Separation of time scales in one-dimensional directed nucleation-growth processes

Paolo Pierobon,[1,*] Judith Miné-Hattab,[2] Giovanni Cappello,[2] Jean-Louis Viovy,[2] and Marco Cosentino Lagomarsino[3,†]

[1]*Institut Curie, Centre de recherche, INSERM U932 Immunité et cancer, 12 rue Lhomond, 75005 Paris, France*
[2]*Institut Curie, Centre de recherche, CNRS UMR168 Laboratoire Physico-Chimie Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France*
[3]*Università degli studi di Milano, Dip. Fisica, via Celoria 16, 20133 Milano, Italy*

Proteins involved in homologous recombination such as RecA and hRad51 polymerize on single- and double-stranded DNA according to a nucleation-growth kinetics, which can be monitored by single-molecule *in vitro* assays. The basic models currently used to extract biochemical rates rely on ensemble averages and are typically based on an underlying process of bidirectional polymerization, in contrast with the often observed anisotropic polymerization of similar proteins. For these reasons, if one considers single-molecule experiments, the available models are useful to understand observations only in some regimes. In particular, recent experiments have highlighted a steplike polymerization kinetics. The classical model of one-dimensional nucleation growth, the Kolmogorov-Avrami-Mehl-Johnson (KAMJ) model, predicts the correct polymerization kinetics only in some regimes and fails to predict the steplike behavior. This work illustrates by simulations and analytical arguments the limitation of applicability of the KAMJ description and proposes a minimal model for the statistics of the steps based on the so-called stick-breaking stochastic process. We argue that this insight might be useful to extract information on the time and length scales involved in the polymerization kinetics.

PACS number(s): 87.15.A−, 87.10.Mn, 02.50.Ey

I. INTRODUCTION

Many proteins in the cell have the ability to bind and polymerize on one-dimensional substrates. Among several examples in biology we find recombination on DNA: the proteins of the RecA superfamily, such as hRad51 in humans, polymerize along single- and double-strand DNA, forming a right handed helix, the so-called "nucleoprotein filaments." This property is essential for the biology of homologous recombination and DNA damage repair [1–3]: mutations in the DNA repair genes have implications in the development of mutation-related diseases (such as the Werner or Bloom syndromes) and cancer [4].

To understand recombination, it is important to have a quantitative characterization of these polymerization processes, of their mechanical implications on DNA, and, eventually, of the microscopic details of their action. Among the basic questions that can be asked regarding this process, an important one concerns the dynamics of formation of the nucleoprotein filaments. Several studies have shown that it is composed of two separate *nucleation* and *growth* phases and have focused on determining the characteristic time scales of these two subprocesses, i.e., the inverse of the kinetic constants for a binding of a monomer to the DNA filament or to an existing polymer. These quantities have the dimension of time and are related to the energy barrier necessary for the binding of the first monomer (nucleation) or of the following ones (growth). Most of the work performed to date concerns the *E. coli* RecA protein, where nucleation rates and polymerization parameters are now well characterized [5–9,34].

A crucial question related to the hRad51 polymerization concerns the elongational and torsional forces involved in

the process [7,8]. For example, it is known that RecA is able to stretch DNA 1.5 times its resting length, which can be obtained *in vitro* only under severe strain [6,10]. Information on the kinetics of this process can be extracted from single-molecule techniques such as magnetic tweezers (for a review, see [11]). These methods allow one to control both the external forces and the topology applied on a single DNA while measuring its extension in the presence of various monomer concentrations in solution. Recently, these techniques have made available new data on the mechanics and the kinetics of polymerization of hRad51 and RecA in real time under varying protein concentration and external load [12–17]. Such experiments monitor the change in length of a dsDNA molecule following the polymerization of RecA or hRad51, while the molecule is being stretched, supercoiled, or kept under torsion. In recent experiments, with varying force and concentration of hRad51 in solution, several different regimes can be observed, including a puzzling steplike growth in time [15].

From the modeling viewpoint, a rich variety of approaches can be taken, as it is clear that the formation of nucleoprotein filaments depends on many details, including the ionic strength of the buffer, the role of ATP (Adenosine-triphosphate), the mechanical and thermodynamic forces controlling the energy barrier for the nucleation and growth, and the torsional stress on DNA. On the other hand, there exists a "standard" model (originally formulated by Kolmogorov, Avrami, Mehl, and Johnson [18–22], henceforth referred to as KAMJ) commonly used to capture the main generic features of nucleoprotein filament polymerization [5]. The KAMJ includes solely the nucleation and growth processes by effective parameters and explores the consequences of combining them.

However, the KAMJ model considers only nondirected polymerization, which is unrealistic for nucleoprotein filaments in most conditions. While this discrepancy might not emerge in bulk measurements, it might play a role in single-

*Corresponding author; paolo.pierobon@curie.fr
†marco.cosentino-lagomarsino@unimi.it

molecule experiments. With the same essential ingredients of the KAMJ model, we investigate the situations where the *directionality* of growth can have a consequence on the typical polymerization behavior. We show how, at different values of the ratio of the nucleation- and growth-characteristic times, the typical time scales of the system separate, leading to the steplike growth of the nucleoprotein filament observed experimentally. This regime cannot be captured by the KAMJ framework, and we will argue that it can possibly be used to extract the effective polymerization and nucleation rates separately and to test for the existence of characteristic polymerization lengths.

The paper is structured as follows. We first discuss the existing experiments on the polymerization of hRad51 and the theory for the nucleation and growth in one dimension customarily used to describe these experiments. In Sec. III, we quantify the breakdown of the standard model using the kinetic Monte Carlo simulations. Subsequently, we introduce a simple model valid in the time-scale separation limit and with unbounded polymerization, which links to a classical stochastic process in statistics, known as "stick breaking." We compute the main observables from this model and compare them to the available data. In Sec. IV, we discuss the implications of our models. We summarize our results in Sec. V.

II. PHENOMENOLOGY: STEPLIKE GROWTH AND VIOLATION OF MEAN-FIELD BEHAVIOR

A. Experimental motivations

The analysis we present in this paper was motivated by the understanding of the phenomenology observed in single-molecule experiments on polymerization of a hRad51 on DNA. However, the ideas can, in principle, be applied to any system where a nucleation-directed polymerization process occurs on one-dimensional templates. These experiments are typically performed with a magnetic tweezer, a device that, by pulling on a paramagnetic bead, allows us to apply a force and to measure the elongation of a single DNA filament with nanometric precision. The setup in Ref. [15] does not allow the measurement of the torque applied to the bead. In practice, the measurements are performed in two extreme conditions: when the filament is free to rotate (the DNA is nicked or single strand) or at infinite torque (when the filament is double strand and anchored to the cover glass). The experiments have measured the real time extension of a nicked dsDNA molecule, applying a constant force of (2 and 6 pN, respectively) and increasing the concentration of hRad51 proteins in solution. Note that although the observables are the ones studied in a bulk enzymology measure, this kind of approach allows higher precision and an insight in the structure and function of the DNA associated protein.

As pointed out in [15] one observes three polymerization regimes [Fig. 2(A)],

(i) linear growth (at [hRad51]=100 nM);

(ii) steplike growth (at [hRad51]=150 nM); and

(iii) exponentially saturating growth (at [hRad51] $\geq$ 200 nM).

The first regime occurs at very low concentrations: after the first nucleus has formed, the growth occurs at constant velocity without new nucleation events. The exponential growth occurs in the opposite case, when the rate-limiting process is the growth, and can be interpreted as the result of an effective random deposition. The intermediate regime is the most interesting. The presence of steps in the polymerization suggests the alternation of slow nucleation events on the filament (the plateau) and fast *unidirectional* polymerization (the steps). In this last process, the nucleoprotein filament grows until it reaches a preexisting polymer or an extremity of the fragment.

In a minimal model for polymerization in one dimension, in the same spirit as the KAMJ model, we neglect a number of details related to the biochemistry of the proteins:

(i) We consider the formation of the filament in the conditions (regarding force, ATP, and salt concentration) where steps were reported. It is known that presence of ATP and Ca^{2+} is necessary for the formation of filaments; however, we chose to incorporate this aspect in the kinetics constants and not in the model.

(ii) The depolymerization is considered negligible. The available data indicate that it is at least two orders of magnitude slower than polymerization (at least on dsDNA); moreover, depolymerization occurs in bursts only when the ATP is hydrolyzed and the first monomer detaches [16].

(iii) The binding location of the hRAD51 monomer to DNA is considered to be random, with uniform distribution. Although it has been argued that the DNA sequence plays a role in the binding constants [23], this choice is motivated by keeping the description as simple as possible.

(iv) Cooperative binding (although observed from many experiments) is not considered and neither is an intrinsic finite length of the monomer (addressed, for example, by the simulations in Ref. [24]). In our effective description, the first feature simply rescales the speed of polymerization, without changing the quantitative behavior of the observables. The second feature (the coverage of a filament by segments of a given length) is a classic problem in statistical mechanics, known as *parking lot problem* [25]; we do not consider this feature (although interesting) since this process is known to show an anomalous relaxation dynamics [26] only at the end of the polymerization curve, where other effects (forces, torque, and sequence of the DNA) are likely to perturb the data considered here.

These hypotheses allow us to describe the formation of nucleoprotein filament as a simple nucleation-growth process.

B. KAMJ model

A general theory for the average kinetics of the simple nucleation-growth process in one dimension has existed since the 1930s, starting with the works of KAMJ, and is customarily applied to nucleation-growth type of polymerization on one-dimensional templates [5,18–22]. It assumes that monomers nucleate on the filament with rate k_n and that, once a nucleus is formed, a filament can grow from it with rate k_g. The number of nuclei m can then be related to the

coverage fraction ϕ by the following equation:

$$\partial_t m = k_n(1 - \phi) - k_g \frac{m^2}{1 - \phi}. \qquad (1)$$

This equation describes a birth-death process where the new nuclei are "born" with a probability proportional to the free fraction of the system and "die" when two growing fronts collide. The last term estimates the rate of front collisions and is given by the product of the density of growth fronts $[m/(1 - \phi)]$ and the number of growing fronts $(k_g m)$.

At the same time, the coverage fraction ϕ is proportional to the number of growing nuclei (assuming that they all grow at the same constant velocity),

$$\partial_t \phi = k_g m. \qquad (2)$$

From these two equations one finds

$$\phi(t) = 1 - e^{-2k_n k_g t^2}, \qquad (3)$$

$$m(t) = k_g t e^{-2k_n k_g t^2}. \qquad (4)$$

Note that this theory is called "mean field" in the sense that only the average behavior is considered, and correlations between growing fronts are neglected. One can prove that the behavior just presented is equivalent to unidirectional growth on fragments upon the substitution $k_g \to k_g/2$. It has been observed [27] that the hypotheses of this model approximate real systems well when the number of growing fragments is large, and the nucleation rate is sufficiently slow that the changes in the covering fraction can be neglected in Eq. (2) (i.e., that $k_g \ll k_n$). At the same time the system has to be sufficiently large so that the coalescence of growth fronts is rare and can be written in the form of the death term in Eq. (1) ($k_n/k_g \ll N$). Hence, as pointed out in Ref. [27], this theory fits the behavior of simulated data well when the time scale and system size allow for the competition between nucleation and growth to emerge

$$1/N \ll k_g/k_n \ll 1. \qquad (5)$$

When these assumptions break down, the underlying stochastic process (which is based on unidirectional rather than bidirectional growth) may become relevant.

A single DNA segment of ∼10 kbp as the one used in a typical magnetic tweezer experiment does not, in general, guarantee the above conditions. Also, the KAMJ theory does not allow to separately extract the two rates from the data but only their product. Separate time scales can be extracted from higher moment analysis (such as dynamic correlations) as shown in [28], but this requires a very large number of realizations.

In Ref. [27] Turner proposed another description based on a single-nucleus approximation. We describe this approach in Appendix A, where we explain how it is complementary to our approach, as it cannot deal with complete time-scale separation. When the two time scales are completely separated, the polymerization curve changes qualitatively, which requires another description, as pointed out in the following sections.
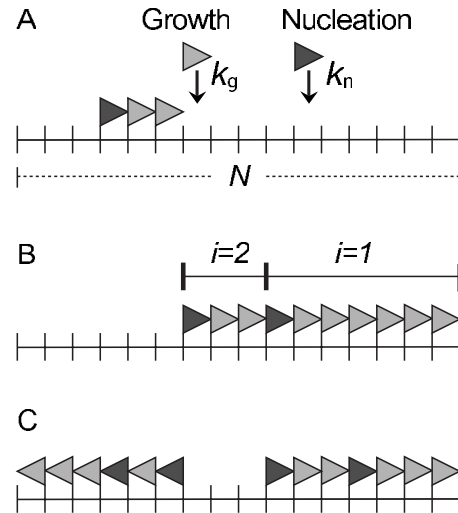

FIG. 1. (A) Scheme of the nucleation-polymerization process. Dark triangles represent the nuclei added with rate k_n and light triangles the polymerizing monomers, added with rate k_g to a nucleus or a polymerizing front. (B) Asymmetric stick-breaking regime. All nuclei can form with only one orientation. In the stick-breaking analogy, each polymer, labeled by i, corresponds to a "stick." (C) Symmetric stick-breaking regime. In this case the nuclei can form with both orientations, and thus the filaments can have both polymerization directions.

III. SIMULATIONS AND RESULTS

A. Time scale separation

To characterize the breakdown of the KAMJ behavior, we simulated the unidirectional nucleation growth on one-dimensional lattices up to a size of $N=10^4$ using a kinetic Monte Carlo–Gillespie algorithm [29]. This scheme first generates a random number to choose from the two possible processes: attach a nucleus at a random site or attach a monomer to an existent growth front. A second random number R is then generated to compute the waiting time for this event, using the Poisson process defined by the relation $\Delta t = -\tau \ln(R)$. The typical time scale τ is updated at each cycle and depends on the number of available sites for nucleation n_n and on the number of growth fronts n_g by the relation $\tau^{-1} = k_g n_g + k_n n_n$. The algorithm is halted when the whole DNA filament is covered. Polymers grow unidirectionally and their orientation is chosen randomly at nucleation. In the asymmetric version of this algorithm the polymers grow only toward one end [see Figs. 1(B) and 1(C)].

For different values of k_n and k_g, we compared the average normalized growth curve $\phi(t)$ with individual realizations of the stochastic process [Fig. 2(B)]. As expected, in the regime in which steplike growth sets in, the single realization can differ dramatically from the average curve. In order to identify the crossover, we defined a *distance* of a realization of the process from the average behavior as the integral

$$D = \int |t(s) - \bar{t}(s)|^2 ds,$$

where $t(s)$ indicates the time in which the realization reaches length s and $\bar{t}(s)$ is the same observable averaged over all
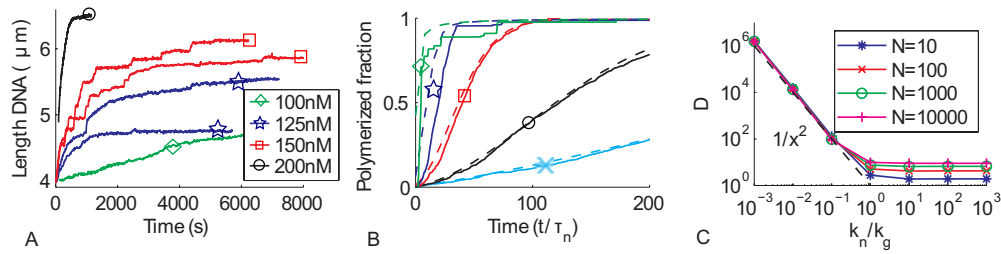
FIG. 2. (Color online) (A) Double strand DNA elongation curves as a function of time for different hRad51 concentrations under a force of 6 pN. Data refer to six experiments in four different protein concentrations (see legend). The elongation monitors the total coverage by nucleoprotein filaments. Steplike growth is clearly visible at a hRad51 concentration of 125 and 150 nM. Maximum DNA elongation varies from experiment to experiment because the attachment point of DNA is not controlled. (B) Comparison of realizations (continuous line) and average behavior (dashed line) extracted from kinetic Monte Carlo simulations. The five curves represent different values of k_g for constant $k_n = 1$, 80 000 ($\diamond$), 8000 ($\star$), 800 ($\square$), 80 ($\bigcirc$), and 8 ($\times$). (C) Violation of mean-field behavior: the quantity D, which measures the distance of a realization from the average behavior (see text), is plotted as a function of the ratio of the nucleation and growth rates for different system sizes; the value of D diverges for $k_g < k_n$ as $1/x^2$.

realization. This distance is small if the single realization well approximates the average one. We call a system that behaves in this way *self-averaging*. As expected, when the two time scales of nucleation and growth are separated, the system does not show a smooth kinetics but rather a steplike evolution, and the single realization differs from the average behavior. In Fig. 2(C) we plot the distance D for different parameter ratios k_n/k_g (we set $k_g = 1$ for simplicity) as a function of different system sizes. The distance D diverges exponentially when the nucleation rate becomes comparable or smaller than growth rate. This clearly identifies a steplike growth regime where the predictions of the KAMJ model breaks down, relevant to the experiments shown in Fig. 2(A).

B. Stick-breaking process and step statistics

In the complete time-scale separation limit, it is possible to model the problem analytically, linking it to a well-studied stochastic process, known in the mathematical literature as the "discrete uniform stick-breaking scheme" [30]. This process is defined by the length of fragments one obtains from infinitely and repeatedly "breaking a stick" of a unitary length (hence the name). To enhance the analogy to our problem, let us consider a one-dimensional template where a protein can attach and nucleate at a constant rate and then can polymerize at infinite velocity only toward one end of the filament [Fig. 1(B)]. The polymerization stops when the end of the filament or an occupied site is reached. We model the template as a lattice with sites $r = 1, \ldots, N$. At each step i a site $\bar{n}_i$ is randomly chosen among the N sites according to a uniform distribution. All the empty sites at its right ($l > n_i$) are covered. A typical time can be associated to each step by supposing that the events occur according to the Poisson statistics, and therefore (following Gillespie's scheme) the waiting time can be extracted from an exponential distribution $P(\Delta t) = \tau \exp(-\Delta t/\tau)$, where the time scale at each step is inversely proportional to the number of empty sites, $\tau = \tau_0/(N - \bar{n}_i)$. In the limit of an infinite fragment and if time is counted with the index of the ith jump, one has the discrete stick-breaking scheme.

In what follows, we study this process with the methods of statistical physics by direct simulation and analytic calcu-

lations, and we derive the observables that are useful in this context, linking to the known theorems from the mathematical literature. The main observable is the fraction of covered template or alternatively the fraction of empty sites at each step, which is related to the probability of having n empty sites at the step i, i.e., $P(n_i)$. Knowing this probability allows us to calculate averages of all the observables. While the distribution for all the steps is known [30], this quantity is hard to manipulate and not very instructive. In Appendix B, we derive with a simple method the probability distributions for the first few fragments, where the expressions remain relatively straightforward.

A first interesting result is that for the statistics of the segment, length does not change if the (directed) polymerization process can proceed in both directions, and the direction is chosen at random at nucleation [Fig. 1(C)]. This is relevant to the experimental system because it refers to what has been observed for hRad51 and RecA polymerizing on dsDNA (on ssDNA they both polymerize in the 3'-to-5' direction).

The equivalence of the two processes can be argued simply as follows. Let us consider the process with randomly directed polymerization. When a particle attaches to the lattice, the polymerization starts *either* toward the left *or* to the right. For every step, this symmetric process can be mapped onto the asymmetric one. Imagine, for example, that the first polymerization occurs toward the right and the second toward the left, then it is sufficient to "move" the left polymerized part to the end of the right polymer and the process would coincide with the previous one, as can be checked directly by enumerating the possible outcomes. This mapping is possible since the conditional probabilities given by Eq. (B1) are uniform for both the systems. It is straightforward to prove that this property also holds if the choice for a particular direction of polymerization is biased. Our simulations confirm this result, as can be seen from Fig. 3(B).

We will mainly deal with four observables: (i) the mean empty portion of the lattice (i.e., naked DNA) filament still available for polymerization at step i $\langle n_i \rangle$, (ii) the marginal probability that a step is smaller than the subsequent one $P(\Delta_i > \Delta_{i-1})$, (iii) the number of jumps S before a filament of length N is covered, and (iv) the fragment distribution.
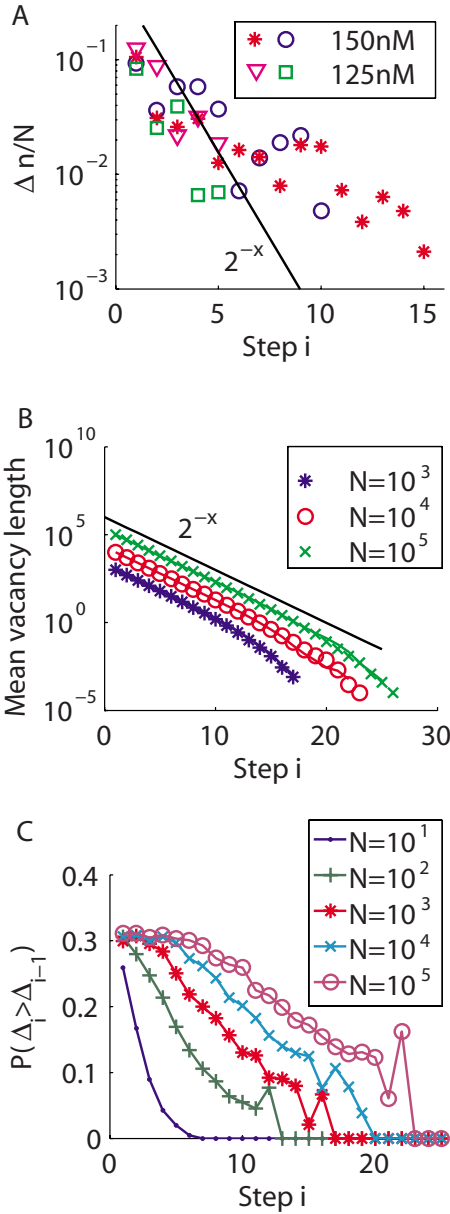
FIG. 3. (Color online) (A) Experimental normalized length of the polymerized segments; the data are compatible with the stick-breaking process for both concentrations of hRad51 (150 and 120 nM). (B) Simulated mean vacancy length according to the stick-breaking process. The process follows a power law identical to the one in A. (C) Probability for a segment to be larger than the previous one, $P(\Delta_i > \Delta_{i-1})$; the mean-field prediction works well for the first steps; finite-size effects appear already after $\ln(N)$ steps.

1. Average steps

The first measurable quantity we consider is the average portion of filament which remains empty at step i, $\langle n_i \rangle$. Intuitively, one expects that this quantity is half the lattice size for the first step, one quarter for the second one, and so on. This corresponds to a mean-field approximation where each step does not influence the subsequent one. This argument leads to an exponential relation

$$\langle n_i \rangle = \frac{1}{2^i}. \tag{6}$$

This approximation is valid for reasonably large systems (large N), as shown in Fig. 3(B), and ceases to apply for the last few steps at any finite size for large i because the steps become small and their discrete character appears. Evidence for this behavior is found also in the data [Fig. 3(A)].

As we mentioned earlier, for the first intervals, it is possible to derive the exact probability distributions (see Appendix B). Computing exactly this quantity we obtain

$$\langle n_1 \rangle = \frac{N+1}{2}, \tag{7}$$

$$\langle n_2 \rangle = \frac{1}{N}\left[H_N + \frac{(N-1)(N^2+2N-4)}{4N} \right], \tag{8}$$

where $H_n \equiv \Sigma_{k=1}^m \frac{1}{k}$ are "harmonic numbers." The limit for $N \to \infty$ of these quantities converges to expression (6).

While the experimental data available to us are limited, we compared them qualitatively with Eq. (6). The results are shown in Fig. 3(A). The predictions agree well with the experimental data sets at hRad51 concentration of 150 mM and less well with the sets at 125 mM, where, however, the step-like regime is less evident [see Fig. 1(A)] and a substantial portion of the DNA fragments was probably adsorbed on the substrate. While this can be regarded as a qualitative experimental indication that the experimentally observed steps have indeed the origin that we hypothesize, new experiments would be needed to establish this.

2. Comparison of subsequent steps

We compute the probability $P(\Delta_i > \Delta_{i-1})$ for the ith step Δ_i to be larger than the previous one, Δ_{i-1}. The intervals are related to the stochastic variable n_i through the relation $\Delta_i = n_{i-1} - n_i$ (with $\Delta_1 = N - n_1$). We propose that $P(\Delta_i > \Delta_{i-1})$ can be a useful observable in experiments, where the steps are measured directly, to quantify how the behavior is close to the stick-breaking regime. This quantity is difficult to determine analytically and is not analyzed directly in the mathematical literature. However, it can be accessed by numerical simulations.

For the first step it is possible to deduce $P(\Delta_i > \Delta_{i-1})$ from a continuum argument as follows. Let us observe that the probability for the second interval to be larger than the first, $P(\Delta_2 > \Delta_1)$, is related to the conditional probability $P(\Delta_2|\Delta_1)$. We can write

$$P(\Delta_2 > \Delta_1) = \int_0^{L/2} P(\Delta_2 > \Delta_1|\Delta_1)P(\Delta_1)d\Delta_1$$

$$= 1 - \ln 2 \approx 30.7\%, \tag{9}$$

where $P(\Delta_2 > \Delta_1|\Delta_1) = (L - 2\Delta_1)/(L - \Delta_1)$, and we have used L to indicate the template size N. This agrees well with simulated data, as shown in Fig. 3(C). The figure also shows that this probability distribution drops drastically after a few steps. The location of the drop depends logarithmically on the template size.
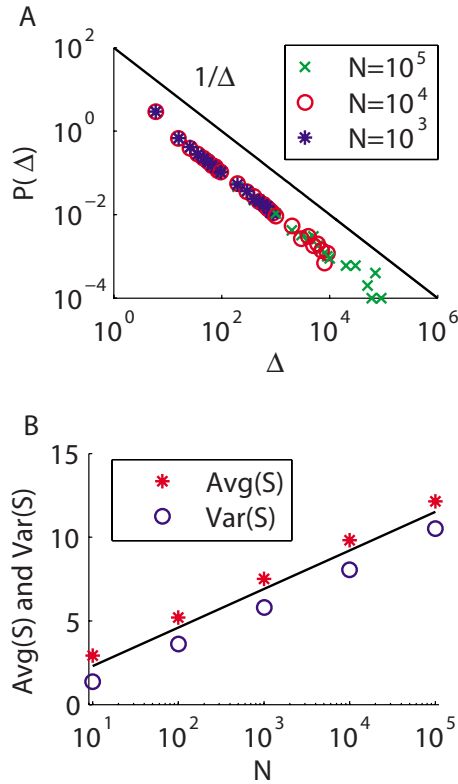
FIG. 4. (Color online) (A) Non-normalized fragment size distribution $P(\Delta)$: according to the stick-breaking process this distribution follows a power law Δ^{-1}. (B) Average and variance of the fragment number for the fully polymerized system as a function of the system size. The stick-breaking process predicts a logarithmic behavior confirmed here.

a. Number of steps. In our simulations, the total number of steps S needed to cover a template of size N scales like $\ln N$ with a variance that is also $\ln N$ [Fig. 4(B)]. This is the subject of a general asymptotic theorem. In Appendix B we derive this result analytically with a mean-field argument. This fact can be used to explain the logarithmic scaling of the drop in $P(\Delta_i > \Delta_{i-1})$. Indeed, this probability is expected to drop dramatically when the size effects become relevant; i.e., the template is "almost full," which happens at a number of steps that is order $\ln N$.

b. Step-size distribution. Finally, we consider the probability distribution function of the size of the steps, regardless of the step index, $P(\Delta)$. As shown in Fig. 4, our direct simulations give a probability distribution with clear power-law tail Δ^{-1}, with a cutoff determined by the template size. This means that no characteristic length scale enters in the system in this regime of time-scale separation. This result is derived in Appendix C with a simple mean-field argument.

IV. DISCUSSION

When polymerization is much faster than nucleation, the model predicts steplike growth on DNA fragments with length of order ~ 10 kbp. A similar behavior has been observed on comparably long filaments in magnetic tweezer experiments [15]. The slopes of the different growth parts of the curve seem to be proportional to the number of growing nuclei and appear to take discrete values. The minimal slope is the single-polymer growth velocity (i.e., $1/k_g$ in our notation), while the nucleation rate can be extracted from the step frequency in single experiments.

The elongational process occurring during polymerization is commonly regarded as a *local* stress; i.e., each monomer locally deforms the DNA. The addition of a monomer costs the same energy at every step (excluding the nucleus formation). The stress does not propagate along the polymer but is shared among the few monomers spiraling between two subsequent base pairs. On the other hand, there may exist stresses other than the local ones, which accumulate with increasing polymer size. For example, they can be caused by twist or mismatches between nucleic acids and binding sites of local oligomers. Also, recent experiments confirm the crystallographic observations that the attachment of hRad51 to a dsDNA induces a 1.5 times elongation on DNA and a negative twist of 65° per monomer (recently measured by Arata *et al.* [31]).

Previous studies [13] have reported a typical length of polymerization of ~ 30 monomers for free DNA templates. A characteristic length for the polymerized filaments can emerge from different phenomena: specific combination of nucleation and growth time scales (as suggested in [13] itself and [24]); configuration-coupled polymerization (e.g., because of the role of template structure); or accumulation of elongation or torsional stress (as described in the introduction). If the results presented in [13] are due to internal stress accumulation they would suggest that, during polymerization, the barrier to add a monomer builds up with nucleofilament length. However, in the experimental curves of Fig. 2 this characteristic length is not apparent. This could be related to the fact that in the experiments [15] the DNA is stretched (at 6 pN) and the energy barrier for the attachment of a monomer could be flattened. By quantifying systematically the step size and distribution under different applied forces, it should be possible to distinguish between a process with an intrinsic length scale and a stick-breaking-like polymerization with no length scale, answering important questions about the microscopic details and the forces involved in nucleoprotein polymerization. In general, it is true that, although single-molecule experiments do not allow to gather a large statistics, they give access to regimes that could not be observed otherwise and they are potentially a valuable tool for discriminating between models.

V. CONCLUSIONS

One-dimensional nucleation-growth processes have been studied extensively for more than a century, and many mathematical features have been worked out exactly. In this work, motivated by single-molecule experiments on the polymerization of a recombination on DNA, we investigate a regime that, to our knowledge, escapes the conventional description application of nucleation-growth models to the problem. In particular, we have shown that the polymerization of a nucleoprotein filament on nicked dsDNA has a regime that cannot

be described by the standard mean-field KAMJ model (where the number of growing filaments is large and the average growth curve corresponds to that of a single molecule). Accordingly, we have proposed a model that can describe this time-scale separation limit. Within this regime, the microscopic process of directed polymerization becomes relevant. This happens when growth is much faster than nucleation and gives rise to steplike polymerization curves on DNA fragments. The regime can have experimental importance because the time scales of polymerization and nucleation are separated and could be measured independently.

In the extreme separation of time scales, we derived both analytical and numerical predictions for some relevant observables. In particular, the mean size of polymerized fragments decreases exponentially, and hence the waiting time between subsequent nucleations increases exponentially, consistently with the available experimental data. Finally, we quantified the behavior for the number of steps as a function of template size, pointing to a scale-free distribution of step sizes. The observables described above (mean size of the fragments, waiting time, step-size distribution, etc.) can be used in experiments to gather evidence for structural parameters such a characteristic scale for nucleoprotein filament lengths. Additional experiments at low concentration will be necessary to validate the model and refine it with more details. On the other hand, our simulations and interpretation indicate that the data issued from this kind of experiments, which are more precise than the bulk measurements, have not yet been fully exploited.

APPENDIX A: SINGLE NUCLEUS APPROXIMATION

In this appendix we derive explicitly some results for the polymerization of a filament starting from a single nucleus. This approximation captures quite well the regime at very low concentrations. In Ref. [27] Turner suggested, correctly, that the kinetic polymerization curve proposed in [5] fits experimental data with only one parameter, containing the two time scales of the problem: the nucleation and the growth rates. In order to separate the time scale, he proposed, therefore, to treat the problem as if only one nucleus could polymerize over a finite-sized filament (at the same velocity in both directions).

According to the model in Ref. [27], the covering fraction can be written as a piecewise function of time t and of the position of the nucleus x,

$$\phi(x,t) = \begin{cases} 2vt & \text{for } 0 < t < \dfrac{x}{v} \text{ and } 0 < x < \dfrac{1}{2} \\[2mm] 1 & \text{for } \dfrac{1-x}{v} < t < \dfrac{1}{v} \text{ and } 0 < x < \dfrac{1}{2} \\[2mm] x + vt & \text{for } \dfrac{x}{v} < t < \dfrac{1-x}{v} \text{ and } 0 < x < \dfrac{1}{2}. \end{cases}$$

(A1)

Integrating over all the possible positions of the nucleus one obtains the average coverage fraction,

$$\langle \phi(t) \rangle_x = \begin{cases} vt - (vt)^2/2 & \text{for } 0 < t < \dfrac{1}{v} \\[2mm] 1 & \text{for } \dfrac{1}{v} < t. \end{cases}$$

(A2)

In the case of a polymer growing only in one direction, it is easy to see that in the same single-nucleus hypothesis, the above equations remain valid upon a substitution $v \to v/2$ (since the velocity is half the velocity of a double polymerization front).

The rates k_n and k_g can, therefore, be extracted by a fit with the analytical function in the first part of the experimental curve and with an *ad hoc* least-squares minimization with the Monte Carlo simulated curve to extract the ratio of the rates. This has been done in Ref. [13], giving reasonable results *in some specific experimental conditions*.

Note that this fit, however, gives reasonable results only if a single realization is "self-averaging" in the sense explained in the present paper.

It is also important to note that this "single-nucleus" approach gives exactly the same result as the KAMJ model for the short-time behavior of the polymerization curve. Indeed, the KAMJ short-time behavior can be recovered from Eq. (A2), supposing that the first step is distributed according to a Poisson distribution. In this case, an average over possible realizations gives

$$\langle \langle \phi(t) \rangle_x \rangle_{t_0} = \int_0^\infty \langle \phi(t - t_0) \rangle_x \exp(-\lambda t_0) \lambda dt_0.$$

(A3)

By computing the asymptotic expansion at short time we find $\langle \langle \phi(t) \rangle_x \rangle_{t_0} \sim_{t \to 0} \lambda v t^2$, meaning that a mean-field computation based on nucleation and growth always gives a quadratic growth (and hence a zero-slope curve for ϕ) at small times, as predicted by the KAMJ model. The finite-derivative short-time behavior highlighted in Ref. [15] is different and relates to the fact that the experiment deals with single realizations of the process. This further justifies the introduction of our description in terms of the stick-breaking process.

APPENDIX B: PROBABILITY DISTRIBUTIONS OF THE FIRST FEW STEPS

In this appendix we give the recipe to compute the probability distribution of the polymer length in the stick-breaking process and the analytical results for the first two polymers.

The conditional probability relating two following steps is the inverse of the available sites in the lattice,

$$P(n_i|n_{i-1}) = \frac{1}{n_{i-1} - 1}. \tag{B1}$$

The total probability can be therefore expressed as a sum of the product of all the conditional probabilities,

$$P(n_i) = \sum_{n_{i-1}=n_i+1}^{N-i+2} \cdots \sum_{n_k=n_{k+1}+1}^{N-k+1} \cdots \sum_{n_1=n_2+1}^{N} P(n_i|n_{i-1})$$
$$\cdots P(n_k|n_{k-1}) \cdots P(n_1|n_0), \tag{B2}$$

where the last probability is simply $P(n_1|n_0) = 1/N$. This expression resembles a path integral that includes the propagator from the step 0 to the step i and hence can be rewritten in a more compact way as

$$P(n_i) = \sum_{(*)} \prod_{k=1}^{i} P(n_k|n_{k-1}), \tag{B3}$$

where the subscript $(*)$ refers to the conditions

$$n_i + 1 \le n_{i-1} \le N - i + 2 n_{k+1} + 1 \le n_k \le \vdots \le N - k + 1 n_2 + 1$$
$$\le n_1 \le N.$$

Equation (B3) can be written as a sum of harmonic products,

$$P(n_i) = \frac{1}{N} \sum_{(*)} \prod_{k=2}^{i} \frac{1}{n_{k-1} - 1}, \tag{B4}$$

which, through a translation of indexes, takes the simpler form,

$$P(n_i) = \frac{1}{N} \sum_{(**)} \prod_{k=1}^{i-1} \frac{1}{n_k}. \tag{B5}$$

where the subscript $(**)$ refers now to the conditions

$$n_i \le n_{i-1} \le N - i + 1 n_{k+1} \le n_k \le N - k n_2 \le n_1 \le N - 1.$$

This sum can now be computed numerically. Using the properties of the harmonic numbers we provide here the analytical expressions for the first three steps,

$$P(n_1) = \frac{1}{N}, \tag{B6}$$

$$P(n_2) = \frac{1}{N}(H_{N-1} - H_{n_2-1}), \tag{B7}$$

$$P(n_3) = \frac{1}{2N}\left[(H_{N-2} - H_{n_3-1})^2 + (H_{N-2,2} - H_{n_3-1,2}) \right.$$
$$\left. + \frac{2}{N-1}(H_{N-2} - H_{n_3-1}) \right]. \tag{B8}$$

Higher order probabilities involve sums that can be written as combination of generalized harmonic numbers $H_{n,r} \equiv \Sigma_{k=1}^{n} 1/k^r$. It is possible to simplify the computation with an approximation that relies on the recurrence of the probabili-

ties. It consists in using only the harmonic numbers of order one and can be used to write a closed expression,

$$P(n_i) \approx \frac{1}{(i-1)!}(H_{N-i+1} - H_{n_i-1})^{i-1}. \tag{B9}$$

This approximation is not completely justified: although it is true that the higher harmonic numbers are negligible, we cannot say anything on the differences. However, a numerical check on systems of $N = 100$ site and up to n_6 gave quite good agreement with exact computations (the mistake made is around 20% and occurs only for extremal values of n_i, i.e., $n_i \sim 1$ and $n_i \sim N$).

Note that the probabilities written above for $i > 1$ assume implicitly that it is possible to attach the ith particle since none of the $(i-1)$th already filled the system (i.e., $n_j \ne 0$ for $j < i$). Hence, for example, $P(n_2)$, as it is written, is not normalized unless we include in the normalization sum also the case $n_2 = 0$ (i.e., the lattice is totally filled right at the first stroke).

By means of these probabilities one can compute the first three average polymer lengths exactly using the definition (with the caveat that the first $i-1$ sites at least need to be occupied by previously attached particles). The formula used to compute Eq. (7) is

$$\langle n_i \rangle = \sum_{n_i=1}^{N-i+1} P(n_i) n_i. \tag{B10}$$

APPENDIX C: MEAN-FIELD ARGUMENT FOR THE NUMBER OF STEPS AND THEIR SIZE DISTRIBUTION

In this appendix, we present a mean-field argument for the scaling of the step number $S \sim \ln N$, with template length N and the asymptotic step-size distribution $P(\Delta) \sim \frac{1}{\Delta}$ for the stick-breaking process. While the two results are established by asymptotic theorems that can be found in the mathematical literature [30], our aim here is to give an intuitive illustration of their derivation. In order to do this, we introduce a process which is equivalent to the stick breaking but uses a different sequential construction for the steps, with which it is simple to derive mean-field results.

Consider first the following process (the Polya urn) [30]. In the first step, a separator, labeled by 0, is placed, for example, on a lattice. In the subsequent steps, a new element, labeled by the step number N, can be added with equal probability in the $(N+1)$ positions that include the $N-1$ interstitial positions between N elements and the two sides. In case an interstitial position is chosen, all the other elements to the left or the right can be imagined to slide by one lattice position. The elements to the left of the separator are called "monomers" and the ones to its right are called "holes." Hence, the probability to add a monomer at size N if n monomers are already present is $P(n,N) = \frac{n}{N+1}$. The above-defined process is equivalent to our first polymerization step.

This can be easily seen by the fact that, at each step or fragment size, the possible (and equiprobable) outcomes are the permutations of the labels $0, \ldots, N$. In particular, for each N, the separator 0 is found uniformly on the available sites, which means that the polymer size (the portion of the lattice occupied by monomers) is chosen with uniform distribution.

Generalizing to multiple polymers, we define a second process (called in the statistics literature "Chinese restaurant process"), where at each step one has two possible moves:

(1) add a monomer to polymer i, with probability

$$p_O^i = \frac{n_i}{N+1},$$

(2) add a new polymer with one monomer, with probability

$$p_N = \frac{1}{N+1},$$

where n_i is the number of monomers of polymer i. The name "Chinese restaurant" comes from the analogy with a process where a new person arriving at a restaurant has probability p_O^i to be added to an "old" table containing already n persons and probability p_N to be added alone to a new table. This process maps exactly on the discrete stick-breaking process because fixing size N and ordering polymers by size, each one follows a Polya urn constrained on the available space left by the previous ones [30].

Using this alternative sequential construction it is simple to argue the results in the large N regime. Considering n_i and S as averaged observables and taking the continuum limit one has the equations

$$\partial_N n_i(N) = \frac{n_i}{N+1} \quad \text{and} \quad \partial_N S(N) = \frac{1}{N+1}. \tag{C1}$$

These equations have to be solved with initial conditions $n_i(N_i) = 1$, where N_i is the size of birth of polymer i. Using the second equation one readily has the first result,

$$S(N) = \log(N+1) \sim \log(N). \tag{C2}$$

To compute the asymptotic of $P(\Delta = j, N)$ it is necessary to evaluate the ratio $S(j, N)/S(N)$. One can use a simple argument along the lines in Ref. [32]. From the solution of the mean-field equation, $j > n_i(N)$ implies $N_i > N^*$, with $N^* = \frac{N-j+1}{j}$, so that the cumulative distribution for the steps can be estimated by computing the ratio of the (average) number of polymers born before size N^* and the number of polymers born before size N, $P[n_i(N) > j] = S(N^*)/S(N)$. $P(\Delta = j, N)$ can be obtained by derivation of this function. For $N, j \to \infty$ and j/n small, we find

$$P(\Delta = j, N) \sim \frac{1}{j}, \tag{C3}$$

which is our second result.

---

[1] S. Kowalczykowski, D. Dixon, A. Eggleston, S. Lauder, and W. Rehrauer, Microbiol. Rev. **58**, 401 (1994).
[2] P. R. Bianco, R. B. Tracy, and S. C. Kowalczykowski, Front. Biosci. **3**, D570 (1998).
[3] S. C. West, Nat. Rev. Mol. Cell Biol. **4**, 435 (2003).
[4] J. Hoeijmakers, Nature (London) **411**, 366 (2001).
[5] G. V. Shivashankar, M. Feingold, O. Krichevsky, and A. Libchaber, Proc. Natl. Acad. Sci. U.S.A. **96**, 7916 (1999).
[6] M. Hegner, S. B. Smith, and C. Bustamante, Proc. Natl. Acad. Sci. U.S.A. **96**, 10109 (1999).
[7] T. van der Heijden, J. van Noort, H. van Leest, R. Kanaar, C. Wyman, N. H. Dekker, N. Dekker, and C. Dekker, Nucleic Acids Res. **33**, 2099 (2005).
[8] R. Galletto, I. Amitani, R. J. Baskin, and S. C. Kowalczykowski, Nature (London) **443**, 875 (2006).
[9] R. Fulconis, J. Miné, A. Bancaud, M. Dutreix, and J.-L. Viovy, EMBO J. **25**, 4293 (2006).
[10] S. Smith, L. Finzi, and C. Bustamante, Science **258**, 1122 (1992).
[11] J. Hilario and S. C. Kowalczykowski, Curr. Opin. Chem. Biol. **14**, 15 (2010).
[12] J. van Mameren, M. Modesti, R. Kanaar, C. Wyman, G. Wuite, and E. Peterman, Biophys. J. **91**, L78 (2006).
[13] M. Modesti, D. Ristic, T. van der Heijden, C. Dekker, J. van Mameren, E. J. G. Peterman, G. J. L. Wuite, R. Kanaar, and C. Wyman, Structure (London) **15**, 599 (2007).
[14] R. Fulconis, A. Bancaud, J.-F. Allemand, V. Croquette, M.

Dutreix, and J.-L. Viovy, Biophys. J. **87**, 2552 (2004).
[15] J. Miné, L. Disseau, M. Takahashi, G. Cappello, M. Dutreix, and J.-L. Viovy, Nucleic Acids Res. **35**, 7171 (2007).
[16] J. van Mameren, M. Modesti, R. Kanaar, C. Wyman, E. Peterman, and G. Wuite, Nature (London) **457**, 745 (2009).
[17] J. Hilario, I. Amitani, R. Baskin, and S. Kowalczycowski, Proc. Natl. Acad. Sci. U.S.A. **106**, 361 (2009).
[18] A. Kolmogorov, Izv. Akad. Nauk SSSR, Ser. Mat. **3**, 335 (1937) [Bull. Acad. Sci. USSR, Phys. Ser. **1**, 355 (1937)].
[19] M. Avrami, J. Chem. Phys. **7**, 1103 (1939).
[20] M. Avrami, J. Chem. Phys. **8**, 212 (1940).
[21] M. Avrami, J. Chem. Phys. **9**, 177 (1941).
[22] W. A. Johnson and P. A. Mehl, Trans. Am. Inst. Min., Metall. Pet. Eng. **135**, 416 (1939).
[23] E. Seitz and S. Kowalczykowski, Nucleic Acids Res. **34**, 2847 (2006).
[24] T. van der Heijden, R. Seidel, M. Modesti, R. Kanaar, C. Wyman, and C. Dekker, Nucleic Acids Res. **35**, 5646 (2007).
[25] J. Evans, Rev. Mod. Phys. **65**, 1281 (1993).
[26] E. Frey and A. Vilfan, Chem. Phys. **284**, 287 (2002).
[27] M. S. Turner, Biophys. J. **78**, 600 (2000).
[28] S. Jun, H. Zhang, and J. Bechhoefer, Phys. Rev. E **71**, 011908 (2005).
[29] D. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[30] J. Placard, in *Ecole d'Eté de Probabilité de Saint-Flour XXXII 2002*, edited by J. Pitman (Springer, Berlin, 2002).
[31] H. Arata, A. Dupont, J. Miné-Hattab, A. Renodon-Cornière,

M. Takahashi, G. Cappello, and J.-L. Viovy, Proc. Natl. Acad. Sci. U.S.A. **106**, 19239 (2009).

[32] L. Barabasi and R. Albert, Science **286**, 509 (1999).

[33] C. Joo, S. McKinney, M. Nakamura, I. Rasnik, and T. Ha, Cell **126**, 515 (2006).

[34] Several works have also been directed toward measuring the size of the critical nucleus, now estimated as four to five RecA monomers [8,33] and five to six human Rad51 (hRad51) monomers [15,24] (other publications [17] report three monomers).