

Molecular information processing in nonequilibrium copolymerizations

David Andrieux and Pierre Gaspard

Center for Nonlinear Phenomena and Complex Systems,

Université Libre de Bruxelles, Code Postal 231, Campus Plaine, B-1050 Brussels, Belgium

We consider general fluctuating copolymerization processes, with or without underlying templates. We study the spatial ordering resulting from the nonequilibrium conditions as well as the information transmitted between the template and the synthesized copolymer. Information transmission turns out to be optimal in the infinite-dissipation limit when no correlations between the monomers are present. By contrast, we show that optimized regimes of information generation may exist when template and correlations effects interact.

I. INTRODUCTION

Information transmission and processing are crucial tasks for many biological functions. The most remarkable information processing system is certainly the genetic apparatus. Genetic information is coded into the DNA molecular structure. Its nucleotide configuration stores the information on the structure and function of subsequently synthesized proteins. The processing of biological information can be discussed in terms of the dynamics of populations associated with the different possible genetic sequences, the populations evolving by replications and mutations into quasi-species [1, 2], as observed *in vitro* in evolution experiments on RNA and viruses [3, 4]. However, DNA information processing occurs at the molecular level where thermal and chemical fluctuations are present. At thermodynamic equilibrium, the reactive events are random and it turns out that no information can be generated or transmitted because of the principle of detailed balancing [5, 6]. In this regard, the origin of a spatial order and information generation in biological systems cannot be accounted for by equilibrium mechanisms, especially due to the combinatorial explosion of possibilities.

On the other hand, recent developments in nonequilibrium thermodynamics show that out-of-equilibrium systems present a spontaneous symmetry breaking under time-reversal. Out of equilibrium, the probabilities of the typical paths become larger than the probabilities of the time-reversed paths. In this sense, the trajectories are more ordered in time than their time-reversed counterparts [7]. The emergence of this dynamical order is a natural consequence of the second law of thermodynamics [7, 8] and was shown to arise down to the nanometer scale [9, 10]. In Ref. [11], it has been shown that this dynamical order can be used to generate spatial order and information in copolymerization processes. Indeed, copolymerization can store in space the dynamical order which is generated in time by the nonequilibrium processes, establishing a coupling from time back to the spatial support of information. The possibility of such transfers of information has already been envisaged at the macroscopic level [12] and was shown in Ref. [11] to arise already at the molecular level. A related effect is that the spatial order or disorder of the copolymer sequence has a direct influence on the thermodynamics of the copolymer growth [11, 13].

These developments have thus shown how information can be generated under nonequilibrium conditions. Yet, it does not reveal how and under which conditions this information generation is efficient or reliable. The purpose of the present paper is thus to further develop the study of copolymerization processes, in particular from the viewpoint of their information transmission capacity. Indeed, random errors during the insertion of monomers lead to an imperfect fidelity in the transcription of the template. If the template encodes information such as the DNA molecule and the genetic code, the transmission of information is measured by the mutual information between the template and the copolymer synthesized. Remarkably, the mutual information appears linked to the dissipation necessary to the copolymerization reactions [11]. Here, we will consider the situation where the copolymers present correlations in the spatial arrangement of the monomers, in which case the information transmission is not always optimal in the infinite-dissipation limit, in contrast to the uncorrelated situation.

In Sec. II, we first introduce the stochastic description as well as the nonequilibrium thermodynamics of such copolymerization processes. In Sec. III, we proceed with the study of growth processes determined by local interactions, with or without an underlying template. In Sec. IV, we explore the case where correlations exist within the copolymer chain, either by first-neighbor interactions or by the existence of a two-state mechanism. Under certain conditions, new behaviors of the information transfer may appear. Conclusions are drawn in Sec. V.

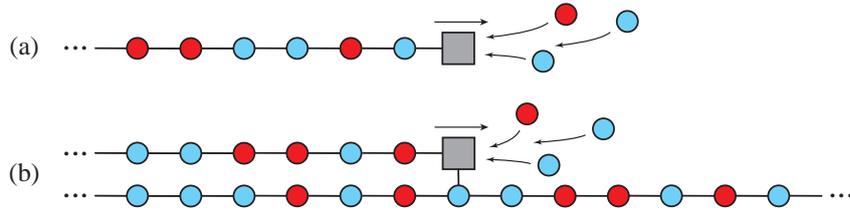


FIG. 1: (a) Free copolymerization without a template. (b) Template-based copolymerization. The circles schematically depict the monomers and the square the catalyst of polymerization.

II. COPOLYMERIZATION PROCESSES

Copolymerization processes can be described at the macroscopic level in terms of the population sizes, i.e., the chemical concentrations of the monomers, dimers, trimers, etc... [14, 15]. Since we are interested in the molecular structure generated during copolymerization, we adopt a mesoscopic description at the scale of a single copolymer chain.

A. Master equation description

We consider a copolymer chain surrounded by a solution containing monomers m . The solution surrounding each single copolymer chain is assumed to be sufficiently large to play the role of a reservoir where the concentrations of the monomers are kept constant. In this situation, an arbitrary chain ω can grow by adding a monomer m or shorten by releasing the last monomer of the chain: $\omega + m \rightleftharpoons \omega m$ (see Fig. 1a).

Such single-molecule processes can be described in the framework of stochastic theory as continuous-time random processes [16, 17]. In this framework, the probability $P_t(\omega)$ to find the copolymer chain ω of finite length at time t is ruled by the evolution equation:

$$\frac{dP_t(\omega)}{dt} = \sum_{\omega'} [P_t(\omega') W(\omega', \omega) - P_t(\omega) W(\omega, \omega')] , \quad (1)$$

where the coefficients $W(\omega, \omega')$ denote the rates of the transitions $\omega \rightarrow \omega'$ [16, 17]. For many processes at fixed pressure and temperature T , the ratio of forward to backward transition rates can be expressed in terms of the free enthalpy $G(\omega)$ of a single copolymer chain ω surrounded by the solution, according to

$$\frac{W(\omega, \omega')}{W(\omega', \omega)} = \exp \frac{G(\omega) - G(\omega')}{k_B T} , \quad (2)$$

where k_B is Boltzmann's constant [18]. The free enthalpy (i.e., Gibbs' free energy) is related to the enthalpy $H(\omega)$ and the entropy $S(\omega)$ of the copolymer chain ω in solution according to the standard thermodynamic relation:

$$G(\omega) = H(\omega) - TS(\omega) . \quad (3)$$

We notice that these quantities as well as the transition rates $W(\omega, \omega')$ depend in general on the concentrations of monomers in the solution surrounding the chain ω [16–18].

The overall thermodynamic entropy of the stochastic system described by the master equation (1) is given by [19]

$$S(t) = \sum_{\omega} P_t(\omega) S(\omega) - k_B \sum_{\omega} P_t(\omega) \ln P_t(\omega) . \quad (4)$$

This entropy varies in time with the probability distribution according to Eq. (1). It is known that these variations dS/dt can be decomposed as

$$\frac{dS}{dt} = \frac{d_e S}{dt} + \frac{d_i S}{dt} \quad (5)$$

into the exchange of entropy with the environment

$$\frac{d_e S}{dt} = \frac{d\langle S \rangle_t}{dt} - k_B \sum_{\omega, \omega'} P_t(\omega) W(\omega, \omega') \ln \frac{W(\omega, \omega')}{W(\omega', \omega)} \quad (6)$$

and the internal production of entropy

$$\frac{d_i \mathcal{S}}{dt} = k_B \sum_{\omega, \omega'} P_t(\omega) W(\omega, \omega') \ln \frac{P_t(\omega) W(\omega, \omega')}{P_t(\omega') W(\omega', \omega)} \geq 0, \quad (7)$$

which is always non-negative in agreement with the second law of thermodynamics [16, 17, 20]. The first term in Eq. (6) is the time derivative of the average entropy $S(\omega)$. The average of any quantity $X(\omega)$ is defined by

$$\langle X \rangle_t \equiv \sum_{\omega} P_t(\omega) X(\omega), \quad (8)$$

and we notice that its time derivative can be written as

$$\frac{d \langle X \rangle_t}{dt} = \sum_{\omega, \omega'} P_t(\omega) W(\omega, \omega') [X(\omega') - X(\omega)], \quad (9)$$

according to the master equation (1). As a consequence of Eq. (2), the second term in Eq. (6) can be expressed as $(1/T)d \langle G \rangle_t / dt$ so that the exchange of entropy is given by

$$\frac{d_e \mathcal{S}}{dt} = \frac{1}{T} \frac{d \langle H \rangle_t}{dt} \quad (10)$$

in terms of the time variation of the mean enthalpy of the system. Therefore, the exchange of entropy during the time interval Δt is related to the heat of copolymerization $\Delta \langle H \rangle$ according to $\Delta_e \mathcal{S} = \Delta \langle H \rangle / T$.

We may also consider a copolymerization process under the influence of a template, which determines the probabilities to add or remove monomers (see Fig. 1b). This is typically the case during DNA transcription or replication processes where the DNA sequence influences the probabilities to add the monomers, favoring the Watson-Crick pairing rule A-T and C-G. The template is thus composed of a chain $\alpha = \alpha_1 \alpha_2 \dots \alpha_l \dots$ and we suppose that the template is described by a stationary statistical distribution $\nu_l(\alpha) = \nu_l(\alpha_1 \alpha_2 \dots \alpha_l)$ which is normalized to unity $\sum_{\alpha} \nu_l(\alpha) = 1$ for all the chains of length l . We expect that the template ensemble average corresponds to the average over a typical realization of the template sequence. The transition rates in (1) will thus depend on the underlying substrate, $W(\omega, \omega' | \alpha)$, as well as the free enthalpy $G(\omega | \alpha)$, the enthalpy $H(\omega | \alpha)$, and the entropy $S(\omega | \alpha)$ of the copolymer ω growing on its template α . In this case, the overall thermodynamic entropy is again given by Eq. (4) so that the previous relations also hold in the case of copolymerization with a template.

B. Relationship between dissipation and information

It is assumed that, for fixed concentrations of monomers, the growth process can go on without limit and possibly reach a regime where the synthesized copolymer has stationary statistical properties [21, 22]. That is, after a long enough time, the probability to have a chain ω at time t will take the form

$$P_t(\omega) = p_t(l) \mu_l(\omega | \alpha), \quad (11)$$

where $p_t(l)$ is the probability distribution of the chain lengths $l \equiv |\omega|$ at time t and $\mu_l(\omega | \alpha) = \mu_l(\omega_1 \omega_2 \dots \omega_l | \alpha)$ denotes the stationary statistical distribution of the chains ω of length l grown on the template α , with the normalization $\sum_{\omega} \mu_l(\omega | \alpha) = 1$ for all l and α . In such a stationary regime, the mean length of the chains is evaluated as $\langle l \rangle_t = \sum_l p_t(l) \times l$ and it grows at the average velocity

$$v = \frac{d \langle l \rangle_t}{dt}. \quad (12)$$

As shown in Ref. [11], the entropy production (7) for these processes can be written as

$$\frac{d_i \mathcal{S}}{dt} = v A \geq 0 \quad (13)$$

with the growth velocity (12) and the thermodynamic force or affinity per monomer

$$A = \varepsilon + D(\text{polymer} | \text{template}). \quad (14)$$

This affinity has two contributions. The first one is the mean driving force given by $\varepsilon \equiv -g/T$ in terms of the free enthalpy per monomer:

$$g \equiv \lim_{l \rightarrow \infty} \frac{1}{l} \sum_{\omega} \mu_l(\omega) G(\omega), \quad (15)$$

where $G(\omega)$ is the free enthalpy of the copolymer ω [11]. The second contribution is the Shannon conditional entropy of the copolymer with respect to the template:

$$D(\text{polymer}|\text{template}) \equiv \lim_{l \rightarrow \infty} -\frac{1}{l} \sum_{\alpha} \nu_l(\alpha) \sum_{\omega} \mu_l(\omega|\alpha) \ln \mu_l(\omega|\alpha), \quad (16)$$

which characterizes the disorder or randomness in the sequence of monomers composing the copolymer [23, 24]. We note that these results do not depend on the specific choice of transition rates so that they are general in this respect.

The conditional entropy between two random objects can also be expressed in terms of the mutual information $I(X, Y) \equiv D(X) - D(X|Y)$ between the two objects [24, 25] so that the affinity (14) is also given by

$$A = \varepsilon + D(\text{polymer}) - I(\text{polymer}, \text{template}), \quad (17)$$

where

$$D(\text{polymer}) = \lim_{l \rightarrow \infty} -\frac{1}{l} \sum_{\omega} \mu_l(\omega) \ln \mu_l(\omega) \quad (18)$$

is the disorder of the copolymer sequence. The mutual information is bounded according to $0 \leq I(X, Y) \leq \min\{D(X), D(Y)\}$ and measures how much the knowledge on one of these objects reduces the uncertainty about the other. If the template distribution corresponds to a coding scheme such as the genetic code, the various errors will have different impacts on the transmitted message depending on the substituted monomer and the error location. As shown by Shannon [23], these aspects are included in the concept of the mutual information I which measures the amount of reliable information that can be transmitted. In this regard, the mutual information characterizes the fidelity of the replication [13]. We thus have the important result that, in this context, information transmission is directly linked to the dissipation according to Eq. (17).

In the special case of free copolymerization (i.e., without template, see Fig. 1a) or when the template does not influence the growth process, the copolymer is independent of the template so that the mutual information vanishes. Accordingly, we have the result

$$A = \varepsilon + D(\text{polymer}) \quad (19)$$

in the case of free copolymerization.

If the copolymer grows and $v > 0$, the affinity is the entropy production per added monomer so that Eq. (14) shows that substantial information can be generated if the mean driving force ε is large enough. The two contributions to the affinity are, on the one hand, the driving force ε given in terms of the mean free enthalpy of copolymerization and, on the other hand, the relative disorder (16) of the sequence of monomers in the chain with respect to the template. This shows in particular that the copolymer can grow by an entropic effect of disorder even when the driving forces are slightly negative. Indeed, the randomness incorporated in the chain has a non-negative contribution in the form of the disorder (16) to the thermodynamic force or affinity per monomer (14), showing how the stored pattern can influence back the dynamics of the system [11, 13]. Equilibrium occurs when the affinity (14) vanishes, $A_{\text{eq}} = 0$, in which case the mean driving force is equal to minus the equilibrium disorder. In this case, no systematic growth is possible and each monomer appears with its equilibrium distribution.

III. LOCAL PROCESSES

In general, the transition rates $W(\omega, \omega')$ depend on the structure and composition of the whole chain ω so that nonlocal or cooperative effects can generate long-range correlations along the chain, thereby reducing the disorder (16) or (18). However, in many cases, the transition rates only depend on the few monomers at the end of the chain, in particular, if the copolymerization process is mainly controlled by the free enthalpy changes of bond formation in the reactions $\omega + m \rightleftharpoons \omega m$.

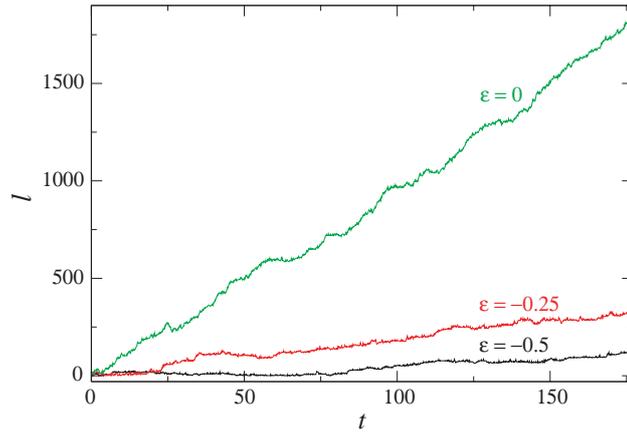


FIG. 2: **Free copolymerization:** Random growth of the copolymer length l versus the time t for different values of the driving force ε . The kinetic constants take the values $k_{+1} = 1, k_{+2} = 100$, while the reversed constants are fixed by $k_{-m} = k_{+m} \exp(-\varepsilon)$.

A. Free copolymerization

In the most local process, the transition rates only depend on the last monomer m which is added to or removed from the chain. In this case, the transition rates are denoted as $W(\omega, \omega m) = k_{+m}$ and $W(\omega m, \omega) = k_{-m}$ and the quantities

$$\varepsilon_m = \ln \frac{k_{+m}}{k_{-m}} \quad (20)$$

represent the driving forces favoring the chain growth. They are given by the free enthalpy changes of the chemical reactions and are measured in units of the thermal energy: $\varepsilon_m = [G(\omega) - G(\omega m)] / (k_B T)$. In this respect, the driving forces incorporate the chemical potentials of the monomers so that they can be varied by changing control parameters such as the chemical concentrations of the different monomers in solution around the copolymer chain. The master equation ruling this process reads

$$\begin{aligned} \frac{d}{dt} P_t(m_1 \cdots m_l) &= k_{+m_l} P_t(m_1 \cdots m_{l-1}) + \sum_{m_{l+1}=1}^M k_{-m_{l+1}} P_t(m_1 \cdots m_{l+1}) \\ &\quad - \left(k_{-m_l} + \sum_{m_{l+1}=1}^M k_{+m_{l+1}} \right) P_t(m_1 \cdots m_l). \end{aligned} \quad (21)$$

Random trajectories corresponding to the master equation (1) can be obtained using Gillespie's algorithm [26], as illustrated in Fig. 2 for different values of the driving force.

If we assume that the statistical distribution of the grown copolymer is stationary as described by Eq. (11), the mean growth velocity can be obtained as

$$v = \sum_m (k_{+m} - k_{-m} \mu_m), \quad (22)$$

as shown in the appendix, where $\mu_m \equiv \mu_1(m)$ denotes the stationary distribution of monomers.

At thermodynamic equilibrium, the detailed balance conditions are satisfied and read

$$P_{\text{eq}}(\omega) k_{+m} = P_{\text{eq}}(\omega m) k_{-m}. \quad (23)$$

In this case, the probability of a polymer configuration is given by

$$P_{\text{eq}}(m_1 \cdots m_l) \sim \exp \left(\sum_{i=1}^l \varepsilon_{m_i} \right), \quad (24)$$

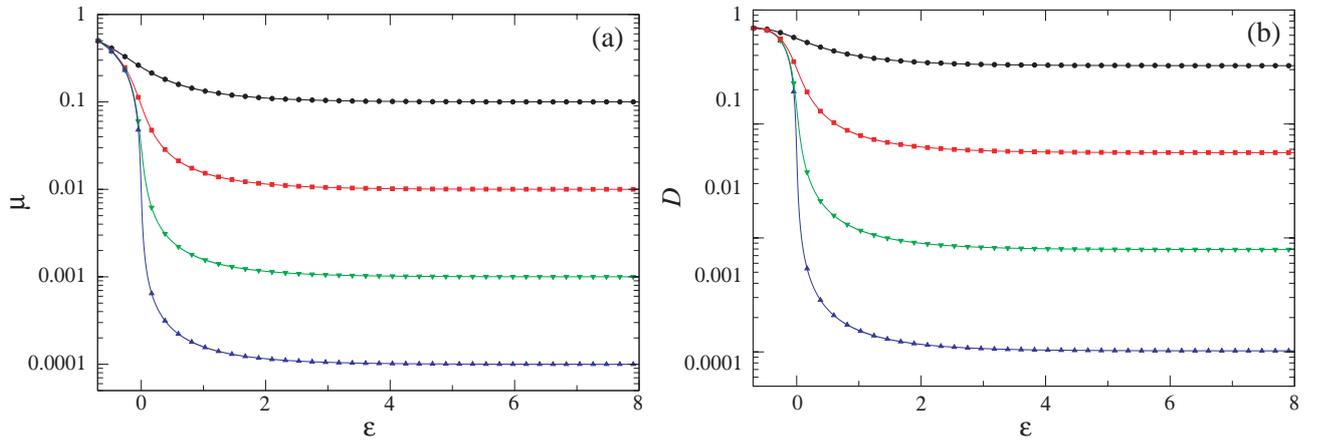


FIG. 3: **Free copolymerization:** (a) Probability μ to observe the monomer 1 as a function of the driving force and for different values of the kinetic constant k_{+2} . (b) Disorder (18) of the copolymer chain as a function of the driving force and for different values of the kinetic constant k_{+2} . In both figures, $k_{+2} = 2 \times 9$ (circle), $k_{+2} = 2 \times 99$ (square), $k_{+2} = 2 \times 999$ (triangle down), and $k_{+2} = 2 \times 9999$ (triangle up). The kinetic constant $k_{+1} = 2$, while the reversed constants are determined as $k_{-m} = k_{+m} \exp(-\varepsilon)$.

so that the equilibrium distribution of monomers is

$$\mu_{m,\text{eq}} = \frac{e^{\varepsilon_m}}{\sum_n e^{\varepsilon_n}}. \quad (25)$$

Therefore, the velocity (22) becomes

$$v = \sum_m k_{+m} \left(1 - \frac{1}{\sum_n e^{\varepsilon_n}} \right) \quad (26)$$

under the condition $\sum_m e^{\varepsilon_m} < 1$, which furthermore guarantees the normalizability of the probability distribution, $\sum_\omega P(\omega) = 1$. Since the growth velocity vanishes at equilibrium with the dissipation, we obtain the following condition

$$\sum_m e^{\varepsilon_m} = 1 \quad \text{at equilibrium.} \quad (27)$$

Accordingly, no stationary distribution is reached at equilibrium and the chain length performs a semi-infinite random walk, $l(t) \sim \sqrt{t}$. If we further assume that there is no free enthalpy difference between the chains ωm , the driving forces are equal to each other, $\varepsilon_m = \varepsilon$ for all m , whereupon equilibrium occurs when the driving force takes the value $\varepsilon_{\text{eq}} = -\ln M$ and the disorder reaches its maximum value $D_{\text{eq}} = \ln M$.

Out of equilibrium, the polymer grows at a positive velocity (22) with the random insertion or deletion of monomers. In these circumstances, the disorder can be reduced at the expense of the driving force ε according to Eq. (19). Indeed, in the stationary regime, the distribution $\mu_m = \mu_1(m)$ of monomers satisfies

$$k_{+m} - k_{-m}\mu_m = v\mu_m \quad (28)$$

with the constraint $\sum_m \mu_m = 1$, as shown in the appendix.

This result is the consequence of Eq. (A6) combined with Eq. (A4) and the fact that, for this local growth process, the monomers are uncorrelated so that Eq. (A9) holds. This result is deduced in the appendix from the master equation for the case of two species of monomers $M = 2$ and verified in the numerical simulations for $M \geq 2$.

The system formed by Eqs. (22) and (28) can be solved to obtain the speed v and the monomer distribution in terms of the kinetic constants. For instance, if $M = 2$, the distribution of monomers is given by $(\mu, 1 - \mu)$ where μ is the positive solution of

$$\mu^2(k_{-1} - k_{-2}) - \mu(k_{+1} + k_{+2} + k_{-1} - k_{-2}) + k_{+1} = 0, \quad (29)$$

that is

$$\mu = \frac{-b + \sqrt{b^2 - 4ac}}{2a} \quad (30)$$

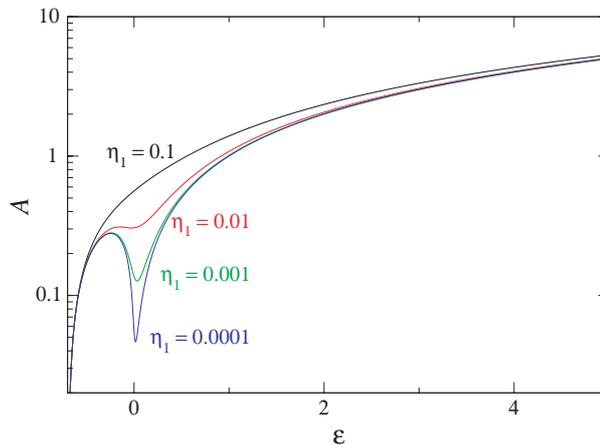


FIG. 4: **Free copolymerization:** Entropy production per monomer $A = \varepsilon + D$ as a function of the driving force ε and for different values of the parameter $\eta_1 \equiv k_{+1}/(k_{+1} + k_{+2})$. The kinetic constants take the values $k_{+1} = 2$ and $k_{-m} = k_{+m} \exp(-\varepsilon)$.

with $a = k_{-1} - k_{-2}$, $b = -(k_{+1} + k_{+2} + k_{-1} - k_{-2})$, and $c = k_{+1}$. The mean value of the driving forces can now be calculated as $\varepsilon = \sum_m \mu_m \varepsilon_m$ by averaging them over the normalized distribution μ_m of monomers.

In Fig. 3a, we depict the probability μ of the monomer 1 as a function of the driving force and for different values of the parameters. We observe that this probability decreases significantly away from equilibrium while the probability $1 - \mu$ of the other monomer increases. As a consequence, the configurational disorder of the chain is reduced under high driving forces, as seen in Fig. 3b. We emphasize that this spatial ordering is purely due to kinetic effects as we do not change the relative stability of the monomers,

$$\frac{k_{+m}}{k_{+n}} = \frac{k_{-m}}{k_{-n}}, \quad (31)$$

for all the values of the driving forces since we have here chosen $\varepsilon_m = \varepsilon$ for all the m 's.

This decrease of the disorder may lead to a non-monotonous increase of the dissipation per monomer (19) as a function of the driving forces. This is illustrated in Fig. 4. The transition between a monotonous and non-monotonous curve occurs around $\varepsilon \simeq 1/85$ for these values of the parameters and ensues from the compromise between the disorder and the driving forces.

In the limit of arbitrarily high driving forces, the monomers attach to but never detach from the chain and the process is totally irreversible. The growth proceeds by the random addition of monomers at the velocity $v = \sum_m k_{+m}$. The insertion probabilities are here given by $\mu_m = k_{+m}/\sum_n k_{+n}$, showing that the chain may be disordered due to the multiplicity of the monomer species composing the copolymer. The disorder reaches the value $D = -\sum_m \mu_m \ln \mu_m$, which is typically smaller than the equilibrium disorder D_{eq} . Therefore, the nonequilibrium drive may contribute to some extent to the ordering of the copolymer.

B. Template-based copolymerization

We here consider the more general case of a growth process influenced by an underlying template. In the most local process, the template only influences the insertion or removal of the last monomer of the chain. We therefore take transition rates of the form $W(\omega, \omega m | \alpha) = k_{+m|n}$ and $W(\omega m, \omega | \alpha) = k_{-m|n}$ if $n = \alpha_{l+1}$ is the template at the position $l + 1$ and ω is a chain of length l . The corresponding driving forces

$$\varepsilon_{m|n} = \ln \frac{k_{+m|n}}{k_{-m|n}} \quad (32)$$

now depend on the substrate as well. Accordingly, in the case (32), the mean driving force is given by $\varepsilon = \sum_{m,n} \nu_1(n) \mu_1(m|n) \varepsilon_{m|n}$, where $\nu_1(n)$ is the distribution of monomers in the template. The transition rates and the driving forces (32) incorporate external conditions such as the chemical concentrations of the nucleotides, the polymerase cofactors, and the products of the polymerization. Accordingly, the driving force ε will typically be a function of such control parameters.

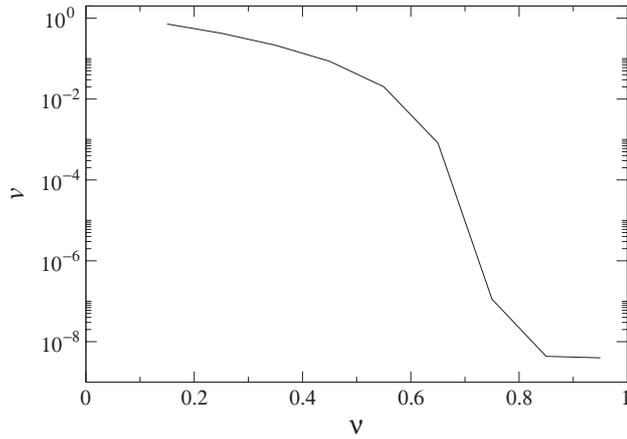


FIG. 5: **Template-based copolymerization:** Velocity of the growth process as a function of the probability ν of the monomer 1 in the template given by a Bernoulli random sequence of uncorrelated monomers 1 and 2 of respective probabilities ν and $1-\nu$. For large value of ν , the process is at equilibrium and presents a vanishingly small speed as the observation time increases. The copolymerization process proceeds with the kinetic constants $k_{+1|1} = 5$, $k_{+2|1} = 1/20$, $k_{+2|2} = 2$, and $k_{+1|2} = 1/500$ with the reversed constants $k_{-m|1} = k_{+m|1}e^{-\varepsilon_1}$, and $k_{-m|2} = k_{+m|2}e^{-\varepsilon_2}$ ($m = 1, 2$), and the driving forces $\varepsilon_1 = -\ln 3$ and $\varepsilon_2 = 1$.

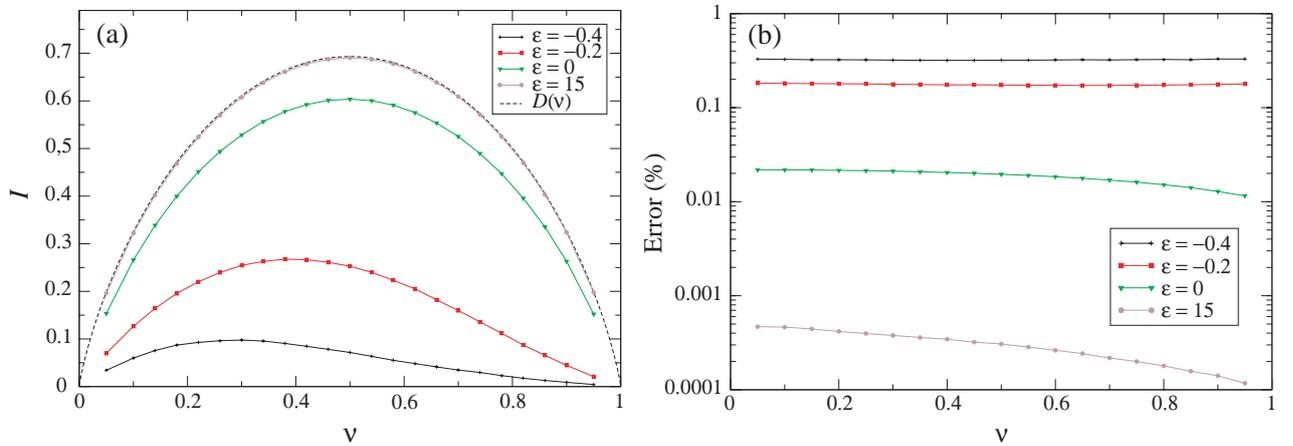


FIG. 6: **Template-based copolymerization:** (a) Mutual information as a function of the probability ν of the monomers 1 in the template and for different values of the driving forces. The mutual information is always lower than the disorder $D(\nu) = -\nu \ln \nu - (1-\nu) \ln(1-\nu)$ of the template (dashed line). (b) Percentage of errors in copying the template. The constants take the values $k_{+1|1} = 1$, $k_{+2|1} = 1/10000$, $k_{+2|2} = 2$, and $k_{+1|2} = 1/1000$, while the reversed kinetic constants are given by $k_{-m|n} = k_{+m|n}e^{-\varepsilon}$ with $m, n = 1, 2$.

A first observation is that the statistical properties of the template can affect the growth properties. In the simplest case where $\varepsilon_m = \varepsilon$ for all the values of m , the detailed balance conditions imply that equilibrium occurs at $\varepsilon = -\ln M$ independently of the template distribution. However, if the values of ε_m are not all equal, the template distribution may determine the growth properties. For instance, Fig. 5 shows that no growth is possible when the template presents a too strong bias towards one monomer.

In the template-based copolymerization, nonequilibrium conditions allow for the transmission of information. More precisely, the polymerization process in the presence of a template can be viewed as a transmission channel in information theory. The capacity of the channel is defined as the maximum of the mutual information over input distributions. Figure 6a shows for instance that the mutual information between the template and the copolymer depends on the template distribution ($\nu, 1-\nu$). The channel capacity is the maximum of the mutual information, which strongly depends on the driving forces while the optimal template distribution is close to $1/2$ at large driving forces and shifts towards smaller values at small driving forces. On the other hand, Fig. 6b shows that the percentage of errors in copying the template strongly decreases as the process is driven away from equilibrium. Depending on the regime of functioning, i.e., on the driving forces, different template distributions can be more effective than other

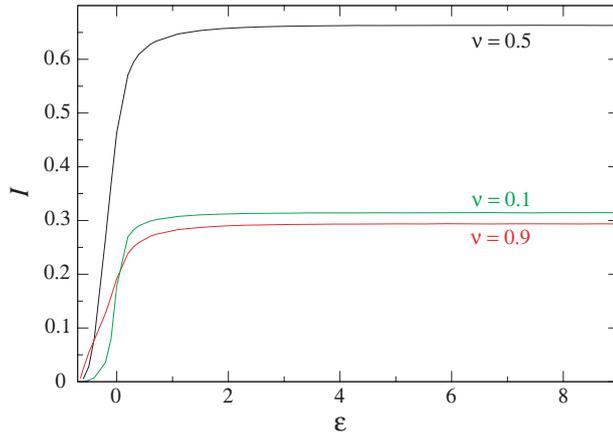


FIG. 7: **Template-based copolymerization:** Mutual information between the copolymer and the template as a function of the driving force and for different templates. The template monomers are distributed as a binary Bernoulli random sequence of distribution $(\nu, 1 - \nu)$. The copolymerization process proceeds with the kinetic constants $k_{+1|1} = 2$, $k_{+2|1} = 1/50$, $k_{+2|2} = 1$, and $k_{+1|2} = 1/2000$. The reversed kinetic constants are given by $k_{-m|n} = k_{+m|n} \exp(-\varepsilon)$ with $m, n = 1, 2$.

for the purpose of information transmission, as seen in Fig. 7. An important observation is that, irrespectively of the template distribution, the transmitted information is an increasing function of the driving force. Hence, the maximum value of the mutual information is reached in the infinite-dissipation limit.

In the infinite-dissipation limit, i.e., when monomers are inserted with rates $k_{+m|n}$ but never detach, the insertion probabilities $\eta_{m|n}$ are given by $\eta_{m|n} = k_{+m|n} / \sum_m k_{+m|n}$ for a monomer m instead of n . Accordingly, the velocity is given by $v_{\max}^{-1} = \sum_n \nu_1(n) / (\sum_m k_{+m|n})$ while the mutual information is $I_{\max} = \sum_{m,n} \eta_{mn} \nu_1(n) \ln[\eta_{nm} / \sum_n \eta_{mn} \nu_1(n)]$.

IV. CORRELATED PROCESSES

In this section, we introduce two types of mechanisms generating correlations within the copolymer chain, namely first-neighbor interactions and the presence of multiple dynamical states.

The correlation function for a chain composed of symbols $\omega_1 \omega_2 \dots \omega_l$ is computed by the formula

$$C(n) = \frac{1}{M} \sum_{m=1}^M \left[\frac{1}{L} \sum_{i=1}^L \chi_m(\omega_i) \chi_m(\omega_{i+n}) - \left(\frac{1}{L} \sum_{i=1}^L \chi_m(\omega_i) \right)^2 \right], \quad (33)$$

where χ is the indicator function: $\chi_m(\omega_i) = 1$ if $\omega_i = m$ and 0 otherwise. We notice that the expression (33) does not introduce any bias due to the choice of a numerical representation of the symbols. The correlation function vanishes for $n \rightarrow \infty$ if the monomers separated by a distance $n \rightarrow \infty$ become statistically independent.

In the presence of positive correlations, the tendency to present a sequence of identical monomers interacts in a non-trivial way with the underlying template, which results in new behaviors of the mutual information generated as a function of the driving forces. For instance, the infinite-dissipation limit is not always optimal, in contrast with the situation of Sec. III. In this case, there exists a finite value of the driving force maximizing the mutual information.

A. Cooperative effects

We first consider the simplest case of cooperative effects, namely a first-neighbor interactions between the monomers. The rate constants now depend on the last monomer inserted in the copolymer chain



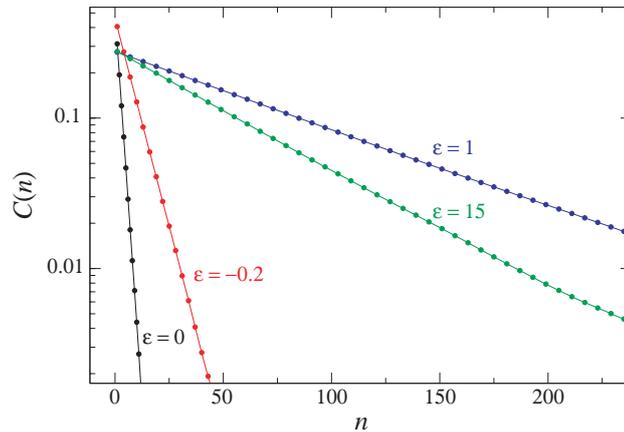


FIG. 8: **Free copolymerization with first-neighbor interactions:** Correlation function (33) for different values of the driving force ε . The constants take the values $k_{+11} = 2$, $k_{+21} = 1/250$, $k_{+22} = 5$, and $k_{+12} = 1/20$. The reversed kinetic constants are given by $k_{-mn} = k_{+mn} \exp(-\varepsilon)$ with $m, n = 1, 2$.

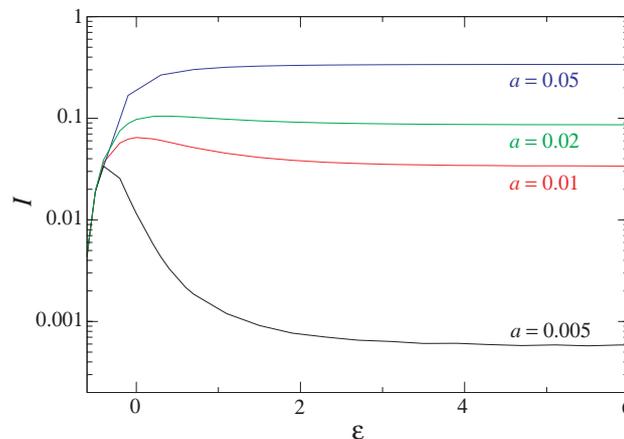


FIG. 9: **Template-based copolymerization with first-neighbor interactions:** Mutual information between the template and the copolymer as a function of the driving force. The template is generated with a distribution of probability $\nu = 0.5$. The kinetic constants take the values $k_{+11|1} = 2$, $k_{+21|1} = 2 \times a$, $k_{+12|1} = 0.04 \times a$, $k_{+22|1} = 0.04$, $k_{+11|2} = 0.05$, $k_{+21|2} = 0.05 \times a$, $k_{+12|2} = 5 \times a$, and $k_{+22|2} = 5$. The parameter a controls the rate of alternance between first neighbors ($a=0.005, 0.01, 0.02, 0.05$). The driving forces are chosen identical for any monomer organization and underlying template: $k_{-xy|y'} = k_{+xy|y'} \exp(-\varepsilon)$.

The copolymerization is supposed to proceed without template. Here, we will focus on purely kinetic effects so that we consider identical driving forces

$$\varepsilon = \ln \frac{k_{+mn}}{k_{-mn}}, \quad (35)$$

irrespective of the monomer organization.

This process presents exponentially decaying correlations as shown in Fig. 8. These correlations vanish at equilibrium and increase with the driving forces before reaching a maximum correlation length in the infinite-dissipation limit.

In the infinite-dissipation limit, the insertion of monomers obeys a first-order Markov chain. For instance, if $M = 2$, its stationary distribution is given by $[p(1), p(2)] = N[k_{21}(k_{11} + k_{12}), k_{12}(k_{22} + k_{21})]$ with the normalization constant $N^{-1} = k_{21}(k_{11} + k_{12}) + k_{12}(k_{22} + k_{21})$. The probability of a sequence is then determined by the Markov chain $p(\omega_1)P_{\omega_1\omega_2} \cdots P_{\omega_{l-1}\omega_l}$, with the transition probabilities $P_{\omega\omega'} = k_{\omega\omega'}/(k_{\omega 1} + k_{\omega 2})$.

We can now consider the case where a template influences the copolymerization process. The transition rates now depend on the substrate as well. In the case the growth process is only influenced by the last template unit, the transition rates can be expressed as $W(\omega m, \omega mn|\alpha) = k_{+mn|n'}$ and $W(\omega mn, \omega m|\alpha) = k_{-mn|n'}$ where $n' = \alpha_{l+2}$ is the template at the position $l + 2$ if ω is a chain of length l . We here consider the case $M = 2$ with a template given

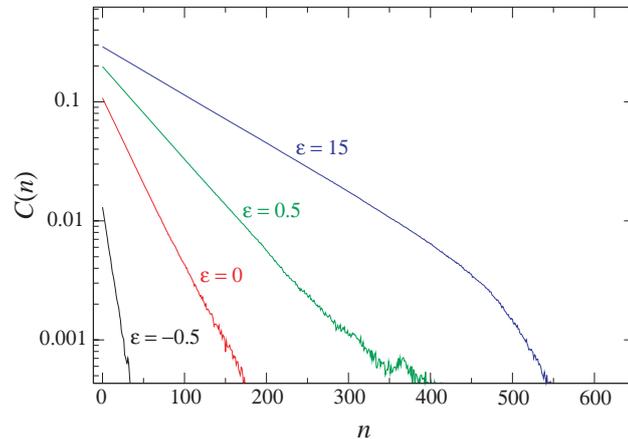


FIG. 10: **Free copolymerization by Coleman-Fox multistate mechanism:** Correlation function (33) for different driving forces. The constants take the values $k_{+1} = 14$, $k_{+2} = 2$, $q_{+1} = 1$, $q_{+2} = 15$, $\lambda_1 = 0.05$, and $\lambda_2 = 0.1$. The reversed kinetic constants are given by $k_{-m} = k_{+m} \exp(-\varepsilon)$ and $q_{-m} = q_{+m} \exp(-\varepsilon)$ with $m = 1, 2$.

by a Bernoulli random sequence of distribution $(\nu, 1 - \nu)$. Figure 9 shows the mutual information as a function of the driving force for different values of the rate of alternance between first neighbors. If the rate of alternance is low and if the driving is too strong, we see that the growing copolymer cannot reliably acquire information about the template. Instead, an optimal regime exists at finite driving force, that is in the low-dissipation regime.

B. Multistate mechanism

We go on and consider a mechanism first studied by Coleman and Fox [22] in the totally irreversible case and for free polymerization. This mechanism assumes the existence of several dynamical states influencing the growth process. For two such dynamical states, the polymerization reactions read



where the subscript 1,2 denotes the dynamical state of the chain. Such processes possess the remarkable property to present a non-Markovian behavior at all orders [22]. The statistical properties of the copolymer generated according to the reaction scheme (36) have been obtained by Coleman and Fox [22] in the infinite-dissipation limit, $k_{-m} = q_{-m} = 0$.

A slightly generalized derivation along the lines of Ref. [11] shows that no other contribution to the entropy with a linear increase in time appears so that the result (14) remains valid for this situation as well. This is expected since the reaction (36c) between the dynamical states is necessarily in equilibrium and therefore does not contribute to the entropy production.

In Fig. 10, we depict the correlation function (33) which is well described in general by combinations of exponentials [22]. The correlations increase with the driving forces before reaching a maximum correlation length in the infinite-dissipation limit. The correlation function vanishes at equilibrium, as it should.

We now consider the case where an underlying template is present. Precisely, we consider the case $M = 2$ with a template given by a Bernoulli random sequence of distribution $(\nu, 1 - \nu)$. The transition rates now read $W(\omega_1, \omega m_1 | \alpha) = k_{+m|n}$, $W(\omega m_1, \omega_1 | \alpha) = k_{-m|n}$, $W(\omega_2, \omega m_2 | \alpha) = q_{+m|n}$, $W(\omega m_2, \omega_2 | \alpha) = q_{-m|n}$, $W(\omega_1, \omega_2 | \alpha) = \lambda_1$, and $W(\omega_2, \omega_1 | \alpha) = \lambda_2$, where $n = \alpha_{l+1}$ is the template at the position $l + 1$ if ω is a chain of length l . The mutual information as a function of the driving force is depicted in Fig. 11 for different template distributions. In spite of the long correlations and in contrast with cooperative effects of the previous subsection, the mutual information between the template and the copolymer appears to be an increasing function of the driving forces for all values of the parameters we investigated.

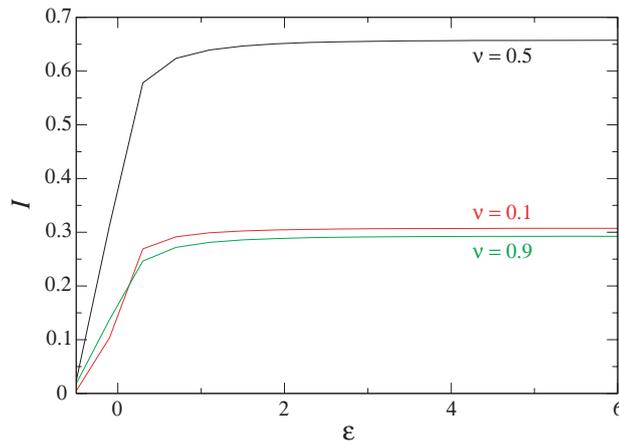


FIG. 11: **Template-based copolymerization by Coleman-Fox multistate mechanism:** Mutual information between the template and the copolymer as a function of the driving force and for different template distributions. The rate constants take the values $k_{1|1} = 10$, $k_{1|2} = 1/60$, $k_{2|1} = 0.1$, $k_{2|2} = 5$, $q_{1|1} = 2$, $q_{2|1} = 0.01$, $q_{1|2} = 0.03$, $q_{2|2} = 15$, $\lambda_1 = 0.05$, and $\lambda_2 = 0.1$.

V. CONCLUSIONS

In this paper, we have studied the nonequilibrium dynamics of copolymerization processes. At the level of single polymer chains, the random insertion of monomers generates a fluctuating random process that can be described in terms of a master equation for the probability distribution of the chain organization.

In this framework, we have shown that a spatial ordering of the copolymer can emerge under nonequilibrium conditions, decreasing the spatial disorder in the configuration of the copolymer. When an underlying template is present, this nonequilibrium ordering leads to the transmission of information [11].

Such copolymer chains can be thought of as information transmission channels, transmitting information with some (monomer-dependent) error rate. It turns out that the properties of such transmission channels depend non-trivially on the reaction rates or the driving forces. Accordingly, different regimes exist depending on the driving forces and the template distribution. Yet, in the local copolymerization process, information generation is always maximum in the infinite-dissipation limit. By contrast, the presence of correlations within the copolymer composition generates new behaviors of this information transmission, in particular with respect to the driving forces. In this case, the complex interplay between template and correlations may give rise to an optimal regime of information generation at finite driving forces, hence at finite dissipation.

The experimental investigation of the properties we have described in this paper would require techniques able to analyze the sequences of monomers in copolymers synthesized out of equilibrium with different driving forces. Such sequencing techniques have been developed for the very special copolymer which is DNA and we believe that other copolymers could be sequenced similarly. For fast analysis of copolymer sequences, new single-molecule techniques are under development, which are based on the translocation of the copolymer through a nanopore between two solutions at different electric and chemical potentials [27]. The detection of each copolymer unit momentarily located in the nanopore could be performed by measuring the ionic current signal through the nanopore [27]. Advances are under way in the molecular engineering of nanopores to allow the differentiated detection of each copolymer unit and eventually the determination of the whole copolymer sequence [28]. Although sequencing techniques are currently developed for DNA, their application to other types of copolymers could be envisaged.

Acknowledgments

D. Andrieux thanks the F.R.S.-FNRS Belgium for financial support. This research is financially supported by the Belgian Federal Government (IAP project “NOSY”) and the “Communauté française de Belgique” (contract “Actions de Recherche Concertées” No. 04/09-312), and the F.R.S.-FNRS Belgium (contract F. R. F. C. No. 2.4577.04).

APPENDIX A: SOLVING THE MASTER EQUATION OF FREE LOCAL COPOLYMERIZATION

If the assumption (11) that the copolymer grows with a stationary statistical distribution of monomers is introduced in the master equation (21), we obtain a hierarchy of equations.

First of all, we sum over all the monomers $m_1 \cdots m_l$ to get the equation for the probability distribution of the lengths:

$$\frac{dp_t(l)}{dt} = a p_t(l-1) - (a+b) p_t(l) + b p_t(l+1) \quad (\text{A1})$$

with the coefficients

$$a = \sum_m k_{+m}, \quad (\text{A2})$$

$$b = \sum_m k_{-m} \mu_1(m). \quad (\text{A3})$$

The linear equation (A1) can be solved by supposing a solution of the form $p_t(l) \sim \exp(st + iql)$. Expanding the eigenvalue s in powers of the wavenumber q , we find $s = -ivq - Dq^2 + O(q^3)$ with the growth velocity

$$v = a - b = \sum_m k_{+m} - \sum_m k_{-m} \mu_1(m) \quad (\text{A4})$$

and the diffusion coefficient

$$D = \frac{a+b}{2} = \frac{1}{2} \sum_m k_{+m} + \frac{1}{2} \sum_m k_{-m} \mu_1(m). \quad (\text{A5})$$

Now, summing over all but the last monomer (i.e., $m_1 \cdots m_{l-1}$), multiplying by the length l , and summing over this latter, we obtain an equation containing terms proportional to the mean length $\langle l \rangle_t \simeq vt$ plus a more slowly growing rest. In the limit $\langle l \rangle_t \rightarrow \infty$, the dominant terms proportional to $\langle l \rangle_t$ should thus be in balance and we get the following equation for the pair probabilities $\mu_2(mn)$:

$$\sum_n k_{-n} \mu_2(mn) - \left(k_{-m} + \sum_n k_{+n} \right) \mu_1(m) + k_{+m} = 0. \quad (\text{A6})$$

After summing over m , this expression vanishes identically, which verifies the consistency at this level. Besides, the pair probabilities must satisfy

$$\sum_n \mu_2(mn) = \mu_1(m). \quad (\text{A7})$$

If we sum over less monomers, further equations can be obtained for the higher-order probabilities:

$$\sum_{m_{l+1}} k_{-m_{l+1}} \mu_{l+1}(\cdots m_{l-1} m_l m_{l+1}) - \left(k_{-m_l} + \sum_n k_{+n} \right) \mu_l(\cdots m_{l-1} m_l) + k_{+m_l} \mu_{l-1}(\cdots m_{l-1}) = 0. \quad (\text{A8})$$

In the case there are two species of monomers, $M = 2$, Eqs. (A6) and (A7) form a system of four equations for the four unknown pair probabilities $\mu_2(mn)$ with $m, n = 1, 2$. This system can be solved and shows that the pair probabilities factorize

$$\mu_2(mn) = \mu_1(m) \mu_1(n), \quad (\text{A9})$$

so that the successive monomers are statistically independent of each other. In this case, Eq. (A6) combines with Eq. (A4) to become Eq. (28) for the velocity and moreover leads to Eq. (29) for the distribution $\mu_1(m)$ of monomers.

[1] M. Eigen and P. Schuster, *The hypercycle - a principle of natural self-organization* (Springer, Heidelberg, 1979).

- [2] M. Eigen, *Steps towards Life: A Perspective on Evolution* (Oxford University Press, Oxford, 1992).
- [3] D. R. Mills, R. L. Peterson, and S. Spiegelman, Proc. Natl. Acad. Sci. USA **58**, 217 (1967).
- [4] C. K. Biebricher and W. C. Gardiner, Biophysical Chemistry **66**, 179 (1997).
- [5] G. N. Lewis, Proc. Natl. Acad. Sci. USA **11**, 179 (1925).
- [6] P. W. Bridgman, Phys. Rev. **31**, 101 (1928).
- [7] P. Gaspard, J. Stat. Phys. **117**, 599 (2004).
- [8] P. Gaspard, C. R. Physique **8**, 598 (2007).
- [9] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, and A. Petrosyan, Phys. Rev. Lett. **98**, 150601 (2007).
- [10] D. Andrieux, P. Gaspard, S. Ciliberto, N. Garnier, S. Joubaud, and A. Petrosyan, J. Stat. Mech. P01002 (2008).
- [11] D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. USA **105**, 9516 (2008).
- [12] G. Nicolis, G. Subba Rao, J. Subba Rao, and C. Nicolis, Generation of spatially asymmetric, information-rich structures in far from equilibrium systems, in: P. L. Christiansen and R. D. Parmentier, Editors, *Structure, coherence and chaos in dynamical systems* (Manchester University Press, Manchester and New York, 1989) pp. 287-299.
- [13] C. Jarzynski, Proc. Natl. Acad. Sci. USA **105**, 9451 (2008).
- [14] A. Goldbeter and G. Nicolis, Biophysik **8**, 212 (1972).
- [15] G. Nicolis and C. Nicolis, *Foundations of Complex Systems* (World Scientific, Singapore, 2007).
- [16] J. Schnakenberg, Rev. Mod. Phys. **48**, 571 (1976).
- [17] G. Nicolis and I. Prigogine, *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
- [18] T. L. Hill, *Free Energy Transduction and Biochemical Cycle Kinetics* (Dover, New York, 2005).
- [19] P. Gaspard, J. Chem. Phys. **120**, 8898 (2004).
- [20] J.-L. Luo, C. Van den Broeck, and G. Nicolis, Z. Phys. B - Condensed Matter **56**, 165 (1984).
- [21] B. D. Coleman and T. G. Fox, J. Polym. Sci. A **1**, 3183 (1963).
- [22] B. D. Coleman and T. G. Fox, J. Chem. Phys. **38**, 1065 (1963).
- [23] C. E. Shannon, Bell System Tech. J. **27**, 379 (1948); *ibid.* **27**, 623 (1948).
- [24] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley, Hoboken, 2006).
- [25] G. J. Chaitin, Toward a mathematical definition of "life", in: R. D. Levine and M. Tribus, Editors, *The Maximum Entropy Formalism* (MIT Press, Cambridge, 1979) pp. 477-498.
- [26] D. T. Gillespie, J. Comp. Phys. **22**, 403 (1976).
- [27] M. Zwolak and M. Di Ventra, Rev. Mod. Phys. **80**, 141 (2008).
- [28] Y. Astier, O. Braha, and H. Bayley, J. Am. Chem. Soc. **128**, 1705 (2006).