Multistate reversible copolymerization of nonMarkovian chains under low conversion conditions

Pierre Gaspard
Center for Nonlinear Phenomena and Complex Systems,
Université Libre de Bruxelles (U.L.B.), Code Postal 231, Campus Plaine, B-1050 Brussels,
Belgium

The reversible kinetics of copolymerization is solved analytically for the multistate mechanism proposed by B. D. Coleman and T. G. Fox [J. Chem. Phys. 38, 1065 (1963)] under low conversion conditions where the concentrations of monomeric species are chemostatted and stay constant in time. Although the rates of this mechanism only depends on the currently attached or detached monomer, the growing macromolecular chain forms a nonMarkovian sequence that is characterized by matrices associated with every monomeric unit composing the sequence. These matrices are obtained by solving the kinetic equations and they determine the growth velocity of the copolymers, the statistical properties of its possible sequences, as well as the thermodynamics of the copolymerization process.

I. INTRODUCTION

In 1963, Coleman and Fox proposed a kinetic mechanism for the growth of polymers with nonMarkovian diastereosequence distributions. Their work concerns the tactority of polymers, i.e., the property that consecutive identical monomeric units composing a polymer may have isotactic or syndiotactic placements, possibly forming atactic sequences. In the proposed mechanism, the growing polymer has two possible reactive states, each with its own stereospecificity for the attachment of monomers. Moreover, the attachment rates are assumed to be independent of previously incorporated monomers. If the polymer always remained in a single reactive state during the growth, the sequence would form a Bernoulli chain. Therefore, the possibility for the reactive state of the growing polymer to randomly change between two or more reactive states modifies the statistical properties of the chain and allows correlations to appear between consecutive monomeric units, generating instead a nonMarkovian chain. In the fully irreversible growth regime where the detachment rates are negligible in front of the attachment rates, Coleman and Fox showed that the nonMarkovian chains generated by this multistate mechanism have sequence probability distributions given by products of matrices.

Now, the multistate mechanism may also be considered for copolymerization processes, in which different species of monomers are incorporated in the growing chain. Moreover, in order to investigate the thermodynamic properties of such processes, the reversed reactions of monomeric detachments should no longer be neglected, as previously studied by numerical simulations in Ref. [1]. In this context, the question arises whether the analytical methods developed by Coleman and Fox can be extended to reversible multistate processes.

The purpose of the present paper is to show that such an extension can indeed be achieved and that the sequence probability distributions are also given by matrix products in the general case of reversible multistate copolymerization processes under low conversion conditions. These latter are required in order to achieve the stationarity of monomer concentrations during the copolymer growth. If stationarity did not hold, the copolymer sequence would depend on the peculiar time variation undergone by the monomer concentrations in the solution, a situation encountered for instance in closed reactors where the population of monomers is depleted as the copolymers are growing until full equilibrium is reached. Instead, it is here supposed with Coleman and Fox that the monomer concentrations stay constant in time, which is the case under low conversion conditions or in an open reactor where the solution is continuously fed with monomers. The process is described using a stochastic approach at the level of a single copolymer. In this approach, the time evolution is considered for the probability to find the copolymer with given sequence, length, and reactive state. The time dependence of this probability is ruled by kinetic equations in terms of the transition rates associated with the different reactions in the process.

The content of the paper is the following.

In Section III the kinetic equations of multistate reversible copolymerization are introduced and solved in the long-time limit in terms of matrices of size equal to the number of reactive states in the mechanism. A matrix is associated with every species of monomeric units in the growing copolymer sequence, as well as with the stationary probability distribution of the different reactive states of the copolymer. The probability of any given sequence is thus obtained as a corresponding matrix product. The mean growth velocity is shown to be given in terms of a velocity matrix, satisfying a self-consistent equation, which plays a central role in the theory. Furthermore, these different matrices determine the probability of any given monomeric subsequence in the bulk of the whole sequence, as well as the mean value, the variance, and the correlation function of any observable quantity defined along the sequence. The results previously obtained in the fully irreversible regime by Coleman and Fox are recovered.

In Section III thermodynamics is developed for mul-
tistate reversible copolymerization in steady growth regimes. In particular, the expression of the entropy production rate is deduced and the consequences of the equilibrium detailed balance conditions are studied.

In Section IV theory is compared with simulations using Gillespie’s algorithm for two illustrative examples: on the one hand, an example where the thermodynamic equilibrium limit exists and, on the other hand, one without equilibrium.

Conclusion and perspectives are given in Section V.

II. KINETICS

A. The kinetic equations

We consider the Coleman-Fox multistate mechanism including the reversed reactions for the copolymerization of chains composed of different species of monomeric units \( m = 1, 2, ..., M \):

\[
\begin{align*}
  m_1 m_2 \cdots m_{l+1}^* + m_l & = m_1 m_2 \cdots m_{l-1} m_l^*, \quad (1) \\
  m_1 m_2 \cdots m_{l+1}^* + m_l & = m_1 m_2 \cdots m_{l-1} m_l^*, \quad (2) \\
  m_1 m_2 \cdots m_{l-1} m_l^* & = m_1 m_2 \cdots m_{l-1} m_l^*, \quad (3)
\end{align*}
\]

in the case of two reactive states, which are here below denoted by \( i = 1, 2 \). The transition rates of the reactions (1) and (2) are \( w_{i,m_i} \) and those of the transitions between the two states are \( w_{1 \rightarrow 2} \) and \( w_{2 \rightarrow 1} \). The attachment rates are proportional to the concentration \( c_m \) of the corresponding monomer in the solution, \( w_{i+m} = k_{i+m} c_m \), with some rate constant \( k_{i+m} \). Instead, the detachment rates, as well as the rates of transitions between the two states, do not depend on the monomer concentrations, \( w_{i-m} = k_{i-m} \) and \( w_{i-j} = k_{ij} \). Under low conversion conditions, we may assume that the monomer concentrations \( c_m \) and thus the rates stay constant in time.

The copolymerization of a single chain is described as a stochastic process in terms of the probabilities \( P_t(m_1 m_2 \cdots m_l, i) \) to find the polymer with the different possible sequences \( m_1 m_2 \cdots m_l \) of length \( l \) and states \( i \) at the time \( t \). These probabilities are ruled by the following infinite hierarchy of coupled kinetic equations:

\[
\frac{d}{dt} P_t(m_1 \cdots m_l, i) = w_{i+m} P_t(m_1 \cdots m_{l-1}, i) \\
+ \sum_{m_{l+1}=1}^{M} w_{i,m_{l+1}} P_t(m_1 \cdots m_{l+1}, i) \\
+ w_{j \rightarrow i} P_t(m_1 \cdots m_l, j) \\
- \left( w_{i-m} + \sum_{m_{l+1}=1}^{M} w_{i+m_{l+1}} + w_{i-j} \right) P_t(m_1 \cdots m_l, i),
\]

where \( j = 2 \) if \( i = 1 \) or \( j = 1 \) if \( i = 2 \). The total probability is conserved in time, \( \sum \omega P_t(\omega) = 1 \), where the sum extends over the sequences \( \omega = m_1 m_2 \cdots m_l \) and the states \( i \).

B. Solving the kinetic equations

The kinetic equations are solved by introducing the following set of probabilities:

\[
\begin{align*}
p_t(i,l) & = \sum_{m_1 \cdots m_l} P_t(m_1 \cdots m_{l-1} m_l, i), \quad (5) \\
p_t(m_1, i,l) & = \sum_{m_1 \cdots m_{l-1}} P_t(m_1 \cdots m_{l-1} m_l, i), \quad (6) \\
p_t(m_{l-1} m_l, i,l) & = \sum_{m_1 \cdots m_{l-2}} P_t(m_1 \cdots m_{l-2} m_{l-1} m_l, i), \quad (7)
\end{align*}
\]

Introducing the notation

\[
a_i \equiv \sum_{m} w_{i+m}, \quad (8)
\]

these probabilities obey the following equations as a consequence of the kinetic equations (4):

\[
\frac{d}{dt} p_t(i,l) = a_i p_t(i,l - 1) \\
+ \sum_{m_{l+1}} w_{i,m_{l+1}} p_t(m_{l+1}, i, l + 1) + w_{j \rightarrow i} p_t(j,l) \\
- \sum_{m_l} w_{i,m_l} p_t(m_l, i, l) - (a_i + w_{i-j}) p_t(i,l), \quad (9)
\]

\[
\frac{d}{dt} p_t(m_l, i,l) = w_{i,m_l} p_t(i,l - 1) \\
+ \sum_{m_{l+1}} w_{i,m_{l+1}} p_t(m_{l+1}, i, l + 1) + w_{j \rightarrow i} p_t(m_l, j,l) \\
- (w_{i-m} + a_i + w_{i-j}) p_t(m_l, i,l), \quad (10)
\]

and equations similar to Eq. (10) for the probabilities \( p_t(m_{l-1} m_l, i,l) \) and the next ones. We note that we recover Eq. (9) by summing Eq. (10) over \( m_l \), and similarly for the further equations in the hierarchy.

Since these equations are linear, their general solution can be written as a linear superposition of solutions of the form

\[
p_t(m_{l-r+1} \cdots m_{l-1} m_l, i) = \exp(s_q t + u q l) \chi_q(m_{l-r+1} \cdots m_{l-1} m_l, i)
\]

with an arbitrary parameter \( -\pi < q \leq +\pi \). The exponential rate \( s_q \) is here expected to have the following form,

\[
s_q = -q v - D q^2 + O(q^3),
\]

where \( v = \sqrt{-1} \), \( v \) is the mean growth velocity of the copolymer chain counted in monomers per second, and \( D \) is the diffusivity of the random drift of the length \( l \). The particular solutions (11) are substituted in Eqs. (9), (10), ... of the hierarchy. We carry out an expansion in
powers of \( q \) around \( q = 0 \) and use the fact that \( s_0 = 0 \) together with the notation
\[
\psi(m_{1-r+1} \cdots m_{l-1} m_l, i) = \chi_0(m_{1-r+1} \cdots m_{l-1} m_l, i).
\]
At order \( q^0 \), the following equations are obtained
\[
0 = -w_{i \rightarrow j} \psi(i) + w_{j \rightarrow i} \psi(j),
\]
\[
0 = w_{i \rightarrow m,i} \psi(i) + \sum_{m_{l+1}} w_{i \rightarrow m_{l+1}, m_{l+1}} \psi(m_{l+1}, i)
- (w_{i \rightarrow m_l} + a_i + w_{j \rightarrow i}) \psi(m_l, i) + w_{j \rightarrow i} \psi(m_l, j),
\]
and equations similar to Eq. (14) for the quantities (13) with the next values of \( r \).
At order \( q^1 \), Eq. (13) give
\[
- w \psi(i) = -a_i \psi(i) - \sum_m w_{i \rightarrow m} \psi(m, i)
- w_{i \rightarrow j} \psi'(i) + w_{j \rightarrow i} \psi'(j),
\]
where \( j = 2 \) if \( i = 1 \) or \( j = 1 \) if \( i = 2 \), \( m = m_{l+1} \), and \( \psi'(i) = d\chi_q(i)/dq_{i=0} \). Moreover, we suppose that the following normalization condition is satisfied:
\[
\psi(1) + \psi(2) = 1.
\]
Summing Eq. (16) over \( i = 1, 2 \), we obtain the relation giving the mean growth velocity
\[
v = \sum_i a_i \psi(i) - \sum_i \psi(m, i) w_{i \rightarrow m}
= \sum_i \psi(i) w_{i \rightarrow m} - \psi(m, i) w_{i \rightarrow m}.
\]

In order to solve the previous equations, we introduce a \( 2 \times 2 \) matrix describing the coupling between the pairs of equations due to the rates \( w_{i \rightarrow j} \) of the transitions between the reactive states
\[
W_0 \equiv \begin{pmatrix} -w_{1 \rightarrow 2} & w_{2 \rightarrow 1} \\ w_{1 \rightarrow 2} & -w_{2 \rightarrow 1} \end{pmatrix},
\]
as well as other \( 2 \times 2 \) matrices with the rates of the attachment and detachment reactions
\[
W_{\pm m} \equiv \begin{pmatrix} w_{1 \pm m} & 0 \\ 0 & w_{2 \pm m} \end{pmatrix},
\]
together with
\[
A \equiv \sum_m W_{\mp m}.
\]
Furthermore, we also define the \( 2 \times 2 \) matrices
\[
\Psi \equiv \begin{pmatrix} \psi(1) & \psi(1) \\ \psi(2) & \psi(2) \end{pmatrix},
\]
\[
\Psi(m_l) \equiv \begin{pmatrix} \psi(m_l, 1) & \psi(m_l, 1) \\ \psi(m_l, 2) & \psi(m_l, 2) \end{pmatrix},
\]
\[
\Psi(m_{l-1} m_l) \equiv \begin{pmatrix} \psi(m_{l-1} m_l, 1) & \psi(m_{l-1} m_l, 1) \\ \psi(m_{l-1} m_l, 2) & \psi(m_{l-1} m_l, 2) \end{pmatrix},
\]
As before, we have that
\[
\sum_{m_l} \Psi(m_l) = \Psi, \]
\[
\sum_{m_{l-1}} \Psi(m_{l-1} m_l) = \Psi(m_l),
\]
\[
\vdots
\]
With these definitions, Eqs. (13), (15), ... of the hierarchy can be rewritten in the following matricial form:
\[
0 = W_0 \cdot \Psi,
\]
\[
0 = W_{+m} \cdot \Psi + \sum_{m_{l+1}} W_{-m_{l+1}} \cdot \Psi(m_{l+1})
- (A + W_{-m_l} - W_0) \cdot \Psi(m_l),
\]
\[
0 = W_{+m_l} \cdot \Psi(m_{l-1}) + \sum_{m_{l+1}} W_{-m_{l+1}} \cdot \Psi(m_{l-1} m_{l+1})
- (A + W_{-m_l} - W_0) \cdot \Psi(m_{l-1} m_l),
\]
\[
\vdots
\]
First, we note that Eqs. (14) and thus (21) can be solved with
\[
\psi(1) = \frac{w_{2 \rightarrow 1}}{w_{1 \rightarrow 2} + w_{2 \rightarrow 1}},
\]
\[
\psi(2) = \frac{w_{1 \rightarrow 2}}{w_{1 \rightarrow 2} + w_{2 \rightarrow 1}},
\]
satisfying the normalization condition (17).
Next, we assume that the following matricial factorization holds,
\[
\Psi(m_{l-r+1} \cdots m_{l-1} m_l) = Y_{m_l} \cdot Y_{m_{l-1}} \cdots Y_{m_{l-r+1}} \cdot \Psi,
\]
in terms of \( 2 \times 2 \) matrices \( Y_m \) to be determined. This factorization is suggested by the results of Coleman and Fox in the fully irreversible regime. Now, substituting the assumption into the equations (28), ... of the hierarchy, we observe that they can all be solved if the matrices \( Y_m \) satisfy the following relation,
\[
0 = W_{+m} + \sum_{m'} W_{-m'} Y_{m'} \cdot Y_{m} - (A + W_{-m} - W_0) \cdot Y_{m}.
\]
Introducing the \( 2 \times 2 \) matrix
\[
V \equiv A - \sum_m W_{-m} \cdot Y_m = \sum_m (W_{+m} - W_{-m} \cdot Y_m),
\]
we see that the solution of Eq. (33) can be written in the form
\[
Y_m = (V - W_0 + W_{-m})^{-1} \cdot V_{+m}.
\]
Inserting Eq. (35) back into Eq. (34), we obtain the self-consistent matrix equation
\[
V = (V - W_0) \cdot \sum_m (V - W_0 + W_{-m})^{-1} \cdot W_{+m}.
\]
that can be solved by direct numerical iteration for the given values of the rates. Finally, the mean growth velocity \[ \nu = \text{tr}(\mathbf{V} \cdot \mathbf{\Psi}) \] (37)

with tr denoting the trace of $2 \times 2$ matrices. In this regard, the matrix (34) is called the velocity matrix. Once this latter is obtained by solving the self-consistent equation (36), we can get the matrices (35), which determine the composition of the copolymer sequences.

Besides, the diffusivity $D$ can be calculated similarly at next order $q^2$. According to the central limit theorem, the length probability distribution will thus be given in the long-time limit by the following Gaussian distribution

$$ p_l(l) \simeq \frac{1}{\sqrt{4\pi D l}} \exp \left[ -\frac{(l-vt)^2}{4Dt} \right]. \quad (38) $$

Therefore, the solution of the coupled kinetic equations can be expressed for $t \to \infty$ as

$$ P_l(m_1 \cdots m_l, i) \simeq p_l(l) \psi(m_1 \cdots m_l, i) \quad (39) $$

in terms of Eq. (38) and the stationary probability distribution

$$ \psi(m_1 \cdots m_l, i) = \sum_{i_{1 \cdots i_l}} \langle \mathbf{Y}_{m_1} \rangle_{i_{1 \cdots i_l}} \cdots \langle \mathbf{Y}_{m_l} \rangle_{i_{1 \cdots i_l}} \langle \mathbf{\Psi} \rangle_{i_{1 \cdots i_l}} \quad (40) $$

to find the sequence $m_1 \cdots m_l$ of length $l$ in the reactive state $i$. The normalization condition

$$ \sum_{m_1 \cdots m_l} \psi(m_1 \cdots m_l, i) = 1 \quad (41) $$

is satisfied because of Eqs. (17), (24), (26), ... Moreover, the stationary probability to find the sequence $m_1 \cdots m_l$ of length $l$ in any reactive state is given by the expression

$$ \mu(m_1 \cdots m_l) = \sum_i \psi(m_1 \cdots m_l, i) $$

$$ = \text{tr} \mathbf{\Psi}(m_1 \cdots m_l) = \text{tr} \langle \mathbf{Y}_{m_1} \cdots \mathbf{Y}_{m_l} \cdot \mathbf{\Psi} \rangle. \quad (42) $$

Since these probabilities cannot be factorized into conditional probabilities as for Markov chains, we conclude that the sequences are nonMarkovian.

All these results can be extended to mechanisms with more than two reactive states, replacing the $2 \times 2$ matrix by corresponding $I \times I$ matrices where $I$ is the number of reactive states.

We note that the results of Ref. 4 are recovered if $w_{i \to j} = 0$, as shown in Appendix A.

C. Mean, variance, and correlation function

In order to satisfy Eqs. (24), (26), ..., we have the property that

$$ \mathbf{R} \cdot \mathbf{\Psi} = \mathbf{\Psi} \quad \text{for} \quad \mathbf{R} \equiv \sum_m \mathbf{Y}_m, \quad (43) $$

implying

$$ \mathbf{R}^n \cdot \mathbf{\Psi} = \mathbf{\Psi} \quad \forall n \in \mathbb{N}. \quad (44) $$

Because of the definition (22), the vector

$$ \mathbf{\xi}_1 \equiv \left( \begin{array}{c} \psi(1) \\ \psi(2) \end{array} \right) \quad (45) $$

is the right-eigenvector of the $2 \times 2$ matrix $\mathbf{R}$ associated with the eigenvalue $\Lambda_1 = 1$. The corresponding left-eigenvector is denoted by the vector $\mathbf{\eta}_1$. We thus have that

$$ \mathbf{R} = \mathbf{\xi}_1 \mathbf{\Lambda}_1^{T} + \mathbf{\xi}_2 \mathbf{\Lambda}_2 \mathbf{\eta}_2^{T} \quad (46) $$

with a second eigenvalue $|\Lambda_2| < 1$ and the biorthonormality condition $\mathbf{\eta}_1^{T} \cdot \mathbf{\xi}_\beta = \delta_{\alpha\beta}$, where the superscript $T$ denotes the transpose.

The mean value of some function $f(m)$ of the monomeric unit $m$ at the location $k$ of the chain is defined by

$$ \langle f(m_k) \rangle = \sum_{m_1 \cdots m_{k-1} \cdots m_l} f(m_k) \mu(m_1 \cdots m_{k-1} \cdots m_l) \quad (47) $$

Using Eqs. (12) and (13), we get

$$ \langle f(m_k) \rangle = \text{tr} \left( \mathbf{R}^{l-k} \cdot \mathbf{F} \cdot \mathbf{R}^{k-1} \cdot \mathbf{\Psi} \right) = \text{tr} \left( \mathbf{R}^{l-k} \cdot \mathbf{F} \cdot \mathbf{\Psi} \right) \quad (48) $$

where

$$ \mathbf{F} \equiv \sum_m f(m) \mathbf{Y}_m, \quad (49) $$

so that the mean value is given by

$$ \langle f \rangle = \lim_{l-k \to \infty} \langle f(m_k) \rangle = \mathbf{\eta}_1^{T} \cdot \mathbf{F} \cdot \mathbf{\xi}_1. \quad (50) $$

The variance $\text{Var}(f) \equiv \langle f^2 \rangle - \langle f \rangle^2$ can be evaluated similarly.

Besides, the correlation function of the function $f$ is defined by

$$ \Gamma(n) \equiv \lim_{l-k \to \infty} \left[ \langle f(m_k) f(m_{k+n}) \rangle - \langle f \rangle^2 \right], \quad (51) $$

where

$$ \langle f(m_k) f(m_{k+n}) \rangle = \sum_{m_1 \cdots m_{k-n} \cdots m_l} f(m_k) f(m_{k+n}) $$

$$ \times \mu(m_1 \cdots m_{k-n} \cdots m_l \cdots m_l), \quad (52) $$

giving $\langle f^2 \rangle$ for $n = 0$, and

$$ \langle f(m_k) f(m_{k+n}) \rangle = \text{tr} \left( \mathbf{R}^{l-k-n} \cdot \mathbf{F} \cdot \mathbf{R}^{n-1} \cdot \mathbf{F} \cdot \mathbf{\Psi} \right) \quad (53) $$

for $n > 0$. Consequently, we have that

$$ \lim_{l-k \to \infty} \langle f(m_k) f(m_{k+n}) \rangle = (\mathbf{\eta}_1^{T} \cdot \mathbf{F} \cdot \mathbf{\xi}_1)^2 $$

$$ + \Lambda_2^{-1} (\mathbf{\eta}_1^{T} \cdot \mathbf{F} \cdot \mathbf{\xi}_2) (\mathbf{\eta}_2^{T} \cdot \mathbf{F} \cdot \mathbf{\xi}_1). \quad (54) $$
Therefore, the correlation function is given by \( \Gamma(0) = \text{Var}(f) \) for \( n = 0 \), and
\[
\Gamma(n) = \Lambda_2^{n-1} (\eta_1^T \cdot F \cdot \xi_2) (\eta_2^T \cdot F \cdot \xi_1)
\] (55)
for \( n > 0 \). Since \( |\Lambda_2| < 1 \), the correlation function decreases exponentially as \( \Gamma(n) \sim \exp(-\gamma n) \) for \( n \to \infty \) with the rate \( \gamma = -\ln |\Lambda_2| \). In the following, we shall use the normalized correlation function, \( C(n) \equiv \Gamma(n)/\Gamma(0) \).

**D. Bulk probabilities**

The bulk probability \( \bar{\mu}(m) \) to find the monomeric unit \( m \) anywhere inside the grown copolymer chain can be obtained as the mean value of the indicator function of the monomeric unit \( m \) given by the corresponding Kronecker symbol: \( f(m_k) = \sigma_m(m_k) = \delta_{m,m_k} \). Therefore, we get
\[
\bar{\mu}(m) = \langle \sigma_m \rangle = \eta_1^T \cdot Y_m \cdot \xi_1.
\] (56)
Similarly, the bulk probability to find the subsequence \( m_1 \cdots m_k \) anywhere in the chain is given by
\[
\bar{\mu}(m_1 \cdots m_k) = \langle \sigma_{m_1} \cdots \sigma_{m_k} \rangle = \eta_1^T \cdot Y_{m_{k}} \cdots Y_{m_{m}} \cdot \xi_1.
\] (57)
If we introduce the \( 2 \times 2 \) matrix
\[
\Upsilon = \xi_1 \eta_1^T,
\] (58)
Eq. (57) reads
\[
\bar{\mu}(m_1 \cdots m_k) = \text{tr} (Y_{m_k} \cdots Y_{m_1} \cdot \Upsilon).
\] (59)
In general, neither the tip probabilities (42), nor the bulk probabilities (56) have an expression that factorizes as for Bernoulli or Markov chains, which confirms the non-Markovian character of the macromolecular chains yielded by multistate copolymerization processes.

**E. Fully irreversible regime**

This regime has been studied in Ref. [1] in the context of polymerization with isotactic and syndiotactic placements, corresponding respectively to \( m = I \) and \( m = S \). Fully irreversible regimes are defined by supposing that the detachment rates are negligible in front of the attachment rates:
\[
\mathcal{W}_{-m} = 0.
\] (60)
Accordingly, the solution of Eq. (36) is given by
\[
\mathbf{V} = \sum_m \mathcal{W}_{+m}
\] (61)
and the mean growth velocity (37) is thus equal to
\[
v = \frac{w_{2 \to 1}}{w_{1 \to 2} + w_{2 \to 1}} \ a_1 + \frac{w_{1 \to 2}}{w_{1 \to 2} + w_{2 \to 1}} \ a_2,
\] (62)
which reads
\[
v = \frac{\lambda_a}{\lambda_a + \lambda_b} \ k_1 [M] + \frac{\lambda_b}{\lambda_a + \lambda_b} \ k_2 [M],
\] (63)
with the notations \( k_i = \sum_{m} k_{i,m} \) \( (i = 1, 2) \), \( \lambda_a = w_{2 \to 1} \), \( \lambda_b = w_{1 \to 2} \), and the monomeric concentration \( c_m = [M] \), showing the equivalence with Eqs. (2.4a)-(2.5) of Ref. [1]. According to Eq. (35), we also have that
\[
\mathbf{Y}_m = (\mathbf{V} - \mathbf{W}_0)^{-1} \cdot \mathbf{W}_m.
\] (64)
In order to compare with the related results of Ref. [1], we introduce the transformation
\[
\mathbf{T} \equiv \begin{pmatrix} w_{2 \to 1} & 0 \\ 0 & w_{1 \to 2} \end{pmatrix} = \begin{pmatrix} \lambda_a & 0 \\ 0 & \lambda_b \end{pmatrix},
\] (65)
such that
\[
\mathbf{W}_0 = \mathbf{T} \cdot \mathbf{W}_0^T \cdot \mathbf{T}^{-1}.
\] (66)
The correspondence with the matrices used in Ref. [1] is established according to
\[
\mathbf{\Phi} \equiv \mathbf{T} \cdot \mathbf{\Upsilon}^T \cdot \mathbf{T}^{-1},
\] (67)
\[
\mathbf{X}_m = \mathbf{T} \cdot \mathbf{\Upsilon}_m \cdot \mathbf{T}^{-1},
\] (68)
for \( m \in \{I, S\} \). A direct calculation using Mathematica shows that
\[
\mathbf{\Phi} \equiv \begin{pmatrix} \phi_1 & \phi_1 \\ \phi_2 & \phi_2 \end{pmatrix},
\] (69)
where
\[
\phi_1 = \frac{k_1 \lambda_a}{k_1 \lambda_a + k_2 \lambda_b}, \quad \phi_2 = \frac{k_2 \lambda_b}{k_1 \lambda_a + k_2 \lambda_b},
\] (70)
with \( k_i = k_{i,+1} + k_{i,+S}, \) which corresponds to Eq. (2.7) of Ref. [1], and
\[
\begin{aligned}
\mathbf{X}_m \downarrow 1 & = \frac{k_{1,+m} \lambda_a + k_{2}[M]}{k_{1} \lambda_a + k_2 \lambda_b + k_1 k_2 [M]}, \\
\mathbf{X}_m \downarrow 2 & = \frac{k_{2,+m} \lambda_b}{k_{1} \lambda_a + k_2 \lambda_b + k_1 k_2 [M]}, \\
\mathbf{X}_m \uparrow 1 & = \frac{k_{1,+m} \lambda_a}{k_{1} \lambda_a + k_2 \lambda_b + k_1 k_2 [M]}, \\
\mathbf{X}_m \uparrow 2 & = \frac{k_{2,+m} \lambda_b}{k_{1} \lambda_a + k_2 \lambda_b + k_1 k_2 [M]},
\end{aligned}
\] (71)-(74)
for \( m \in \{I, S\} \), which corresponds to Eqs. (3.9a)-(3.9d) of Ref. [1]. Therefore, the results of Ref. [1] are precisely recovered in the fully irreversible regime.

III. THERMODYNAMICS

A. Entropy production

For isothermal-isobaric stochastic processes ruled by kinetic equations
\[
\frac{d}{dt} P_t(\omega) = \sum_{\omega' \neq \omega} [P_t(\omega') W(\omega' \to \omega) - P_t(\omega) W(\omega \to \omega')],
\] (75)
where \( W(\omega \to \omega') \) are the transition rates between the coarse-grained states \( \omega \) and \( \omega' \), the link to thermodynamics is established by using the relations

\[
\frac{W(\omega \to \omega')}{W(\omega' \to \omega)} = e^{\beta[G(\omega) - G(\omega')]}
\]

(76)
giving the ratio of the rates of opposite transitions in terms of Gibbs’ free energies \( G(\omega) \) associated with the coarse-grained states \( \omega \), while \( \beta = (k_B T)^{-1} \) is the inverse temperature expressed in terms of the temperature \( T \) and Boltzmann’s constant \( k_B \). Gibbs’ free energy \( G(\omega) \) is related to the enthalpy \( H(\omega) \) and the entropy \( S(\omega) \) of the coarse-grained state \( \omega \) by \( G(\omega) = H(\omega) - T S(\omega) \). Furthermore, the overall thermodynamic entropy of the statistical sample described by the probability distribution \( P_t(\omega) \) is defined by

\[
S_t = \sum_{\omega} P_t(\omega) [S(\omega) - k_B \ln P_t(\omega)].
\]

(77)

In general, the time derivative \( dS_t/dt \) of the entropy can be separated into the rate of entropy exchange with the environment of the growing copolymer

\[
d_eS = \frac{1}{T} \frac{d\langle H \rangle_t}{dt}
\]

(78)
expressed in terms of the mean enthalpy \( \langle H \rangle_t = \sum_{\omega} P_t(\omega) H(\omega) \), and the rate of entropy production

\[
\frac{dS_t}{dt} - \frac{d_eS}{dt} = -\frac{1}{T} \frac{d\langle G \rangle_t}{dt} - k_B \frac{d}{dt} \sum_{\omega} P_t(\omega) \ln P_t(\omega) \geq 0,
\]

(79)
which is always non-negative in accordance with the second law of thermodynamics.

These considerations apply in particular to the kinetic equations (1) where the coarse-grained state \( \omega \) is defined by the sequence \( m_1 \cdots m_l \) in the reactive state \( i \).

In the regime of steady growth where the mean growth velocity is positive \( v > 0 \), the entropy production rate (79) can be evaluated using the probability distribution (39) that is the solution of the kinetic equations in the long-time limit (110). On the one hand, the mean value of the Gibbs free energy is given by \( \langle G \rangle_t \simeq \langle f \rangle_t g \) in terms of the mean length \( \langle l \rangle_t \simeq vt \), so that

\[
\frac{d\langle G \rangle_t}{dt} = v g
\]

(80)
with the mean Gibbs free energy per monomeric unit

\[
g = \lim_{l \to \infty} \frac{1}{l} \sum_{m_1 \cdots m_l} \psi(m_1 \cdots m_l, i) G(m_1 \cdots m_l, i).
\]

(81)

On the other hand, Eq. (39) also implies that

\[
-\frac{d}{dt} \sum_{\omega} P_t(\omega) \ln P_t(\omega) = v D
\]

(82)
with the sequence disorder per monomeric unit

\[
D \equiv \lim_{l \to \infty} -\frac{1}{l} \sum_{m_1 \cdots m_l} \psi(m_1 \cdots m_l, i) \ln \psi(m_1 \cdots m_l, i),
\]

(83)
which is always a non-negative quantity, \( D \geq 0 \). Therefore, the entropy production rate can be expressed as

\[
\frac{dS}{dt} = k_B v A \geq 0,
\]

(84)
with the dimensionless entropy production per monomeric unit, called the affinity,

\[
A \equiv \varepsilon + D,
\]

(85)
the dimensionless free-energy driving force defined in terms of Eq. (83) and the thermal energy \( k_B T \) as

\[
\varepsilon \equiv -\frac{\varepsilon}{k_B T},
\]

(86)
and the sequence disorder defined by Eq. (83).

Computing the time derivative of the mean Gibbs free energy \( \langle G \rangle_t \) with the kinetic equations (75), we get

\[
\frac{d\langle G \rangle_t}{dt} = \sum_{\omega \neq \omega'} P_t(\omega) W(\omega \to \omega') [G(\omega') - G(\omega)]
\]

\[
= -k_B T \sum_{\omega \neq \omega'} P_t(\omega) W(\omega \to \omega') \ln \frac{W(\omega \to \omega')}{W(\omega' \to \omega)},
\]

(87)
where we used Eq. (76). According to Eqs. (39) and (80), we find that the free-energy driving force is here given by

\[
\varepsilon = \frac{1}{v} \sum_{m, i} \psi(i) w_{i, m} - \psi(m, i) w_{i, -m} \ln \frac{w_{i, +m}}{w_{i, -m}}.
\]

(88)
We notice that the velocity is similarly expressed by Eq. (113).

If the detachment rates are negligible in front of the attachment rates, the free-energy driving force (88) is arbitrarily large, so that the entropy production rate becomes infinite in the fully irreversible regime.

B. The equilibrium limit

The state of thermodynamic equilibrium is identified by the principle of detailed balance, which requires that

\[
w_{1 \to 2} P_{eq}(m_1 \cdots m_l, 1) = w_{2 \to 1} P_{eq}(m_1 \cdots m_l, 2),
\]

(89)
\[
w_{1, +m} P_{eq}(m_1 \cdots m_l - 1, 1) = w_{1, -m} P_{eq}(m_1 \cdots m_l - m, 1),
\]

(90)
\[
w_{2, +m} P_{eq}(m_1 \cdots m_l - 1, 2) = w_{2, -m} P_{eq}(m_1 \cdots m_l - m, 2).
\]

(91)
This imposes the following constraints on the attachment and detachment rates:

\[
\frac{w_{1, +m}}{w_{1, -m}} = \frac{w_{2, +m}}{w_{2, -m}}
\]

(92)
for all \( m = 1, 2, \ldots, M \). If the attachment rates are proportional to the corresponding monomeric concentrations according to the mass action law, we thus have that

\[
\frac{w_{i,+m}}{w_{i,-m}} = \frac{k_{i,+m}}{k_{i,-m}} = e^{\varepsilon^m}
\]  

(93)

for all \( m = 1, 2, \ldots, M \), independently of the reactive state \( i \). These constraints must be satisfied for the existence of a thermodynamic equilibrium limit.

Now, taking \( l = 1 \) and \( m = m_i \) in Eqs. (90)–(91), and using Eq. (93), we get

\[
\psi_{\text{eq}}(m, i) = e^{\varepsilon^m} \psi_{\text{eq}}(i)
\]  

(94)

for \( i = 1, 2 \). Therefore, the equilibrium tip probabilities are given by

\[
\mu_{\text{eq}}(m) = \sum_i \psi_{\text{eq}}(m, i) = e^{\varepsilon^m}
\]  

(95)

because of Eq. (17). Since the probability distribution (95) should also be normalized to unity, we must have

\[
\sum_{m=1}^{M} e^{\varepsilon^m} = 1 \quad \text{at equilibrium.}
\]  

(96)

Moreover, Eq. (92) for \( r = 1 \) gives \( \psi(m, i) = \sum_j (Y_m)_{ij} \psi(j) \), while the probabilities \( \psi(i) \) define the vector \( \xi_1 \), so that

\[
Y_m \cdot \xi_1 = e^{\varepsilon^m} \xi_1 \quad \text{at equilibrium.}
\]  

(97)

As a consequence, the sequence probability (92) factorizes as

\[
\mu_{\text{eq}}(m_1 \cdots m_l) = \prod_{k=1}^{l} \mu_{\text{eq}}(m_k),
\]  

(98)

showing that the sequences form a Bernoulli chain in the special limit where the equilibrium conditions (99)–(11) are satisfied. Because of Eq. (90), the equilibrium bulk probabilities of the monomeric units are also given by

\[
\bar{\mu}_{\text{eq}}(m) = e^{\varepsilon^m}.
\]  

(99)

Furthermore, the mean growth velocity is equal to zero in the equilibrium limit. Indeed, Eq. (99) has the matrix form \( \mathbf{W}_+ = e^{\varepsilon^m} \mathbf{W}_- \), so that Eq. (54) implies that \( \mathbf{V} : \Psi = 0 \) at equilibrium and the mean growth velocity (57) is thus vanishing. Accordingly, we find that the equilibrium free-energy driving force (88) can be expressed as

\[
\varepsilon_{\text{eq}} = \sum_{m=1}^{M} \varepsilon_m \bar{\mu}_{\text{eq}}(m)
\]  

(100)

in terms of the equilibrium bulk probabilities (99). The equilibrium value of the affinity (57) is also equal to zero, which implies that the equilibrium value of the sequence disorder is given by

\[
D_{\text{eq}} = -\varepsilon_{\text{eq}},
\]  

(101)

in agreement with the result that the copolymer is a Bernoulli chain in the equilibrium limit. As required, the entropy production rate is thus equal to zero at equilibrium.

Besides, the correlation function (55) reduces to \( \Gamma(n) = \text{Var}(f) \delta_{n,0} \) in the equilibrium limit. Indeed, we have that \( \bar{\eta}_1^T \cdot \mathbf{F} \cdot \xi_1 = \langle f \rangle_{\text{eq}} \bar{\eta}_1^T \cdot \xi_1 = 0 \) where \( \langle f \rangle_{\text{eq}} = \sum_m f(m) \bar{\mu}_{\text{eq}}(m) \), because of Eq. (97) and the biorthonormality condition \( \bar{\eta}_1^T \cdot \xi_1 = 0 \) between the left- and right-eigenvectors associated with different eigenvalues. There is thus no statistical correlation in the sequences, as expected since they form Bernoulli chains in the equilibrium limit.

IV. ILLUSTRATIVE EXAMPLES

In the following examples, \( M = 2 \) monomeric species \( A \) for \( m = 1 \) and \( B \) for \( m = 2 \) are considered. The theoretical predictions are calculated with the methods of Appendix B and the growth is simulated using Gillespie’s algorithm (11,12) as described in Appendix C.

A. Example with thermodynamic equilibrium

The parameters of this example are taken as

\[
k_{1,+A} = 2, \quad k_{1,+B} = 4, \quad k_{2,+A} = 4, \quad k_{2,+B} = 2,
\]

\[
k_{1,-A} = 1, \quad k_{1,-B} = 6, \quad k_{2,-A} = 2, \quad k_{2,-B} = 3,
\]

\[
k_{12} = 1, \quad k_{21} = 2, \quad c_B = 1.
\]  

(102)

Here, the concentration \( c_A \) is the nonequilibrium control parameter. This set of parameter values satisfies the condition (53) for the existence of an equilibrium limit at \( c_A = 1/6 \). Accordingly, the thermodynamic equilibrium limit is reached at the concentration where the mean growth velocity vanishes.

For this example, the tip probabilities (40) have been computed using theory and simulations. The results are compared at the concentration value \( c_A = 5 \) in Table II, showing the tip probabilities \( \psi(i), \psi(m_1, i), \psi(m_{l-1}m_l, i) \), and \( \psi(m_{l-2}m_{l-1}m_l, i) \). On the one hand, these probabilities are obtained in theory by solving Eq. (92) by iterations to get the velocity matrix that is next used to find the matrices (68) and thus the tip probabilities (40). On the other hand, the growth is simulated during the time lapse \( l = 100 \) with Gillespie’s algorithm to generate a sample of \( 10^7 \) sequences. With this sample, the probabilities \( p_l(i, l), p_l(m_1, i, l), p_l(m_{l-1}m_l, i, l) \), and \( p_l(m_{l-2}m_{l-1}m_l, i, l) \) are first computed by statistics to next obtain the tip probabilities by summing over the
TABLE I. Growth of a copolymer in the conditions (102) with $c_A = 5$: Comparison between the values of the tip probabilities [40] obtained by theory with $10^4$ iterations of Eq. (36) and numerically by Gillespie’s algorithm generating $10^6$ copolymer sequences, each of total time $t = 100$. Under the chosen conditions, the mean growth velocity is $v = 14.58$, the free-energy driving force $\varepsilon = 1.8623$, the sequence disorder $D = 0.4534$, and the affinity $A = \varepsilon + D = 2.3157$. The mean length of the sequences is thus equal to $\langle l \rangle = vt = 1458$. The same simulation data are used for Fig. 1. The numbers are rounded off to five decimal digits.

| $m_{-2}$ | $m_{-1}$ | $m_1$ | $i$ | theory | simulation |
|-----------|-----------|-------|-----|--------|------------|
| 1         | 0.66667   | 0.66671 |
| 2         | 0.33333   | 0.33329 |
| A         | 1         | 0.52263 | 0.52262 |
| B         | 1         | 0.14403 | 0.14409 |
| A         | 2         | 0.30048 | 0.30048 |
| B         | 2         | 0.03285 | 0.03281 |
| A A       | 1         | 0.41483 | 0.41494 |
| B A       | 1         | 0.10781 | 0.10768 |
| A B       | 1         | 0.11328 | 0.11331 |
| B B       | 1         | 0.03076 | 0.03078 |
| A A       | 2         | 0.26832 | 0.26832 |
| B A       | 2         | 0.03217 | 0.03216 |
| A B       | 2         | 0.02889 | 0.02888 |
| B B       | 2         | 0.00396 | 0.00393 |
| A A A     | 1         | 0.33315 | 0.33337 |
| B A A     | 1         | 0.08168 | 0.08158 |
| A B A     | 1         | 0.08518 | 0.08512 |
| B B A     | 1         | 0.02262 | 0.02257 |
| A A B     | 1         | 0.09018 | 0.09020 |
| A B B     | 1         | 0.02309 | 0.02311 |
| A B B     | 1         | 0.02422 | 0.02422 |
| B B B     | 1         | 0.00654 | 0.00655 |
| A A A     | 2         | 0.23791 | 0.23788 |
| B A A     | 2         | 0.03041 | 0.03044 |
| A B A     | 2         | 0.02785 | 0.02786 |
| B B A     | 2         | 0.00431 | 0.00430 |
| A A B     | 2         | 0.02532 | 0.02534 |
| B A B     | 2         | 0.00357 | 0.00355 |
| A B B     | 2         | 0.00336 | 0.00333 |
| B B B     | 2         | 0.00060 | 0.00059 |

Different values of the length $l$ to get $\psi(i) = \sum_l p_i(i, l)$, $\psi(m_l, i) = \sum_l p_l(m_l, i, l)$, etc. In Table I, we see the excellent agreement between theory and simulations in this example. The non-Markovian character of these probability distribution is also confirmed by these results. For the same data, Fig. 1 shows the probabilities $p_i(i, l)$ and $p_l(m_l, i, l)$ as a function of the length $l$ in comparison with the theoretical prediction [48-50]. The probability distributions should be Gaussian after a long enough time according to the central limit theorem. Here also, agreement is observed between theory (lines) and simulations (data points).

Furthermore, the different quantities of interest have been investigated as a function of the concentration $c_A$.

Figure 2 depicts the mean growth velocity [57] and the bulk probability [56] to find the monomeric unit $A$ anywhere inside the grown chain versus the concentration $c_A$. Again, the data points show the simulation results with Gillespie’s algorithm and the lines the theoretical predictions. As the concentration $c_A$ increases, the composition of the copolymer in monomeric units $A$ also increases, as seen in Fig. 2.
by Eq. \( \text{[54]} \) in units where \( k_B = 1 \).

In this example, the velocity is vanishing at the equilibrium concentration \( c_A = 1/6 \). At this special value where the equilibrium detailed balance conditions \( \text{[50]}-\text{[51]} \) are satisfied, the copolymer sequence forms a Bernoulli chain according to Eq. \( \text{[52]} \) where the tip probabilities \( \text{[53]} \) take the values \( \mu(A) = e^{\gamma A} = 1/3 \) and \( \mu(B) = e^{\gamma B} = 2/3 \), since \( \varepsilon_A = \ln(1/3) \) and \( \varepsilon_B = \ln(2/3) \) for the parameter values \( \text{[52]} \). As a consequence of Eq. \( \text{[53]} \), the bulk probabilities take the same values, which is confirmed by the value \( \bar{\mu}(A) = 1/3 \) observed in Fig. 2 at the concentration \( c_A = 1/6 \) where \( v = 0 \). Therefore, the equilibrium values of the free-energy driving force \( \text{[100]} \) and the sequence disorder \( \text{[101]} \) are here equal to \( D_{\text{eq}} = -\varepsilon_{\text{eq}} = \ln(3/2^{2/3}) = 0.63651 \). The affinity and the entropy production rate are thus vanishing at the concentration \( c_A = 1/6 \), as required if equilibrium is reached.

In Fig. 3 the free-energy driving force is vanishing \( (\varepsilon = 0) \) at the concentration \( c_A = 0.70485 \), where \( v = 2.1147 \), \( A = D = 0.68956 \), and \( dS/dt = \gamma D = 1.4582 \). At this concentration, entropy is only produced due to sequence disorder. In the concentration range \( 1/6 < c_A < 0.70485 \), the free-energy driving force is negative, so that the copolymer growth is driven by the entropic effect of sequence disorder. For \( c_A > 0.70485 \), the free-energy driving force is positive and the growth is driven by free energy. For large values of the concentration, the free-energy driving force becomes dominant over sequence disorder: \( \varepsilon \gg D \). In this strongly irreversible regime where the attachment rates dominate, the entropy production rate can be evaluated with Eq. \( \text{[58]} \) as

\[
\frac{dS}{dt} \approx v \varepsilon \approx \sum_{m,i} \psi(i) k_{i,m} c_m \ln \frac{k_{i,m} c_m}{k_{i,-m}},
\]

in units where \( k_B = 1 \). For \( c_A \gg c_B \) in the example \( \text{[102]} \), the entropy production rate increases as \( dS/dt \approx (8 c_A/3) \ln(2 c_A) \) with the concentration \( c_A \), as seen in Fig. 3.

Figure 4 compares the normalized correlation function \( C(n) = \Gamma(n)/\Gamma(0) \) of the quantity

\[
f(m) = \begin{cases} 1 & \text{if } m = A, \\ 2 & \text{if } m = B, \end{cases}
\]

calculated in theory with Eq. \( \text{[53]} \) in terms of the eigenvalue \( \Lambda_2 \) of the matrix \( \text{[3]} \) and in simulations with Eq. \( \text{[4]} \), showing good agreement. For \( n = 0 \), the normalized correlation function is equal to unity. For \( n > 0 \), the correlation function drops to significantly lower values and it decays exponentially at the rate \( \gamma = -\ln \Lambda_2 \). As observed in Fig. 4, the correlation function decays faster and faster as the concentration \( c_A \) decreases, which is consistent with the absence of statistical correlations in the equilibrium limit where the sequences form Bernoulli chains.
B. Example without thermodynamic equilibrium

The parameters of this example are taken as

\[ k_{1,+A} = 1, \quad k_{1,+B} = 2, \quad k_{2,+A} = 4, \quad k_{2,+B} = 1, \]
\[ k_{1,-A} = 5, \quad k_{1,-B} = 3, \quad k_{2,-A} = 6, \quad k_{2,-B} = 2, \]
\[ k_{12} = k_{21} = 1, \quad c_B = 1. \] (105)

This other set of parameter values is not compatible with the existence of an equilibrium limit.

For this example, the mean growth velocity \( \langle v \rangle \) and the bulk probability \( \bar{\mu}(A) \) for the monomeric unit A are depicted in Fig. 5 as functions of the concentration \( c_A \) and the thermodynamic quantities in Fig. 6. Again, there is excellent agreement between the simulation results (data points) and the theoretical predictions (lines).

Here, the mean growth velocity vanishes at the critical concentration \( c_A = 0.94233 \). However, the system remains out of equilibrium, as seen in Fig. 6 showing that, at the critical concentration, the entropy production rate remains positive at the value \( dS/dt \simeq 0.39736 \), while both the affinity \( A \) and the free-energy driving force \( \varepsilon \) diverge as \( A \simeq \varepsilon \simeq 0.39736/v \) in consistency with the vanishing of the mean growth velocity. At the critical concentration, the sequence disorder takes the value \( D \simeq 0.68 \). Since the free-energy driving force remains positive in this example, the growth is always driven by free energy and there is here no regime of disorder-driven growth.

As in the previous example, the free-energy driving force becomes dominant over sequence disorder, \( \varepsilon \gg D \), for large values of the concentration \( c_A \) where the attachment rates dominate. In this strongly irreversible regime, the entropy production rate increases with the concentration \( c_A \) as \( dS/dt \simeq 2.5c_A \ln(0.524c_A) \) according to Eq. (103), as observed in Fig. 6.

For the example (105), the normalized correlation function \( C(n) \) versus \( n \) for different values of the concentration \( c_A \) is shown in Fig. 7, where the data points depict the values of Eq. (103) for the simulations and the lines the theoretical functions calculated with Eq. (55). Here also, there is good agreement between theory and simulations with

FIG. 5. Growth of a copolymer in the conditions (105): The mean growth velocity \( v \) and the bulk probability \( \bar{\mu}(A) \) of the monomeric unit A, versus the concentrations \( c_A \). The data points show the results of simulations with Gillespie’s algorithm generating a sequence after \( 10^6 \) jumps of the algorithm for every value of the concentration \( c_A \). The lines depict the theoretical predictions of Eqs. (57) and (50).

FIG. 6. Growth of a copolymer in the conditions (105): The mean growth velocity \( v \), the free-energy driving force \( \varepsilon \), the sequence disorder \( D \), the affinity \( A \), and the entropy production rate \( dS/dt \), versus the concentrations \( c_A \). The data points are obtained with a statistics of \( 10^6 \) sequences generated by Gillespie’s algorithm running over the total time \( t = 200 \). The lines show the theoretical predictions.

FIG. 7. Growth of a copolymer in the conditions (105): The correlation function \( C(n) \) versus \( n \) for different values of the concentration \( c_A \). The data points show the simulation results of Eq. (103) using a long sequence generated by \( 10^6 \) jumps of Gillespie’s algorithm. The total length of the sequence ranges from \( L \simeq 1.6 \times 10^7 \) for the lowest concentration value \( c_A = 2 \) to \( L \simeq 6.5 \times 10^7 \) for the largest one \( c_A = 10 \). The lines depict the theoretical predictions of Eq. (55).
Gillespie’s algorithm. Again, the decay rate is slower at larger than smaller values of the concentration \( c_A \).

V. CONCLUSION AND PERSPECTIVES

In this paper, theory is developed for the kinetics and thermodynamics of multistate reversible copolymerization processes, in which the growing copolymer may undergo transitions between several reactive states controlling the monomeric attachment and detachment rates and these rates are supposed to be independent of previously incorporated monomeric units. In this case, if the copolymer remained in a single reactive state, its sequences would form a Bernoulli chain. However, as a consequence of the transitions between several reactive states, the growing copolymer sequences instead form a nonMarkovian chain. The probability distribution of the sequences are given in terms of products of matrices associated with every monomeric species and of size equal to the number of reactive states in the mechanism. In general, the chain is nonMarkovian because the matrix products cannot be factorized as for Bernoulli or Markov chains. The matrices of the theory also determine the mean growth velocity, the bulk probabilities of monomeric subsequences, the statistical correlation functions along the sequences, as well as the thermodynamic quantities. In particular, the entropy production rate is shown to be given in terms of the mean growth velocity, the free-energy driving force, and the sequence disorder, confirming the results of Ref. 10.

Two illustrative examples are used to compare theory with simulations using Gillespie’s algorithm. For the first example, the rate constants are compatible with the existence of a thermodynamic equilibrium limit. In this example, the entropy production rate vanishes together with the mean growth velocity at the equilibrium value of the tuned monomeric concentration. At this marginal concentration where the equilibrium detailed balance conditions are satisfied, the chain reduces to a Bernoulli chain. Close to equilibrium, there exists a regime where the growth is driven by the entropy effect of sequence disorder, as predicted in Ref. 10. For the second example, the rate constants are not compatible with the existence of an equilibrium limit. Accordingly, the entropy production rate keeps a positive value at the critical concentration where the mean growth velocity vanishes and the free-energy driving force diverges together with the affinity. In both examples, the correlation functions are observed to decay exponentially, as predicted by theory. Excellent agreement is found between theory and simulations using Gillespie’s algorithm.

All these results also concern multistate reversible homopolimerization processes yielding sequences with different tacticity. Accordingly, they show that the theory of Coleman and Fox14 can be extended from fully irreversible to reversible multistate mechanisms, allowing us to include the detachment of monomers beside their attachment in our models. The present theory of multistate reversible copolymerization under low conversion conditions can be further extended to ultimate and penultimate multistate mechanisms where the attachment and detachment rates also depend on previously incorporated monomeric units. For such other kinetics, the sequences form Markov chains instead of Bernoulli chains14,15 if the copolymer remains in a single reactive state, but nonMarkovian chains if transitions occur between several reactive states. The extension of the theory to template-directed copolymerization can also be considered.

ACKNOWLEDGMENTS

This research is financially supported by the Université libre de Bruxelles (ULB) and the Fonds de la Recherche Scientifique -FNRS under the Grant PDR T.0094.16 for the project “SYMSTATPHYS”.

Appendix A: The case of kinetics generating Bernoullian chains

In this appendix, we show that we recover previous results for kinetics generating Bernoulli chains if either the transition rates \( w_{1\rightarrow 2} \) or \( w_{2\rightarrow 1} \) is vanishing.

If \( w_{1\rightarrow 2} = 0 \), the chain stays in the reactive state \( i = 1 \), so that \( \psi(1) = 1 \) and \( \psi(2) = 0 \). Therefore, the mean growth velocity \( w_{1\rightarrow 2} \) is equal to \( v = v_{11} + v_{21} \). However, in Eq. (36), the matrices \( W_{-m} \) and \( W_{+m} \) are diagonal and \( W_0 \) is upper triangular. Consequently, the iteration (36) converges towards an upper triangular matrix \( V \), so that \( v_{21} = 0 \). The mean growth velocity is thus given by the solution \( v = v_{11} \) of the self-consistent equation

\[
1 = \sum_m \frac{w_{i, -m}}{v_{11} + w_{i, -m}},
\]

with \( i = 1 \), as for the growth of a Bernoulli chain staying in this reactive state.

Similarly, if \( w_{2\rightarrow 1} = 0 \), the chain stays in the reactive state \( i = 2 \), so that \( \psi(1) = 0 \) and \( \psi(2) = 1 \). In this case, the mean growth velocity \( w_{2\rightarrow 1} \) is equal to \( v = v_{12} + v_{22} \). In Eq. (36), the matrix \( W_0 \) is now lower triangular, so that the iteration (36) converges towards a lower triangular matrix \( V \) and thus \( v_{12} = 0 \). Accordingly, the mean growth velocity is now given by the solution \( v = v_{22} \) of the self-consistent equation (A1) with \( i = 2 \), as for the growth of a Bernoulli chain staying in this other reactive state.

Furthermore, if \( w_{1\rightarrow 2} = w_{2\rightarrow 1} = 0 \), the matrix (11) is equal to zero and the matrix equation (36) reduces to

\[
V = V \cdot \sum_m (V + W_{-m})^{-1} \cdot W_{+m},
\]

which is diagonal and decouples into the two previous self-consistent equations (A1) for the mean growth ve-
locities of the two types of copolymers, which are thus growing independently of each other.

Appendix B: Numerical methods in theory

If there are \( I \) reactive states, the matrices are of size \( I \times I \). In the reported examples, we take \( I = 2 \). The central equation of the theoretical framework to be solved is Eq. (30) for the velocity matrix. Its solution is obtained by iteration starting from an initial positive matrix \( V_0 \). The right-hand side of Eq. (30) gives the next iterate, which is reinserted in the right-hand side, and so on and so forth. This iterative scheme converges towards the solution. In the examples, \( 10^4 \) iterations are used to obtain the solution. Thereafter, the mean growth velocity can be calculated with Eq. (37) and the solution. In the examples, \( 10^4 \) iterations are used to obtain the solution. Thereafter, the mean growth velocity and the free-energy driving force:

\[
i \rightarrow j, \quad t \rightarrow t + \Delta t(\omega \rightarrow \omega'), \quad l \rightarrow l + \Delta l(\omega \rightarrow \omega'), \quad E \rightarrow E + \ln [W(\omega \rightarrow \omega')/W(\omega' \rightarrow \omega)]. \tag{C4}
\]

After many jumps and the elapsed time reaching the value \( t \), the mean growth velocity and the free-energy driving force are respectively evaluated by \( v \simeq l/t \) and \( \varepsilon \simeq E/t \). The determination of the sequence disorder requires the computation of the tip probabilities \( \psi(m_1 \cdots m_l, i) \) with a large enough statistics, as specified in the figure captions. In the simulations, the tip probabilities are obtained for \( l = 1, 2, 3 \) and the value of sequence disorder is approximated by \( D \simeq D^{(3)} - D^{(2)} \) in terms of the corresponding Shannon sequence entropies \( H \). The affinity is thus calculated with Eq. (88) and the entropy production rate with Eq. (84). The correlation function is obtained for the function \( f(l) \) by using

\[
\Gamma(n) = \frac{1}{L} \sum_{k=1}^{L} [f(m_k) - \langle f \rangle] [f(m_{k+n}) - \langle f \rangle], \tag{C5}
\]

where \( L \) is the total length of the long sequence used in the computation and \( \langle f \rangle = (1/L) \sum_{k=1}^{L} f(m_k) \) is the mean value of the function \( f(m) \).

1B. D. Coleman and T. G. Fox, J. Chem. Phys. 38, 1065 (1963).
2B. D. Coleman and T. G. Fox, J. Am. Chem. Soc. 85, 1241 (1963).
3B. D. Coleman and T. G. Fox, J. Polym. Sci. A 1, 3183 (1963).
4D. Andrieux and P. Gaspard, J. Chem. Phys. 130, 014901 (2009).
5L. Kryven, Y. R. Zhao, K. B. McAuley, and P. Iedema, Chem. Eng. Sci. 177, 491 (2018).
6M. Seware and C. L. Perrin, Macromolecules 18, 528 (1985).
7R. Szymanski, Makromol. Chem. 188, 2605 (1987).
8R. Szymanski, Prog. Polym. Sci. 17, 917 (1992).
9A. Blokhuis and D. Lacoste, J. Chem. Phys. 147, 094905 (2017).
10D. Andrieux and P. Gaspard, Proc. Natl. Acad. Sci. USA 105, 9516 (2008).
11D. T. Gillespie, J. Comput. Phys. 22, 403 (1976).
12D. T. Gillespie, J. Phys. Chem. 81, 2340 (1977).
13R. Wolfram, Mathematica, 2nd edition (Addison-Wesley Publishing Company, Redwood City CA, 1991).
14P. Gaspard and D. Andrieux, J. Chem. Phys. 141, 044908 (2014).
15P. Gaspard, J. Stat. Phys. 164, 17 (2016).