PHYSICAL REVIEW E **86**, 031126 (2012)

Stochastic analysis of complex reaction networks using binomial moment equations

Baruch Barzel[1,2] and Ofer Biham[1]
[1]*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel*
[2]*Center for Complex Network Research, Northeastern University, Boston, Massachusetts 02115, USA*

(Received 29 August 2011; published 20 September 2012)

The stochastic analysis of complex reaction networks is a difficult problem because the number of microscopic states in such systems increases exponentially with the number of reactive species. Direct integration of the master equation is thus infeasible and is most often replaced by Monte Carlo simulations. While Monte Carlo simulations are a highly effective tool, equation-based formulations are more amenable to analytical treatment and may provide deeper insight into the dynamics of the network. Here, we present a highly efficient equation-based method for the analysis of stochastic reaction networks. The method is based on the recently introduced binomial moment equations [Barzel and Biham, Phys. Rev. Lett. **106**, 150602 (2011)]. The binomial moments are linear combinations of the ordinary moments of the probability distribution function of the population sizes of the interacting species. They capture the essential combinatorics of the reaction processes reflecting their stoichiometric structure. This leads to a simple and transparent form of the equations, and allows a highly efficient and surprisingly simple truncation scheme. Unlike ordinary moment equations, in which the inclusion of high order moments is prohibitively complicated, the binomial moment equations can be easily constructed up to any desired order. The result is a set of equations that enables the stochastic analysis of complex reaction networks under a broad range of conditions. The number of equations is dramatically reduced from the exponential proliferation of the master equation to a polynomial (and often quadratic) dependence on the number of reactive species in the binomial moment equations. The aim of this paper is twofold: to present a complete derivation of the binomial moment equations; to demonstrate the applicability of the moment equations for a representative set of example networks, in which stochastic effects play an important role.

DOI: 10.1103/PhysRevE.86.031126 PACS number(s): 02.50.Fz, 05.10.Gg, 82.40.Qt

I. INTRODUCTION

Reaction networks appear in many scientific contexts, including chemistry [1], biology [2–4], and ecology [5]. These systems are composed of a set of reactants, such as chemical compounds, proteins, or animal species, which undergo a set of reactions, such as chemical binding, protein-protein interactions, and predation. The dynamics of such networks can be characterized by the time dependent copy numbers or concentrations of the interacting species. In many cases, the systems are of macroscopic dimensions and the interacting species appear in large quantities. Under these conditions, the law of large numbers applies and fluctuations in the population sizes or copy numbers are negligible. As a result, these systems can be analyzed using rate equation models that incorporate the law of mass-action kinetics within the mean-field approximation. In this formulation, the discrete nature of the interacting species and the fluctuations in their copy numbers are ignored.

Consider a reaction network in a microscopic system, such as a biological cell or a small accretion surface. The copy numbers of reactive atoms and molecules in such a system may be small and exhibit large fluctuations [6–10]. As a result, rate equations fail and the simulation of these reactions requires stochastic methods based on the master equation [11,12]. The master equation can be solved either by direct numerical integration [13–15] or by Monte Carlo (MC) simulations [16–18]. The master equation provides the probability distribution of the copy numbers of the interacting species, from which the average population sizes can be obtained. In certain cases, the master equation can be solved using a generating function [19–22]. The set of coupled ordinary differential equations is then transformed into a single partial differential equation for the generating function. This equation is solved using numerical methods. In a few cases, it can be solved analytically [23,24]. The master equation can also be approximated using the Fokker-Planck equation [25]. This is a partial differential equation in which the population sizes of the reactive species are represented by continuous variables [26]. The methods described above are useful for simple reaction networks, which involve a small number of interacting species. However, as the number of species increases, the number of equations in the master equation quickly proliferates. This makes the master equation infeasible for complex networks including a large number of interacting species [13,14].

MC methods enable the analysis of larger and more complex networks. In particular, the Gillespie algorithm [16,17] has become the method of choice for the simulation of stochastic chemical, biological, and ecological networks. This is a kinetic Monte Carlo approach, namely, an algorithm that generates "paths" of the stochastic process [27]. The basic idea is simple. At each time step, the next move is drawn from all possible processes that may take place at that point, where each step is endowed with a suitable rate. The time elapsed until the next move is drawn from a Poisson distribution, the average of which is determined by these rates. After each move, the list of available processes is updated and their new rates are evaluated. This method is remarkably efficient in integrating over an exponentially large phase space. The availability of ever more powerful computers has enabled researchers to simulate networks of increasing size and complexity and to analyze them in greater detail.

Improved variants of the algorithm and related methods were developed [28–35]. Some of these improvements are based on the the tau-leaping method [30,36] in which several reactions take place in a single time interval. Other methods are based on the distinction between fast and slow processes, which are treated differently [37–39]. MC methods suffer from several drawbacks. In order to extract the expectation values of desired moments and correlations, one needs to perform statistical analyses over large amounts of noisy data. This may become impractical for highly complex systems. Also, the combination of fast and slow processes tends to reduce the efficiency of the simulations.

Recently, we have developed an approach for the dimensional reduction of the master equation, referred to as the multiplane method. This reduction extends the applicability of the direct integration of the differential equations. The multiplane method, which was developed in the context of chemical reactions on surfaces, provides a significant reduction in the number of equations [40–42]. In the multiplane method, one breaks the network into a set of maximal fully connected subnetworks (maximal cliques). Lower-dimensional master equations are constructed for the marginal probability distributions associated with these cliques, with suitable couplings between them. This reduces the number of equations and extends the feasibility of the stochastic analysis. In the reduced multiplane method, one breaks the network into minimal cliques, including only pairs of interacting nodes, lowering even further the number of equations [43].

Another recently developed method, based on moment equations, provides an efficient analysis of stochastic reaction networks [44–46]. This method exhibits crucial advantages over the current simulation methodologies. As opposed to the exponential proliferation of the master equation, the number of moment equations rises only polynomially with the number of reactive species. In fact, for a wide range of conditions, the number of equations is reduced to merely one equation for each reactive species (node) and one equation for each reaction (edge). Moreover, for any given network, the moment equations, up to second order moments, can be constructed using a diagrammatic scheme [46]. This enables one to automate the construction of the set of equations, a feature which is essential in the case of complex networks. Furthermore, the moment equations are linear in terms of the moments. Thus, the stability and convergence properties can be easily controlled and the steady state solution can be obtained by standard algebraic methods. In the original formulation of the moment equations beyond a single species [45,46], only first and second order moments were included. It was shown that even at this rather low order, the equations are still accurate for a surprisingly wide range of conditions. However, in this formulation, only simple chemical processes of the form $X_1 + X_2 \rightarrow X_3$ were accounted for. The moment equations were used for the analysis of a reaction network that takes place on the surfaces of interstellar dust grains and leads to the formation of water ice and methanol through cascades of hydrogen addition reactions [45,46]. In this network, water ice is formed by the hydrogenation of oxygen atoms, through the reactions $\mathrm{H} + \mathrm{O} \rightarrow \mathrm{OH}$ and $\mathrm{H} + \mathrm{OH} \rightarrow \mathrm{H_2O}$. Water ice on interstellar grains also forms by other reaction channels, through the hydrogenation of $\mathrm{O_2}$ and $\mathrm{O_3}$ molecules [47–49]. An interesting feature of these networks is the competition between the hydrogenation processes and the reaction $\mathrm{H} + \mathrm{H} \rightarrow \mathrm{H_2}$ over the limited supply of hydrogen atoms. The kinetic properties of these networks are highly sensitive to the physical conditions and particularly to the temperature of the grains.

In a broader perspective, reaction networks may exhibit a more complicated behavior. Certain reactions may lead to several products and may take different reaction paths. First order reactions, where a single reactant dissociates into a set of other species, may also be included. In addition, processes such as catalysis, various forms of regulation, predation, competition for limited resources, as well as positive and negative feedback, give rise to a richer and more complicated dynamics in these interaction networks. In the formulation presented in Refs. [44–46], the moment equations can not account for some of these processes. Also, while in the context of chemical networks reactions involving more than two reactants are often ignored, in a broader context, the inclusion of multibody reactions is sometimes important [50]. Moreover, even for the simple reactions considered in Refs. [45,46], there are cases where moments of order higher than two must be included for the moment equations to provide meaningful results. The problem is that the equations for such higher order moments lose their standard form. The diagrammatic formulation no longer applies, and the resulting equations are tediously elaborate.

In this paper, we present a generalized formulation of the moment equations, written in terms of the binomial moments, introduced in Ref. [51]. The binomial moments, as opposed to ordinary moments, have a simple and intuitive physical meaning, which captures the structure of the reactions in the network. They lead to a transparent and easily writable form of the equations. They also enable a simple and highly efficient closure condition for the equations. As opposed to other formulations of the moment equations [52,53] which include only first and second moments, here the truncation can be extended to moments of any desired order, without the need for repeated summations over the master equation.

The aim of this paper is twofold: first, to present the complete derivation of the binomial moment equations from the master equation; second, to demonstrate the applicability of the moment equations to several representative examples of model networks. The example networks are ordered with increasing levels of complexity. The first network consists of a single species and a simple annihilation reaction of n species. The second network involves two species and a reversible dimerization reaction. The third example is a complex reaction network that involves 10 species, which undergo a complex set of reactions. These networks include a large variety of processes that appear in realistic systems and exemplify the structure and dynamics of the binomial moment equations that describe these systems.

The paper is organized as follows. In Sec. II, we present the basic concepts of reaction networks and their modeling using rate equations. The master equation for such networks is presented in Sec. III. In Sec. IV, we derive the binomial moment equations from the master equation. In Sec. V, the binomial moment equations are applied to the analysis of several stochastic networks of increasing complexity. The results are summarized and discussed in Sec. VI.

II. REACTION NETWORKS

Consider a chemical reaction network, consisting of J molecular species X_i, $i = 1,\ldots,J$. The molecules may undergo reactions of the form

$$\sum_{i=1}^{J} n_i X_i \to \sum_{j=1}^{J} m_j X_j, \tag{1}$$

where the stoichiometric coefficients n_i and m_i are integers. The order of the reaction is given by $n = \sum_{i=1}^{J} n_i$, namely, the number of reactants on the left-hand side of Eq. (1). In the context of chemical reactions, one often omits reactions of order higher than $n = 2$. However, in the formulation presented below there is no such limitation. The reaction appearing in Eq. (1) can be expressed in a vector form as

$$\vec{n} \to \vec{m}, \tag{2}$$

namely, a configuration of molecules with stoichiometric coefficients given by $\vec{n}$, reacts to form a new configuration given by $\vec{m}$. These vectors are J dimensional. The rate constant for the configuration $\vec{n}$ to react is given by $T_{\vec{n}}$ (s^{-1}). Certain reactions may take several reaction paths with different probabilities or branching ratios. The probability that the reaction will result in the configuration $\vec{m}$ is denoted by $P_{\vec{n}}^{\vec{m}}$. These probabilities satisfy $\sum_{\vec{m}} P_{\vec{n}}^{\vec{m}} = 1$. Thus, the rate constant for the reaction $\vec{n} \to \vec{m}$ to occur is given by $T_{\vec{n}} P_{\vec{n}}^{\vec{m}}$.

The mean-field rate equations are expressed in terms of the average population sizes $\langle N_i \rangle$. The total reaction rate for the reaction $\vec{n} \to \vec{m}$ (for all choices of $\vec{m}$) is proportional to the number of different ways one can gather a set of reactants given by $\vec{n}$. For a macroscopic system with an average of $\langle N_i \rangle$ copies of X_i, where $\langle N_i \rangle \gg n_i$, this number is well approximated by $\prod_{i=1}^{J} \langle N_i \rangle^{n_i} / n_i!$. The rate equations for the time dependent average population sizes take the form

$$\frac{d\langle N_i \rangle}{dt} = \sum_{\vec{n}\vec{m}} m_i T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \left(\prod_{j=1}^{J} \frac{\langle N_j \rangle^{n_j}}{n_j!} \right) - \sum_{\vec{n}} n_i T_{\vec{n}} \left(\prod_{j=1}^{J} \frac{\langle N_j \rangle^{n_j}}{n_j!} \right), \tag{3}$$

where $i = 1,\ldots,J$. The first term accounts for positive contributions of the reactions to $\langle N_i \rangle$, while the second term accounts for the negative contributions. For each reaction of the form $\vec{n} \to \vec{m}$, the population of X_i is reduced by n_i molecules and replenished by m_i molecules. Equation (3) provides accurate results as long as the system is large and the average population sizes satisfy $\langle N_i \rangle \gg 1$. However, for small systems, where the average population sizes are of order unity or less, fluctuations in the populations of the reactive species become significant. These fluctuations are not accounted for by the rate equations, and stochastic methods are required.

III. MASTER EQUATION

In the stochastic analysis, one describes the microscopic state of the system by the vector $\vec{N} = (N_1,\ldots,N_J)$, where N_i is the number of copies of species X_i. In the following, we introduce a combinatorial approach, which allows us to express the master equation and the moment equations in a transparent form. Let $\vec{v} = (v_1,\ldots,v_J)$ be a vector of integers. It can be expressed as a linear combination of the J basis vectors $\vec{e}_i = (0,\ldots,v_i = 1,\ldots,0)$. We denote by $\mathcal{Q}_{\vec{v}}$ a combination of molecules consisting of exactly v_i copies of the species X_i.

Consider a reaction network in the state $\vec{N}$. The number of $\mathcal{Q}_{\vec{v}}$ combinations that exist in this system is given by

$$W(\vec{N},\vec{v}) = \binom{\vec{N}}{\vec{v}}, \tag{4}$$

where

$$\binom{\vec{N}}{\vec{v}} = \prod_{i=1}^{J} \binom{N_i}{v_i}, \tag{5}$$

and $\binom{N_i}{v_i}$ is the binomial coefficient. To illustrate the motivation for this definition, consider the reaction $\vec{n} \to \vec{m}$. It occurs at a rate proportional to $T_{\vec{n}} P_{\vec{n}}^{\vec{m}}$, and to the number of $\mathcal{Q}_{\vec{n}}$ combinations which are present in the system, given by $W(\vec{N},\vec{n})$.

In writing the master equation, we use the increment operator defined by $\mathcal{S}^{\vec{u}} f(\vec{N}) = f(\vec{N} + \vec{u})$, where f is any function of the state $\vec{N}$. For example, the reaction $\vec{n} \to \vec{m}$ can be written as $\vec{N} \to \mathcal{S}^{\vec{m}} \mathcal{S}^{-\vec{n}} \vec{N}$. Let $P(\vec{N})$ represent the time dependent probability for the system to be in the state $\vec{N}$. The master equation for this system takes the form

$$\frac{dP(\vec{N})}{dt} = \sum_{\vec{n},\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} [\mathcal{S}^{\vec{n}} \mathcal{S}^{-\vec{m}} - \mathcal{I}] W(\vec{N},\vec{n}) P(\vec{N}), \tag{6}$$

where $\mathcal{I}$ is the identity operator. In this equation, one sums over all the reactions $\vec{n} \to \vec{m}$ which may take place in the system. These reactions yield a positive contribution to $P(\vec{N})$ if the state of the system is $\vec{N} + \vec{n} - \vec{m}$, and a negative contribution to $P(\vec{N})$ if the system is in the state $\vec{N}$. The truncation of the master equation is achieved by setting upper cutoffs C_i, $i = 1,\ldots,J$, such that $P(\vec{N}) = 0$ if $N_i > C_i$ for any value of i. The number of coupled equations is thus $N_E = \prod_{i=1}^{J} (C_i + 1)$. This number grows exponentially with the number of reactive species J. This feature severely limits the applicability of the master equation to complex reaction networks [13,14].

IV. BINOMIAL MOMENT EQUATIONS

A more compact description of stochastic reaction networks can be obtained using moment equations [52]. These equations, derived by tracing over the master equation, consist of ordinary differential equations for the time derivatives of the moments $\langle N_1^{k_1} \ldots N_J^{k_J} \rangle$, where k_i are integers. The order of a certain moment is given by $k = \sum_{i=1}^{J} k_i$. The difficulty with the moment equations is that higher order moments appear on the right-hand sides of these equations. To obtain a closed set of equations, one needs to apply a suitable truncation scheme, in which the higher order moments are expressed in terms of low order moments [44–46,52,53]. In practice, the truncation is typically done at the level of third order moments, namely, only first and second order moments are taken into account. Making the truncation at higher orders turns out to be difficult even for relatively simple networks.

Here, we present a different formulation of the moment equations, based on the binomial moments, introduced in Ref. [51]. The binomial moments are linear combinations of the moments of $P(\vec{N})$ that capture the structure of the reactions that take place in the system. Due to this property, the equations we obtain for these binomial moments take a simple and transparent form up to any desired order and can be easily constructed. Unlike the ordinary moments, the binomial moments tend to decrease with increasing order. This enables us to use a particularly simple truncation scheme where moments beyond a specified order are neglected.

Consider a reaction network described by the probability distribution $P(\vec{N})$. The binomial moment $\langle W_{\vec{v}} \rangle$ is defined as the average number of combinations of the form $\mathcal{Q}_{\vec{v}}$ that appear in the system. It is given by

$$\langle W_{\vec{v}} \rangle = \sum_{\vec{N}} W(\vec{N},\vec{v}) P(\vec{N}). \tag{7}$$

To understand the meaning of the binomial moments, consider the case where $\vec{v} = \vec{e}_i$. Here, the corresponding binomial moment $\langle W_{\vec{e}_i} \rangle$ is the average number of combinations consisting of a single copy of the species X_i. This is simply the average population size $\langle N_i \rangle$. In case that $\vec{v} = \vec{e}_i + \vec{e}_j$, the corresponding moment is $\langle W_{\vec{v}} \rangle = \langle N_i N_j \rangle$, which stands for the average number of X_i-X_j pairs present in the system. Note that for $\vec{v} = 2\vec{e}_i$, the corresponding binomial moment is $\langle W_{\vec{v}} \rangle = (\langle N_i^2 \rangle - \langle N_i \rangle)/2$. Similarly, the number of X_i triplets is given by $\langle W_{\vec{v}} \rangle = (\langle N_i^3 \rangle - 3\langle N_i^2 \rangle + 2\langle N_i \rangle)/6$, where $\vec{v} = 3\vec{e}_i$. The order of the binomial moment $\langle W_{\vec{v}} \rangle$ is

$$k = \sum_{i=1}^{J} v_i. \tag{8}$$

The physical meaning of a binomial moment $\langle W_{\vec{v}} \rangle$ of order k is the average number of k-plets with a configuration given by $\vec{v}$ in the system.

Consider a system in the state $\vec{N} = \vec{n}$ undergoing the reaction $\vec{n} \to \vec{w}$. This reaction generates w_i new X_i molecules, thus adding $\binom{\vec{w}}{\vec{m}}$ new $\mathcal{Q}_{\vec{m}}$ combinations to the system. The total rate by which reactions of the form $\vec{n} \to \vec{w}$ (for all possible choices of $\vec{w}$) contribute to the number of $\mathcal{Q}_{\vec{m}}$ combinations is thus given by

$$B_{\vec{n}}^{\vec{m}} = \sum_{\vec{w}} T_{\vec{n}} P_{\vec{n}}^{\vec{w}} \binom{\vec{w}}{\vec{m}}. \tag{9}$$

To obtain the binomial moment equations, we use Eq. (7) to express the time derivatives of the moments $\langle W_{\vec{v}} \rangle$ in the form

$$\frac{d\langle W_{\vec{v}} \rangle}{dt} = \sum_{\vec{N}} W(\vec{N},\vec{v}) \frac{dP(\vec{N})}{dt}. \tag{10}$$

The time derivative of the probability $P(\vec{N})$ is taken from the master equation [Eq. (6)]. In order to apply the summation, we first introduce the combinatorial equalities

$$\begin{aligned} \mathcal{S}^{\vec{u}} \binom{\vec{N}}{\vec{v}} &= \sum_{\vec{w}} \binom{\vec{N}}{\vec{v}-\vec{w}} \binom{\vec{u}}{\vec{w}}, \\ \mathcal{S}^{-\vec{u}} \binom{\vec{N}}{\vec{v}} &= \binom{\vec{N}}{\vec{u}}^{-1} \binom{\vec{v}+\vec{u}}{\vec{u}} \binom{\vec{N}}{\vec{v}+\vec{u}}, \\ \binom{\vec{N}}{\vec{n}} \binom{\vec{N}}{\vec{v}} &= \sum_{\vec{w}} \binom{\vec{n}}{\vec{w}} \binom{\vec{v}+\vec{w}}{\vec{n}} \binom{\vec{N}}{\vec{v}+\vec{w}}. \end{aligned} \tag{11}$$

The proofs of these equalities are provided in the Appendix. Inserting the time derivative of $P(\vec{N})$ into Eq. (10) leads to

$$\frac{d\langle W_{\vec{v}} \rangle}{dt} = \sum_{\vec{n},\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \sum_{\vec{N}} \left\{ \binom{\vec{N}}{\vec{v}} [\mathcal{S}^{\vec{n}} \mathcal{S}^{-\vec{m}} - \mathcal{I}] \binom{\vec{N}}{\vec{n}} P(\vec{N}) \right\}. \tag{12}$$

By changing the order of terms in the summation and switching the signs of the increment operators, we obtain

$$\frac{d\langle W_{\vec{v}} \rangle}{dt} = \sum_{\vec{n},\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \sum_{\vec{N}} \left\{ \binom{\vec{N}}{\vec{n}} P(\vec{N}) [\mathcal{S}^{-\vec{n}} \mathcal{S}^{\vec{m}} - \mathcal{I}] \binom{\vec{N}}{\vec{v}} \right\}. \tag{13}$$

Using the three equalities of Eq. (11), we write

$$\begin{aligned} \frac{d\langle W_{\vec{v}} \rangle}{dt} = \sum_{\vec{n},\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \sum_{\vec{N}} \Bigg\{ P(\vec{N}) \Bigg[\sum_{\vec{w}} \binom{\vec{m}}{\vec{w}} \binom{\vec{v}+\vec{n}-\vec{w}}{\vec{n}} \\ \times \binom{\vec{N}}{\vec{v}+\vec{n}-\vec{w}} - \sum_{\vec{w}} \binom{\vec{n}}{\vec{w}} \binom{\vec{v}+\vec{w}}{\vec{n}} \binom{\vec{N}}{\vec{v}+\vec{w}} \Bigg] \Bigg\}. \end{aligned} \tag{14}$$

We now collect all the terms, and apply the summation over $\vec{N}$ to express the right-hand side in terms of the binomial moments. The equation takes the form

$$\begin{aligned} \frac{d\langle W_{\vec{v}} \rangle}{dt} &= \sum_{\vec{n},\vec{w}} \left[\sum_{\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \binom{\vec{m}}{\vec{w}} \right] \left[\binom{\vec{v}+\vec{n}-\vec{w}}{\vec{n}} \langle W_{\vec{v}+\vec{n}-\vec{w}} \rangle \right] \\ &\quad - \sum_{\vec{n},\vec{w}} \left(\sum_{\vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \right) \binom{\vec{n}}{\vec{w}} \binom{\vec{v}+\vec{w}}{\vec{n}} \langle W_{\vec{v}+\vec{w}} \rangle. \end{aligned} \tag{15}$$

Using Eq. (9), and the fact that $\sum_{\vec{m}} P_{\vec{n}}^{\vec{m}} = 1$, the resulting binomial moment equation is

$$\begin{aligned} \frac{d\langle W_{\vec{v}} \rangle}{dt} = \sum_{\vec{n},\vec{m}} \Bigg[B_{\vec{n}}^{\vec{m}} \binom{\vec{v}+\vec{n}-\vec{m}}{\vec{n}} \langle W_{\vec{v}+\vec{n}-\vec{m}} \rangle \\ - T_{\vec{n}} \binom{\vec{n}}{\vec{m}} \binom{\vec{v}+\vec{m}}{\vec{n}} \langle W_{\vec{v}+\vec{m}} \rangle \Bigg], \end{aligned} \tag{16}$$

where, for the sake of convenience, we exchanged the vectors $\vec{m}$ and $\vec{w}$ in the summation. The first term accounts for positive contributions of the reactions to $\langle W_{\vec{v}} \rangle$, while the second term accounts for the negative contributions. We first refer to the positive contributions. Consider a single combination $\mathcal{Q}_{\vec{v}+\vec{n}-\vec{m}}$. There are on average $\langle W_{\vec{v}+\vec{n}-\vec{m}} \rangle$ such combinations in the system. For each one of these combinations, the reaction

$\vec n \to \vec m$ produces a new $Q_{\vec v}$ combination, so $\langle W_{\vec v}\rangle$ increases. The rate in which it increases is given by the product of the rate constant $B_{\vec n}^{\vec m}$ and the binomial coefficient $\binom{\vec v+\vec n-\vec m}{\vec n}$. We now refer to the negative contributions. Consider the combination $Q_{\vec v+\vec m}$, which undergoes a reaction of the form $\vec n \to \vec w$, for any possible choice of $\vec w$. The overall rate of these reactions is given by $T_{\vec n}\binom{\vec v+\vec m}{\vec n}$. Each time such a reaction takes place, a single combination $Q_{\vec n}$ is removed from the system. The removed combination $Q_{\vec n}$ can be decomposed into $Q_{\vec m}$ and $Q_{\vec n-\vec m}$. Note that there are $\binom{\vec n}{\vec m}$ different possibilities to perform this decomposition of $Q_{\vec n}$. When $Q_{\vec m}$ is removed, the combination $Q_{\vec v+\vec m}$ is replaced by $Q_{\vec v}$. This $Q_{\vec v}$ combination is then eliminated by the subsequent removal of $Q_{\vec n-\vec m}$.

Equation (16) is not in a closed form because higher order moments appear on the right-hand side of each equation. However, the binomial moments tend to decrease as their order increases. To demonstrate this feature, consider a binomial moment $\langle W_{\vec v}\rangle$ of order k. It represents the average number of appearances of a certain combination $Q_{\vec v}$ which consists of k molecules in the system [Eq. (8)].

In small systems, where the average copy numbers are small, $\langle W_{\vec v}\rangle$ tends to decrease as k increases. To demonstrate this feature, consider the binomial moment

$$\langle W_{\vec v}\rangle = \sum_{\substack{N_1=0\\ \vdots \\ N_J=0}}^{\infty} \binom{N_1}{v_1}\binom{N_2}{v_2}\cdots\binom{N_J}{v_J}P(\vec N). \tag{17}$$

The product of the binomial coefficients on the right-hand side vanishes for all cases in which $N_i < v_i$, for at least one value of i. Therefore, we can rewrite the sum as

$$\langle W_{\vec v}\rangle = \sum_{\substack{N_1=v_1\\ \vdots \\ N_J=v_J}}^{\infty} \binom{N_1}{v_1}\binom{N_2}{v_2}\cdots\binom{N_J}{v_J}P(\vec N). \tag{18}$$

For a moment $\langle W_{\vec v}\rangle$ of order k, the sum in Eq. (7) includes only terms for which $\sum_{i=1}^{J} N_i \geqslant k$. In the small system limit, $P(\vec N)$ quickly decreases as the N_i's are increased. Therefore, the binomial moments quickly decrease as their order increases.

This property enables us to use the following truncation scheme. We choose a cutoff C such that $\langle W_{\vec v}\rangle$ is set to zero whenever $k > C$. The number of different binomial moments of order k is given by $\binom{k+J-1}{k}$. Thus, the number of binomial moment equations, after the truncation, is

$$N_E = \sum_{k=1}^{C}\binom{k+J-1}{k}. \tag{19}$$

While the number of equations in the master equation grows exponentially with J, the number of binomial moment equations scales only polynomially with J. Moreover, in practice one can obtain accurate results using surprisingly low values of the cutoffs. In the earlier version of this formulation, it was shown that for a broad range of conditions, the cutoff $C = 2$ is sufficient [45,46]. In this case, the equations include first order moments that account for the average copy numbers and second order moments that account for the number of pairs of species X_i and X_j in the system. In cases where a cutoff of $C = 2$ is not sufficient, one may raise the cutoff until accurate results are obtained. In the small system limit, the required cutoff is usually low. In the large system limit, stochastic equations are no longer required and can be replaced by the rate equations. Somewhat related closure conditions appear in transport-limited reactions, where the local densities are low. In these systems, the three-particle correlation functions are expressed in terms of two-particle correlation functions [54].

To obtain a deeper understanding of the binomial moment equations, we compare them with the rate equations (3) in the limit of large populations. The binomial moment equations for the first moments, which represent the average population sizes of the different reactive species, take the form

$$\frac{d\langle W_{\vec e_i}\rangle}{dt} = \sum_{\vec n,\vec m} w_i T_{\vec n} P_{\vec n}^{\vec m}\langle W_{\vec n}\rangle - \sum_{\vec n} n_i T_{\vec n}\langle W_{\vec n}\rangle. \tag{20}$$

In the deterministic limit, one can write

$$\langle W_{\vec v}\rangle = \left\langle\binom{\vec N}{\vec v}\right\rangle \simeq \prod_{i=1}^{J}\left\langle\binom{N_i}{v_i}\right\rangle. \tag{21}$$

For large populations, this can be further approximated by

$$\langle W_{\vec v}\rangle \simeq \prod_{i=1}^{J}\frac{\langle N_i\rangle^{v_i}}{v_i!}. \tag{22}$$

Under these conditions, the binomial moment equations for the first moments [Eq. (20)] are reduced to the form of the rate equations [Eq. (3)].

V. APPLICATIONS AND RESULTS

To demonstrate the applicability of the binomial moment equations, we consider several examples. The first example is a simple reaction network involving a single reactive species undergoing an n-body annihilation reaction of the form $X + X + \cdots + X \to \varnothing$. The second example describes a dimerization-dissociation reaction in which the resulting dimer species is also included in the system. In the third example, we show the applicability of the moment equations to a complex network consisting of many reactive species and an intricate pattern of reactions between them.

A. Annihilation processes

Consider a system which includes a single reactive species X. Here, the state of the system is described by the scalar value N, which denotes the number of copies of X molecules present in the system. The reactions in this system, presented in the form $\vec n \to \vec m$, are $0 \to 1$ (addition of a single molecule, or $\varnothing \to X$) and $n \to 0$ (annihilation of a $Q_{\vec n}$ combination consisting of n molecules, or $X + X + \cdots + X \to \varnothing$). The rates are T_0 for the addition process and T_n for the annihilation process. Each reaction in this system exhibits a single reaction path, namely, $P_0^1 = P_n^0 = 1$. The only nonzero components of $B_{\vec n}^{\vec m}$ are $B_0^0 = T_0$, $B_0^1 = T_0$, and $B_n^0 = T_n$. The rate equation

[Eq. (3)] for this system takes the form

$$\frac{d\langle N\rangle}{dt} = T_0 - nT_n\frac{\langle N\rangle^n}{n!}, \tag{23}$$

where $\langle N\rangle$ is the time dependent average population size of the species X. The first term describes the formation of X molecules and the second term accounts for the n-body reactions. The steady state population size is obtained by setting the time derivative on the left-hand side to zero. The result is

$$\langle N\rangle = [(n-1)!S]^{\frac{1}{n}}, \tag{24}$$

where $S = T_0/T_n$. Since $\langle N\rangle \propto S^{1/n}$, we refer to $S^{1/n}$ as the system size. The reaction rate under steady state conditions is given by

$$R = T_n\frac{\langle N\rangle^n}{n!} = \frac{T_0}{n}. \tag{25}$$

This solution is valid as long as the population size is large, and the system features deterministic behavior. These conditions are satisfied in the limit where $S \gg 1$. However, in the limit where $S \ll 1$, the average population size becomes small, and stochastic effects become important.

In the stochastic regime, the dynamics must be described by stochastic methods based on the master equation. The master equation for this system takes the form

$$\begin{aligned}\frac{dP(N)}{dt} &= T_0\,[P(N-1) - P(N)]\\ &\quad + T_n\left[\binom{N+n}{n}P(N+n) - \binom{N}{n}P(N)\right].\end{aligned} \tag{26}$$

The binomial moments for this system are $\langle W_v\rangle$, which represent the average number of $\mathcal{Q}_v$ combinations present in the system. The equation for the moment $\langle W_v\rangle$ is

$$\frac{d\langle W_v\rangle}{dt} = T_0\langle W_{v-1}\rangle - \sum_{m=0}^{n-1} T_n\binom{n}{m}\binom{v+m}{n}\langle W_{v+m}\rangle. \tag{27}$$

The first term of the equation accounts for the formation process. When this process occurs, at a rate T_0, each combination $\mathcal{Q}_{v-1}$, of which there are on average $\langle W_{v-1}\rangle$, turns into a combination $\mathcal{Q}_v$, thus contributing to $\langle W_v\rangle$. The second term accounts for the n-body annihilation process. In this process, a combination $\mathcal{Q}_{v+m}$ is stripped of n molecules. As long as $m < n$, this will result in the elimination of a single $\mathcal{Q}_v$ combination. The rate for this to occur is proportional to the rate constant T_n and to the number of $\mathcal{Q}_n$ combinations, which is given by $\binom{v+m}{n}$. Finally, the binomial prefactor $\binom{n}{m}$ accounts for the number of ways to choose the *first* m molecules to be removed, after which the remaining $n - m$ molecules are withdrawn from the resulting $\mathcal{Q}_v$ combination.

In the stochastic limit, where $S \ll 1$, we may assign a strict cutoff on the binomial moment equations of $C = n$. According to this cutoff, binomial moments of order higher than n are set to zero. The meaning of this cutoff is that we assume that the appearance of combinations with more than n copies of the species X in the system is rare. This assumption is valid as long as the system is sufficiently small. Note that this is the minimal cutoff one can take without completely terminating the reaction process. The resulting truncated binomial moment equations are

$$\begin{aligned}\frac{d\langle W_1\rangle}{dt} &= T_0 - nT_n\langle W_n\rangle,\\ \frac{d\langle W_2\rangle}{dt} &= T_0\langle W_1\rangle - T_n\binom{n}{2}\langle W_n\rangle,\\ &\vdots\\ \frac{d\langle W_n\rangle}{dt} &= T_0\langle W_{n-1}\rangle - T_n\langle W_n\rangle.\end{aligned} \tag{28}$$

The average population size is given by the binomial moment $\langle W_1\rangle$, and the reaction rate is given by $R = T_n\langle W_n\rangle$. Under steady state conditions, the binomial moment equations yield

$$\langle W_1\rangle = \frac{n-1}{2}, \quad R = \frac{T_0}{n}. \tag{29}$$

Note that the reaction rate R is obtained directly from the first equation in Eqs. (28). This means that extending the truncation to higher order moments, adding more equations, will not affect the value of R. Thus, the reaction rate, as obtained from the moment equations is exact, regardless of the size of the system. It also agrees with the rate equation result [Eq. (25)]. In contrast, in the stochastic regime, the result for the first moment $\langle W_1\rangle$, representing the average population size, does not agree with the result of the rate equation [Eq. (24)]. In this regime, where $T_n \gg T_0$ ($S \ll 1$), the reactions dominate the behavior of the system. Thus, molecules are formed at a constant rate of T_0, and once there are n molecules present in the system, a reaction occurs almost instantaneously, and sweeps the system clean of molecules. The distribution $P(N)$ is uniform as long as $N < n$, and vanishes for $N \geqslant n$. The average population size $\langle W_1\rangle$ is thus given by Eq. (29).

The binomial moment equations also enable us to obtain the variance of the population $\sigma^2 = \langle N^2\rangle - \langle N\rangle$. Using the binomial moments, this can be expressed as

$$\sigma^2 = 2\langle W_2\rangle + \langle W_1\rangle - \langle W_1\rangle^2. \tag{30}$$

Using Eq. (28), we find that $\sigma^2 = (n^2 - 1)/12$.

The results obtained above for $\langle W_1\rangle$, in which the binomial moment equations were truncated at $C = n$, are valid only in the limit $S \ll 1$. To extend the domain of validity of the results, one needs to add more equations, accounting for higher order moments. Raising the cutoff of the binomial moment equations to $C = n + 1$ provides the first order correction to Eq. (29). With this truncation, the binomial moment equations become

$$\begin{aligned}\frac{d\langle W_1\rangle}{dt} &= T_0 - nT_n\langle W_n\rangle,\\ \frac{d\langle W_2\rangle}{dt} &= T_0\langle W_1\rangle - T_n\binom{n}{2}\langle W_n\rangle - T_n n(n+1)\langle W_{n+1}\rangle,\\ &\vdots\\ \frac{d\langle W_{n+1}\rangle}{dt} &= T_0\langle W_n\rangle - T_n(n+1)\langle W_{n+1}\rangle.\end{aligned} \tag{31}$$

Here, the steady state population size is

$$\langle W_1\rangle = \frac{n-1}{2} + S, \tag{32}$$

which includes a first order correction in terms of S. The variance obtained from the moment equations with this truncation is

$$\sigma^2 = \frac{n^2-1}{12} + \left(\frac{3n-1}{2}\right) S - S^2. \tag{33}$$

The next order correction can be obtained by adding the equation for $\langle W_{n+2}\rangle$. This leads to

$$\begin{aligned}\langle W_1\rangle &= \frac{n-1}{2} + \frac{(n+1)T_0}{(N+1)T_n + nT_0}\\ &\simeq \frac{n-1}{2} + S - \frac{n}{n+1}S^2.\end{aligned} \tag{34}$$

The analysis above demonstrates the fact that unlike the ordinary moment equations, in the binomial moment equations it is straightforward to add higher order moments until sufficiently accurate results are obtained.

In Fig. 1, we present results for a system that exhibits a three-body annihilation reaction, namely, $n = 3$. We show the average population size $\langle W_1\rangle$ versus S, as obtained from the binomial moment equations for different cutoffs. The master equation results are also shown (solid line). For a cutoff of $C = 3$ (circles), the results of the binomial moment equations coincide with those of the master equation in the limit of $S \ll 1$. As the cutoff is raised to $C = 4$ (squares), $C = 5$ (triangles), and $C = 6$ (diamonds), the range of validity of the binomial moment equations is extended. Finally, for a cutoff of $C = 30$ (+), the binomial moment equation results agree with the master equation for the entire range of parameters that is displayed. The rate equation results (dashed line) are accurate only in the limit where $S \gg 1$, but show significant deviations in the stochastic regime, where $S < 1$. Note that already for $C = 6$ a smooth convergence to the rate equation results is obtained. Thus, in practice, for this system, it is sufficient to include merely five or six equations. If the system requires more equations than that, it is already safe to use the rate equations. To appreciate the efficiency of the binomial moment equations, we focus on the limit where $S \gg 1$, shown on the right-hand side of Fig. 1. In this limit, the average population reaches approximately $\langle W_1\rangle = 60$. Still, the binomial moment equations achieve accurate results even though the cutoff was set only at $C = 30$. This exemplifies that using the binomial moment equations with the closure condition proposed here allows one to choose very restrictive cutoffs, and obtain valid results even with a relatively small amount of equations.

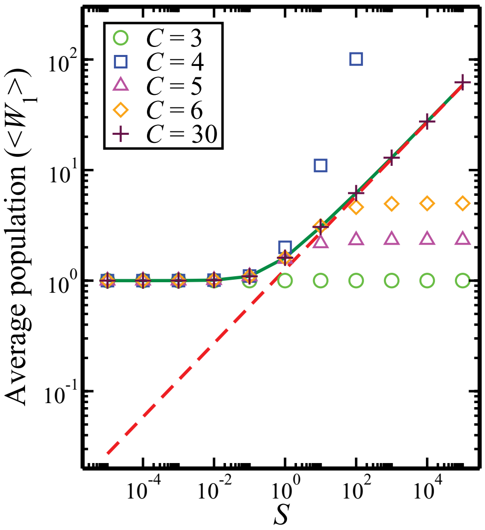

FIG. 1. (Color online) The average population size $\langle W_1\rangle$ vs the system size S for a three-body annihilation reaction as obtained from the binomial moment equations with a cutoff of $C = 3$ (circles), $C = 4$ (squares), $C = 5$ (triangles), $C = 6$ (diamonds), and $C = 30$ (+). For small systems, the results are in perfect agreement with those obtained from the master equation (solid line). By raising the cutoff, we obtain accurate results from the moment equations for a wider range of system sizes. Beyond the cutoff of $C = 5$ or 6, the rate equations can be used (dashed line). A cutoff of 30 is sufficient to cover the entire displayed range. In the small system limit, the rate equations deviate significantly, and the binomial moment equations are needed.

B. Dimerization reaction

The next example we consider is the dimerization-dissociation system. This system includes two species, X_1 and X_2, undergoing the following reaction:

$$X_1 + X_1 \rightleftharpoons X_2. \tag{35}$$

Using the vector notation, this system is described by two reactions: $2\vec{e}_1 \to \vec{e}_2$ with the rate constant $T_{2\vec{e}_1} = \alpha$, and $\vec{e}_2 \to 2\vec{e}_1$ with the rate constant $T_{\vec{e}_2} = \beta$. The rate equations for this system take the form

$$\begin{aligned}\frac{d\langle N_1\rangle}{dt} &= -\alpha\langle N_1\rangle^2 + 2\beta\langle N_2\rangle,\\ \frac{d\langle N_2\rangle}{dt} &= \frac{1}{2}\alpha\langle N_1\rangle^2 - \beta\langle N_2\rangle,\end{aligned} \tag{36}$$

where $\langle N_1\rangle(t)$ and $\langle N_2\rangle(t)$ are the average population sizes of X_1 and X_2, respectively. Note that the total number of individual copies of X_1, whether single or dimerized, is conserved. This is expressed in the rate equations by the conservation law $dN_0/dt = 0$, where $N_0 = \langle N_1\rangle + 2\langle N_2\rangle$ is the overall amount of X_1 molecules, determined by the initial conditions of the system. The time dependent solution of Eq. (36) can be obtained by separation of variables. It is given by

$$\langle N_1\rangle(t) = \frac{-\beta+\omega}{2\alpha} + \frac{\omega e^{-\omega(t-t_0)}}{\alpha(1-e^{-\omega(t-t_0)})}, \tag{37}$$

where $\omega = \sqrt{\beta^2 + 4\alpha\beta N_0}$ and t_0 is determined by the initial conditions of the system. The relaxation time is $\tau = 1/\omega$, and the steady state populations are

$$\langle N_1\rangle = \frac{-\beta+\omega}{2\alpha}, \quad \langle N_2\rangle = \frac{2\alpha N_0 + \beta - \omega}{4\alpha}. \tag{38}$$

As long as $N_0 \gg 1$, the system obeys the law of mass action kinetics, and the rate equations apply. However, when N_0 is of order unity, the system is stochastic and the rate equations are no longer reliable. To account for the stochasticity, we refer to the moment equations. The nonvanishing components of $B^{\vec{m}}_{\vec{n}}$ are $B^{0}_{2\vec{e}_1} = \alpha$, $B^{\vec{e}_2}_{2\vec{e}_1} = \alpha$, $B^{0}_{\vec{e}_2} = \beta$, $B^{\vec{e}_1}_{\vec{e}_2} = 2\beta$, and $B^{2\vec{e}_1}_{\vec{e}_1} = \beta$. The binomial moment equations for the dimerization-dissociation system take the form

$$\begin{aligned}\frac{d\langle W_{\vec{v}}\rangle}{dt} = {} & \alpha\binom{\vec{v}+2\vec{e}_1-\vec{e}_2}{2\vec{e}_1}\langle W_{\vec{v}+2\vec{e}_1-\vec{e}_2}\rangle + 2\beta\binom{\vec{v}+\vec{e}_2-\vec{e}_1}{\vec{e}_2} \\ & \times\langle W_{\vec{v}+\vec{e}_2-\vec{e}_1}\rangle + \beta\binom{\vec{v}+\vec{e}_2-2\vec{e}_1}{\vec{e}_2}\langle W_{\vec{v}+\vec{e}_2-2\vec{e}_1}\rangle \\ & - \left[\alpha\binom{\vec{v}}{2\vec{e}_1} + \beta\binom{\vec{v}}{\vec{e}_2}\right]\langle W_{\vec{v}}\rangle - 2\alpha\binom{\vec{v}+\vec{e}_1}{2\vec{e}_1}\langle W_{\vec{v}+\vec{e}_1}\rangle.\end{aligned} \tag{39}$$

Here, the truncation of the equations derives from the conservation of N_0, leading to $\langle W_{\vec{v}}\rangle = 0$ for any vector $\vec{v}$ where $v_1 + 2v_2 > N_0$. Note that this truncation is exact since for a system with overall N_0 copies of X_1 there can not be any combinations $Q_{\vec{v}}$ that do not satisfy the above condition.

For $N_0 = 2$, there are three nonvanishing binomial moments $\langle W_{\vec{e}_1}\rangle$, $\langle W_{2\vec{e}_1}\rangle$, and $\langle W_{\vec{e}_2}\rangle$. The equations for these moments are

$$\begin{aligned}\frac{d\langle W_{\vec{e}_1}\rangle}{dt} &= 2\beta\langle W_{\vec{e}_2}\rangle - 2\alpha\langle W_{2\vec{e}_1}\rangle, \\ \frac{d\langle W_{2\vec{e}_1}\rangle}{dt} &= \beta\langle W_{\vec{e}_2}\rangle - \alpha\langle W_{2\vec{e}_1}\rangle, \\ \frac{d\langle W_{\vec{e}_2}\rangle}{dt} &= \alpha\langle W_{2\vec{e}_1}\rangle - \beta\langle W_{\vec{e}_2}\rangle.\end{aligned} \tag{40}$$

Here, two conservation laws emerge. The first conservation law is $\langle W_{\vec{e}_1}\rangle + 2\langle W_{\vec{e}_2}\rangle = N_0$, which is equivalent to the overall conservation of the copy number of X_1. The second conservation, which is valid only for $N_0 = 2$, is $\langle W_{2\vec{e}_1}\rangle + \langle W_{\vec{e}_2}\rangle = 1$. It reflects the fact that any removal of X_1 pairs from the system must result in the addition of an X_2 dimer and vice versa. Using these two conservation laws, Eq. (40) reduces to a single equation for $\langle W_{\vec{e}_1}\rangle$:

$$\frac{d\langle W_{\vec{e}_1}\rangle}{dt} = 2\beta - (\alpha+\beta)\langle W_{\vec{e}_1}\rangle. \tag{41}$$

This equation gives rise to the time dependent solution

$$\langle W_{\vec{e}_1}\rangle = \frac{2\beta}{\alpha+\beta} + Ae^{-(\alpha+\beta)t}. \tag{42}$$

The relaxation time is $\tau = 1/(\alpha+\beta)$, and the steady state solution is

$$\langle W_{\vec{e}_1}\rangle = \frac{2\beta}{\alpha+\beta}, \quad \langle W_{2\vec{e}_1}\rangle = \frac{\beta}{\alpha+\beta}, \quad \langle W_{\vec{e}_2}\rangle = \frac{\alpha}{\alpha+\beta}. \tag{43}$$

The variance σ_i^2, $i = 1,2$, in the populations of X_1 and X_2 can be obtained using Eq. (30). It is given by

$$\sigma_1^2 = \frac{4\alpha\beta}{(\alpha+\beta)^2}, \quad \sigma_2^2 = \frac{\alpha\beta}{(\alpha+\beta)^2}. \tag{44}$$

The stochastic solution is different from that obtained from the rate equations. This is because the limitations imposed on this system due to the discrete nature of the molecules are overlooked by the deterministic analysis.

To obtain a deeper insight on the effects of the discreteness in this system, we solve the binomial moment equations for $N_0 = 3$. The equations for the five nonvanishing moments are

$$\begin{aligned}\frac{d\langle W_{\vec{e}_1}\rangle}{dt} &= 2\beta\langle W_{\vec{e}_2}\rangle - 2\alpha\langle W_{2\vec{e}_1}\rangle, \\ \frac{d\langle W_{2\vec{e}_1}\rangle}{dt} &= 2\beta\langle W_{\vec{e}_1+\vec{e}_2}\rangle + \beta\langle W_{\vec{e}_2}\rangle - \alpha\langle W_{2\vec{e}_1}\rangle - 6\alpha\langle W_{3\vec{e}_1}\rangle, \\ \frac{d\langle W_{3\vec{e}_1}\rangle}{dt} &= \beta\langle W_{\vec{e}_1+\vec{e}_2}\rangle - 3\alpha\langle W_{3\vec{e}_1}\rangle, \\ \frac{d\langle W_{\vec{e}_2}\rangle}{dt} &= \alpha\langle W_{2\vec{e}_1}\rangle - \beta\langle W_{\vec{e}_2}\rangle, \\ \frac{d\langle W_{\vec{e}_1+\vec{e}_2}\rangle}{dt} &= 3\alpha\langle W_{3\vec{e}_1}\rangle - \beta\langle W_{\vec{e}_1+\vec{e}_2}\rangle,\end{aligned} \tag{45}$$

and the three conservation laws that emerge are

$$\begin{aligned}\langle W_{\vec{e}_1}\rangle + 2\langle W_{\vec{e}_2}\rangle = N_0, \quad \langle W_{3\vec{e}_1}\rangle + \langle W_{\vec{e}_1+\vec{e}_2}\rangle = 1, \\ 4\langle W_{\vec{e}_1+\vec{e}_2}\rangle + 2\langle W_{2\vec{e}_1}\rangle - \langle W_{\vec{e}_1}\rangle = 3.\end{aligned} \tag{46}$$

Here, the solution under steady state conditions takes the form

$$\langle W_{\vec{e}_1}\rangle = \frac{3\alpha^2+6\alpha\beta+3\beta^2}{3\alpha^2+4\alpha\beta+\beta^2}, \quad \langle W_{\vec{e}_2}\rangle = \frac{3\alpha^2+3\alpha\beta}{3\alpha^2+4\alpha\beta+\beta^2}. \tag{47}$$

Consider the limit where $\alpha \gg \beta$. In this limit, the system is dominated by the process of dimerization. Deterministically, the system reaches a state where $N_1 \ll N_2$, and $N_2 \simeq N_0/2$ [Eq. (38)]. However, in this limit, for $N_0 = 3$, the stochastic solution yields $\langle W_{\vec{e}_1}\rangle \simeq \langle W_{\vec{e}_2}\rangle \simeq 1$. For small values of N_0, the effect of parity plays an important role in the behavior of the system. Even if the dimerization process is dominant, if the total number of monomers is odd, there will inevitably remain at least a single monomer in the system at all times.

In Fig. 2(a), we show the monomer to dimer ratio $\eta = \langle W_{\vec{e}_1}\rangle/\langle W_{\vec{e}_2}\rangle$, as obtained from the binomial moment equations (circles) versus N_0 under steady state conditions. The results are in perfect agreement with those obtained from the master equation (solid line). The effect of parity is clearly observed by the sawtooth structure of the graph. The rate equation results (dashed line) do not account for this effect. For large values of N_0, the moment equation results coincide with the rate equations. The parameters used here are $\alpha = 2$ and $\beta = 1$. In Fig. 2(b), we present σ_i^2/N_0 versus N_0, as obtained form the moment equations, for $i = 1$ (circles) and for $i = 2$ (squares). The results are in perfect agreement with those obtained from the master equation (solid lines). As the system enters the deterministic regime (large values of N_0), the variance becomes small compared to the size of the system (parametrized by N_0).

C. Complex reaction network

Consider the more complex reaction network, presented in Fig. 3. This network consists of processes involving single molecules as well as pairs of molecules. The network includes 10 reactive species, 3 zero order reactions, 14 first order reactions, and 12 second order reactions. The zero order reactions lead to the formation of X_1, X_2, and X_3 molecules, where $P_0^{\vec{e}_i} = 1/3$ for $i = 1$, 2, and 3. The rest of the species are formed via first and second order reactions. The first order reactions include the degradation of each of the reactive

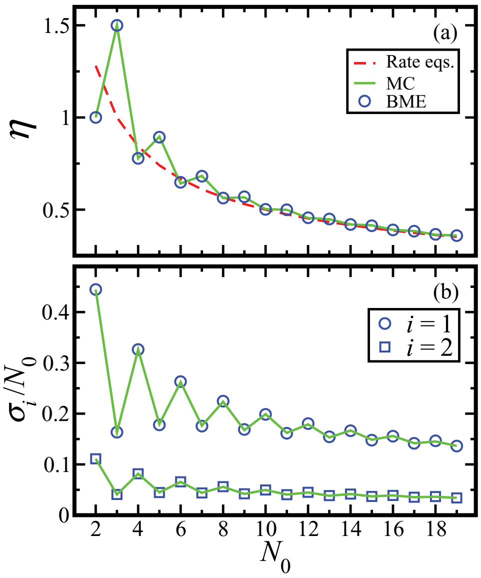

FIG. 2. (Color online) (a) The monomer to dimer ratio in the dimerization-dissociation reaction vs N_0, as obtained from the binomial moment equations (circles). The results are in perfect agreement with those obtained from MC simulations (solid line). The sawtooth shape of the graph reflects the parity effect of this system. This effect is overlooked by the deterministic analysis based on the rate equations (dashed line). (b) The variance (normalized) σ_i/N_0 vs N_0 as obtained from the binomial moment equations for the monomer population (circles), and for the dimer population (squares). The results are in perfect agreement with those obtained from MC simulations (solid lines).

species, and the dissociation of X_6 and X_7. The four first order reactions involving X_6 are given by $\vec{e}_6 \to 0$ (degradation), $\vec{e}_6 \to 2\vec{e}_1 + \vec{e}_4$, $\vec{e}_6 \to \vec{e}_4 + \vec{e}_7$, and $\vec{e}_6 \to \vec{e}_1 + \vec{e}_5$, where $P^0_{\vec{e}_6} = 0.999$ and $P^{2\vec{e}_1+\vec{e}_4}_{\vec{e}_6} = P^{\vec{e}_4+\vec{e}_7}_{\vec{e}_6} = P^{\vec{e}_1+\vec{e}_5}_{\vec{e}_6}/2 = 0.00025$. The two first order reactions involving X_7 are $\vec{e}_7 \to 0$ and $\vec{e}_7 \to 2\vec{e}_1$, where $P^0_{\vec{e}_7} = 0.999$, $P^{2\vec{e}_1}_{\vec{e}_7} = 0.001$. The second order reactions involve all the pairs of nodes connected by edges. The reaction of X_1 and X_4 includes three reaction paths $\vec{e}_1 + \vec{e}_4 \to \vec{e}_5$, $\vec{e}_1 + \vec{e}_4 \to \vec{e}_3 + \vec{e}_7$, and $\vec{e}_1 + \vec{e}_4 \to 2\vec{e}_1 + \vec{e}_3$, where $P^{\vec{e}_5}_{\vec{e}_1+\vec{e}_4} = 0.25$, $P^{\vec{e}_3+\vec{e}_7}_{\vec{e}_1+\vec{e}_4} = 0.5$, and $P^{2\vec{e}_1+\vec{e}_3}_{\vec{e}_1+\vec{e}_4} = 0.25$. The paths for the reaction of X_5 and X_6 are $\vec{e}_5 + \vec{e}_6 \to 5\vec{e}_1 + 2\vec{e}_3$, $\vec{e}_5 + \vec{e}_6 \to \vec{e}_5 + \vec{e}_6$, and $\vec{e}_5 + \vec{e}_6 \to \vec{e}_9$, with the probabilities 1/4, 1/4, and 1/2, respectively. This means that when a pair of X_5 and X_6 molecules encounter each other, they either dissociate into their fundamental components, remain unchanged, or combine to form the molecule X_9. To characterize the size of the system, we introduce the parameter S. The rate of zero order reactions is taken to be proportional to the size of the system, while the rate of second order reactions is inversely proportional to the size of the system [55]. We thus set $T_0 = \gamma S$, $T_{\vec{e}_i} = \delta_i$, and $T_{\vec{e}_i+\vec{e}_j} = a_{ij}/S$.

For this reaction network, setting the cutoff at $C = 3$ for all the species, one obtains $(3+1)^{10} \approx 10^6$ equations in the master equation. In contrast, the same cutoff set in the binomial moment equations yields only 285 equations [Eq. (19)]. A lower cutoff of $C = 2$ will result in 65 equations in the binomial moment equations, compared with approximately 6×10^4 equations in the master equation.

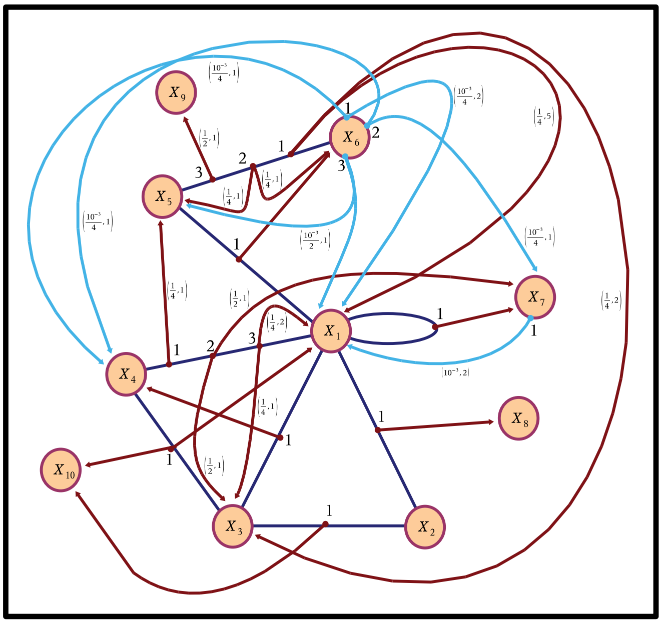

FIG. 3. (Color online) Schematic illustration of a chemical reaction network. The nodes represent the different reactive species, and the edges represent the reactions between them. Arrows are drawn from the edges to the reaction products. First order reactions are represented by arrows emerging from the reacting nodes. In cases where more than a single reaction path exists, or where the stoichiometric coefficient of the product is not one, the path probability and the relevant stoichiometric coefficient appear in parentheses by the arrows.

In Fig. 4, we show the population sizes of the species $X_1, \ldots, X_9$ versus the system size S, under steady state conditions, as obtained from the binomial moment equations with a cutoff of $C = 2$ (circles). For small systems, the results are in perfect agreement with those obtained from the master equation (solid lines). Note that in this example the master equation was solved using MC simulations [16,17]. Direct integration of the master equation is infeasible in this case due to the large number of equations.

For large systems, the stochastic results converge to those obtained from the rate equations (dashed lines). For most species, even in the limit of $S \gg 1$, the results of the binomial moment equations coincide with those obtained from the master equation. This is despite the fact that the number of equations is dramatically reduced. In fact, for certain species, it is sufficient to choose a cutoff of $C = 2$ to account for the abundances in the entire range of system sizes. In the case that the the cutoff of $C = 2$ is insufficient, one may raise the cutoff to 3 or 4 until a smooth convergence to the deterministic regime is obtained. Results obtained for a cutoff of $C = 3$ are included for $\langle N_1 \rangle$, $\langle N_2 \rangle$, and $\langle N_3 \rangle$ (+). Consider the population size of the species X_9. In order for this species to be produced, a pair of X_5 and X_6 must be present in the system. However, the only contribution to the formation of such pairs comes from the presence of triplets. For instance, when the combination $\mathcal{Q}_{\vec{e}_1+\vec{e}_4+\vec{e}_6}$ translates into $\mathcal{Q}_{\vec{e}_5+\vec{e}_6}$ through the second order reaction $\vec{e}_1 + \vec{e}_4 \to \vec{e}_5$. This pair can not be produced by any other process involving combinations of a lower order. Thus, a cutoff lower than $C = 3$ in the binomial

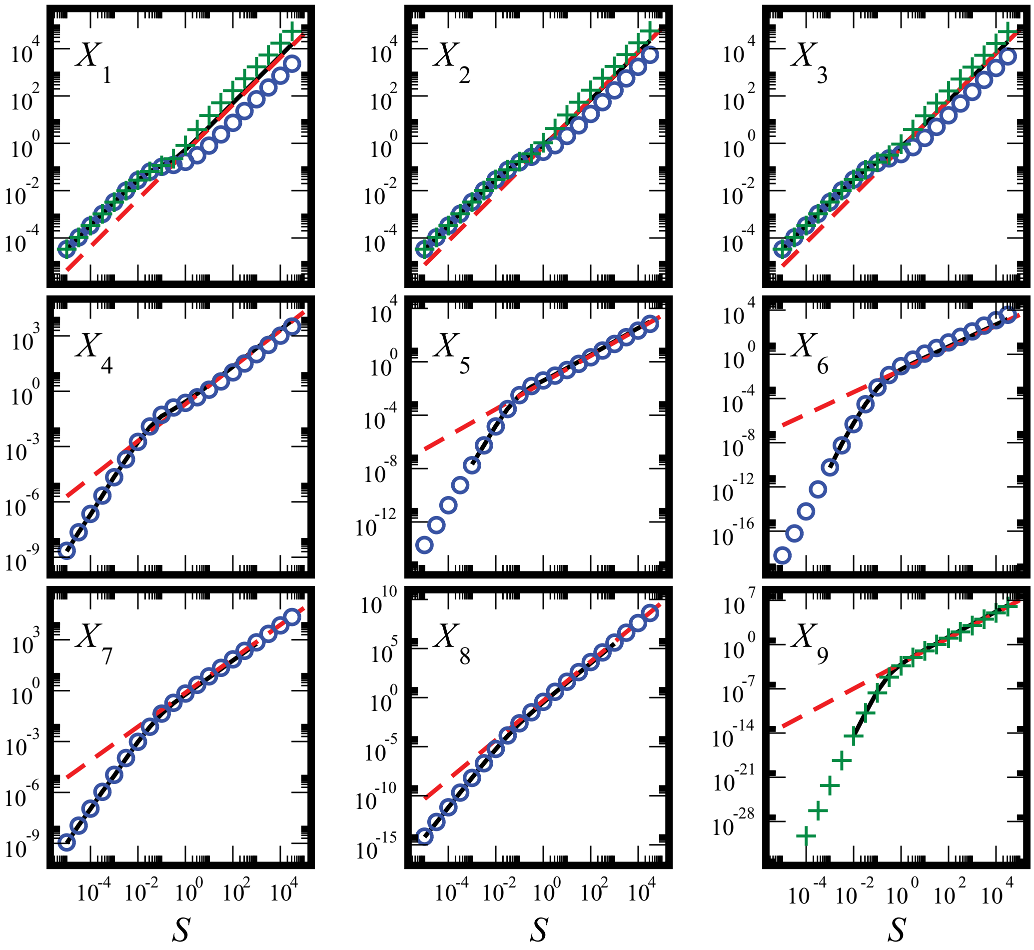

FIG. 4. (Color online) The average population sizes $\langle W_{\tilde{e}_i} \rangle$, $i = 1, \ldots, 9$, of the reactive species in the network shown in Fig. 3 versus the system size parameter S as obtained from the binomial moment equations (circles). The results were obtained with a cutoff of $C = 2$. In the limit where S is small, the results are in perfect agreement with those obtained from the master equation (solid lines). The rate equation results (dashed lines) show significant deviations in this limit. Here, for certain species, the moment equation results deviate in the limit of large systems. For these species, we include results obtained with a higher cutoff of $C = 3$ (+).

moment equations will not enable the production of X_9 even in the stochastic limit. This is an example of a case where third order moments are required, even when populations are small, and even when no reactions of higher order than two are included. In such cases, the ability to extend the moment equations to higher cutoffs becomes crucial.

In this system, when $S \ll 1$, several reactions become very slow and can be considered as rare events. For instance, the production of X_9 molecules requires the reaction of X_5 and X_6. However, since the abundances of these molecules are very small in the stochastic limit, their simultaneous appearance in the system rarely occurs. These rare reactions are difficult to account for by MC simulations, due to the exceedingly long run times required. This is the reason that in the stochastic limit there appear no results for $\langle N_5 \rangle$, $\langle N_6 \rangle$, and $\langle N_9 \rangle$ obtained from the MC simulations. In contrast, the binomial moment equations account for these rare processes without any limitation.

In Fig. 5, we focus on the transition between the stochastic regime and the deterministic regime. We show the convergence of the binomial moment equations to the deterministic results of the rate equations (dashed lines) for the average populations of X_1, X_2, and X_3. Results are shown for cutoff assignments of $C = 2$ (circles), $C = 3$ (+), $C = 4$ (triangles), $C = 5$ (×), and $C = 6$ (squares). Already at $C = 5$ or 6 the convergence to the deterministic results is smooth. When a higher cutoff is required, one can reliably use the rate equations.

VI. SUMMARY AND DISCUSSION

We have presented an equation-based formulation of stochastic reaction networks using the binomial moment equations. This formulation enables us to efficiently analyze complex reaction networks and evaluate the time dependent copy numbers and reaction rates as well as other moments of interest. The crux of the method lies in the definition of the binomial moments. These moments capture the essence of the combinatorics that governs the reaction rates, reflecting the stoichiometric structure of the reactions. As a result, the equations take a transparent form and it is easy to include moments up to any desired order. This is in contrast with ordinary moment equations, where it is often difficult to write the equations for moments of order higher than two. The binomial moments enable one to use a simple and convenient

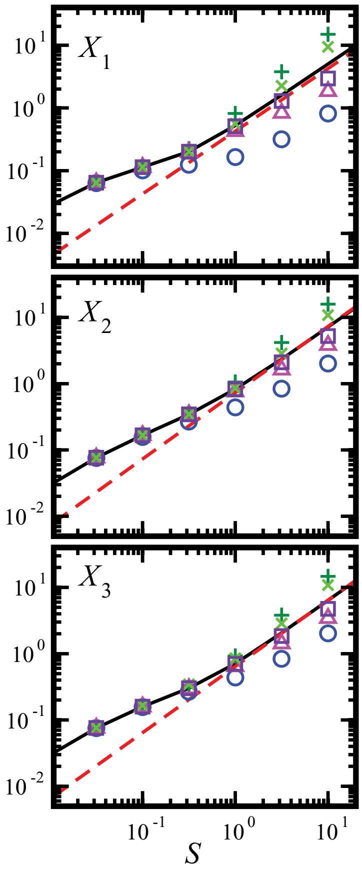

FIG. 5. (Color online) The transition between the stochastic regime $S < 1$ and the deterministic regime $S > 1$. Results are shown for the binomial moments $\langle W_{\vec{e}_1} \rangle$, $\langle W_{\vec{e}_2} \rangle$, and $\langle W_{\vec{e}_3} \rangle$ vs the system size S, as obtained from the moment equations with a cutoff of 2 (circles), 3 (+), 4 (triangles), 5 (×), and 6 (squares). As the cutoff is raised, the convergence to the rate equation results (dashed line) becomes smoother. When a cutoff higher than 6 is required, one enters the range of validity of the rate equations. The master equation results are also shown (solid lines).

truncation scheme in which all moments above a specified order are set to zero. This closure scheme is fully controlled and determined by the maximal number of particles allowed to reside simultaneously in the system, which is clearly limited by the system size.

The binomial moment equations display some important advantages over commonly used approaches such as direct integration of the master equation and MC methods. Direct integration of the master equation is applicable only for very small networks because the number of equations increases exponentially with the number of species in the network. The most commonly used approach to the simulation of stochastic reaction networks is the Monte Carlo method [16,17]. The Monte Carlo method is easy to implement and is of very general applicability. It provides numerical realizations of specific paths of the stochastic process. From these paths, one can evaluate the population sizes of the different species and the reaction rates. The data obtained from the Monte Carlo method can also be used in order to calculate temporal correlation functions. However, to obtain accurate results, the Monte Carlo method needs to produce large amounts of statistical data. This is difficult in systems that combine fast and slow processes or exhibit rare events.

As opposed to MC simulations, the moment equations can efficiently account for rare events and for systems that combine fast and slow processes covering a broad range of time scales. Moreover, unlike the rate equations, the moment equations are linear. This feature guarantees the stability of the differential equations, and in some simple cases also enables an analytical time dependent solution [56]. The steady state solution of the moment equations can be obtained by simply solving a set of coupled linear algebraic equations.

In the case of reaction networks that can be solved using only first and second moments, such as the hydrogen addition reactions presented in Refs. [45,46], the binomial moment equations do not exhibit significant advantages over the ordinary moment equations. However, the analysis of stochastic networks often requires higher moments, either due to the structure of the network (which may include reactions that involve more than two species) or due to the kinetic parameters. In the latter case, some parts of the network may be in the deep stochastic limit, while other parts may be in some intermediate regime. As a result, only some of the species appear in very low copy numbers. In such cases, the binomial moment equations have a crucial advantage. These equations can be constructed up to high order moments. Moreover, they exhibit very good convergence due to the fact that the moments decrease as a function of their order.

The binomial moment equations are designed to apply in the limit of small systems, where the copy numbers of the reactive species are low, and fluctuations are important. In some cases, the equations maintain their accuracy even when the system is large, despite the low cutoffs (such as $C = 2$). If this is not the case, the cutoffs should be raised until accurate results are obtained. In practice, when simulating a complex network, one may need to use both the rate equations and the moment equations. In the small system limit, where the copy numbers are low, the results of the moment equations should be used. As the large system regime is approached, the cutoffs used in the moment equations should be raised accordingly. Once a good convergence of the moment equation results to the rate equation results is obtained, one no longer needs to use the moment equations. In most cases, this convergence occurs when the cutoffs are not much higher than 3 or 4.

For the sake of completeness, we also refer to the possibility to analyze stochastic reaction networks using the chemical Langevin equations [11]. These equations are essentially rate equations to which noise terms are added, turning them into stochastic differential equations. These equations are suitable for the analysis of systems that are affected by external noise. In the reaction networks considered here, the fluctuations are intrinsic. They result from the stochastic nature of the reactions and from the low copy numbers of the reactants. Under these conditions, it is crucial to account for the discrete nature of the reactants. The problem with the Langevin approach is that unlike the moment equations, it does not account for the discreteness. Once the discreteness is lost, it is not possible to evaluate the reaction rates correctly. Therefore, the Langevin approach is not suitable for these stochastic reaction networks. Moreover, the Langevin equations are based on the mean-field approximation in the sense that the reaction rates are expressed in terms of the first moments and their products. This is unlike the moment equations in which higher moments are taken into account.

In recent years, there has been much progress in the study of complex interaction networks. The topological structure of chemical, biological, and ecological networks was extracted using large sets of empirical data. An important theoretical challenge is to study the dynamics of the reaction networks. The aim of such studies is to evaluate the population sizes of the interacting species as well as their fluctuations and correlations. The binomial moment equations provide an effective set of tools for describing the dynamics of these systems. They open the way to systematic studies of large and complex stochastic networks beyond the feasibility limit of existing methods. Moreover, as an equation-based paradigm, it is amenable to analytical treatments that are expected to provide crucial insight about the networks.

ACKNOWLEDGMENTS

B.B. thanks the Hoffman Leadership and Responsibility program at the Hebrew University for support. This work was supported by the U.S.-Israel Binational Science Foundation.

APPENDIX: COMBINATORIAL IDENTITIES

In this appendix, we prove three combinatorial identities that are used in the paper.

Identity 1:

$$\mathcal{S}^{\vec{u}}\binom{\vec{N}}{\vec{v}} = \sum_{\vec{w}}\binom{\vec{N}}{\vec{v}-\vec{w}}\binom{\vec{u}}{\vec{w}}. \tag{A1}$$

Proof. For convenience we introduce a vectorial generalization of the exponential function, in the form

$$x^{\vec{N}} = \prod_{i=1}^{J} x^{N_i}. \tag{A2}$$

The binomial expansion of this function takes the form

$$(1+x)^{\vec{N}} = \sum_{\vec{n}}\binom{\vec{N}}{\vec{n}}x^{\vec{n}}. \tag{A3}$$

We use the equality

$$(1+x)^{\vec{N}+\vec{u}} = (1+x)^{\vec{N}}(1+x)^{\vec{u}}, \tag{A4}$$

and take the binomial expansion of both sides. Comparing the coefficients of $x^{\vec{v}}$, one obtains

$$\binom{\vec{N}+\vec{u}}{\vec{v}} = \sum_{\vec{w}}\binom{\vec{N}}{\vec{v}-\vec{w}}\binom{\vec{u}}{\vec{w}}, \tag{A5}$$

which proves Identity 1.

Identity 2:

$$\mathcal{S}^{-\vec{u}}\binom{\vec{N}}{\vec{v}} = \binom{\vec{N}}{\vec{u}}^{-1}\binom{\vec{v}+\vec{u}}{\vec{u}}\binom{\vec{N}}{\vec{v}+\vec{u}}. \tag{A6}$$

Proof. We first prove this identity for scalars. The right-hand side of (A6) can be written as

$$\frac{(N-u)!u!}{N!}\cdot\frac{(v+u)!}{v!u!}\cdot\frac{N!}{[N-(v+u)]!(v+u)!}, \tag{A7}$$

which can be easily reduced to

$$\binom{N-u}{v}, \tag{A8}$$

in agreement with Identity 2. The generalization to vectors is straightforward. The left-hand side of Eq. (A6) can be expressed as a product by

$$\binom{\vec{N}-\vec{u}}{\vec{v}} = \prod_{i=1}^{J}\binom{N_i-u_i}{v_i}. \tag{A9}$$

Expressing the right-hand side as a product, one observes that the equality is satisfied for each pair of corresponding terms and thus it is satisfied for the entire product.

Identity 3:

$$\binom{\vec{N}}{\vec{n}}\binom{\vec{N}}{\vec{v}} = \sum_{\vec{w}}\binom{\vec{n}}{\vec{w}}\binom{\vec{v}+\vec{w}}{\vec{n}}\binom{\vec{N}}{\vec{v}+\vec{w}}. \tag{A10}$$

Proof. We first prove this identity for scalars. The right-hand side of (A10) can be written as

$$\sum_{w}\binom{\vec{n}}{\vec{w}}\frac{N!}{(N-v-w)!(v+w-n)!n!}, \tag{A11}$$

which by factoring out $\binom{N}{v}$ becomes

$$\binom{N}{v}\sum_{w}\binom{N-v}{w}\binom{v}{n-w}. \tag{A12}$$

Using (A1), one can substitute the sum in the above expression by $\binom{N}{n}$ to obtain the left-hand side of (A10). As above, the generalization to vectors is straightforward.

---

[1] O. N. Temkin, A. V. Zeigarnik, and D. G. Bonchev, *Chemical Reaction Networks: A Graph-Theoretical Approach* (CRC Press, Boca Raton, 1996).

[2] A. R. Peacocke, *An Introduction to the Physical Chemistry of Biological Organization* (Oxford University Press, New York, 1989).

[3] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall, Boca Raton, 2006).

[4] B. O. Palsson, *Systems Biology, Properties of Reconstructed Networks* (Cambridge University Press, Cambridge, 2006).

[5] J. D. Murray, *Mathematical Biology* (Springer, Berlin, 1989).

[6] S. B. Charnley, A. G. G. M. Tielens, and S. D. Rodgers, Astrophys. J. **482**, L203 (1997).

[7] H. H. McAdams and A. Arkin, Proc. Natl. Acad. Sci. USA **94**, 814 (1997).

[8] P. Caselli, T. I. Hasegawa, and E. Herbst, Astrophys. J. **495**, 309 (1998).
[9] M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nat. Rev. Genet. **6**, 451 (2005).
[10] R. Lande, S. Engen, and B. E. Sæther, *Stochastic Population Dynamics in Ecology and Conservation* (Oxford University Press, New York, 2003).
[11] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
[12] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2004).
[13] T. Stantcheva, V. I. Shematovich, and E. Herbst, Astron. Astrophys. **391**, 1069 (2002).
[14] T. Stantcheva and E. Herbst, Mon. Not. R. Astron. Soc. **340**, 983 (2003).
[15] O. Biham, I. Furman, V. Pirronello, and G. Vidali, Astrophys. J. **553**, 595 (2001).
[16] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
[17] D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
[18] S. B. Charnley, Astrophys. J. **509**, L121 (1998).
[19] D. A. McQuarrie, J. Chem. Phys. **38**, 433 (1963).
[20] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell, J. Chem. Phys. **40**, 2914 (1964).
[21] M. Assaf and B. Meerson, Phys. Rev. Lett. **97**, 200602 (2006).
[22] M. Assaf and B. Meerson, Phys. Rev. E **75**, 031122 (2007).
[23] N. J. B. Green, T. Toniazzo, M. J. Pilling, D. P. Ruffle, N. Bell, and T. W. Hartquist, Astron. Astrophys. **375**, 1111 (2001).
[24] O. Biham and A. Lipshtat, Phys. Rev. E **66**, 056103 (2002).
[25] E. W. Montroll, J. Chem. Phys. **14**, 202 (1946).
[26] A. Kolmogorov, Mathem. Annalen **104**, 415 (1931).
[27] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon, Oxford, 1999).
[28] Y. Cao, D. T. Gillespie, and L. R. Petzold, J. Chem. Phys. **124**, 044109 (2006).
[29] M. A. Gibson and J. Bruck, J. Phys. Chem. **104**, 1876 (2000).
[30] D. T. Gillespie, J. Chem. Phys. **115**, 1716 (2001).
[31] T. Tian and K. Burrage, J. Chem. Phys. **121**, 10356 (2004).
[32] B. Munsky and M. Khammash, J. Chem. Phys. **124**, 044104 (2006).
[33] M. F. Pettigrew and H. Resat, J. Chem. Phys. **126**, 184101 (2007).
[34] H. Salis and Y. N. Kaznessis, J. Chem. Phys. **123**, 214106 (2005).
[35] S. MacNamara, K. Burrage, and R. B. Sidje, Multiscale Model. Simul. **6**, 1146 (2007).
[36] D. T. Gillespie, Annu. Rev. Phys. Chem. **58**, 35 (2007).
[37] E. L. Haseltine and J. B. Rawlings, J. Chem. Phys. **117**, 6959 (2002).
[38] J. Puchalka and A. M. Kierzek, Biophys. J. **86**, 1357 (2004).
[39] S. MacNamara, A. M. Bersani, K. Burrage, and R. B. Sidje, J. Chem. Phys. **129**, 095105 (2008).
[40] A. Lipshtat and O. Biham, Phys. Rev. Lett. **93**, 170601 (2004).
[41] B. Barzel, O. Biham, and R. Kupferman, Phys. Rev. E **76**, 026703 (2007).
[42] B. Barzel, O. Biham, and R. Kupferman, Multiscale Model. Simul. **6**, 963 (2007).
[43] B. Barzel, O. Biham, R. Kupferman, A. Lipshtat, and A. Zait, Phys. Rev. E **82**, 021117 (2010).
[44] A. Lipshtat and O. Biham, Astron. Astrophys. **400**, 585 (2003).
[45] B. Barzel and O. Biham, Astrophys. J. Lett. **658**, L37 (2007).
[46] B. Barzel and O. Biham, J. Chem. Phys. **127**, 144703 (2007).
[47] A. G. G. M. Tielens and W. Hagen, Astron. Astrophys. **114**, 245 (1982).
[48] S. Ioppolo, H. M. Cuppen, C. Romanzin, E. F. van Dishoeck, and H. Linnartz, Astrophys. J. **686**, 1474 (2008).
[49] H. M. Cuppen, S. Ioppolo, C. Romanzin, and H. Linnartz, Phys. Chem. Chem. Phys. **12**, 12077 (2010).
[50] J. Schellenberger, J. O. Park, T. M. Conrad, and B. O. Palsson, BMC Bioinf. **11**, 213 (2010).
[51] B. Barzel and O. Biham, Phys. Rev. Lett. **106**, 150602 (2011).
[52] D. A. McQuarrie, J. Appl. Prob. **4**, 413 (1967).
[53] C. A. Gómez-Uribe and G. C. Verghese, J. Chem. Phys. **126**, 024109 (2007).
[54] J.-C. Lin, C. R. Doering, and D. ben-Avraham, Chem. Phys. **146**, 355 (1990).
[55] I. Lohmar, J. Krug, and O. Biham, Astron. Astrophys. **504**, L5 (2009).
[56] B. Barzel and O. Biham, Phys. Rev. E **80**, 031117 (2009).