Learning heterogenous reaction rates from stochastic simulations

Ariana Torres-Knoop
SURF, Hoog Catharijne, Moreelsepark 48, 3511 EP Utrecht, the Netherlands

Ivan Kryven*
Mathematical Institute, Utrecht University, Budapestlaan 6, 3508 TA Utrecht, Netherlands
and Centre for Complex Systems Studies, 3584 CE Utrecht, Netherlands

(Received 28 February 2021; accepted 10 April 2021; published 4 May 2021)

Reaction rate equations are ordinary differential equations that are frequently used to describe deterministic chemical kinetics at the macroscopic scale. At the microscopic scale, the chemical kinetics is stochastic and can be captured by complex dynamical systems reproducing spatial movements of molecules and their collisions. Such molecular dynamics systems may implicitly capture intricate phenomena that affect reaction rates but are not accounted for in the macroscopic models. In this work we present a data assimilation procedure for learning nonhomogeneous kinetic parameters from molecular simulations with many simultaneously reacting species. The learned parameters can then be plugged into the deterministic reaction rate equations to predict long time evolution of the macroscopic system. In this way, our procedure discovers an effective differential equation for reaction kinetics. To demonstrate the procedure, we upscale the kinetics of a molecular system that forms a complex covalently bonded network severely interfering with the reaction rates. Incidentally, we report that the kinetic parameters of this system feature peculiar time and temperature dependences, whereas the probability of a network strand to close a cycle follows a universal distribution.

DOI: 10.1103/PhysRevE.103.052402

I. INTRODUCTION

How do we deduce chemical rate constants from observations? On the macroscopic scale, where concentrations of chemical compounds are deterministic quantities, this question was answered by Arrhenius, who linked the reaction rate constants with slopes and intersection points of the concentration related profiles. Microscopic systems, for instance, living cells [1,2], micropores [3], and those used for *in silico* computer experiments [4–8], typically have a small reaction volume, and therefore, the corresponding reaction rates may feature stochastic fluctuations that are not accounted for in the Arrhenius theory. Other assumptions of the Arrhenius theory, such as the well-mixed environment, Boltzmann's *Stosszahlansatz*, the absence of memory, and noncooperation of particles, may lead to artifacts even in the case of macroscopic systems. If such artifacts occur [6,9,10], the reaction rate constants appear to be time dependent. For example, irreversible polymerization leads to progressively growing molecules, and therefore, each reaction firing changes the conditions of the system and, consequently, the reaction rates [11,12]. Molecular networks pose an especially severe case: their physical properties evolve considerably in the course of the assembly process, and the latter may undergo various types of phase transitions [6–8]. As an illustration of how strong such changes can be, Fig. 1 depicts the formation of a percolating molecular network that significantly limits the mobility of all species.

Molecular dynamics (MD) simulations [5] describe the evolution of a complex system by solving the equation of motion for each molecule and do not require reaction rate constants as input. For the purpose of this paper, we view the outcome of such simulations as large streams of data that implicitly contain information about the rates. Provided the reaction rates are extracted from these time series, the rates may be used as input for large-scale models, hence enabling a multiscale paradigm. Among such macroscopic models are ordinary differential equations for species concentrations, the chemical master equation, the Langevin equation, the stochastic simulation algorithm, and other Monte Carlo methods [13].

While the foundation of reaction rates is frequently discussed in the literature [14–17], this paper takes a phenomenological view and develops a practical method for inferring reaction rate parameters from noisy microscopic observations as given by, for example, molecular dynamics simulations.

II. CHEMICAL RATE EQUATION

Consider a system that consists of N chemical species reacting via M reactions. Each species may be represented by multiple particles, which is indicated by the particle count vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^\top$, where x_i are the numbers of copies that species i is represented by. We thus have $\sum_{i=1}^{N} x_i$ particles in total. The reactive interactions that occur between these species can be modeled using three levels of mathematical description [13]: the equation of motion, stochastic process, and rate equation.

The rate equations are ordinary differential equations (ODEs) that instead of species counts x_i govern the evolution

*i.kryven@uu.nl

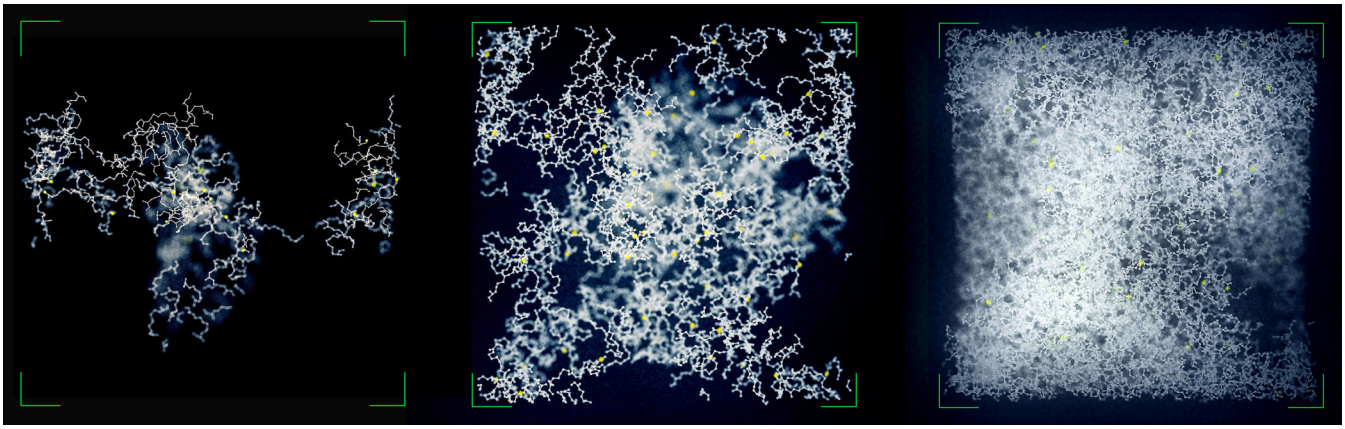

FIG. 1. Time snapshots of the carbon skeleton of the largest cluster in the diacrylate network as given by molecular simulations suggest that the reaction rates may considerably slow down during the course of polymerization. Left to right: 20%, 30%, and 80% of reaction progress as measured by the double bond conversion χ.

of their the molar concentrations $\boldsymbol{c} = (c_1, c_2, \ldots, c_N)^\top$, with

$$c_i = \frac{x_i}{VN_A}, \tag{1}$$

where the volume $V \to \infty$ and x_i are assumed to scale in such a way that keeps the pressure constant and N_A is Avogadro's constant. In the general case of M reactions, the ODEs are given by

$$c_i'(t) = \sum_{j=1}^{M} k_j S_{i,j} \boldsymbol{c}^{\boldsymbol{\nu}_j}(t), \;\; i = 1, 2, \ldots, N, \tag{2}$$

where k_j are the reaction rate constants, $\boldsymbol{\nu}_j$ are binary vectors defining the participation of species i in reaction j, and the vector power $\boldsymbol{c}^{\boldsymbol{\nu}} = c_1^{\nu_1} c_1^{\nu_2} \cdots c_N^{\nu_N}$ is evaluated in an element-wise manner. Matrix $\boldsymbol{S}$ has size $N \times M$ and is composed of stoichiometric vectors as its rows. For example, $S_{i,j} = 1$ if the ith species is the product of the jth reaction, $S_{i,j} = -1$ if it is a nonunique reactant, $S_{i,j} = -1/2$ if it is the only reactant in the second order reaction, and $S_{i,j} = 0$ is a nonparticipant.

The intuition behind Eq. (2) becomes clearer after considering the following example. Consider a system that consists of three chemical species, A, B, and C, having particle counts $\#\mathrm{A} = x_1$, $\#\mathrm{B} = x_2$, and $\#\mathrm{C} = x_3$ and reacting via the following mechanism:

$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C. \tag{3}$$

By defining species concentrations with Eq. (1), we arrive at the following set of ODEs:

$$\begin{aligned} c_3' &= k_1 c_1 c_2 - k_2 c_3, \\ c_2' &= -k_1 c_1 c_2 + k_2 c_3, \\ c_1' &= -k_1 c_1 c_2 + k_2 c_3, \end{aligned} \tag{4}$$

where k_i are the rate constants. In order to see that Eq. (4) is the special case of Eq. (2) it is sufficient to substitute

$$S = \begin{pmatrix} -1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}^\top, \quad \boldsymbol{\nu}_1 = (1, 1, 0)^\top,$$
$$\boldsymbol{\nu}_2 = (0, 0, 1)^\top.$$

One can see that the elements of $\boldsymbol{\nu}_1$ sum up to 2, which indicates that $j = 1$ is a first order reaction, whereas the elements of $\boldsymbol{\nu}_2$ sum up to 1, indicating that the reaction order of $j = 2$ is 2.

III. STOCHASTIC RATE EQUATION

We will now introduce a stochastic rate equation that operates with discrete particle counts x_i as opposed to continuous concentrations used in (2). Suppose that all elements of the species count vector are large, $\boldsymbol{x} \gg 0$, and in small time increments τ these values undergo a small relative change. Let $\boldsymbol{z} = (z_1, z_2, \ldots, z_M)^\top$ be the column vector of reaction firings observed during time interval τ. We also assume that the dynamics is a heterogenous renewal process, that is, the elements of $\boldsymbol{z}$ are independent Poisson random variables: $z_j \sim \mathrm{Poiss}[\lambda_i \boldsymbol{x}^{\boldsymbol{\nu}_i} \tau]$, $j = 1, \ldots, M$, which, when combined with reaction stoichiometry $\boldsymbol{S}$, provides the update vectors for species counts $\boldsymbol{x}$ at a given time interval. By iterating $\tau_l = t_l - t_{l-1}$ over all discrete time intervals, one recovers the whole evolution trajectory of species count vector $\boldsymbol{x}_l$ for $l = 1, \ldots, L$:

$$\begin{aligned} \boldsymbol{x}_l &= \boldsymbol{x}_{l-1} + \boldsymbol{S}\boldsymbol{z}_{l-1}, \\ \boldsymbol{z}_l &\sim \left(\mathrm{Poiss}\left[\lambda_1 \boldsymbol{x}_l^{\boldsymbol{\nu}_1} \tau_l\right], \ldots, \mathrm{Poiss}\left[\lambda_M \boldsymbol{x}_l^{\boldsymbol{\nu}_M} \tau_l\right]\right)^\top, \end{aligned} \tag{5}$$

where coefficients λ_i are (time-dependent) parameters related to reaction rates k_i. Appendix A sketches the derivation of Eq. (5) and explains the relationship between rate constants k_i and λ_i. This equation resembles an implementation of the τ-leaping method [18] and can be regarded as an N-dimensional random walk on species count numbers.

Stochastic process (5), although practical, relies upon the system being well mixed, memoryless, and noncooperative, among other assumptions. We suggest that this can be partially remedied for by inferring the time-dependent coefficients λ_i from molecular simulations that do not suffer from these issues.

IV. DATA ASSIMILATION PROCEDURES

In this section we assume that the empirical trajectories of the species counts $\tilde{\boldsymbol{x}}_l$ and the counts of all reaction firings $\tilde{\boldsymbol{z}}_l$ are known. We solve the inverse problem for estimating parameters $\lambda_1, \ldots, \lambda_M$, which may depend on time. Namely, we propose several statistical inference methods, so-called maximum likelihood estimators (MLEs), for estimating effective reaction rates λ_i that can be readily used in the stochastic model (5) or ODEs (2). The source code implementing the estimators (6)-(11) is provided [19].

Constant rate estimator. Assuming that the stochastic rates λ_j do not depend on time, the following estimates hold:

$$\lambda_j = \frac{\langle \tilde{z}_{j,l} \rangle}{\langle \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle}, \quad \mathrm{var}(\lambda_j) = \frac{\lambda_j^2}{L \langle \tilde{z}_{j,l} \rangle}, \tag{6}$$

where

$$\langle x_l \rangle := \frac{1}{L} \sum_{l=1}^{L} x_l$$

denotes the time-average and $\mathrm{var}(\lambda_j)$ refers to the asymptotic variance of this estimator, which may be used to derive the confidence intervals. See Appendix B for the derivations.

Moving-average rate estimator. The following estimators yield rates in the form of a time series:

$$\lambda_{j,l} = \frac{\langle \tilde{z}_{j,l} \rangle_s}{\langle \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle_s}, \quad \mathrm{var}(\lambda_{l,j}) = \frac{\lambda_{j,l}^2}{(2s+1) \langle \tilde{z}_{j,l} \rangle_s}, \tag{7}$$

where

$$\langle x_l \rangle_s := \frac{1}{2s+1} \sum_{l=l-s}^{l+s} x_l$$

represents the moving average with window size s. See Appendix C for the derivations.

Exponential rate estimator. Consider the following ansatz for the parameters of process (5):

$$\lambda_j(t) = \lambda_{j,0} e^{-\alpha_j t}. \tag{8}$$

The estimators for the coefficients are given by

$$\lambda_{j,0} = \frac{\langle \tilde{\boldsymbol{z}}_{j,l} \rangle}{\langle e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle}$$

and

$$\alpha_j = -\ln \omega_j,$$

where $\omega_j \in [0,1]$ are the unique roots of $\langle (t_l \langle \tilde{\boldsymbol{z}}_{j,l} \rangle - \langle \tilde{\boldsymbol{z}}_{j,l} t_l \rangle) \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \omega_j^{t_l} \rangle = 0$ for $j = 1, \ldots, M$. The variances of the exponents are given by

$$\mathrm{var}(\alpha_j) = \frac{1}{L \lambda_{j,0} \langle t_l^2 e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle},$$

and those of the prefactor are given by

$$\mathrm{var}(\lambda_j) = \mathrm{var}(\lambda_{j,0}) = \frac{\lambda_{j,0}^2}{L \langle \tilde{\boldsymbol{z}}_{j,l} \rangle}.$$

See Appendix D for the derivations.

Exponential polynomial rate estimator. Assume that the reaction rate parameters that appear in the random walk model (5) have an exponential dependence on time of the form

$$\lambda_j(t) = e^{-p_j(t)}, \tag{9}$$

where $p_j(t) = \alpha_{j,0} + \alpha_{j,1} t + \alpha_{j,2} t^2 + \cdots + \alpha_{j,s} t^S$ is a polynomial of order S. For each j, the estimators of $\alpha_{j,s}$ are found from the system of S algebraic equations:

$$\left\langle \left(e^{-p_j(t_l)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l - \tilde{\boldsymbol{z}}_{j,l} \right) t_l^s \right\rangle = 0, \ s = 0, \ldots, S, \tag{10}$$

and the variances of the rates' logarithms are given by

$$\mathrm{var}[\ln \lambda_j(t)] = \frac{1}{L} \boldsymbol{b}^\top \boldsymbol{H}_j^{-1} \boldsymbol{b}, \tag{11}$$

where $\boldsymbol{H}_j$ are $(S+1) \times (S+1)$ matrices with elements

$$(\boldsymbol{H}_j)_{k,s} = \left\langle e^{-p_j(t_l)} t_l^k t_l^s \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \right\rangle$$

and $\boldsymbol{b} = (1, t, t^2, \ldots, t^S)^\top$. In fact, one can replace time t in MLE (9) with any monotonous function of time that tracks the progress of the chemical system, for example, the conversion of an important species. The derivations are given in Appendix E.

Model selection. There are two parameters describing the quality of the estimate that may be used when choosing the best MLE and, in the case of the polynomial estimator, when choosing the polynomial order. A small variance implies that the system is large enough to derive consistent estimates with a given estimator. A small residual implies that the estimator explains observed data. In order to rationally determine the best order of the polynomial for approximation, we propose to minimize two qualities simultaneously: the variance and residual.

V. EXAMPLE: RATES OF NETWORK FORMATION

In this section, we illustrate the application of the estimators in a real world example. We infer the reaction rates of polymer network formation as captured by the MD simulations illustrated in Fig. 1 and show how to replace these computationally expensive MD simulations with a simple system of ODEs that are valid on arbitrary large timescales.

System setup. Our *microsystem* [7] is as follows: 2000 diacrylate molecules confined in a 7.52×10^{-25} m^3 simulation box with periodic boundary conditions and integrated in time up to 10^{-8} s in the *NPT* ensemble. Initially, 5% of all monomers are set to be active (bearing radicals), and the activation energy of the reaction has been reduced to speed up the simulations. The true kinetic parameters can be recovered by an appropriate unbiasing procedure (see Ref. [7] for a discussion). This microsystem is confronted with the *macrosystem* that reflects the desired real world target: 4.7 mol of monomer units (which is of the order of 10^{24} particles), polymerized under continuous initialization that maintains a steady concentration of radicals at 10^{-4} $\frac{\mathrm{mol}}{\mathrm{L}}$ (e.g., photopolymerization). We investigate the rates of the two most important species: Vinyl groups (V) and a radicals (R) that react via two reaction channels, propagation and termination, respectively:

$$\begin{aligned} V + R &\to R, \\ R + R &\to \varnothing. \end{aligned} \tag{12}$$

TABLE I. Inferred reaction rate parameters for 1,6-Hexanediol diacrylate (HDDA) polymerization as given by the constant MLE. Confidence intervals indicate two standard deviations.

T (K)	Propagation k_1 ($\frac{\text{mol}}{\text{Ls}}$)	Termination, k_2 [$\frac{\text{mol}}{\text{Ls}}$]
200	14.55 ± 0.2802	$1.282\,10^6 \pm 4.488\,10^5$
250	792.3 ± 15.04	$7.865\,10^6 \pm 2.387\,10^6$
300	17733.0 ± 268.7	$2.113\,10^7 \pm 5.492\,10^6$
350	97422.0 ± 1385.0	$2.494\,10^7 \pm 5.964\,10^6$
400	$4.276\,10^5 \pm 5301.0$	$3.106\,10^7 \pm 6.898\,10^6$
450	$1.682\,10^6 \pm 23599.0$	$8.03\,10^7 \pm 1.890\,10^7$
500	$3.644\,10^6 \pm 43900.0$	$6.258\,10^7 \pm 1.336\,10^7$
550	$1.031\,10^7 \pm 1.391\,10^5$	$1.237\,10^8 \pm 2.871\,10^7$
600	$1.64\,10^7 \pm 1.944\,10^5$	$1.515\,10^8 \pm 3.577\,10^7$

This mechanism is characterized by

$$S = \begin{pmatrix} -1 & 0 \\ 0 & -1/2 \end{pmatrix}, \quad \boldsymbol{\nu}_1 = (1,1), \ \boldsymbol{\nu}_2 = (0,2),$$

which in combination with molecular dynamics data $\tilde{\boldsymbol{x}}_l$ and $\tilde{\boldsymbol{z}}_l$ provides enough information to apply the rate estimators. Since the activation energy E_a has been reduced in the microsystem, we use the decomposition of the rate

$$k(t) = A(t)e^{-E_a/(RT)} \tag{13}$$

and perform the inference solely for preexponential factor $A(t)$, which is expected to be the most sensitive to the interferences from the network formation. Here, T denotes the temperature, and R is the gas constant. To recover the rate coefficient $k(t)$, Eq. (13) should be supplied with $E_{a,1} = 31.02\,\frac{\text{kJ}}{\text{mol}}$ for propagation and $E_{a,2} = 8.673\,\frac{\text{kJ}}{\text{mol}}$ for termination reactions (activation energies from the RMGPY database [20]).

Estimated kinetic rates. Table I reports the constant rate estimations obtained with Eq. (6). These estimates correspond to the prefactors indicated by the horizontal lines in Figs. 2(a)–2(d).

According to the variance analysis given in Appendix F, the exponential polynomial estimator was found to yield optimal estimates using fourth order polynomials for the propagation reaction and third order ones for the termination, that is,

$$\begin{aligned} k_1(\chi, T) &= A_1(t)e^{-\frac{E_{a,1}}{RT}} \\ &= Ce^{-(\alpha_{1,4}\chi^4+\alpha_{1,3}\chi^3+\alpha_{1,2}\chi^2+\alpha_{1,1}\chi+\alpha_{1,0})}e^{-\frac{E_{a,1}}{RT}} \end{aligned} \tag{14}$$

and

$$\begin{aligned} k_2(\chi, T) &= A_2(t)e^{-\frac{E_{a,2}}{RT}} \\ &= Ce^{-(\alpha_{2,3}\chi^3+\alpha_{2,2}\chi^2+\alpha_{2,1}\chi+\alpha_{2,0})}e^{-\frac{E_{a,2}}{RT}}, \end{aligned} \tag{15}$$

where the scaling constant is $C = VN_a = 452.93\,\frac{\text{L}}{\text{mol}}$. Instead of time t, we characterize the progress of the network formation by

$$\chi(t) = \frac{\#V(0) - \#V(t)}{\#V(0)}. \tag{16}$$

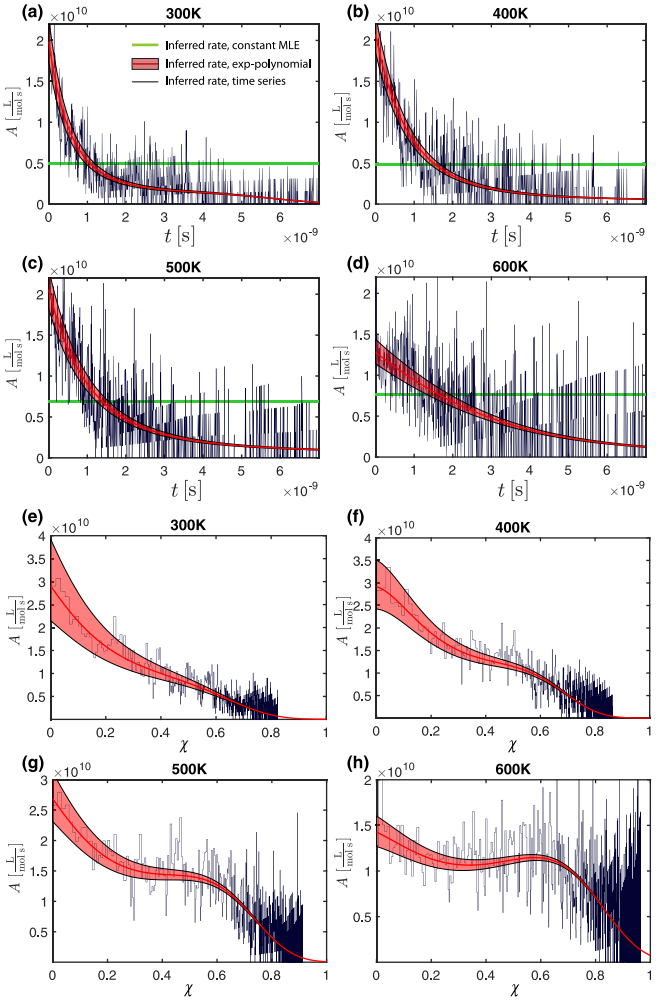

FIG. 2. (a)–(d) Inferred reaction rate prefactors $A(t)$ from a single MD trajectory. Horizontal lines represent the constant estimator, Eq. (6), and bands represent the fourth order exponential polynomial estimator, Eq. (9). Solid lines correspond to the time series estimator (7). The margins indicate two-standard-deviation confidence. (e)–(h) Inferred prefactors $A(\chi)$ with time series and exponential polynomial estimators shown. All panels share the same legend.

This quantity is also known as the bond conversion in chemistry or occupancy probability in the theory of percolation. The coefficients are given in Table II.

Nonlinear rate behavior. Figure 2 presents the values of $A(t)$ and $A(\chi)$ inferred from single MD trajectories for different temperatures of polymerization T. Independent of T, both $A(t)$ and $A(\chi)$ strongly decrease throughout the reaction progress. This complex behavior can be possibly explained by the fact that the system undergoes two phase transitions that may not necessarily coincide: the transition from disconnected clusters to a spanning network (the percolation transition [21]) and the transition from a liquid or resinlike to solid or glassy state (the glass transition [8]). Thus in total, we have four distinct domains in the T-χ phase space: Ω_{00}, which is viscous with no network; Ω_{10}, which is glassy with no network; Ω_{01}, which is rubbery and has a network; and Ω_{11}, which is glassy and has a network. As shown in Fig. 3(a), the partition of the phase space into these domains indicates

TABLE II. The coefficients for the optimal order exponential polynomial MLEs.

T (K)	Propagation rate					Termination rate			
	$\alpha_{1,4}$	$\alpha_{1,3}$	$\alpha_{1,2}$	$\alpha_{1,1}$	$\alpha_{1,0}$	$\alpha_{2,3}$	$\alpha_{2,2}$	$\alpha_{2,1}$	$\alpha_{2,0}$
200	78.806	−67.177	16.329	2.931	−16.479	163.830	−156.050	43.836	−16.623
250	115.780	−132.800	48.777	−1.957	−17.741	56.364	−63.904	27.499	−17.383
300	38.288	−38.096	12.209	2.036	−17.787	89.115	−98.233	36.041	−17.526
350	17.323	−14.024	2.348	2.447	−16.529	41.583	−47.910	19.160	−16.423
400	31.335	−36.106	12.000	0.990	−17.584	84.355	−103.340	37.832	−17.272
450	15.275	−14.874	3.435	1.506	−16.412	25.825	−30.223	13.987	−16.426
500	13.135	−11.238	0.775	1.969	−17.098	47.405	−66.980	28.765	−16.719
550	16.446	−20.363	7.150	0.065	−16.451	49.559	−65.727	25.621	−16.464
600	15.189	−16.538	3.135	1.164	−16.178	39.276	−54.521	22.272	−15.959

that the topological transition occurs around $\chi_c \approx 0.2$ independently of temperature, whereas the critical value of χ for glass transition is a function of T.

By color-coding the points in the profiles of $A(\chi)$ depending on which domain they belong to, Fig. 3(b) reveals that increasing T has opposite effects on A below and above the topological phase transition: increased temperature inhibits the value of prefactor A for $\chi < \chi_c$ and promotes this value for $\chi > \chi_c$. Moreover, the collisions in a network are governed by different mechanisms than collisions in the ideal gas: the shortest path between species embedded in a network becomes the most important factor that explains the collision rates, which, in turn, is independent of temperature or pressure. To emphasize the universal dependence of the system's geometry on the topology we compute the return probability of the shortest path in the network when it closes a chordless cycle (a so-called topological hole [8]). The probability that a polymer chain closes a chordless cycle of length n is typically derived from the return probability of a random walk that models the chain's geometry; however, the exact definition of this random walk is a topic of debate [22–25]. As shown in Fig. 4(a), the empirical probability that a network strand closes a cycle is universal and can be asymptotically related to Flory's expression for the self-avoiding random walk,

$$p \sim n^{-3/2} e^{-\frac{3}{2}n^{-1} - \alpha n^{1/2}},$$

where the chain stiffness parameter $\alpha = 1.2$ was found by fitting. The fact that the return probability does not depend on temperature is exclusive to networks since the latter feature more geometrically constrained configurations compared to loose chains.

Upscaling. The most important applied implication of the rate inference is that one can use this procedure to perform predictions with accuracy close to that of molecular simulations but on the macroscopic scale. Since all kinetic parameters are derived from the particle potentials, as encoded by the force field, such predictions can be almost parameter free. In order to perform the predictions, one models the reaction mechanism (12) with ODEs (1) that are supplied with the inferred expressions of $A(\chi)$, where χ is given by Eq. (16). Figures 4(b)–4(d) illustrate this principle: Fig. 4(b) compares MD data with the stochastic and ODE models, still

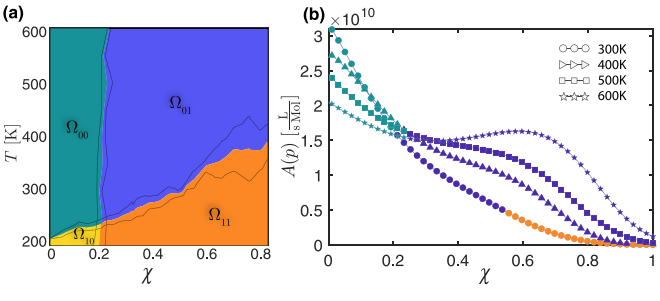

FIG. 3. (a) The T-χ phase space: Ω_{00}, viscous, no network; Ω_{10}, glassy, no network; Ω_{01}, rubbery, network; Ω_{11}, glassy, network. See [8] for the computational procedure. The solid lines mark the one-standard-deviation confidence interval around the domain boundaries. (b) Inferred profiles of $A(\chi)$ show that the polymerization temperature has opposite effects on the reaction rates in different domains, Ω_{00}, Ω_{10} and Ω_{01}, Ω_{11}. The colors code the domain of the phase space.

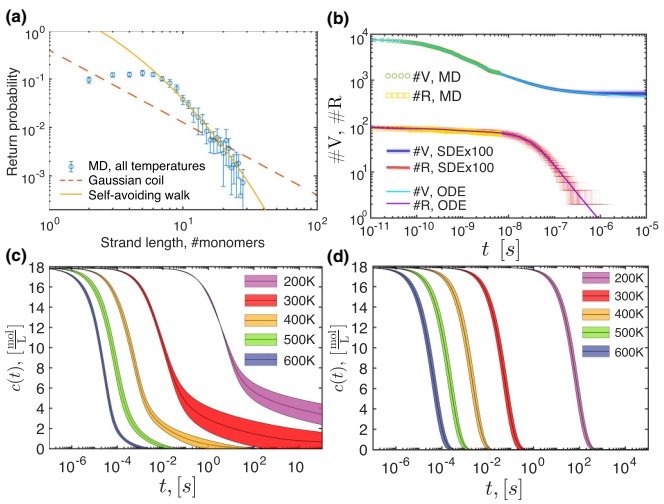

FIG. 4. (a) The empirical probability of a network strand to close a cycle compared to Flory's self-avoiding random walk and Gaussian coil. Error bars correspond to one standard deviation. (b)–(d) The upscaling procedure: (b) molecular simulations versus learned SDEs and ODEs with exponential polynomial coefficients, (c) macroscopic ODEs with exponential polynomial coefficients, and (d) macroscopic ODEs with constant coefficients.

in the microsystem, whereas Figs. 4(c) and 4(d) present the upscaled results as given by the ODEs with inferred rates for the macrosystem up to $t = 100$ s. Note that representing the rates as exponential polynomial functions of χ [Fig. 4(c)] as opposed to constant rates [Fig. 4(d)] is essential to capture the kinetic slowdown that is induced by the jamming and is especially pronounced at low temperatures.

VI. CONCLUSION

We proposed a solution of the inverse problem to Gillespie's stochastic simulation algorithm [26]: Using the empirical counts of molecular species, we recovered the reaction rate parameters that drive the kinetics. From the point of view of molecular dynamics, a reaction rate is an emergent phenomenon of many reactive particles, and our method allows one to extract the effective kinetic parameters from such simulations. Assuming that the inferred parameters are scale invariant, we show that the results of reactive molecular simulations may be upscaled in such a way that they become descriptive at the macroscopic scale.

Molecular simulations of many reaction-driven macroscopic phenomena are already on the way; see, for example, the studies on crystallization [4,27], self-assembly [28], aggregation [29], separation [30], and polymerization [7,9,31], and the concept of ordinary differential equations that learn from molecular simulations may facilitate the discovery of new macroscopic laws and improving existing kinetic models for these phenomena. As a proof of concept, we applied the method to diacrylate polymerization to reveal an intricate phenomenological dependance of the kinetic parameters on temperature and time in this system and postulated that these dependencies are induced by the complex evolution of the underlaying network. With this example we demonstrated that it is possible to model the transition between freely interacting spices and a dense network with ordinary differential equations with nonlinear coefficients. We expect that combining such MD-informed kinetic ODEs with random graphs [32–34] may result in accurate macroscopic models that also predict network related phenomena.

ACKNOWLEDGMENTS

The authors are grateful to S. Woutersen for providing critical comments on an early version of the manuscript. I.K. acknowledges the support from the research program VENI under Project No. 639.071.511, and A.T.-K. acknowledges support from the PREDAGIO Project. Both projects were financed by the Netherlands Organisation for Scientific Research (NWO).

APPENDIX A: DERIVATION OF THE STOCHASTIC RATE EQUATION

Consider a system that consists of a single molecule undergoing a first order reaction. If the reaction firing probabilities are independent and proportional to waiting time, the probability that time t passes until this molecule reacts is given by an exponential random variable with parameter λ:

$$\mathbb{P}[t \in [\tau, \tau + \mathrm{d}\tau]] = \lambda e^{-\lambda\tau}.$$

We refer to this fact as $t \sim \mathrm{Exp}[\lambda]$, also known as the "exponential clock" [35]. If instead, we have $x_1 = \#\mathrm{A}$ independent molecules of the same species, the time until the first reaction firing within this set of molecules is given by

$$t \sim \inf\{\underbrace{\mathrm{Exp}[\lambda], \ldots, \mathrm{Exp}[\lambda]}_{x_1 \text{ times}}\} \sim \mathrm{Exp}[x_1\lambda]. \tag{A1}$$

Here, we made use of the standard result about the minimum of multiple exponential random variables [35]. Since t is again an exponential random variable, its expected value is given by $\mathbb{E}[t] = (\lambda x_1)^{-1}$, which gives the characteristic time between reaction firings. Thus, the reaction rate r (the amount of substance per volume per time) is given by

$$r = \frac{1}{\mathbb{E}[t]}\frac{1}{VN_A} = \frac{x_1\lambda}{VN_A} = \lambda c(t) = kc(t), \tag{A2}$$

where the last equality derives from the fact that $c_i = \frac{x_i}{VN_A}$, where N_A is Avogadro's number. Hence, Eq. (A2) settles the relationship between the stochastic rate λ and the rate constant k for first order reactions:

$$k = \lambda. \tag{A3}$$

The rates of second order reactions are dependent on a coincidence of two events: (1) the two reactants collide in the correct configuration, and (2) together they undergo a first order reaction. We thus have a two-stage process:

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB} \to \mathrm{C}, \tag{A4}$$

where AB is an intermediate that represents the species that collided but have not reacted. According to Arrhenius theory, the first stage settles on an equilibrium: the number of AB is a constant fraction of the total number of couple combinations:

$$\#\mathrm{AB} = \mathcal{A}x_1x_2.$$

Since $\mathrm{AB} \to \mathrm{C}$ is a first order mechanism, it features the stochastic rate λ' as given by Eq. (A2). Consequently, one writes the time until the first reaction firing as

$$t \sim \mathrm{Exp}[\lambda'\mathcal{A}x_1x_2] = \mathrm{Exp}[\lambda x_1x_2], \;\; \lambda = \lambda'\mathcal{A}, \tag{A5}$$

which, after applying similar transformations to Eq. (A2), gives the approximation for the second order reaction rate:

$$r = \frac{1}{\mathbb{E}[t]}\frac{1}{VN_A} = \frac{\lambda x_1x_2}{VN_A} = \lambda VN_Ac_1(t)c_2(t) = kc_1(t)c_2(t).$$

Hence, for second order reactions we have

$$k = \lambda VN_A. \tag{A6}$$

Note that if a second order reaction takes place between members of the same species, then the number of couples $\#\mathrm{AA} = \frac{1}{2}x_1(x_1 - 1)$, and therefore, $k \approx \frac{1}{2}\lambda VN_A$. More generally, if the jth reaction (of arbitrary order now) is isolated, the waiting time that passes before the reaction firing is $t \sim \mathrm{Exp}[\lambda_j \boldsymbol{x}^{\boldsymbol{\nu}_j}]$, and by analogy to Eq. (A1), the time until the earliest event in the case of multiple competing reactions is given by $t \sim \inf_j \mathrm{Exp}[\lambda_j \boldsymbol{x}^{\boldsymbol{\nu}_j}] \sim \mathrm{Exp}[\sum_j \lambda_j \boldsymbol{x}^{\boldsymbol{\nu}_j}]$. Moreover, the probability that this is the jth reaction is given by $\mathbb{P}[j] = \frac{\lambda_j \boldsymbol{x}^{\boldsymbol{\nu}_j}}{\sum_i \lambda_i \boldsymbol{x}^{\boldsymbol{\nu}_i}}$. When iterated over multiple time steps, the latter two sampling rules yield the stochastic process (5).

APPENDIX B: CONSTANT MLE

We consider the general setting in which the time intervals $\tau_l = t_l - t_{l-1},\ l = 1, \ldots, L$ need not be equispaced. Let $\lambda_j(t) = \lambda_j = \text{const}$; then the rates of the Poisson random variables from Eq. (5) are given by

$$\lambda_j \boldsymbol{x}_l^{\boldsymbol{\nu}_j} \tau_l,\ l = 1, \ldots, L.$$

Therefore, the probability to observe configuration $\tilde{\boldsymbol{x}}_l, \tilde{\boldsymbol{z}}_l$ on time intervals τ_l is given by

$$\prod_{l=1}^{L}\prod_{j=1}^{M} e^{-\lambda}\frac{\lambda^y}{y!}\Big|_{\substack{y=\tilde{z}_{j,l}\\ \lambda=\lambda_j\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l}},$$

and taking a logarithm of this product gives the log likelihood of the entire ensemble of data:

$$f(\lambda_1, \ldots, \lambda_M) = \sum_{l=1}^{L}\sum_{j=1}^{M}(-\lambda + y\ln\lambda - \ln y!)\Big|_{\substack{y=\tilde{z}_{j,l}\\ \lambda=\lambda_j\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l}}, \tag{B1}$$

which has the following derivatives:

$$\frac{\partial f}{\partial\lambda_j} = -\sum_{l=1}^{L}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l + \frac{1}{\lambda_j}\sum_{l=1}^{L}\tilde{\boldsymbol{z}}_{j,l} = -L\langle\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l\rangle + \frac{1}{\lambda_j}L\langle\tilde{\boldsymbol{z}}_{j,l}\rangle,$$

where $\langle x_l\rangle := \frac{1}{L}\sum_{l=1}^{L} x_l$. By equating this derivative to zero, one obtains expressions for λ_j:

$$\lambda_j = \frac{\langle\tilde{\boldsymbol{z}}_{j,l}\rangle}{\langle\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l\rangle},\ j = 1, \ldots, M. \tag{B2}$$

In order to give an estimate of the variance of these parameters, $\text{var}(\lambda_1, \ldots, \lambda_n)$, we make use of the asymptotic normality property of this MLE and write

$$\text{var}(\lambda_1, \ldots, \lambda_n) = -[\text{Hess}f(\lambda_1, \ldots, \lambda_n)]^{-1}, \tag{B3}$$

where $\text{Hess}f(\lambda_1, \ldots, \lambda_n) := \frac{\partial^2 f}{\partial k_i \partial k_j}$ is the Hessian matrix. Evaluating this variance estimate for Eq. (B1) results in a diagonal covariance matrix, so that

$$\text{var}(\lambda_j) = \frac{\lambda_j^2}{L\langle\tilde{\boldsymbol{z}}_{j,l}\rangle}. \tag{B4}$$

APPENDIX C: MOVING-AVERAGE MLE

For this estimator we require time intervals τ_l to be equispaced. Consider a modification of the previous case in which for every $l = 1, \ldots, L$ the parameter $\lambda(t_l)$ is calculated from a local snippet of the data $\tilde{\boldsymbol{x}}_{l'}, \tilde{\boldsymbol{z}}_{l'}$, where $l' = l - s, \ldots, l + s$. Here, $s = 1, 2, \ldots$ plays the role of a regularity parameter. We obtain the following log-likelihood function for $\lambda_{j,l}$:

$$\begin{aligned} f(\lambda_{1,1}, \ldots, \lambda_{M,L}) &= \sum_{l=l-s}^{s+l}\sum_{j=1}^{M}(-\lambda + y\ln\lambda - \ln y!)\Big|_{y=\tilde{z}_{j,l}\lambda=\lambda_{j,l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l} \\ &= \sum_{l=l-s}^{s+l}\sum_{j=1}^{M}\left[-\lambda_{j,l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l + \tilde{\boldsymbol{z}}_{j,l}\ln(\lambda_{j,l}) + \tilde{\boldsymbol{z}}_{j,l}\ln\left(\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l\right) - \ln(\tilde{\boldsymbol{z}}_{j,l}!)\right], \end{aligned}$$

with the derivatives

$$\frac{\partial f}{\partial\lambda_{j,l}} = -\sum_{l=l-s}^{s+l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l + \frac{1}{\lambda_{j,l}}\sum_{l=l-s}^{s+l}\tilde{\boldsymbol{z}}_{j,l} = (2s+1)\left(\frac{1}{\lambda_{j,l}}\langle\tilde{\boldsymbol{z}}_{j,l}\rangle - \langle\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l\rangle\right),$$

where $\langle x_l\rangle_s := \sum_{l=l-s}^{l+s} x_l$ is the moving average. By equating these derivatives to zero, one obtains expressions for $\lambda_{j,l}$:

$$\lambda_{j,l} = \frac{\langle\tilde{\boldsymbol{z}}_{j,l}\rangle_s}{\langle\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l\rangle_s}. \tag{C1}$$

By following an analogous derivation to that of Eq. (B4), one also obtains the estimate for the variance:

$$\text{var}(\lambda_{j,l}) = \frac{\lambda_{j,l}^2}{(2s+1)\langle\tilde{\boldsymbol{z}}_{j,l}\rangle_s}. \tag{C2}$$

APPENDIX D: EXPONENTIAL MLE

We consider the following ansatz:

$$\lambda_j(t) = \lambda_{j,0}e^{-\alpha_j t}. \tag{D1}$$

By plugging $y = \tilde{\boldsymbol{z}}_{j,l}$ and $\lambda = \lambda_{j,0}e^{-\alpha_j t_l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l$ into the log-likelihood function, we obtain

$$\begin{aligned} &f(\alpha_{1,0}, \ldots, \alpha_{M,0}, \alpha_1, \ldots, \alpha_n) \\ &\quad= \sum_{l=1}^{L}\sum_{j=1}^{M}(-\lambda + y\ln\lambda - \ln y!) \\ &\quad= \sum_{l=1}^{L}\sum_{j=1}^{M}\big(-\lambda_{j,0}e^{-\alpha_j t_l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l + \tilde{\boldsymbol{z}}_{j,l}\ln\lambda_{j,0} - \tilde{\boldsymbol{z}}_{j,l}\alpha_j t_l \\ &\quad\quad - \ln\tilde{\boldsymbol{z}}_{j,l}!\big). \end{aligned} \tag{D2}$$

By equating to zero the partial derivatives with respect to $\lambda_{j,0}$, we obtain

$$\frac{\partial f}{\partial\lambda_{j,0}} = -\sum_{l=1}^{L}e^{-\alpha_j t_l}\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}\tau_l + \frac{1}{\lambda_{j,0}}\sum_{l=1}^{L}\tilde{\boldsymbol{z}}_{j,l} = 0,$$

and consequently,

$$\lambda_{j,0} = \frac{\langle \tilde{\boldsymbol{z}}_{j,l} \rangle}{\langle e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle}. \tag{D3}$$

In a similar fashion, we compute the derivatives with respect to α_j and equate them to zero to obtain

$$\frac{\partial f}{\partial \alpha_j} = \lambda_{j,0} \sum_{l=1}^{L} t_l e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l - \sum_{l=1}^{L} \tilde{\boldsymbol{z}}_{j,l} t_l = \lambda_{j,0} L \langle t_l e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle - L \langle \tilde{\boldsymbol{z}}_{j,l} t_l \rangle = 0.$$

Plugging Eq. (D3) in to the latter equality gives

$$\frac{\langle \tilde{\boldsymbol{z}}_{j,l} \rangle}{\langle e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle} \langle t_l e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle - \langle \tilde{\boldsymbol{z}}_{j,l} t_l \rangle = 0,$$

and since $\langle e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle > 0$, one can multiply by this quantity on both sides to obtain

$$\langle (t_l \langle \tilde{\boldsymbol{z}}_{j,l} \rangle - \langle \tilde{\boldsymbol{z}}_{j,l} t_l \rangle) \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \omega_j^{t_l} \rangle = 0, \ \omega_j \in [0,1], \tag{D4}$$

with $\alpha_j = -\ln \omega_j$. If each of these transcendental equations has a unique real root $\omega_j \in [0,1]$, the MLE (D1) has a minimum at α_j. Equation (D4) can be solved numerically by, for example, the bisection method. As a special case, when $t_l = hl$, $l = 1, 2, \ldots, L$ are equispaced, Eqs. (D4) become polynomial equations. For each j, $\alpha_j = -\frac{1}{h} \ln y$, where

$$\sum_{l=1}^{L} a_l y^l = 1 \tag{D5}$$

and $a_l = (l \langle \tilde{\boldsymbol{z}}_{j,l} \rangle - \langle \tilde{\boldsymbol{z}}_{j,l} l \rangle) \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j}$. This equation can be solved numerically by reformulating it as the eigenvalue problem for the companion matrix.

Analogous to Eq. (B3), the variances of $\lambda_{j,0}$ and α_j can be computed from the Hessian matrices of the corresponding log-likelihood functions. These matrices are not diagonal; however, at $t = 0$ we have $\lambda_{j,0} e^{-\alpha_j t} = \lambda_{j,0}$, and therefore,

$$\mathrm{var}(\lambda_j) = \mathrm{var}(\lambda_{j,0}) = \frac{\lambda_{j,0}^2}{L \langle \tilde{\boldsymbol{z}}_{j,l} \rangle}.$$

In a similar fashion, when $t \gg 1$, $\lambda_{j,0} e^{-\alpha_j t} = e^{(\frac{1}{t} \ln \lambda_{j,0} - \alpha_j) t} \approx e^{-\alpha_j t}$, and $\mathrm{var}(\alpha_j) = \frac{1}{L \lambda_{j,0} \langle t_l^2 e^{-\alpha_j t_l} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l \rangle}$.

APPENDIX E: EXPONENTIAL POLYNOMIAL MLE

In this estimator we assume the ansatz

$$\lambda_j(t) = e^{-p_j(t)}, \tag{E1}$$

where

$$p_j(t) = \alpha_{j,0} + \alpha_{j,1} t + \alpha_{j,2} t^2 + \cdots + \alpha_{j,s} t^S.$$

By plugging $y = \tilde{\boldsymbol{z}}_{j,l}$ and $\lambda = \lambda_j(t) \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l = e^{-p_j(t)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l$ into the log-likelihood function, we obtain

$$f(\alpha_{1,0}, \ldots, \alpha_{M,s}) = \sum_{l=1}^{L} \sum_{j=1}^{M} (-\lambda + y \ln \lambda - \ln y!) = L \big\langle - e^{-p_j(t_l)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l - \tilde{\boldsymbol{z}}_{j,l} p_j(t_l) + \tilde{\boldsymbol{z}}_{j,l} \times \ln \left(\tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \right) + \tilde{\boldsymbol{z}}_{j,l} \ln \tau_l + \ln(\tilde{\boldsymbol{z}}_{j,l}!) \big\rangle,$$

which has derivatives $\frac{\partial f}{\partial \alpha_{j,s}} = L \langle e^{-p_j(t_l)} t_l^s \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l - \tilde{\boldsymbol{z}}_{j,l} t_l^s \rangle$. We obtain MS equations that define $\alpha_{j,s}$ by equating these derivatives to zero:

$$\big\langle \big(e^{-p_j(t_l)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l - \tilde{\boldsymbol{z}}_{j,l} \big) t_l^s big \rangle = 0.$$

As in the preceding case, the variance analysis is performed by computing the Hessian matrix of the log-likelihood function:

$$\frac{\partial^2 f}{\partial \alpha_{j_1,s_1} \partial \alpha_{j_2,s_2}} = \begin{cases} -L \langle e^{-p_j(t_l)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l t_l^{s_1} t_l^{s_2} \rangle & \text{if } j_1 = j_2, \\ 0 & \text{if } j_1 \neq j_2, \end{cases}$$

so that $\mathrm{var}(\alpha_{j,1}, \alpha_{j,2}, \ldots, \alpha_{j,S}) = \frac{1}{L} \boldsymbol{H}^{-1}$, where

$$H_{k,s} = \big\langle e^{-p_j(t_l)} \tilde{\boldsymbol{x}}_l^{\boldsymbol{\nu}_j} \tau_l t_l^k t_l^s \big\rangle.$$

Moreover, this covariance matrix translates into the total variance of the rate parameter logarithm in the following way:

$$\mathrm{var}[\ln \lambda_j(t)] = \mathrm{var} \left(\sum_{s=0}^{S} \alpha_{j,s} t^s \right) = \frac{1}{L} \boldsymbol{b}^\top \boldsymbol{H}^{-1} \boldsymbol{b}, \tag{E2}$$

where $\boldsymbol{b} = (1, t, t^2, \ldots, t^S)^\top$.

APPENDIX F: VARIANCE ANALYSIS AND MODEL SELECTION

We consider exponential polynomial estimator (9) with conversion $\chi(t) = \frac{\#V(0) - \#V(t)}{\#V(0)}$ as the time variable. In Fig. 5 we explore how different polynomial orders $S = 0, \ldots, 6$ influence the inferred profiles of the rate prefactor $A(\chi)$ and their corresponding confidence intervals. To quantify the quality of the exponential polynomial estimator we calculate the residual: $r = \int_0^1 |\ln \lambda_j(\chi) - \ln \lambda_j^*(t)| d\chi$, where $\lambda_j^*(t)$ is given by the time-series estimator (9). Generally speaking, the higher the order of the polynomial is, the smaller the values of r are. Yet this is not the case for the variance of r, which has a tendency to increase with the polynomial order (the trend can also be seen in Fig. 5). Employing the fact that $\mathrm{var}(r) = \int_0^1 \mathrm{var}[\ln \lambda_j(\chi)] d\chi^2$, we find the upper bound of the confidence interval to be $c = r + 2\sqrt{\mathrm{var}(r)}$. The optimal polynomial order is then defined as the order that yields the smallest value of c. Figure 6(a) shows that the residual indeed tends to decrease with increasing polynomial order, whereas Fig. 6(b) shows that there is an optimal saddle point, $S = 4$, at which the confidence interval is the smallest in most of the MD trajectories. We can also see from Fig. 6(a) that the accuracy increases around 5–10 fold when we use the fourth order estimator as opposed to constant one, the zeroth order. A similar analysis for the termination reaction reveals an optimal order of $S = 3$ [see Figs. 6(c) and 6(d)]. We therefore report the inferred rate coefficients using the fourth order polynomial for the propagation and the third order polynomial for the termination reaction.

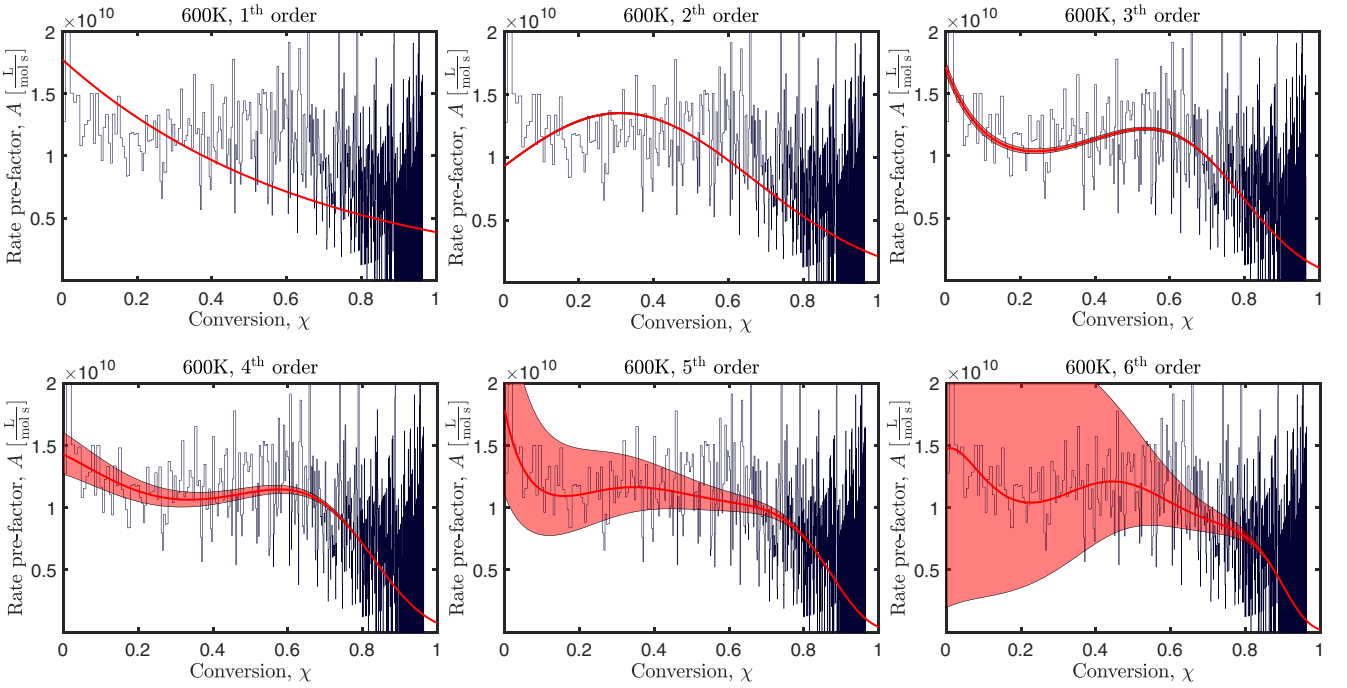

FIG. 5. Conversion-dependent prefactors as estimated with MLEs of various orders [the red line plus 2σ confidence intervals]. The effective time-series prefactor is given for reference (the black line). The optimal balance between small residual and high certainty corresponds to order 4.

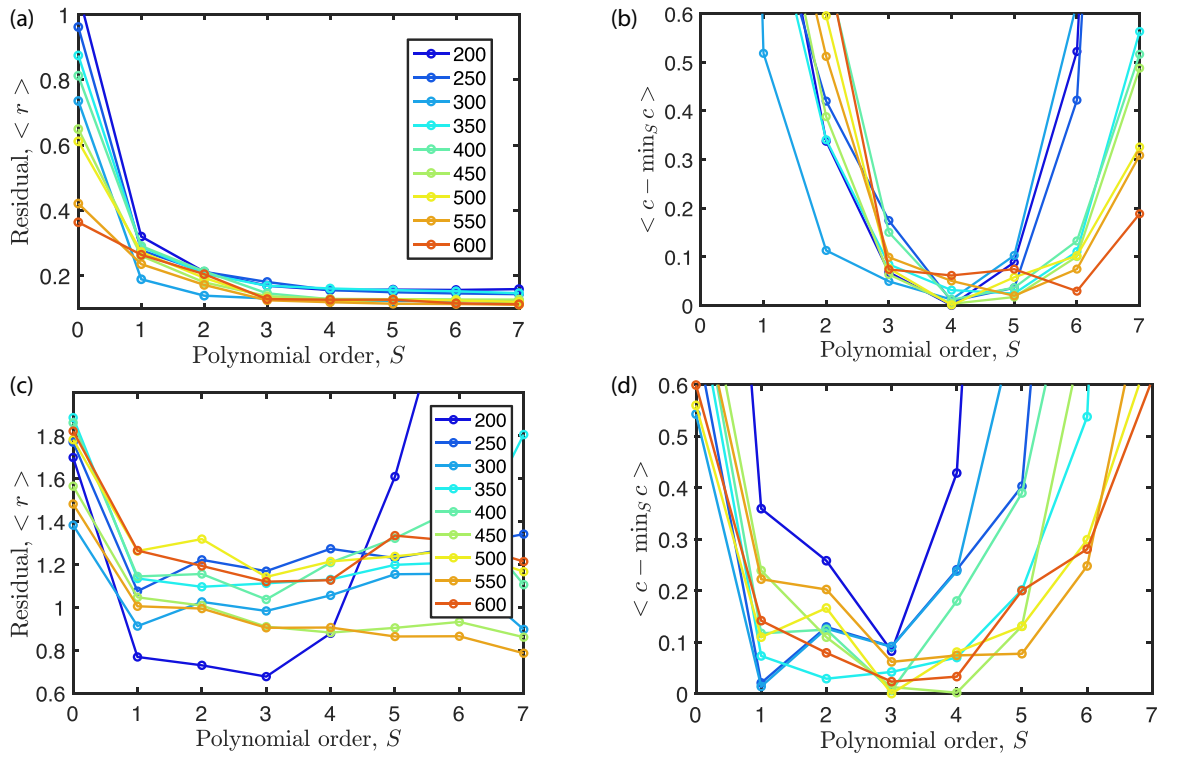

FIG. 6. The effect of the polynomial order S of the MLE estimation of 1,6-Hexanediol diacrylate (HDDA) rates. (a) and (b) Propagation reaction. (c) and (d) Termination reaction. (a) and (c) The estimator residual r as a function of S. (b) and (d) The upper bound c of the residual confidence interval as a function of S. The color scheme indicates the simulation temperature.

[1] T. S. Gardner, C. R. Cantor, and J. J. Collins, Nature (London) **403**, 339 (2000).
[2] I. Kryven, S. Röblitz, and C. Schütte, BMC Syst. Biol. **9**, 67 (2015).
[3] S. Branciamore, E. Gallori, E. Szathmáry, and T. Czárán, J. Mol. Evol. **69**, 458 (2009).
[4] M. Matsumoto, S. Saito, and I. Ohmine, Nature (London) **416**, 409 (2002).
[5] K. Farah, F. Müller-Plathe, and M. C. Böhm, ChemPhysChem. **13**, 1127 (2012).
[6] A. K. Omar and Z.-G. Wang, Phys. Rev. Lett. **119**, 117801 (2017).
[7] A. Torres-Knoop, I. Kryven, V. Schamboeck, and P. Iedema, Soft Matter **14**, 3404 (2018).
[8] A. Torres-Knoop, V. Schamboeck, N. Govindarajan, P. D. Iedema, and I. Kryven, arXiv:2010.08796 [Communications Materials (to be published)].
[9] V. F. Scolari, G. Mercy, R. Koszul, A. Lesne, and J. Mozziconacci, Phys. Rev. Lett. **121**, 057801 (2018).
[10] S. Ciarella, F. Sciortino, and W. G. Ellenbroek, Phys. Rev. Lett. **121**, 058003 (2018).
[11] C. Decker, B. Elzaouk, and D. Decker, J. Macromol. Sci. Chem. **33**, 173 (1996).
[12] T. R. Rooney and R. A. Hutchinson, Industrial Eng. Chem. Res. **57**, 5215 (2018).
[13] D. J. Higham, SIAM Rev. **50**, 347 (2008).
[14] F. P. Buff and D. J. Wilson, J. Chem. Phys. **32**, 677 (1960).
[15] G. H. Weiss, J. Stat. Phys. **42**, 3 (1986).
[16] P. Hänggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[17] P. G. Bolhuis and G. Csányi, Phys. Rev. Lett. **120**, 250601 (2018).
[18] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[19] Source code, https://github.com/ikryven/RateInference.
[20] RMGPY kinetic database, https://rmg.mit.edu.
[21] I. Kryven, Nat. Commun. **10**, 404 (2019).
[22] M. Rubinstein and R. H. Colby, *Polymer Physics* (Oxford University Press, Oxford, UK, 2003).
[23] M. Lang, ACS Macro Lett. **7**, 536 (2018).
[24] R. Wang, A. Alexander-Katz, J. A. Johnson, and B. D. Olsen, Phys. Rev. Lett. **116**, 188302 (2016).
[25] H. D. Rozenfeld, J. E. Kirk, E. M. Bollt, and D. Ben-Avraham, J. Phys. A **38**, 4589 (2005).
[26] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[27] H. Niu, P. M. Piaggi, M. Invernizzi, and M. Parrinello, Proc. Natl. Acad. Sci. USA **115**, 5348 (2018).
[28] J. R. Perilla, B. C. Goh, C. K. Cassidy, B. Liu, R. C. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, Curr. Opin. Struct. Biol. **31**, 64 (2015).
[29] A. K. Buell, J. R. Blundell, C. M. Dobson, M. E. Welland, E. M. Terentjev, and T. P. J. Knowles, Phys. Rev. Lett. **104**, 228101 (2010).
[30] J. Smrek and K. Kremer, Phys. Rev. Lett. **118**, 098002 (2017).
[31] S. Sarkar and S. Lin-Gibson, Adv. Theory Simul. **1**, 1800028 (2018).
[32] I. Kryven, Phys. Rev. E **94**, 012315 (2016).
[33] I. Kryven, J. Math. Chem. **56**, 140 (2018).
[34] V. Schamboeck, P. D. Iedema, and I. Kryven, Sci. Rep. **10**, 14627 (2020).
[35] P. Del Moral and S. Penev, *Stochastic Processes: From Applications to Theory* (Chapman and Hall/CRC, London, UK, 2017).