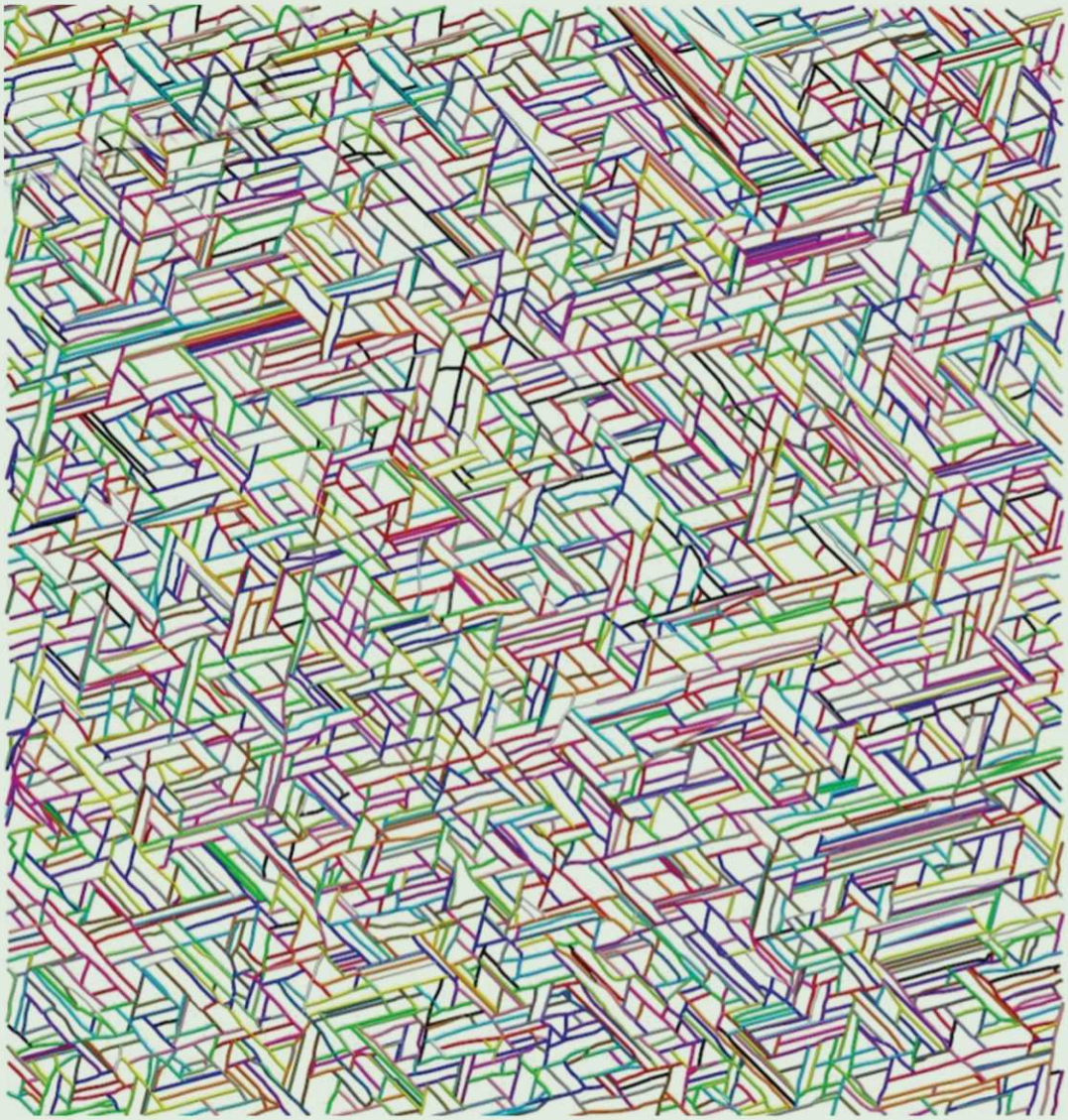
PHYSICAL REVIEW LETTERS

Articles published week ending 11 JULY 2014



Published by
American Physical Society

APS physics

Volume 113, Number 2

Kinetic Signature of Fractal-like Filament Networks Formed by Orientational Linear Epitaxy

Wonmuk Hwang*
Departments of Biomedical Engineering and Materials Science & Engineering, Texas A&M University, College Station, Texas 77845, USA and School of Computational Sciences, Korea Institute for Advanced Study, Seoul 130-722, Korea

Esma Eryilmaz
Department of Physics, Texas A&M University, College Station, Texas 77845, USA
(Received 22 February 2014; revised manuscript received 22 April 2014; published 8 July 2014)

We study a broad class of epitaxial assembly of filament networks on lattice surfaces. Over time, a scale-free behavior emerges with a 2.5–3 power-law exponent in filament length distribution. Partitioning between the power-law and exponential behaviors in a network can be used to find the stage and kinetic parameters of the assembly process. To analyze real-world networks, we develop a computer program that measures the network architecture in experimental images. Application to triaxial networks of collagen fibrils shows quantitative agreement with our model. Our unifying approach can be used for characterizing and controlling the network formation that is observed across biological and nonbiological systems.

DOI: 10.1103/PhysRevLett.113.025502 PACS numbers: 81.16.Fg, 81.16.Rf, 87.10.Rt, 87.16.Ka

Fractal patterns are vast in nature [1]. Because of their scale-free character, relevant feature sizes follow a power-law distribution. Disparate systems exhibiting the same power law often share the same underlying physical feature. Characterizing systems or phenomena according to their power-law exponent is thus an important starting point for understanding them [2–4]. In reality, the scale-free behavior does not continue indefinitely, for which the finite rate of pattern formation process plays a role. Knowledge of such limiting factors can provide additional information about the system. Here, we apply this idea to a new class of filament assembly process on surfaces. The growth direction of a filament is guided by the underlying substrate lattice, and growth stops by steric encounter with other filaments on the substrate (Fig. 1). This process is observed in a broad range of systems, including collagens [5–8], various β-sheet forming peptides [9–14], and organic nanofibers [15–20]. Commonly used substrates include mica, graphite, and KCl, where assembly occurs either in solution (mainly biofilaments) or via vapor deposition (organic nanofibers). The filament network can be triaxial, biaxial, rectangular [Figs. 1(a)–1(c)], or uniaxial, depending on the symmetry of the substrate lattice and the degree to which nascent filaments (nuclei) can rotate on the surface and align with the axes of the lattice [8]. Rotational motion can be controlled, e.g., by the buffer that affects the strength and anisotropy of the interaction between the molecule and the substrate [8,21]. When directional bias is not strong, an isotropic network results [Fig. 1(d)] [22–26].

In these template-directed [7,9,12] or epitaxylike [8,10,11,14] assemblies, the molecular subunits forming a filament are often much greater than the lattice unit cell [8]. In the case of physisorbed films, lattice-imposed ordering of such incommensurate structures is known as "orientational epitaxy" [18,21,30,31]. We thus call the directional growth of filaments on substrates orientational linear epitaxy (OLE). Nanoscopically and chemically defined layers generated by OLE have biomedical [5,9,27,32] and optoelectronic [20,28] applications. While previous studies related to OLE focus on the directional guidance mechanism by the lattice [18,20,21,30,31], morphogenesis of the network as a whole has not been investigated in detail.

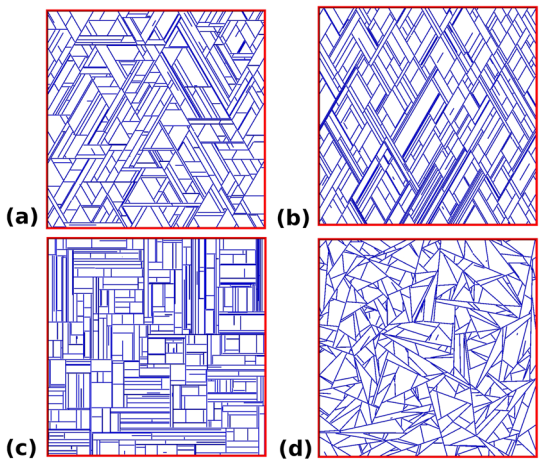


FIG. 1 (color online). Four types of OLE networks that have been experimentally observed: (a) Triaxial [8–15,18,19,21,27], (b) biaxial [5,6], (c) rectangular [16,17,28], and (d) isotropic [22–26]. (a),(b) Triaxial and biaxial form on hexagonal lattices, notably on mica. (c) Rectangular forms on cubic lattices such as KCl. (d) Istotropic forms in the absence of orientational bias. See also Supplemental Material, videos [29].

In building a model of OLE, as similar networks form in diverse systems, we keep only essential features as generally done in models of nonequilibrium statistical mechanics: Filaments nucleate with rate n (number of nucleation events per unit time in a square substrate of width w), at which the growth direction is randomly assigned based on the orientational symmetry of the system. After nucleation, filaments grow with rate g (number of subunits of size σ added to the end of a filament per unit time). Growth stops when a filament end encounters another filament. Figure 1 shows examples generated with $n = g = w = 1$, and $\sigma = 0.002$, for four different types of orientational symmetry observed in experiments (see also videos in the Supplemental Material [29]).

In reality, filament nucleation may result from interactions among surface-diffusing monomers or multimers, details of which are lumped into the nucleation rate n. Similarly, although filament elongation is expected to occur mainly by subunit addition, other pathways such as annealing between growing filaments may also contribute. Such details are system dependent, whose net effect is parametrized by the growth rate g. So a subunit in our model does not necessarily correspond to an actual molecule but instead it represents a filament segment by which the kinetic parameters are defined. Our model also assumes an infinite reservoir of subunits available for nucleation and filament growth. In the case when a finite amount of subunits are used, only the formation of short filaments at later stages will be affected by depletion whereas most parts of the network, as in Fig. 1, that grew while there were abundant subunits, remain nearly intact. To a first approximation, our model thus captures key elements of the OLE process for which there can be a wide range of experimental scenarios.

The OLE process can be divided conceptually into three stages: (i) early nucleation where filaments grow on mostly empty substrate, (ii) encounters among the early-formed filaments, and (iii) self-similar network formation. Stage (i) lasts longer for smaller n (sparser filaments) and smaller g (slower elongation). In stage (ii), filaments are still sparsely distributed and growth terminates mostly by random encounters. In stage (iii), the substrate is divided into polygons possessing a given orientational symmetry, which are divided by new filaments into progressively smaller ones in a roughly self-similar manner.

We use a mean-field theory (MFT) for stage (i) since filaments grow independently. Let L be the filament length (number of subunits). The master equation for its distribution $P_L(t)$ at time t is

$$\dot{P}_1 = n - gP_1,$$
$$\dot{P}_L = g(P_{L-1} - P_L), \quad (L > 1). \tag{1}$$

Here, n describes the increase in P_1 by nucleation. The gain (gP_{L-1}) and loss ($-gP_L$) terms are due to subunit addition. The solution to Eq. (1) is given asymptotically (Supplemental Material [29])

$$P_L(t) \simeq \frac{n}{2g}\mathrm{erfc}\left(\frac{L-gt}{\sqrt{2gt}}\right), \tag{2}$$

where $\mathrm{erfc}(x) = (2/\sqrt{\pi})\int_x^\infty e^{-s^2}ds$ is the complementary error function. Equation (2) saturates to (n/g) for short filaments ($L \ll gt$), and the growing front ($L \sim gt$) has speed g. Stochastic simulation without steric interaction between filaments agrees well with Eq. (2) (Supplemental Material, Fig. S.1 [29]).

To consider encounters among filaments in stage (ii), we add second-order interaction terms:

$$\dot{P}_1 = n - (g + aS)P_1,$$
$$\dot{P}_L = gP_{L-1} - (g + aS)P_L \quad (L > 1), \tag{3}$$

where a is the encounter rate and $S \equiv \sum_m P_m$ is the total filament number density. Equation (3) has an exponentially distributed steady state solution (Supplemental Material [29])

$$P_L = \frac{n}{g}e^{-L\ln(1+\sqrt{na}/g)}. \tag{4}$$

The exponential nature of the distribution remains even if we use the total subunit number density (total length of filaments) $M \equiv \sum_{L=1}^\infty LP_L$ instead of S in Eq. (3) (Supplemental Material [29]).

Filaments interact more strongly in stage (iii), so we use stochastic simulation. Among the model parameters, we fix $n = 1$ and vary g (the choice for n only affects the time scale). Also, for better statistics, we use a small subunit size $\sigma = 10^{-4}$ while $w = 1$ (Fig. 2). After the MFT-like growth of initially nucleated filaments [Fig. 2(a), $t \lesssim 1200$], shorter filaments start to increase in number as their growth becomes limited [Fig. 2(a), $t \gtrsim 2000$]. Similar to Eq. (4), P_L for long filaments is exponential, e^{-kL} [Fig. 2(b), inset]. Later on, P_L becomes a power law as shorter filaments progressively fill smaller areas with similar geometric features determined by the orientational symmetry of the network ($P_L \sim L^{-b}$). The power law extends to smaller lengths with greater t or g [dashed arrows in Figs. 2(a) and 2(b)].

Other types of networks in Figs. 1(b)–1(d) behave similarly although the time to reach the self-similar regime differs (Supplemental Material, Fig. S.2 [29]). The power law exponent b appears universal, about 2.5 for smaller lengths and 3 for longer lengths before the exponential distribution [Fig. 2(c); $4 \lesssim \log(L) \lesssim 7$]. Interestingly, a similar power-law exponent has been observed in planar fragmentation [33,34]. Regarding each filament in our model as a crack on a two-dimensional plate, OLE has features analogous to fragmentation [34,35]. The fractal

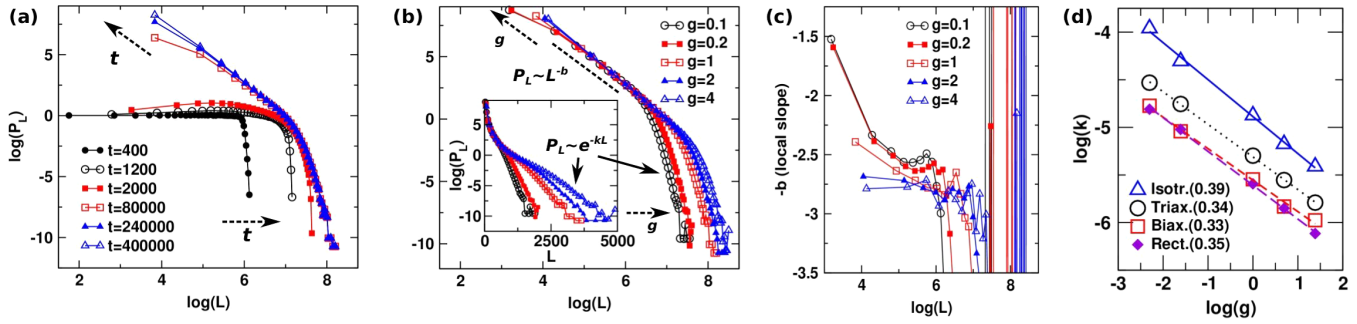


FIG. 2 (color online). Length distribution of OLE networks. Measurements were made over 100–500 runs in each case. (a) Time evolution of P_L for the triaxial network ($g = 1$). For $t \lesssim 1200$ (early nucleation regime), P_L follows the MFT solution, Eq. (2). For example, the front of P_L moves with speed $g = 1$, so that at $t = 400$, it is located at $\log(400) = 6.0$. As filaments encounter ($t \gtrsim 2000$), P_L becomes "frozen," starting from longer filaments and, subsequently, to shorter ones (dashed arrows). (b) Near-steady state profiles at $t = 4 \times 10^5$. Dependence of P_L on g (dashed arrows) is analogous to the t dependence in (a). Inset: log-normal plot showing exponential distribution of longer filaments. (c) Local slopes of the curves in (b). For the intermediate range of L where the power law holds, $b = 2.5 - 3$. The power law behavior becomes more prominent with greater g as this is similar to a longer-time case. Similar behaviors are observed for the networks in Figs. 1(b)–(d) (Supplemental Material, Fig. S.2 [29]). (d) Log-log plot of k vs g, indicating $k \sim g^{-1/3}$ across different networks. Numbers in legends are slopes of linear fits. The prefactor in this relation (k at $g = 1$) correlates with the propensity for encounter between filaments, which is the greatest for the isotropic network, followed by triaxial, and the smallest for biaxial and rectangular networks that behave very similarly.

dimension D of a fragmentation network is related to b by $b = D + 1$ [34]. Since b approaches 3 with t or g, $D \to 2$. This means that, for length scales much greater than those of empty areas available for filament growth, the network effectively fills the whole substrate, becoming a two-dimensional compact object [3].

The inverse length k of the exponential decay follows another universal power law, $k \sim g^{-1/3}$ [Fig. 2(d)]. To see if similar behavior is observed without steric interaction between filaments, we modified the stochastic model. As for MFT, filaments do not interact in space, but instead they are capped with probability $1 - e^{-aS}$ each time when a subunit attempts to add to an uncapped filament end. The capping probability increases with S and a, as in Eq. (3). This case has $b \simeq 2$, consistent with the relation $b = D + 1$, since without steric interaction, the model is one-dimensional (Supplemental Material, Fig. S.3 [29]). Furthermore, longer filaments show an exponential decay with $k \sim g^{-1/2}$ (Supplemental Material, Fig. S.3 [29]). Comparing with Fig. 2(d), it appears that $k \sim g^{-1/b}$. Further studies are needed to test the validity of this relation.

To apply our model to real-world networks, we developed a Computer-Aided Feature Extraction (CAFE) program that identifies individual filaments in experimental images, and applied CAFE to an atomic force microscopy (AFM) image of a triaxial network of collagen fibrils (Figs. 3(a) and 3(b); Supplemental Material [29]). While recognizing all visible filaments without any error is difficult due to the inherent conformational variation of collagen fibrils and various noises, CAFE recognizes the majority of them. The resulting filament orientational distribution P_θ has three peaks separated by ~60°, reflecting the hexagonal phlogopite mica lattice [Fig. 3(c)]. Asymmetry in P_θ is likely caused by the distortion of the surface-exposed lattice [36]. A power law profile was observed in P_L with exponent $b = 2.61$ [Fig. 3(d), solid circles]. While the exponent agrees well with that of our stochastic model, the power law spans only a narrow range of L and the crossover to exponential distribution is unclear. To further elucidate the condition in which the network formed, we sought simulation parameters that recapitulate the network in Fig. 3(a). Filament directions were selected with probability proportional to the peaks of P_θ in Fig. 3(c). CAFE identified 2886 filaments of length greater than 1 μm (about 10 pixels; filaments shorter than this were ignored due to their low aspect ratio). Simulations were thus run until there were approximately 3000 filaments longer than 1 μm. To implement dimensional quantities, the system size was set to $w = 65$ (μm) that matches the AFM image [Fig. 3(a)]. We used $n = 5$ (this choice only affects the relevant time scale; see below) and varied g until the resulting P_L agrees with that from CAFE, which yielded $g \simeq 0.25$–0.75 [Fig. 3(d)]. An example simulation at $g = 0.50$ indeed resembles the AFM image [Fig. 3(e)].

In simulation, the total number of filaments including those shorter than 1 μm was about 4500. As Fig. 3(a) was taken after 5 h incubation, 1/3 min is an appropriate time unit for our model (time for nucleating 5 filaments). With a 50-nm subunit size (σ) used in our simulation, the elongation rate is estimated as $50g \simeq 38$–113 (nm/min). Without interruption, a filament would reach 22.7 μm (at 75.5 nm/min) after 5 h, which is above the maximum observed in AFM, 18.4 μm, supporting the consistency of our analysis (the maximum length with steric interruption should be shorter than 22.7 μm). In dimensionless terms, if we scale the system, subunit sizes, and time to match those

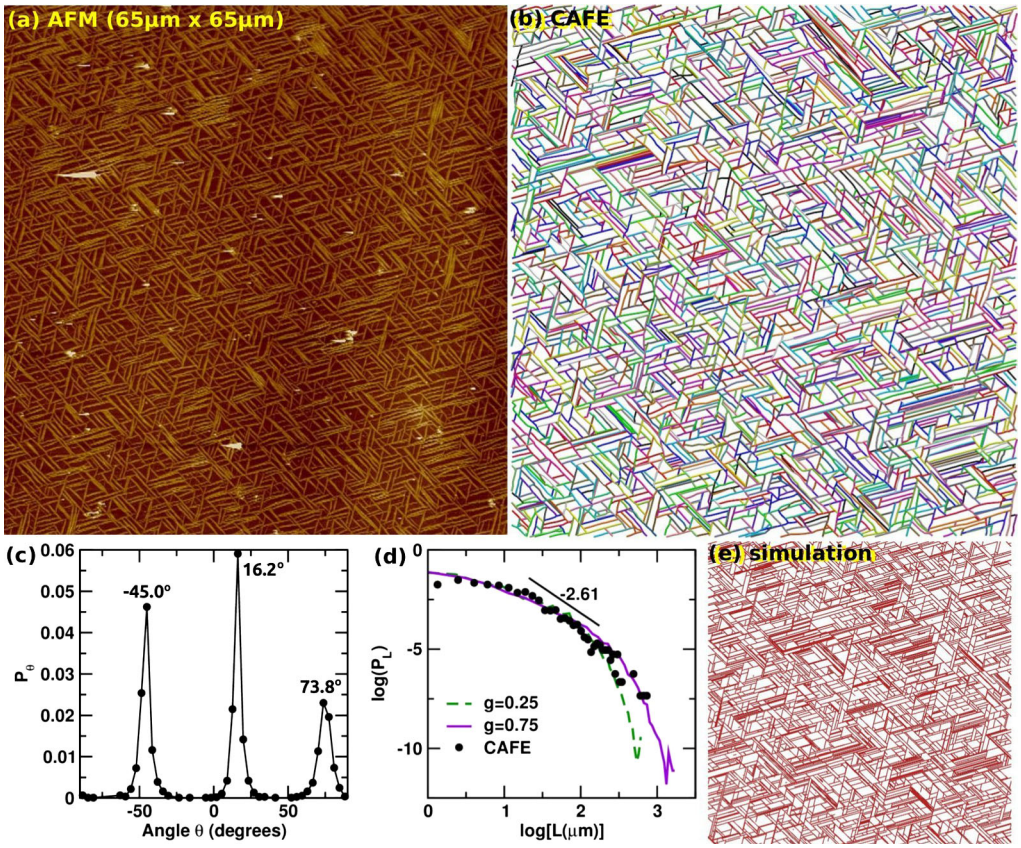


FIG. 3 (color online). Analysis of a triaxial collagen network. (a) AFM scan. (b) Corresponding *in silico* network extracted via CAFE. Filaments are individually recognized and colored randomly. (c) Filament angle distribution. Angle θ is measured relative to the horizontal direction in (a). The asymmetry in P_θ is visible in (a). (d) Length distribution. Solid circle: From the network in (b). Dashed and solid lines: stochastic simulation with $g = 0.25$ and 0.75, respectively (each averaged over 50 runs). (e) An example simulation with $g = 0.50$, which looks qualitatively similar to the AFM image in (a).

in Fig. 2, the corresponding parameters are $n = 1$, $g = 0.39 - 1.2$, and $t = 4500$. At this time, a system with $g = 1$ [Fig. 2(a)] will be undergoing transition between the early encounter to self-similar regimes so that the transition between exponential and power law in P_L is not clearly established, as observed in Fig. 3(d). The semibiaxial character of the collagen network [Fig. 3(c)] may also contribute to the slow convergence to the self-similar regime due to a lower chance of encounter between filaments (Supplemental Material, Fig. S.2 [29]).

In Fig. 3(b), the total filament length in the network is 9.9 mm. Since a collagen fibril in the AFM image is approximately 3-nm thick [5,8] and about 100-nm wide, with a 1.5-nm diameter and 300-nm length for a single hydrated collagen molecule, a 65×65 μm^2 network consists of $\sim 5.6 \times 10^6$ collagen molecules. If we assume the same level of adsorption for the entire ~ 1 cm^2 mica disc, about 1.3×10^{11} collagen molecules will be in fibrils. By comparison, a 50-μl aliquot of 5-μg/ml collagen solution used in experiment contains 5.5×10^{11} collagen molecules (molecular weight = 272 kDa). Therefore, depletion of collagen monomers during the experiment is unlikely to play a major role, which supports our assumption of constant filament nucleation and growth rates.

For comparison, we generated another network using half the collagen concentration (2.5 μg/ml) while keeping other conditions the same. A similar analysis combining CAFE and stochastic simulation reveals changes in the filament orientational symmetry and reduction in the elongation rate to $30 - 90$ nm/min (Supplemental Material, Fig. S.5 [29]), demonstrating the sensitivity of our analysis.

Question arises regarding whether real-world OLE networks can possess a clear self-similar regime. In principle, although this can be achieved with a prolonged incubation time and possibly replenishing molecules in the buffer to prevent depletion, the finite size of a molecule limits the smallest length scale that the network can reach. In the case of collagen, a single molecule is 300-nm long, which is already comparable to the 1-μm length of the shortest fibril identified in Fig. 3. An alternative approach would be to increase the growth rate g compared to the nucleation rate n [Fig. 2(c)]. For a given type of molecule, however, since both rates are determined by intermolecular interaction and

also interaction with the substrate, it would be difficult to control n and g independently. Thus, while the patterns as shown in Fig. 1 are universally observed, partitioning between exponential and self-similar regimes in the long-time limit may be a characteristic for a given system.

In summary, our model of OLE provides a general framework for understanding sterically limited filament assembly on surfaces. The finite assembly kinetics leaves a characteristic signature in network morphology as seen by a transition from exponential to universal power-law behavior. Together with CAFE (cf., Figs. 3 and S.5), our model can detect the subtle influence of the experimental condition on the network, which will be useful for analyzing a wide range of experimental systems exhibiting OLE.

We thank Sidney Redner for helpful discussion about fragmentation theory. We used the Texas A&M Materials Characterization Facility for AFM imaging.

*hwm@tamu.edu

[1] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Macmillan, London, 1983).
[2] P.-G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, NY, 1979).
[3] H. E. Stanley and N. Ostrowsky, *On Growth and Form: Fractal and Non-Fractal Patterns in Physics* (Nijhoff, Boston, MA, 1986).
[4] A.-L. Barabási and H. E. Stanley, *Fractal Concepts in Surface Growth* (Cambridge University Press, Cambridge, England, 1995).
[5] D. A. Cisneros, J. Friedrichs, A. Taubenberger, C. M. Franz, and D. J. Müller, Small **3**, 956 (2007).
[6] R. W. Loo and M. C. Goh, Langmuir **24**, 13 276 (2008).
[7] M. Sun, A. Stetco, and E. F. Merschrod S., Langmuir **24**, 5418 (2008).
[8] W. W. Leow and W. Hwang, Langmuir **27**, 10 907 (2011).
[9] C. L. Brown, I. A. Aksay, D. A. Saville, and M. H. Hecht, J. Am. Chem. Soc. **124**, 6846 (2002).
[10] F. Zhang, H.-N. Du, Z.-X. Zhang, L.-N. Ji, H.-T. Li, L. Tang, H.-B. Wang, C.-H. Fan, H.-J. Xu, Y. Zhang, J. Hu, H.-Y. Hu, and J.-H. He, Angew. Chem., Int. Ed. **45**, 3611 (2006).
[11] Á. Karsai, Ü. Murvai, K. Soós, B. Penke, and M. S. Z. Kellermayer, Eur. Biophys. J. **37**, 1133 (2008).
[12] V. L. Sedman, E. Kasotakis, X. Chen, S. Allen, C. J. Roberts, A. Mitraki, and S. J. B. Tendler, Protein Peptide Lett. **18**, 268 (2011).
[13] J. S. Jeong, A. Ansaloni, R. Mezzenga, H. A. Lashuel, and G. Dietler, J. Mol. Biol. **425**, 1765 (2013).
[14] S.-G. Kang, H. Li, T. Huynh, F. Zhang, Z. Xia, Y. Zhang, and R. Zhou, ACS Nano **6**, 9276 (2012).
[15] V. V. Prokhorov, D. V. Klinov, A. A. Chinarev, A. B. Tuzikov, I. V. Gorokhova, and N. V. Bovin, Langmuir **27**, 5879 (2011).
[16] T. Haber, R. Resel, A. Andreev, M. Oehzelt, D.-M. Smilgies, and H. Sitter, J. Cryst. Growth **312**, 333 (2010).
[17] D.-M. Smilgies, N. Boudet, B. Struth, Y. Yamada, and H. Yanagi, J. Cryst. Growth **220**, 88 (2000).
[18] F. Balzer, L. Kankate, H. Niehus, and H. G. Rubahn, Proc. SPIE Int. Soc. Opt. Eng. **5925**, 59 250A (2005).
[19] C. Simbrunner, D. Nabok, G. Hernandez-Sosa, M. Oehzelt, T. Djuric, R. Resel, L. Romaner, P. Puschnig, C. Ambrosch-Draxl, I. Salzmann, G. Schwabegger, I. Watzinger, and H. Sitter, J. Am. Chem. Soc. **133**, 3056 (2011).
[20] C. Simbrunner, Semicond. Sci. Technol. **28**, 053001 (2013).
[21] J. Chun, J.-L. Li, R. Car, I. A. Aksay, and D. A. Saville, J. Phys. Chem. B **110**, 16 624 (2006).
[22] C. Whitehouse, J. Fang, A. Aggeli, M. Bell, R. Brydson, C. W. G. Fishwick, J. R. Henderson, C. M. Knobler, R. W. Owens, N. H. Thomson, D. A. Smith, and N. Boden, Angew. Chem., Int. Ed. **44**, 1965 (2005).
[23] H. Rapaport, Supramolecular chemistry **18**, 445 (2006).
[24] H. Yang, S.-Y. Fung, M. Pritzker, and P. Chen, J. Am. Chem. Soc. **129**, 12 200 (2007).
[25] W. Hwang, B.-H. Kim, R. Dandu, J. Capello, H. Ghandehari, and J. Seog, Langmuir **25**, 12 682 (2009).
[26] Y.-D. Liu, F.-X. Wang, Y. Xiao, and G.-B. Pan, Mater. Lett. **68**, 178 (2012).
[27] C. R. So, Y. Hayamizu, H. Yazici, C. Gresswell, D. Khatayevich, C. Tamerler, and M. Sarikaya, ACS Nano **6**, 1648 (2012).
[28] H. Yanagi and T. Morikawa, Appl. Phys. Lett. **75**, 187 (1999).
[29] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.113.025502 for supplemental texts, figures, and videos.
[30] A. D. Novaco and J. P. McTague, Phys. Rev. Lett. **38**, 1286 (1977).
[31] R. Pynn, Nature (London) **281**, 433 (1979).
[32] I. Cherny and E. Gazit, Angew. Chem., Int. Ed. **47**, 4062 (2008).
[33] O. Bour, P. Davy, C. Darcel, and N. Odling, J. Geophys. Res. **107**, ETG4-1 (2002).
[34] P. Davy, R. Le Goc, C. Darcel, O. Bour, J.-R. De Dreuzy, and R. Munier, J. Geophys. Res. **115**, B10411 (2010).
[35] P. L. Krapivsky, S. Redner, and E. Ben-Naim, *A Kinetic View of Statistical Physics* (Cambridge University Press, Cambridge, England, 2010).
[36] Y. Kuwahara, Phys. Chem. Miner. **28**, 1 (2001).

Supplemental Material

Kinetic signature of fractal-like filament networks formed by orientational linear epitaxy

Wonmuk Hwang† and Esma Eryilmaz‡

†Departments of Biomedical Engineering and Materials Science & Engineering
Texas A&M University, College Station, TX 77845, U.S.A.
School of Computational Sciences, Korea Institute for Advanced Study, Seoul, Korea 130-722
‡Department of Physics, Texas A&M University, College Station, TX 77845, U.S.A.

S.1 Mean-field theory of independently growing filaments

We solve Eq. 1 with the initial condition $P_L = 0$ for all L at time $t = 0$. Let the total number of filaments be $S(t) = \sum_{L=1}^{\infty} P_L(t)$. Summing Eq. 1 over L leads to $S = nt$, which simply indicates that S increases by filament nucleation. The total number of subunits is $M(t) = \sum_{L=1}^{\infty} LP_L(t)$. Multiplying L to respective terms in Eq. 1, and summing yields $\dot{M} = n + gS$, so that $M = nt + \frac{gn}{2}t^2$.

Eq. 1 can be solved for individual P_L by using the generating function $G(s,t) = \sum_{L=1}^{\infty} s^L P_L(t)$ (s is a continuous variable). Once $G(s,t)$ is found, we can get $P_L(t) = \frac{1}{L!}\frac{\partial^L G}{\partial s^L}|_{s=0}$. Multiplying the L-term in Eq. 1 by s^L and summing, we get $\dot{G} = sn + g(s-1)G$, hence

$$G(s,t) = \frac{sn}{g(s-1)}(e^{g(s-1)t} - 1). \tag{S.1}$$

One can verify $G(1,t) = S(t)$, and $\frac{\partial}{\partial s}G(s,t)|_{s=1} = M(t)$. To get $P_L(t)$, we change variable for G to $u \equiv s-1$ in Eq. S.1, and use the notation $h(u) = 1 + 1/u$. We also rescale time by the growth rate, $t' \equiv tg$, and the nucleation rate, $n' \equiv n/g$:

$$G(u,t') = n'h(u)(e^{t'u} - 1) \tag{S.2}$$

With u, we have $P_L = \frac{1}{L!}\frac{\partial^L G}{\partial u^L}|_{u=-1}$. In Eq. S.2, note that

$$\begin{aligned} h(-1) &= 0 \\ \frac{\partial^L h}{\partial u^L}|_{u=-1} &= -L! \quad (L > 0) \end{aligned} \tag{S.3}$$

These give

$$\begin{aligned} \frac{\partial^L}{\partial u^L}\left(h(u)e^{t'u}\right) &= e^{t'u}\left(\frac{\partial}{\partial u} + t'\right)^L h(u) \\ &= e^{t'u}\sum_{m=0}^{L}\frac{L!}{m!(L-m)!}t'^m\frac{\partial^{L-m}}{\partial u^{L-m}}h \\ &\overset{u=-1}{=} -e^{-t'}\sum_{m=0}^{L-1}\frac{L!}{m!}t'^m. \end{aligned} \tag{S.4}$$

When we set $u = -1$ in the above summation, the term $m = L$ does not contribute ($h(-1) = 0$), which was dropped in the last line of Eq. S.4. Restoring t and g, we get

$$P_L(t) = \frac{n}{g}\left(1 - e^{-gt}\sum_{m=0}^{L-1}\frac{(gt)^m}{m!}\right) \tag{S.5}$$

Eq. S.5 can be rewritten as

$$P_L(t) = \frac{n}{g}\left(1 - e^{-gt}[e^{gt} - \sum_{m=L}^{\infty}\frac{(gt)^m}{m!}]\right)$$
$$= \frac{n}{g}e^{-gt}\sum_{m=L}^{\infty}\frac{(gt)^m}{m!}. \tag{S.6}$$

For $L \gg 1$, we can take continuum limit by applying Stirling's formula, $\log m! \simeq m\log m - m$, to each term on the right hand side of Eq. S.6:

$$\ln\frac{(gt)^m}{m!} \simeq m\ln gt - m\ln m + m$$
$$\simeq gt - \frac{(m-gt)^2}{2gt} \tag{S.7}$$

which yields Eq. 2:

$$P_L(t) \simeq \frac{n}{g}\sqrt{\frac{1}{2\pi gt}}\int_L^{\infty} dm\, e^{-(m-gt)^2/2gt}$$
$$= \frac{n}{2g}\mathrm{erfc}\left(\frac{L-gt}{\sqrt{2gt}}\right), \tag{S.8}$$

The prefactor in Eq. S.8 ensures the steady state value $P_L(t\to\infty) = \frac{n}{g}$, which can be seen by setting $\dot{P}_L = 0$ in Eq. 1. To test this result, we disabled interaction between filaments or with the boundary in our stochastic simulation and carried out 1000 runs for each combination of n and g. The resulting length distribution at different times match well with Eq. S.8 (Fig. S.1).

S.2 Mean-field theory with interaction between filaments

This case is described by Eq. 3. Adding the equations from $L = 1$ to ∞ gives the rate equation for the total filament number density:

$$\dot{S} = n - aS^2 \tag{S.9}$$

which has the solution

$$S(t) = \sqrt{\frac{n}{a}}\tanh([na]^{1/2}t). \tag{S.10}$$

In steady state, $S = \sqrt{n/a}$. Inserting this into Eq. 3 gives

$$P_L = \frac{ng^{L-1}}{(g+\sqrt{na})^L} = \frac{n}{g}e^{-L\ln(1+\sqrt{na}/g)}, \tag{S.11}$$

which is Eq. 4.

Instead of S in Eq. 3, if the rate of encounter among filaments is proportional to the total number of subunits $M = \sum_L LP_L$, following a procedure similar to that in Sec. S.1,

$$\dot{M} = n + gS - aM^2$$
$$\dot{S} = n - aMS \tag{S.12}$$

The steady-state solution $M_\infty \equiv M(t\to\infty)$ satisfies

$$\frac{gn}{aM_\infty} = aM_\infty^2 - n. \tag{S.13}$$

For any positive values of n g, and a, there is a unique solution for $M_\infty > 0$. Replacing S in Eq. 3 with M_∞ and setting $\dot{P}_L = 0$ yields $P_L = \frac{n}{g}e^{-L\ln(1+aM_\infty/g)}$, which is again exponential.

S.3 Computer-Aided Feature Extraction (CAFE)

We made the CAFE program to be script-based so that it is applicable to the analysis of other types of images having different features without compiling the source code. CAFE was written in the C++ language. The output of CAFE is written in the CHARMM coordinate file format [1, 2] so that existing molecular rendering softwares can be used for visualizing the network. Figs. 3b and S.5b were rendered using VMD [3]. Below we outline the main concepts by using a small image as an example (Fig. S.4). A detailed explanation of individual algorithms will be published elsewhere.

S.3.1 Bead assignment

This is the initial filtering of the pixel data (Fig. S.4a to Fig. S.4b). A simple binary thresholding approach does not yield desirable outcome in cases when pixel intensities of filaments vary across the image or when filaments are densely populated. We thus developed the following local adaptive thresholding method:

1. Use a small window of a given size to scan across the image, translated by one pixel at a time. In each step during the scan, calculate the average pixel intensity $\bar{I}$, standard deviation $\sigma(I)$, and the 'center of mass' $\mathbf{r}_{cm}$ of the window weighted by pixel intensity.

2. Add a bead to $\mathbf{r}_{cm}$ if the following three conditions are met:

   - $\mathbf{r}_{cm}$ lies within 1 pixel from the geometrical center $\mathbf{r}_0$ of the window.
   - The pixel intensity at $\mathbf{r}_0$ is greater than $\bar{I} + h\sigma(I)$, where h is a preset threshold.
   - $\bar{I}$ is greater than a low intensity cutoff I_0.

   In this study, we used a 6-pixel wide square window, $h = 0.3$, and $I_0 = 20$. By comparison, the original AFM images were 677-pixel wide and the maximum pixel intensity was 128.

3. If a bead is assigned to a window, pixels within that window are not considered again for assigning other beads in subsequent scans.

S.3.2 Bond assignment

Bonds are assigned between beads that are closer than a cutoff distance R_c. For a given image, R_c is determined by the width of a typical filament. We used $R_c = 4.5$ pixels in this study. Non-integer values are allowed because the bead's position (intensity-weighted center of mass of a window during the bead assignment step) is given as real numbers (Fig. S.4c).

S.3.3 Constructing the filament network

Whereas bonds and beads as assigned above are simply coarse-grained representation of the pixel data, identifying filaments requires extracting topological information from the image, which is the most challenging part of the CAFE procedure. Our main strategy is progressive refinement of the network, to work on regions where filaments are readily recognized, then use the orientation and persistence of these filaments to analyze the remaining regions.

1. *Bond clean-up.* In CAFE, a filament is assigned to a linear chain of beads connected by single bonds. In the initially constructed bond-bead network (Fig. S.4c), a bead can form bonds with multiple surrounding beads. If some of these bonds have angles less than a cutoff ($\frac{\pi}{4}$ in this study), only the shortest bond is kept. Also, if three beads connected by bonds are straight (two bonds connecting the beads form an angle greater than $\frac{3}{4}\pi$), all other bonds to the middle bead are removed. Procedures such as these convert regions with relatively isolated filaments in the original image into linear chains of beads, which are identified as filaments, as in Fig. S.4d.

2. *Filament extension.* Due to noises and conformational variation, a filament in the original image may be recognized as multiple filament segments in the above step. We thus connect pairs of filaments whose ends are closer than a certain cutoff distance (on the order of R_c) and also are similar in slope. Likewise, an isolated bead is connected to the end of a filament if it is located within a cutoff distance from the end and lies close to the local extension line of the filament's end. After joining filaments and adding beads, filaments are smoothed to get rid of local kinks or wiggles. The filament extension procedure is applied multiple times using different values of cutoffs (Fig. S.4e).

3. *Cluster processing.* The above steps do not work well on regions where, *e.g.*, filaments are very closely distributed or several filaments converge (thin blue bonds in Fig. S.4d–f). We call such regions as clusters. In most cases, within a cluster, beads are spread over an area such as the one on the upper right side of Fig. S.4e. In this case, local connectivity among beads provides little information about the organization of filaments within the cluster. Thus, among the filaments identified above, those that surround the cluster are utilized. Similar to the case for joining two closely separated filaments, if two filaments surrounding a cluster closely match when extended, they are joined through the cluster, which can be most clearly seen by comparing the upper right sides of Fig. S.4e and Fig. S.4f. The remaining clusters after this procedure are smaller, and similar filament assignment procedures are applied to them. The final network is shown in Fig. S.4g.

An overlay of the original image and the final network from CAFE (Fig. S.4h) shows that, although most filaments are identified, there are occasional mismatches. As long as the mismatches are not a large fraction of the network, they do not affect our measurement in any major way.

S.4 Experimental method

Collagen networks were formed based on our previously developed protocol [4]. Briefly, type-I collagen from rat tail tendon (BD Biosciences) was diluted to 5 μg/ml in a buffer containing 200 mM KCl, 30 mM Na_2HPO_4, and 10 mM KH_2PO_4. Immediately after preparing the solution, a 50-μl aliquot was deposited on freshly cleaved phlogopite mica (a generous gift from Frank Balzer at University of Southern Denmark). During incubation, mica was placed in a moisture-controlled chamber to prevent drying. After 5 h, the sample was gently rinsed 4 times each with 70-μl deionized water, then dried. Drying does not affect the network formed on mica [4]. For imaging, we used the Bruker Dimension Icon AFM in soft tapping mode.

S.5 Comparative analysis of triaxial collagen networks formed under different experimental conditions

We tested whether CAFE can distinguish between networks formed under slightly different experimental conditions. Fig. 3a was obtained with 5 μg/ml collagen solution, whereas Fig. S.5a was with 2.5 μg/ml. In the latter case, collagen fibrils appear slightly more aligned, but otherwise it is difficult to find any major difference between the two images by visual inspection only. They are both 677 pixels on a side. Different scan sizes (65 μm in Fig. 3a and 51 μm in Fig. S.5a) were determined during AFM imaging while searching for clean areas on mica that are relatively free of debris or undissolved clusters.

We applied CAFE to construct the corresponding network (Fig. S.5b). The resulting orientational distribution has three peaks at the same angles as before (Fig. 3c *vs.* Fig. S.5c), but the location of the minor peak moved from 73.8° to −45.0°. Since the phlogopite mica is cleaved in each experiment, a newly exposed lattice surface may have different direction of distortion [5]. Similar to Fig. 3, the two major orientations (16.2° and 73.8°) visually match with those in Fig. S.5a. In the measured filament length distribution P_L, its slope in log-log plot around the middle of the distribution is −2.68, again consistent with our model (Fig. S.5d).

There were 2576 filaments of length greater than 1 μm. Although this is smaller than the corresponding number (2886) in Fig. 3, considering difference in scan sizes, the filament number density per unit area is

actually higher, 0.99 filaments/μm^2, while Fig. 3a has 0.68 filaments/μm^2. At first, a higher filament density under a lower collagen concentration appears puzzling. However, the lower concentration reduces the filament growth rate, thus it may allow more filaments to nucleate in a given area. To test this possibility, we carried out stochastic simulations. To match with Fig. S.5a, we used a system size 51×51 μm^2, and correspondingly rescaled the nucleation rate to $n = 5 \times (51/65)^2 = 3.08$. This makes the two simulations in Fig. 3 and Fig. S.5 to have the same nucleation rate per unit area. The orientation of a nucleating filament was selected with probability proportional to the peak values in P_θ (Fig. S.5c). Simulations were run until there were more than about 2500 filaments longer than 1 μm. Good agreement with experiment was achieved for $g = 0.15 \sim 0.45$, which indeed reveals a lower growth rate than the case with a higher collagen concentration (Fig. 3d *vs.* Fig. S.5d). An example of the simulated network with $g = 0.25$ again resembles the AFM image (Fig. S.5e).

Using these results, as for Fig. 3, we can estimate the time scale and elongation rate in physical units. In our simulation, there were a total of 3700 filaments (including those with length less than 1 μm). Since the incubation time is 5 h (300 min), the time unit for the simulation is $\frac{300}{3700} \times 3.08 = 0.25$ (min). For $\sigma = 50$ nm, the elongation rate is $50g/0.25 \simeq 30 \sim 90$ (nm/min), which is indeed lower than the case with 5 μg/ml collagen solution (38$\sim$113 nm/min). In order to increase the accuracy of the estimate, taking more AFM images at different time points would be helpful. On the other hand, *in situ* measurement of the elongation process by AFM is difficult due to the slow scan rate. Also note that Figs. 3d and S.5d are not simple fits, but matching of data by actual simulation. Incrementally varying g and carrying out simulation in each case to minimize the root-mean-square deviation between experimental data and simulation would provide a better estimate for g. However, the analyses presented here should be sufficient for the purpose of demonstration.

References

[1] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem*, **4**, 187–217.

[2] Brooks, B. R., Brooks III, C. L., Mackerell Jr., A. D., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M., and Karplus, M. (2009) CHARMM: the biomolecular simulation program. *J. Comput. Chem.*, **30**, 1545–1614.

[3] Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.

[4] Leow, W. W. and Hwang, W. (2011) Epitaxially guided assembly of collagen layers on mica surfaces. *Langmuir*, **27**, 10907–10913.

[5] Kuwahara, Y. (2001) Comparison of the surface structure of the tetrahedral sheets of muscovite and phlogopite by AFM. *Phys. Chem. Miner.*, **28**, 1–8.

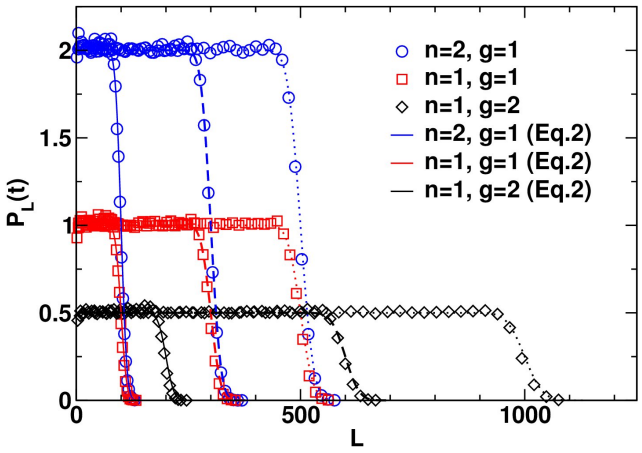


Figure S.1: Comparison between stochastic simulations of independently growing filaments and mean-field theory. Symbols: Histogram of filament lengths averaged over 1000 runs for each case. Lines: Corresponding plots of Eq. 2, at $t = 100$ (solid lines), 300 (dashed), and 500 (dotted). The plateau of the distribution follows the steady-state value $\frac{n}{g}$, and its growing front is located at gt.

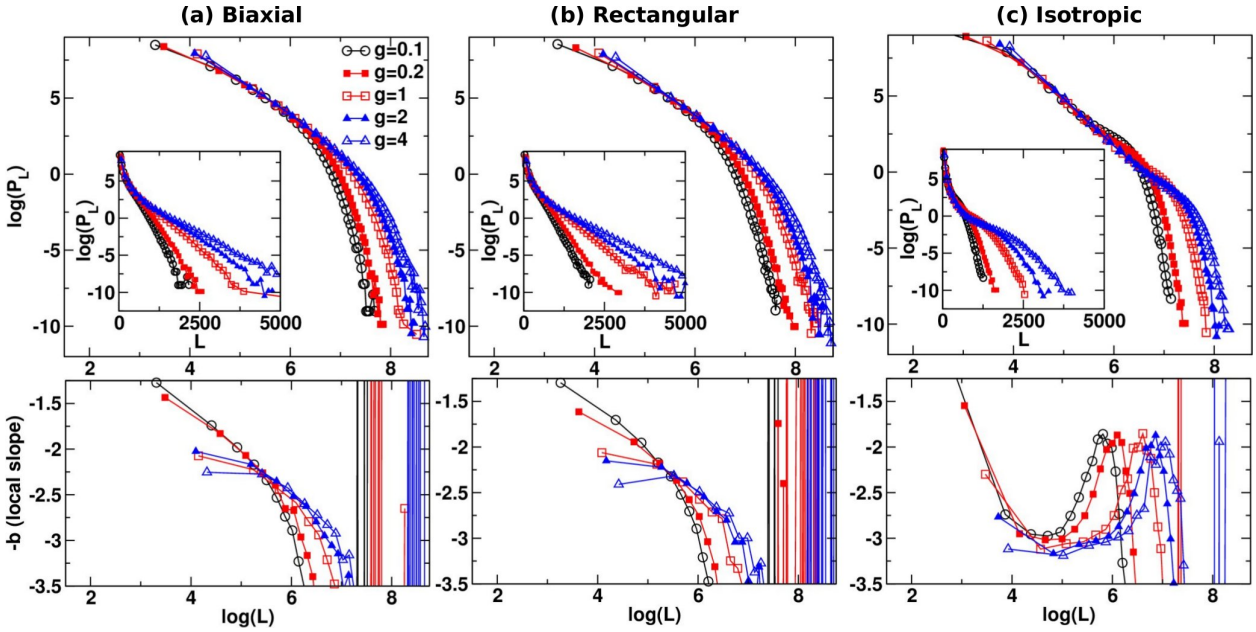


Figure S.2: Length distribution in stochastic simulations of (a) biaxial, (b) rectangular, and (c) isotropic networks. All model parameters are the same as in Fig. 2 except in filament orientations. Measurements were made at $t = 4\times10^5$, averaged over 150 to 600 independent runs for each g. Lower panels are local slopes of the log-log plots. Compared to Fig. 2, the range of L where power law holds differs. For biaxial and rectangular networks, the power law exponent b changes almost continuously, reflecting the fact that confinement of filaments is not as strong as in the triaxial network, and transition from exponential to power-law in the length distribution is not as clear. For greater g (faster growth rate), b varies more slowly with L, and the range of L in which b lies between 2.5 and 3 expands. Considering the similarity between the dependence on t and g (Fig. 2a,b), it is expected that the power law behavior will become clearer at later times. However, for longer simulation, a larger system or a smaller subunit size need to be used to avoid finite-size effect. In (c), b is close to 3 for $\log(L)$ between 4 and 6. As discussed in the paper, the fractal dimension is $D = b - 1 \simeq 2$, suggesting that the network appears to be a compact 2-dimensional object. However, a narrow region of $\log(L)$ right before the exponential distribution has $b \simeq 2$, indicating that the network behaves like a 1-dimensional system. The range of L over which this occurs shrinks with g. The origin for this behavior of the isotropic network is unclear.

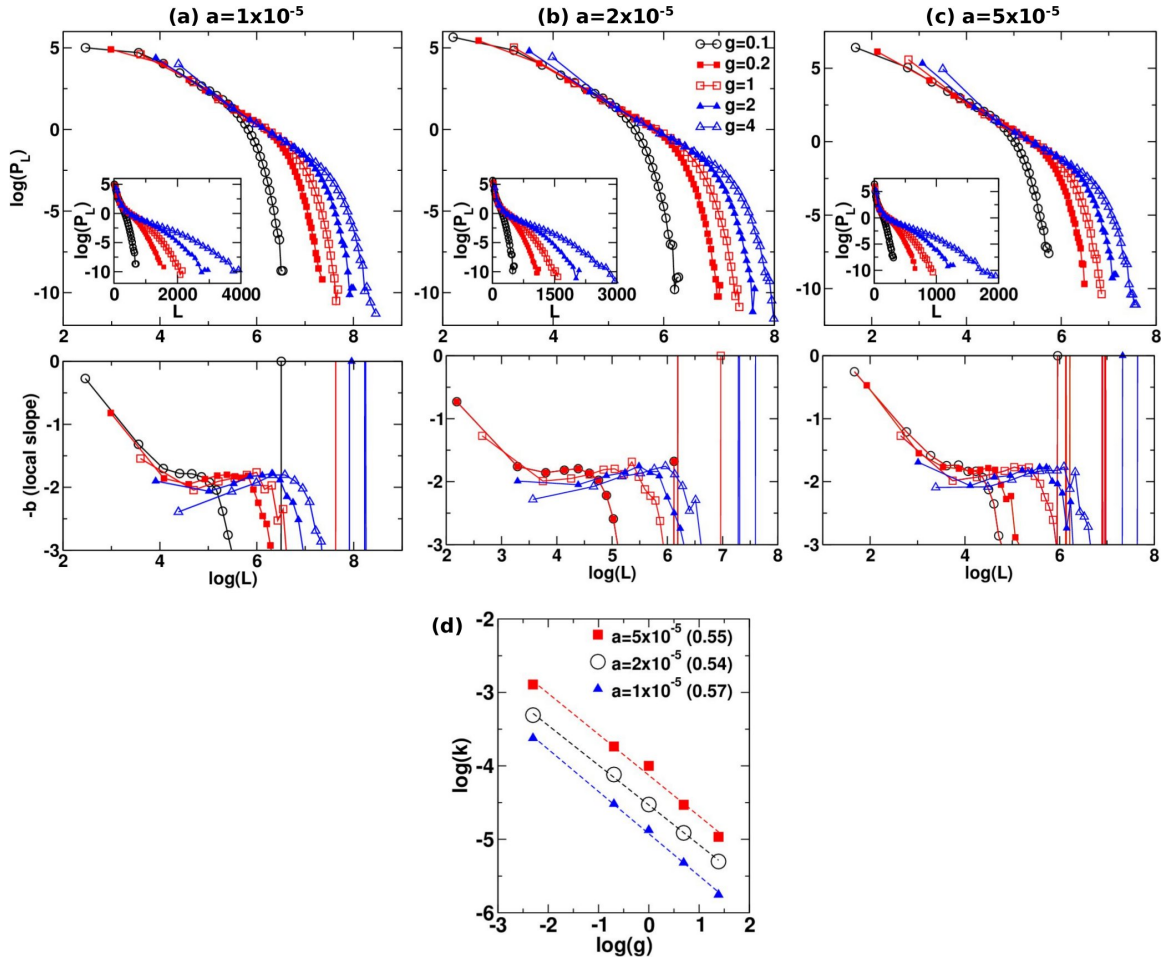


Figure S.3: Results for a modified stochastic model without filament interaction but a growing end of a filament capped instead with probability $1 - e^{-aS}$ (S: total number of filaments). Measurements were made at $t = 1 \times 10^4$, and results are averaged between 500–1000 runs. Other simulation parameters are the same as in Fig. 2. (a)–(c). P_L in log-log scale (top row) and its local slope (bottom row) for three different values of a: (a) $a = 1 \times 10^{-5}$, (b) $a = 2 \times 10^{-5}$, and (c) $a = 5 \times 10^{-5}$. The local slope stays around -2 for the intermediate range of L, indicating a 1-dimensional nature of the process in the absence of steric interaction. Inset: P_L in log-normal scale. (d) Inverse of the decay length k in the exponential distribution as a function of g. Numbers in legends are slopes of linear fits (*cf.*, Fig 2d). The prefactor of the relation (k at $g = 1$) increases with a, as the characteristic length scale of the exponential decay decreases with higher capping probability. This behavior is similar to the case in Fig. 2d where different orientational symmetry of the network affects the likelihood of encounter between filaments.

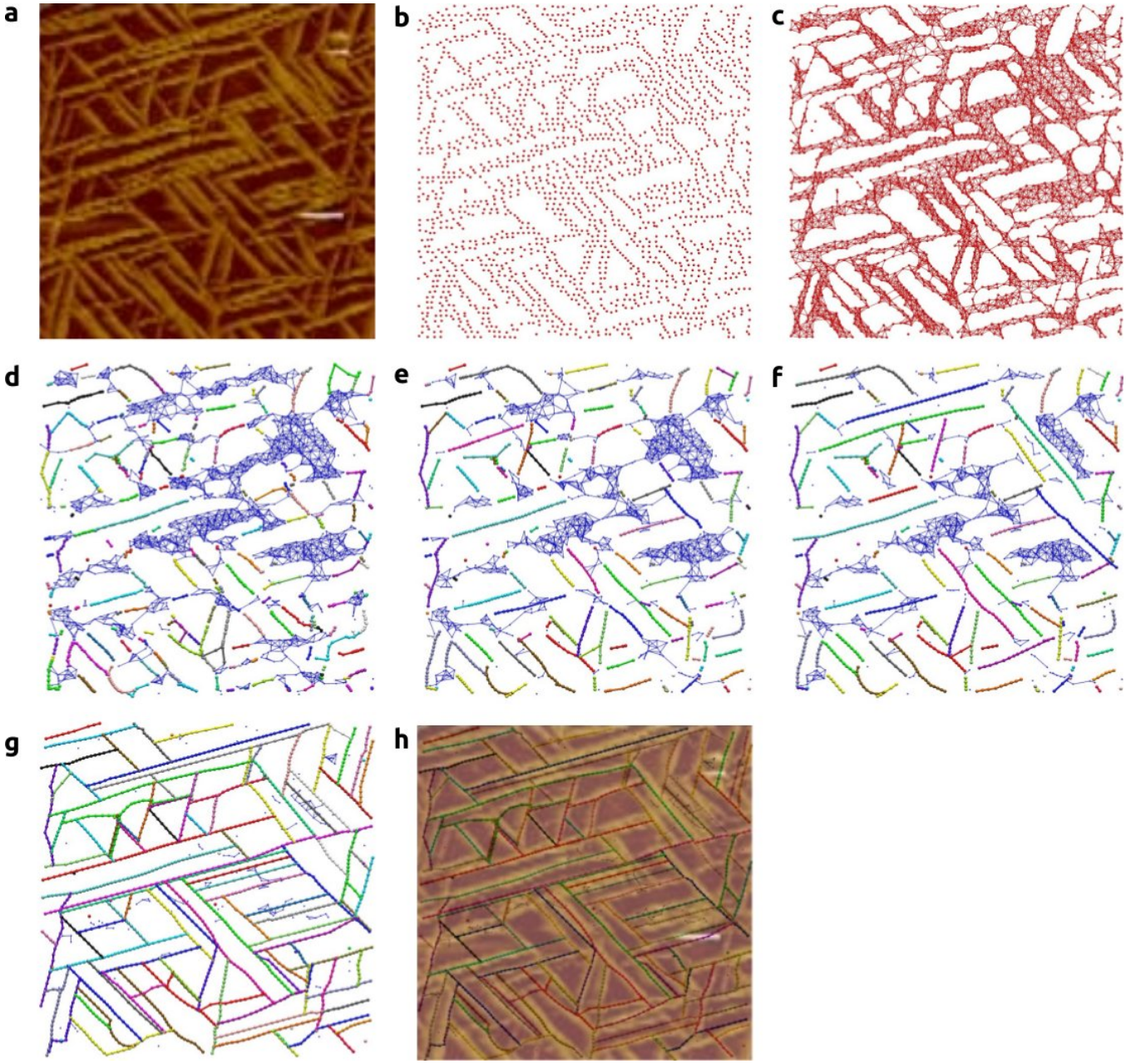


Figure S.4: Snapshots of intermediate steps of CAFE applied to a 128-pixel wide portion of Fig. 3. (a) Original image. (b) Bead assignment. (c) Bond assignment. (d) Initial identification of filaments. Colors are randomly assigned to identified filaments. Regions where filaments are not identified (clusters), are in thin blue bonds. (e) Filament extension. Clusters reduce in size or disappear as a result. (f) An intermediate step during cluster processing. (g) Final network. Residual beads and bonds that are not part of any filament are excluded from our analysis (thin blue dots and bonds). (h) Overlay of the original image (a) and the final network (g).

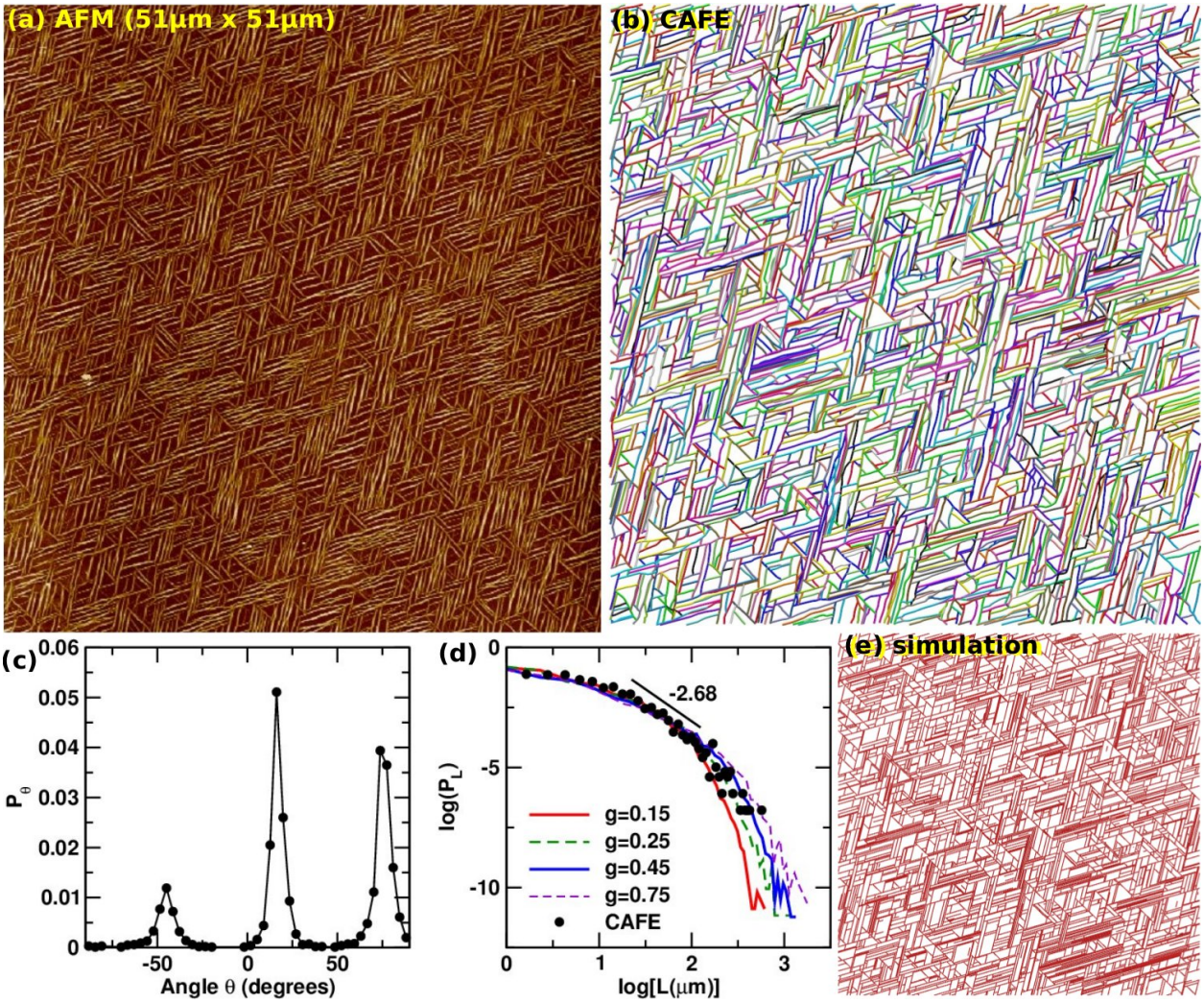


Figure S.5: Analysis of a collagen network formed at 2.5-μg/ml concentration. Other experimental conditions are the same as in Fig. 3a. (a) AFM image. (b) Network identified by CAFE. (c) Filament angle distribution. (d) Length distribution. Compared to Fig. 3d, simulations with $g = 0.15 \sim 0.45$ better match the data, reflecting a slower growth rate with lower collagen concentration. Distributions for $g = 0.25$ and 0.75 are shown for comparison. (e) A snapshot of the simulation with $g = 0.25$.