}_{22})$ and $\lambda_{\pm,a}$ is its corresponding eigenvalue pair. This is a simple yet very important result. It tells us that the deformation modes depend only on the eigenvalues $\lambda_{\pm,a}$, and not on the shape of the subunits. Furthermore, this means that non of the deformation modes interact with each other. We can also expect to see oscillation behavior from this equation when some (or all) eigenvalues are complex.

3.3 Numerical Methods

Although we have obtained a general solution for this model, the boundary conditions are complicated and a closed-form solution is not accessible to us. To extract the information from this model, we have to instead rely on the numerical method. Namely,

[1] In general, even if 0 is contained in the support of the eigenvalue distribution of a random matrix, the probability that any eigenvalue is exactly 0 is 0. The matrix is almost surely diagonalizable.

the linear equation-solving routine that we have used to find the configuration with the minimum deformation energy function.

It is more convenient to solve a single matrix equation with all position vectors and linear terms packed together in one Nl-vector as well as the spring constant matrix packed into one $Nl \times Nl$ matrix. After taking the derivative with respect to $\mathbf{X}_i$ we only have to solve the generalized force balance equation for the whole system

$$\mathbf{M}'\mathbf{X}' = -\mathbf{B}', \tag{3.21}$$

where

$$\begin{aligned}
\mathbf{X}' &= \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_N \end{pmatrix}, \ \mathbf{B}' = \begin{pmatrix} \mathbf{B}^L - d\mathbf{L}^\intercal\mathbf{1} \\ \mathbf{B}^L + \mathbf{B}^R + d\left(\mathbf{L} - \mathbf{L}^\intercal\right)\mathbf{1} \\ \vdots \\ \mathbf{B}^L + \mathbf{B}^R + d\left(\mathbf{L} - \mathbf{L}^\intercal\right)\mathbf{1} \\ \mathbf{B}^R + d\mathbf{L}\mathbf{1} \end{pmatrix}^\intercal, \\
\mathbf{M}'_{Nl\times Nl} &= \begin{pmatrix} \mathbf{M}_{11} & \mathbf{M}_{12} & & & & \\ \mathbf{M}_{12} & \mathbf{M}_{11}+\mathbf{M}_{22} & \mathbf{M}_{12} & & & \\ & & \ddots & & & \\ & & \mathbf{M}_{12} & \mathbf{M}_{11}+\mathbf{M}_{22} & \mathbf{M}_{12} & \\ & & & \mathbf{M}_{12} & \mathbf{M}_{22} & \end{pmatrix}.
\end{aligned} \tag{3.22}$$

There are many algorithms optimized to solve a system of linear equations such as the conjugate gradient method. However, the matrix of interests in our case is very generic and is not full rank. The more reliable way that we have chosen to work with is using the least square method to find the best-fit solution.

4 Warm-up models

4.1 Generalized trapezoid model

We want to preserve the dimension and the shape of subunits while increasing the number of vertices so that we can isolate this effect. For this purpose, we choose the spring constant matrices as follows

$$\mathbf{J} = \mathbf{K} = \frac{1}{8}\begin{pmatrix} 0 & k^c & & 0 \\ k^c & 0 & \ddots & \\ & \ddots & \ddots & k^c \\ 0 & & k^c & 0 \end{pmatrix}, \ \mathbf{L} = \frac{1}{2}\begin{pmatrix} k & & 0 \\ & \ddots & \\ 0 & & k \end{pmatrix}. \tag{4.1}$$

The numerical factors are fixed so that it matches the convention in (2.2). The rest position of the vertices are

$$\begin{aligned} Y_a^R &= (a-1)\frac{\epsilon}{l-1}, \\ Y_a^L &= 1 + \epsilon - Y_a^R \end{aligned} \tag{4.2}$$

These choices reproduce the simple trapezoid model when $l = 2$. Now each subunit is a trapezoid with l vertices on non-parallel edges. The springs only connect two adjacent particles on the same edge, and only two particles in the same row of two edges as shown in Fig.4.

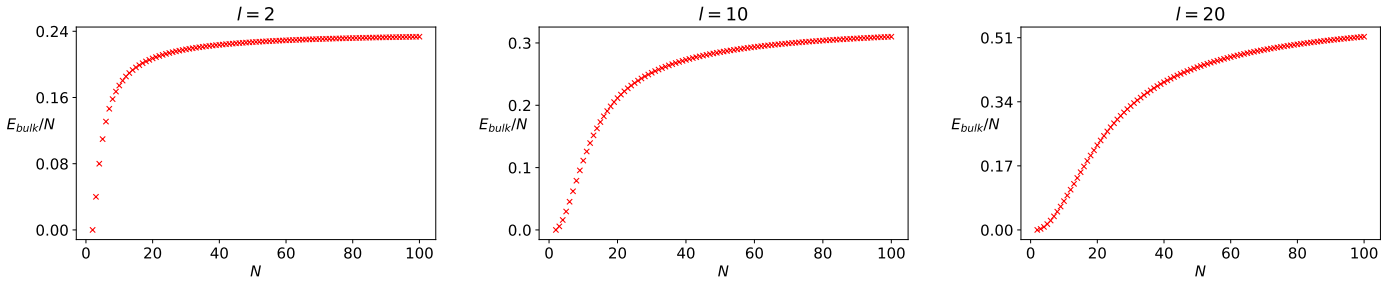


Fig. 5: Deformation energy function per particle E_{bulk}/N plotted as a function of N at unit spring constant $k = k_c = 1$. The energy steadily increases before it saturates at a large value of N. As l increases, the energy function becomes more non-monotonic and saturates at a larger value of N. This effect becomes very apparent at $l = 20$.

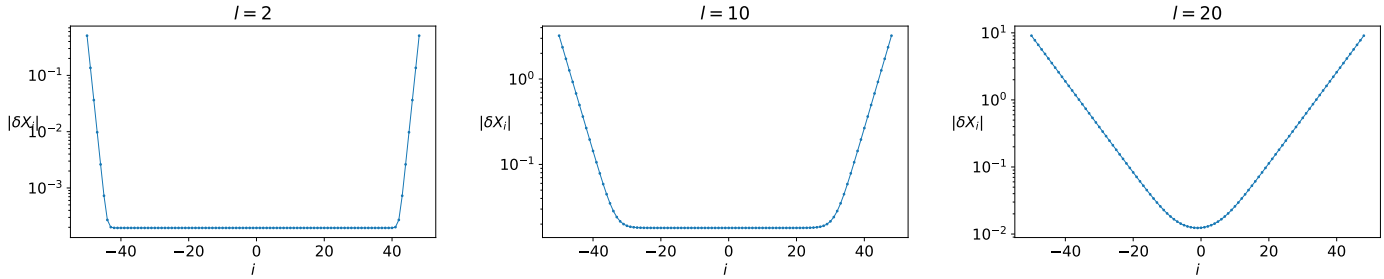


Fig. 6: Deformation norm $|\delta X_i|$ as a function of position i at different values of l plotted in semi-log scale. The spring constant is fixed at $k = k_c = 1$. As the value of l increases, the boundary layer width decreases in size. At $l = 20$ the boundary layer is wider than half the size of the aggregate so the bulk region is hidden.

In the plot of deformation energy per particle Fig.5, we can see that having more vertices makes the subunit more "flexible" by decreasing the slope at the beginning and slows down the saturation, which means a larger aggregate is needed to reach the same amount of deformation energy. This is distinct from changing ϵ in the simple 1D SLA model, which only scales the curve, not stretching the curve horizontally.

The theme of subunits with more vertices being more flexible is also observed in the plot of deformation norm $|\delta X_i|$ in Fig.6. We start at $l = 2$ to make a retrospective comparison with the relative deformation (2.4). In contrast to δ_i, the deformation norm $|\delta \mathbf{X}_i|$ is positive for all values of i but the overall behavior (exponential decay/growth, and the middle plateau) of the two are the same apart from this. By increasing the number of vertices, we increase the width of the boundary layer; the area in the plots where we observe exponential behavior in the deformation norm. The boundary layer increases from 7 particles at $l = 2$ to 20 particles at $l = 10$, and more than 50 particles at $l = 20$, where the bulk region is completely hidden by the boundary layer. Thus, in the aggregate composed of subunits with more vertices, there is less frustration at the same number of subunits, which is in agreement with the deformation energy profile. From this, we expect that larger l will lead to larger aggregates.

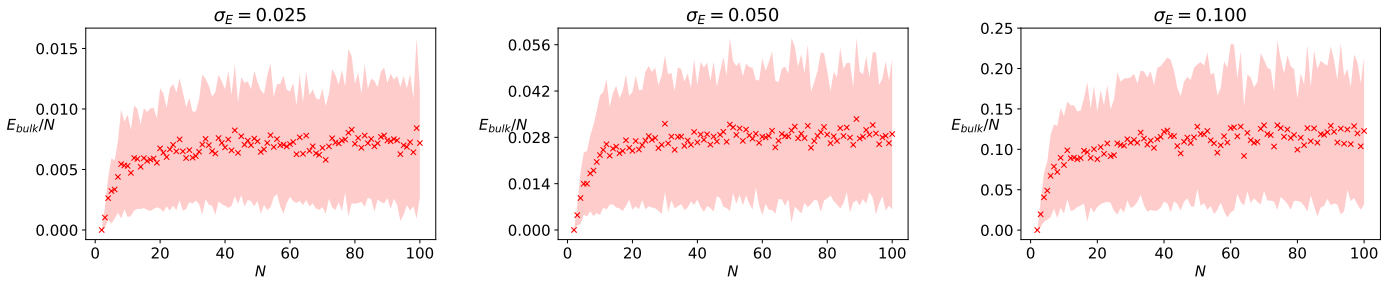


Fig. 7: E_{bulk}/N as a function of N averaged over 100 realizations for an aggregate formed from rectangular particles with $l = 5$ vertices whose rest length are taken from Gaussian distribution with standard deviation $\sigma_E = 0.025, 0.05$, and 0.1. The spring constant is fixed so that $k = k_c = 1$. The shaded area is 1 standard deviation.

4.2 Random flat particles

Having understood the effect of the number of vertices, we move on to a more generic scenario of particles with rough interfaces. We aim to obtain quantitive explanations of the effect of randomness in subunit shape on the deformation energy.

For the first generalization, we keep the antisymmetric subunit shape and keep the spring constant matrices introduced in 4.1. The vertex rest positions are instead identically and independently sampled from a Gaussian distribution with zero mean and variance σ_E:

$$\begin{aligned} Y_a^L &= \mathcal{N}\left(0, \sigma_\epsilon^2\right) \\ Y_a^R &= 1 - Y_a^R. \end{aligned} \tag{4.3}$$

In Fig.7, we observed that the larger the variance of the distribution is, the larger the deformation energy of the aggregate. This is the consequence of the edge of the subunits being more rugged, so there each spring has to deform more when two particles are connected resulting in more frustration in the system. From the simulations, the deformation energy approximately saturates at 0.007, 0.028, and 0.115 as σ_E increases. This result agrees with the simple scaling

$$E_{\text{bulk}} \propto \sigma_\epsilon^2, \tag{4.4}$$

which follows from the fact that the deformation energy is proportional to the vertex position squared as we can see in (3.1).

The behaviors of deformation energy observed here will be the reference to which we can compare a more complex model. One might imagine that a more complex model might change the function shape of the deformation energy per subunit.

5 Fully random SLA

In this section, we will introduce the model for the aggregate with more realistic subunits in 5.1 – having both its shape and its spring constant randomized. After that, we will look into the deformation profile and deformation energy of this model. We shall see that this model gives rise to a new oscillation behavior. This oscillation behavior will be inspected in more detail in 5.2. Finally, we shall discuss in detail the numerical findings of deformation energy from different modes of deformation in 5.3. We will see

the decay modes of deformation give the deformation energy of a similar form to that observed in the random flat model, while the oscillation modes introduce large fluctuations into the deformation energy. Last but not least, we comment on the similarity between the deformation energy of aggregates made with subunits with different l and the possibility of the universal size of aggregates from this model.

5.1 Model definition

We shall keep the randomization scheme of the rest positions of the vertices described in the previous random flat model. In addition to that, we also take the elements of the spring constant matrices from a probability distribution. The physical constraints are that $\mathbf{J}$ and $\mathbf{K}$ have to be symmetric and positive definite – so that the energy function is convex. To this end we sample $\mathbf{J}$, $\mathbf{K}$, and $\mathbf{L}$ from Wishart ensemble. A matrix $\mathbf{A}$ taken from Wishart ensemble has its element defined as the following

$$A_{ij} = \frac{1}{m}\sum_{a=1}^{m} \mathbb{X}_{ia}\mathbb{X}_{ja}, \tag{5.1}$$

where $\mathbb{X}$ is an element of an $n \times m$ matrix whose elements are independently identically distributed from a Gaussian distribution of unit variance. The eigenvalues distribution of Wishart ensemble depend on the ratio $n/m \equiv \lambda_W$. From this definition, we can see that Wishart ensemble satisfies both conditions we require. We assume that sampling spring constants from a random matrix ensemble mimics a physical SLA system where subunits have similar properties but differ in detail because of the presence of noise in the environment.

In Fig.8 we observed that, on top of the previously observed exponential decay of the boundary layer, the bulk of the aggregate shows a novel oscillation behavior. These two behaviors may or may not appear together, depending on the eigenvalues of the matrices. These behaviors are shown in Fig.8 to quantitatively illustrate the deformation in the aggregate. For the purely decay behavior, the aggregate has a well-defined boundary layer, where the subunits relax into their neutral shapes. In the bulk, the interfaces are packed and are invariantly in the frustrated shape. On the other hand, oscillation behavior doesn't display any boundary layer. In this case, one can picture the interfaces here as being nudged around inside the bulk, and never to be fully compressed into their frustrate shapes. The combination of both modes produces somewhat of a mix between oscillations inside the bulk and small boundary layers near both ends. We have chosen here an example where the boundary layers are relatively large. Many other instances of the matrices from the same Wishart ensemble parameter have a much smaller, almost invisible, boundary layer.

The source of this oscillation can be traced back to the eigenvalues of $-\mathbf{M}_{12}^{-1}(\mathbf{M}_{11}+\mathbf{M}_{22})$, which we shall denote by w. For $|w| < 2$, w produces a pair of eigenvalues with unit norm (see appendix B), making these deformation modes purely oscillatory. Mathematically, it suggests that the oscillation is limited to when the eigenvalues w fall into this narrow range. However, we note that we have consistently run into this behavior when we run our simulations. Clearly, for the oscillation to appear, it has to both exist and be selected by the boundary condition. This selection mechanism is, however, much more complex and out of the scope of this report. We shall focus instead only on a setting where these complex eigenvalues arise.

We note that outside of the range $|w| < 2$, w always produces a pair of real eigenvalues, which correspond to purely decay modes. This eliminates the possibility of any other deformation modes.

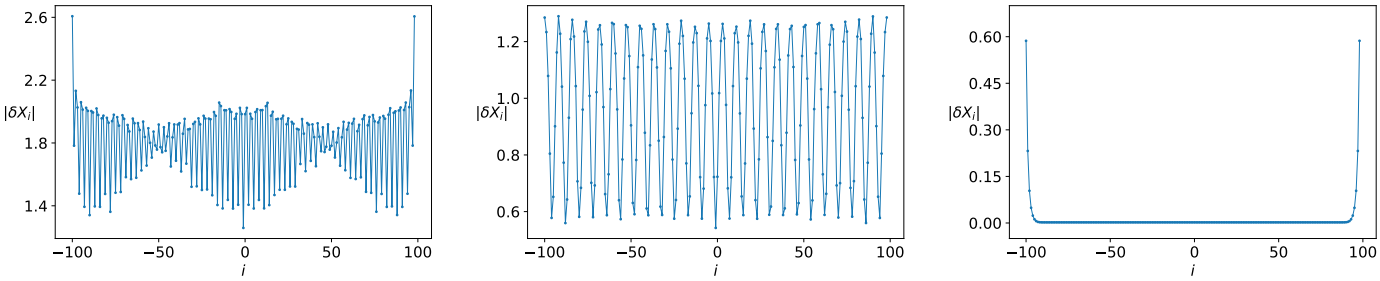


Fig. 8: Deformation norm of a fully random model for $l = 20$, $N = 200$ when there are both oscillation and exponential decay behavior (left), when there is only oscillation behavior (middle), and when there is only exponential decay behavior (right). Note that the amplitude of the oscillation and the decay are of the same order.

One way to control the deformation behavior of this model is to heuristically tune the parameter of the Wishart ensemble to get different ratios of the number of oscillating and decaying modes. However, the relation between eigenvalue distributions and random matrices is usually not straightforward, and we would gain very little physical intuition in that manner. We shall instead turn to a simpler model to gain some simple explanations behind this oscillation behavior.

5.2 Simple model for oscillation

In order to better our understanding of the oscillation modes, we shall come back to our simplest model with some modifications.

We consider the subunit with sprint constant matrices

$$\mathbf{J} = \mathbf{K} = \frac{1}{2}\begin{pmatrix} 0 & a \\ a & 0 \end{pmatrix}, \ \mathbf{L} = \begin{pmatrix} b & c \\ c & b \end{pmatrix}. \tag{5.2}$$

We shall let the rest positions of the vertices remain as random variables from the Gaussian distribution described in (4.3). Although we can fix the shape of the subunits, this oscillation behavior originates purely from the eigenvalues of $-\mathbf{M}_{12}^{-1}(\mathbf{M}_{11} + \mathbf{M}_{22})$, w, and is thus unaffected by the shape of the subunits. Essentially, this is a simple (random) trapezoid particle, with extra diagonal springs. We will show that these additional elements allow the oscillation modes.

It can be shown that the eigenvalues in this setting are

$$w_{1,2} = 2, \ \frac{2}{b^2 - c^2}\left((a+b+c)^2 - a^2\right). \tag{5.3}$$

To get an oscillation mode, we want $|w_2| < 2$. In practice, one can try to understand it numerically. From Fig.9, we can see that, at $a = 1$, oscillation can happen when three parameters don't have the same sign. Within this condition, oscillation is observed in a wide range of parameter values.

Physically, this suggests that oscillation modes arise from the existance of "anti-springs", non-physical springs whose role is to encourage displacement from the neutral position. With only the existence of ordinary springs, the interface can just stay straight to minimally stretch all springs. However, this picture changes when there are two kinds of springs. For example, in the case of $a > 0$ and $c < 0 \ll b$ – approximately trapezoid subunit with two anti-springs connecting two vertices on the same layer, the

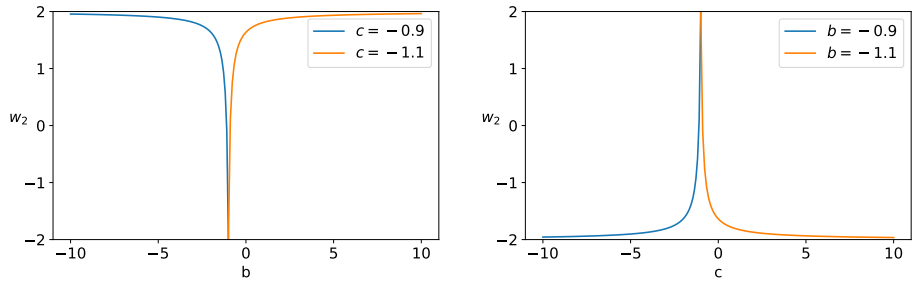


Fig. 9: w_2 as a funtion of b (left) and c (right) at $a = 1$ fixed. In each plot, we further fix the other parameter to demonstrate the range of varying parameters in which $|w_2| < 2$ is accessible. w_2 in both plots diverges when b or c approaches $-a$. Further numeric results show that a dictates where w_2 diverges. When c is fixed, $w_2 = 2$ is approached from below, and when b is fixed, $w_2 = -2$ is approached from above.

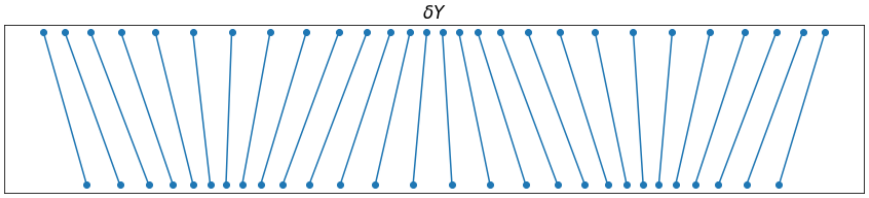


Fig. 10: A schematic illustration of an oscillation mode in the simple model descibed in (5.2) with $a > 0$ and $c < 0 \ll b$. Each dot represents a vertex of the subunits, and the line connecting them between two layers is the interface of subunits. In this configuration, the vertices stay closer/further away from vertices on neighbor interfaces than their rest distance, but approximately maintain the rest distance to the other vertex on the same interface. Since the anti-springs that connect the vertices between two layers is stronger than the springs that connect vertices on the same interface, any overall shift of either layer can further lower the total deformation energy leading to instability.

energy is higher when all interfaces are parallel than when vertices on the same interface deviate from the rest distance from vertices on the neighbor interfaces but maintain the rest distance from the other vertex on the same interface. This can be achieved by an oscillating pattern shown in the right of Fig.10. We note that anti-springs are indeed non-physical and introduce instability to the system – any shift of either all top or bottom layer nodes can furthur decrease the energy – and thus does not align with our initial goal to find a configuration of the system that minimizes the energy. While this simple model does not meet our criteria, we nonetheless take it as a possible explanation for the oscillation modes.

5.3 Numerical results for deformation energy

Having understood the deformation behavior in the aggregate, we are now ready to return to our main path: to attempt to describe the size of aggregates of fully random subunits in equilibrium.

In the introduction of this section, we have argued that there are two distinct kinds of deformation modes: oscillation and exponential decay. From (3.20), any deformation of an aggregate can be decomposed to a linear combination of these two, so it is useful to first consider their behavior separately. For this purpose, we artificially tune the Wishart parameter to find ensembles in which there is mostly oscillation (the

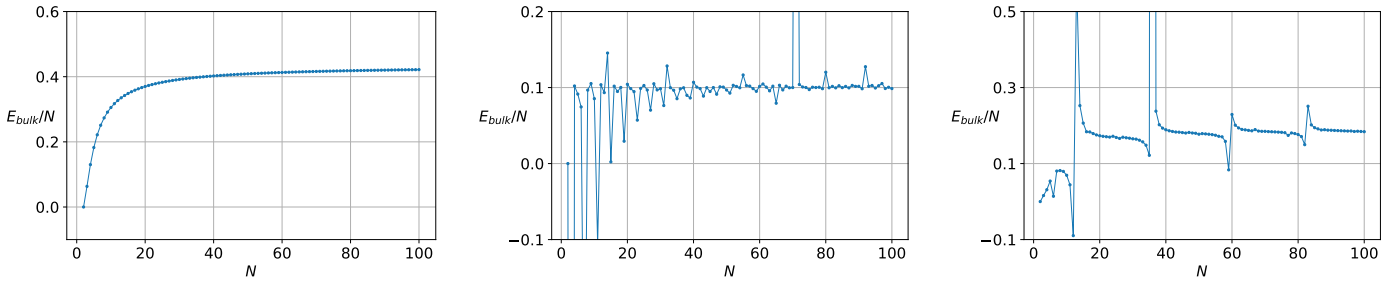


Fig. 11: Deformation energy per subunit as a function of the number of subunits at $l = 10$ when decay modes dominate (left), oscillation modes dominate (middle), and when both modes are approximately equally present (right). The energy function profile in the regime of decay modes is similar to that from the random flat particle model. On the other hand, we observe dips when oscillations are present. When there are both modes, the intervals of dips appear cleaner than when there are only oscillation mode. This is because there are fewer oscillation modes competing with each other in such a small system. We expect the energy fucntion to take a more complex shape as l inceases.

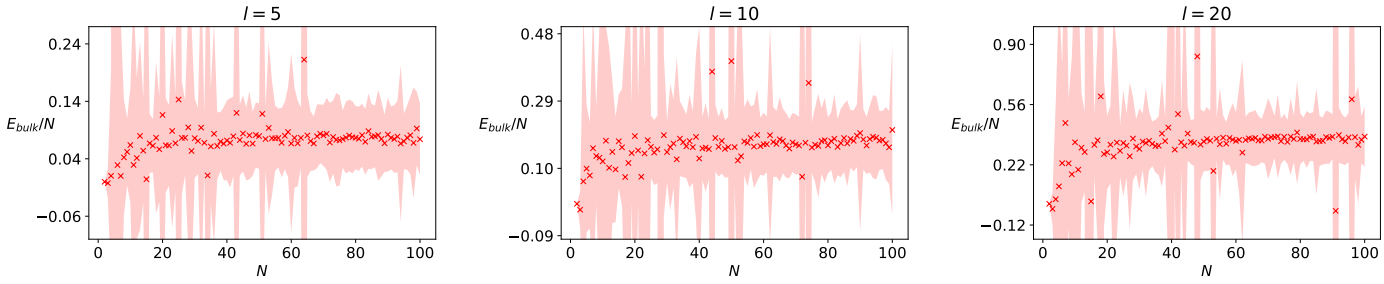


Fig. 12: Deformation energy per subunit as a function of the number of subunits in the middle regime where both decay and oscillation modes equally contribute. The plot is an average of 100 realizations and the shaded area is 1 standard deviation. When l is varied, the amplitudes of the energy functions change but not their functional shape.

probability of finding $|w| < 2$ is close to 1), mostly decay (close to 0), equal combination of both (close to 0.5) (see appendix C for the exact definition of each setting). A single realization from each ensemble is shown in Fig. 11 to illustrate qualitative behaviors of each deformation mode.

When the aggregates have the decay modes as the majority, the average energy function profile looks qualitatively similar to that of the case of generalized trapezoid subunits. Indeed, their deformations come from the same mechnism. Instead, when the oscillation modes are the majority, we observe that the overall shape of the deformation energy is rather flat. On top of that, there are dips in the deformation enegy. We associate these dips with resonant-like behavior of the deformation. This is not surprising since complex-valued eigenvalues w introduce period into the system through their phases. This period of deformation is given by $2\pi/\varphi$, where φ is the phase of the (complex) eigenvalue w. The dips in the deformation energy are positioned at intervals of this period. When there are many complex eigenvalues, we observe that the period whose value is closet to integer dictates the interval of the dips. Finally, when both modes are equally present, we observed a build-up of the energy function for small aggregate similar to what have previously been observed in simpler models. Unlike in

the mostly oscillation case, the initial increase lasts for around 10 subunits, confirming the existence of the build-up region. Moreover, we also observe dips which are associated to the oscillation modes. Interestingly, the deformation energy per subunit of the oscillation mode is half of that of the decay mode. In the real-world system, these dips might act as a soft lock to the size of the aggregate, allowing only aggregate length in certain multiplicity.

A natural question one can ask oneself is whether the increase in the number of vertices slows down the saturation of deformation energy as we have seen in the case of random flat subunits in section 4.2. To answer this, we look at the average of the deformation energy over the realizations of the ensemble with the same parameters. This allows us to croase grain the details of the energy and focus on the typical shape of the function. In Fig.12, we see that the deformation energy has the same profile when l is varied, e.g. large variance and small build-up region. In fact, all average build-up regions seem to be approximate of the same size: around 10 subunits. Large fluctuations in the plot suggest that the deformation profile in each realization can vary hugely.

It is also worth noting that the build-up region of $l = 20$ fully random model when decay modes dominate, in Fig.11, roughly has the width of about 10 subunits. This is roughly the same width as that of the random flat subunits model in Fig.7. This information suggests that the randomness in vertices positions destroys build-up suppression that we have observed in the generalized trapezoid model in Fig.5.

From these results, we expect that the size of the aggregate that minimizes total energy would be a universal quantity independent of l and spring constants since Γ is independent of the details of the subunit. However, we cannot rule out its dependency on the shape of subunits yet since the rest shape of the subunits are sampled from Gaussian distribution in all of our simulations.

6 Discussion

Pathological amyloids are formed by when misfit protein molecules come aggregate together. This phenomenon can be modeled as an open boundary SLA where particles come together to form frustrated aggregates, which in turn, costs deformation energy and limits the aggregate size. In this report, we successfully demonstrate how deformation occurs in aggregate with complex subunits as well as discover a novel oscillation behavior in the bulk of aggregate. This is the first to understand the deformation energy of an aggregate made of complex subunits, which will allow us to understand how the size of an aggregate is determined through the structure of its subunits. Although the original goal of this study is not reached, we have provided a qualitative description of the deformation energy which is a first crucial step to explore the rest of the model.

When the subunit is complex, we observe that the deformation can be classified into decay and oscillation modes. The decay modes are similar to the deformation modes previously observed in simple particles and can be accessed through a wide range of parameters. On the other hand, the oscillation mode can only be accessed through a small range of eigenvalues of $-\mathbf{M}_{12}^{-1}(\mathbf{M}_{11} + \mathbf{M}_{22})$. The emergence of the oscillation mode would provide a point of contact to experiments; to see if this oscillation mode is actually observed in real-world physical systems. It also would be interesting to ask if there exists any mechanism in nature to encourage this oscillation behavior. We also observed that the deformation energy takes on similar functional shapes across different subunit sizes l. This suggests a universality of aggregate size. However, simulations with higher resolutions are needed in order to quantify this.

It would be interesting to test whether what we have observed here holds in a less constrained system. This will allow us to close the gap between mathematical models and real-world protein assembly. For example, one might consider instead asymmetric subunits. One immediate change would be that the interfaces will no longer be able to relax to a flat shape inside the bulk of an aggregate since $\mathbf{B}^L$ no longer cancels with $\mathbf{B}^R$. Another generalization is to see if the distribution of vertex rest positions will affect the functional profile of deformation energy. The most obvious generalization, but perhaps the hardest to be done, is to consider this model in three dimensions, where real-world proteins reside. This will be undoubtedly hard to tackle mathematically, but, when done, will provide tremendous insight into how the SLA of real-world proteins might behave and what new behavior we could expect to see.

It is also interesting to understand the mechanism of how the deformation mode is selected by the boundary conditions i.e. how the boundary conditions assign values to the deformation coefficient $c_{\pm,a}$ in (3.20). We note that we have observed that it is not necessary for the eigenvalue with the largest norm to dictate the deformation in the boundary layer of an aggregate. However, we need to extend the current calculation in order to explain this phenomenon.

Lastly, the connection between the build-up region and the boundary layer width suggests a direct connection between the deformation energy function and the eigenvalues w. It would be interesting to see if one can write energy as some scaling law of w.

Appendix A Moore-Penrose Inverse

Moore-Penrose Inverse arises when one applies the least square method to a non-consistent system of linear equations [15]. It is a way to extract the closest solution when there is no exact solution. However, despite its ubiquitous utility, it is not easy to get an intuition from its mathematical definition. In this section, we shall illustrate the physical meaning of the Moore-Penrose inverse inspired by [16].

We consider a generic system of linear equations

$$\mathbb{A}\mathbf{x} = \mathbb{B}, \tag{A.1}$$

where $\mathbb{A}$ is a $m \times n$ matrix with $m \geq n$, $\mathbf{x}$ is an n-vector, and $\mathbb{B}$ is an m-vector. If $\mathbb{A}$ is full-rank and invertible, then $\mathbf{x}$ is simply $\mathbb{A}^{-1}\mathbb{B}$. In this case, $\mathbb{B}$ can be decomposed into the sum of column vectors of $\mathbb{A}$ (denoted by $\mathbb{A}_i$) with $\mathbf{x}$ being the corresponding coefficient vector

$$\mathbb{B} = \sum_i x_i \mathbb{A}_i. \tag{A.2}$$

Mathematically, $\mathbb{B}$ lives in the space spanned by $\mathbb{A}$, denoted by $S(\mathbb{A})$.

On the other hand, if the ordinary inverse of $\mathbb{A}$ doesn't exist, then $\mathbb{B}$ simply doesn't live in $S(\mathbb{A})$. We can, however, try to find the closest approximation of $\mathbb{B}$ that lives in $S(\mathbb{A})$. Intuitively, this approximation is the orthogonal projection of $\mathbb{B}$ onto $S(\mathbb{A})$. In column vector basis, the projector onto $S(\mathbb{A})$ is simply $\mathbb{A}^\intercal$; after this transformation is applied, all components of $\mathbb{B}$ orthogonal to $S(\mathbb{A})$ is sent to zero. The process of trying to approximate the closest solution then corresponds to solving

$$\mathbb{A}^\intercal\mathbb{A}\mathbf{x}^* = \mathbb{A}^\intercal\mathbb{B}. \tag{A.3}$$

Now, $\mathbb{A}^\intercal\mathbb{A}$ is a full-rank $n \times n$ matrix that is now invertible. The approximate solution is then given by

$$\mathbf{x}^* = \left(\mathbb{A}^\intercal\mathbb{A}\right)^{-1}\mathbb{A}^\intercal\mathbb{B}. \tag{A.4}$$

We now arrive at $(\mathbb{A}^\intercal\mathbb{A})^{-1}\,\mathbb{A}^\intercal$, the celebrated Moore-Penrose inverse.

In our case, $\mathbb{A}$ is a non-invertible square matrix so $\mathbb{A}^\intercal\mathbb{A}$ is not invertible. The construction of the Moore-Penrose inverse has to be done via singular value decomposition instead; by replacing each singular value with its inverse while leaving 0. However, the physical intuition from this example still holds.

It is important to note that this approximation treats the orthogonal complement of $\mathbb{B}$ as if they don't exist. This is the reason why we have to explicitly isolate the linear shift d in (3.1). Otherwise, this effect would be lost after the Moore-Penrose inverse is applied.

Appendix B Eigenvalues and Eigenvectors of a 2×2 block matrix

Here, we shall consider eigenvalues and eigenvectors of a $2\times$ block matrix that appear in (3.11). In general, for a 2×2 block matrix of the structure

$$\boldsymbol{\mathcal{A}} = \begin{pmatrix} \mathbf{0} & \mathbb{I} \\ -\mathbb{I} & \mathbf{A} \end{pmatrix}, \tag{B.1}$$

the eigenvalue of $\boldsymbol{\mathcal{A}}$ can be found in terms of the eigenvalues of $\mathbf{A}$ by the following arguement.

Let $\begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix}$ be an eigenvector of $\boldsymbol{\mathcal{A}}$ with the corresponding eigenvalue λ and $\mathbf{AV} = k\mathbf{V}$. We have that

$$\begin{aligned} \mathbf{V} &= \lambda\mathbf{U} \\ -\mathbf{U} + k\mathbf{V} &= \lambda\mathbf{V}. \end{aligned} \tag{B.2}$$

Solving for k gives

$$k = \frac{1}{\lambda} + \lambda \quad \Rightarrow \quad \lambda_\pm = \frac{k \pm \sqrt{k^2 - 4}}{2} = \frac{1}{\lambda_\mp} \tag{B.3}$$

This means

$$\boldsymbol{\mathcal{A}} \begin{pmatrix} \mathbf{U} \\ \mathbf{V} \end{pmatrix} = \begin{pmatrix} \mathbf{V} \\ \lambda\mathbf{V} \end{pmatrix}.$$

Note also that λ_+ and λ_- arise from the same eigenvalue k and thus we also have the relation for a pair of corresponding eigenvectors of $\boldsymbol{\mathcal{A}}$

$$\mathbf{v}_\pm \propto \begin{pmatrix} \mathbf{V} \\ \lambda_\pm\mathbf{V} \end{pmatrix}. \tag{B.4}$$

Another implication of (B.3) is that for $|k| < 2$, we have that $|\lambda_\pm| = 1$. These unit norm complex eigenvalues correspond to purely oscillation modes in the fully random model.

Appendix C Modified Wishart distribution

In order to investigate the behavior of different deformation modes, we need to artificially modify the Wishart distribution so that the eigenvalue distributions are contained in the desired region.

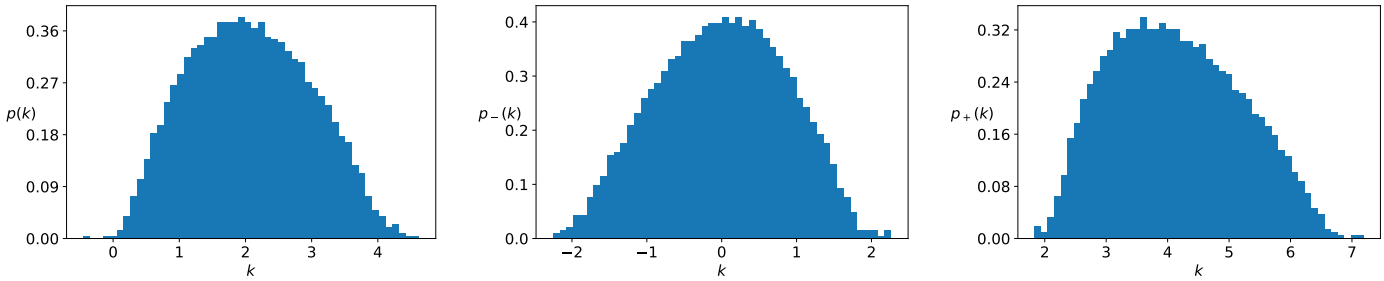


Fig. 13: Left: the distribution of k when $\mathbf{J}$ is sampled from normal Wishart ensemble with $\lambda = 0.02$. Middle: the distribution of k when $\mathbf{M_{11}}$ and $\mathbf{M_{22}}$ are computed from $\mathbf{J}^*_-$ and $\mathbf{K}^*_-$. Right: $\mathbf{J}^*_+$ and $\mathbf{K}^*_+$. All of the histograms are taken at $N = 2000$. Note that not only the shapes but also the centers change when $\mathbf{J}$ and $\mathbf{K}$ are modified. This reflects very complex relations between random matrices and their eigenvalues.

Our goal is to modify $\mathbf{J}$, $\mathbf{K}$, and $\mathbf{L}$ in order to control the eigenvalue distribution of $-\mathbf{M}_{12}^{-1}(\mathbf{M}_{11}+\mathbf{M}_{22})$, again, denoted by k to be inside or outside the region between -2 and 2. In general, this is a very complicated task, as the matrix in the consideration is an extremely complicated function of $\mathbf{J}$, $\mathbf{K}$, and $\mathbf{L}$. Fortunately, we found that at the Wishart ensemble parameter $\lambda = 0.02$, the distribution of k denoted by p(k) is roughly centered around 2 with half of it inside the region $[-2, 2]$. We discovered that this distribution is shifted to the left (or the right), by computing $\mathbf{M_{11}}$ and $\mathbf{M_{22}}$ from $\mathbf{J}^*_+ = \mathbf{K}^*_+ = \mathbf{J} + \mathbb{1}/2N$ (or from $\mathbf{J}^*_- = \mathbf{K}^*_- = \mathbf{J} - \mathbb{1}/2N$), where N is the matrix size, while leaving $\mathbf{L}$ as it is. We shall denote the resultant distribution from adding $p_+(k)$, and subtracting $p_-(k)$. Numerically, at $N = 2000$, we found that

$$\begin{aligned} \int_{-2}^{2} dk\, p(k) &= 0.480 \\ \int_{-2}^{2} dk\, p_+(k) &= 0.003 \\ \int_{-2}^{2} dk\, p_-(k) &= 0.992. \end{aligned} \tag{C.1}$$

These values confirm that the choice of parameters and modifications is suitable to describe three different cases in section 5.3. The shapes of these distributions are shown in Fig.13.

Technically, this modification to $\mathbf{J}$ and $\mathbf{K}$ is not physical, since they can change the positive definiteness of these matrices, which results in changes in the convexity of the energy landscape. However, this does not concern us as this setting only serves as a playground where we isolate and investigate each kind of different mode of deformation independently from each other. These modes can naturally arise in a more physical setting as well and, in our cases of interest, they don't interact with each other. Thus, studying each mode separately in artificial settings provides accurate information that also applies to the physical model as well.

References

[1] Alberts, B., Johnson, A., Lewis J., Raff, M., Roberts, K., & Walter, P. (2002). *Molecular Biology of the Cell* (Garland Science, New York), fourth edition.

[2] Yadav, S., Sharma, A. K., & Kumar, P. (2020). *Nanoscale Self-Assembly for Therapeutic Delivery. Frontiers in Bioengineering and Biotechnology.* https://doi.org/10.3389/fbioe.2020.00127

[3] Pochan, D., & Scherman, O. A. (2021). *Introduction: Molecular Self-Assembly.* Chemical Reviews. https://doi.org/10.1021/acs.chemrev.1c00884

[4] Hagan, M. F., & Grason, G. M. (2021). *Equilibrium mechanisms of self-limiting assembly.* Reviews of Modern Physics. https://doi.org/10.1103/revmodphys.93.025008

[5] Eisenberg, D., & Jucker, M. (2012). The Amyloid State of Proteins in Human Diseases. Cell. https://doi.org/10.1016/j.cell.2012.02.022

[6] Haass, C., & Selkoe, D. (2007). Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. Nat Rev Mol Cell Biol 8, 101–112. https://doi.org/10.1038/nrm2101

[7] Ray, S. et al (2020). α-Synuclein aggregation nucleates through liquid–liquid phase separation. Nature Chemistry. https://doi.org/10.1038/s41557-020-0465-9

[8] Knowles, T., Vendruscolo, M., & Dobson, C. (2014). *The amyloid state and its association with protein misfolding diseases.* Nat Rev Mol Cell Biol 15, 384–396. https://doi.org/10.1038/nrm3810

[9] Chiti, F., & Dobson, C. M. (2017). *Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade.* Annual Review of Biochemistry. https://doi.org/10.1146/annurev-biochem-061516-045115

[10] Eaton, W. A., & Hofrichter, J. (1990). Sickle Cell Hemoglobin Polymerization. Elsevier EBooks. https://doi.org/10.1016/s0065-3233(08)60287-9

[11] Turner, M. S. et al (2003). Twisted Protein Aggregates and Disease: The Stability of Sickle Hemoglobin Fibers. Physical Review Letters. https://doi.org/10.1103/physrevlett.90.128103

[12] Boyer, D. R., Na, M., & Sawaya, M. R. (2021). Why amyloid fibrils have a limited width. Cold Spring Harbor Laboratory - BioRxiv. https://doi.org/10.1101/2021.07.02.450971

[13] Lenz, M., & Witten T. A. (2017). *Geometrical frustration yields fibre formation in self-assembly.* Nature Phys 13, 1100–1104. https://doi.org/10.1038/nphys4184

[14] Le Roy, H., Terzi, M. M., & Lenz, M. work in progress

[15] Baksalary, O.M., Trenkler, G. (2021). *The Moore–Penrose inverse: a hundred years on a frontline of physics research.* EPJ H 46, 9. https://doi.org/10.1140/epjh/s13129-021-00011-y

[16] Zanetti, M. *Geometrical meaning of the Moore-Penrose pseudo inverse.* http://massimozanetti.altervista.org/files/mydocs/geometricalMeaningMoorePenrose.pdf