

Binomial moment equations for stochastic reaction systems

Baruch Barzel and Ofer Biham

Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel

Abstract

A highly efficient formulation of moment equations for stochastic reaction networks is introduced. It is based on a set of binomial moments that capture the combinatorics of the reaction processes. The resulting set of equations can be easily truncated to include moments up to any desired order. The number of equations is dramatically reduced compared to the master equation. This formulation enables the simulation of complex reaction networks, involving a large number of reactive species much beyond the feasibility limit of any existing method. It provides an equation-based paradigm to the analysis of stochastic networks, complementing the commonly used Monte Carlo simulations.

PACS numbers: 05.10.Gg, 82.40.-g, 82.40.Qt, 02.50.Fz, 02.50.Ga

Stochastic reaction networks appear in many natural systems, from grain-surface chemistry in the interstellar medium [1] to biochemistry in cells [2, 3] and to ecological systems [4]. In order to characterize these networks one wishes to evaluate the time dependent populations of the reactive species. Due to the large fluctuations in stochastic systems, the rate equations fail and stochastic methods [5, 6] are required. These methods are based on the master equation, which can be solved either by direct integration or by Monte Carlo simulations [7, 8]. The problem with these methods is that they quickly become infeasible as the number of reactive species increases. More specifically, the number of equations in the master equation proliferates exponentially with the number of species, making the stochastic analysis difficult even for empirical networks of moderate size [9, 10]. A common approach is to derive moment equations by tracing over the master equation [11]. In these equations one directly computes different moments related to the populations of the reactive species. The problem is that higher order moments appear on the right hand side of the equations, making it difficult to obtain a closed set of equations. In most closure schemes, only the first and second moments are included [12].

In this Letter we present a general and highly effective formulation of moment equations, expressed in terms of the *binomial moments*, defined below. These are linear combinations of the ordinary moments, that capture the structure of the reactions and provide much physical insight. As a result, this formulation enables the inclusion of moments up to any desired order. The equations are linear and their number is dramatically reduced compared to the master equation, enabling the analysis of complex reaction networks.

Consider a reaction network, consisting of J reactive species, X_i , $i = 1, \dots, J$. These species undergo reactions of the form



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 it is common to omit reactions of order higher than $n = 2$. However, in the formulation presented below there is no such limitation. The reactions presented in Eq. (1) can be expressed in a vector form as $\vec{n} \rightarrow \vec{m}$, namely a set of molecules with stoichiometric coefficients \vec{n} , reacts to form a new set \vec{m} . The rate constant for the

molecular configuration \vec{n} to react is given by $T_{\vec{n}}$ (s^{-1}). Certain reactions may take several 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 $\vec{n} \rightarrow \vec{m}$ is $T_{\vec{n}} P_{\vec{n}}^{\vec{m}}$.

In the stochastic analysis one describes the state of the system by the vector $\vec{N} = (N_1, \dots, N_J)$, where N_i is the number of copies of species X_i . Below we introduce a combinatorial approach, which will later allow us to express the master equation and the moment equations in a transparent form. Let $\vec{v} = (v_1, \dots, v_J)$ be a vector of integers. It can be expressed as a linear combination of the basis vectors, $\vec{e}_1, \dots, \vec{e}_J$, where $\vec{e}_i = (0, \dots, v_i = 1, \dots, 0)$. We denote by $\mathcal{Q}_{\vec{v}}$ a combination of molecules consisting of exactly v_i copies of the species X_i . The number of such combinations that exist in a system at the state \vec{N} is given by

$$\mathbf{W}(\vec{N}, \vec{v}) = \binom{\vec{N}}{\vec{v}}, \quad (2)$$

where $\binom{\vec{N}}{\vec{v}} = \prod_{i=1}^J \binom{N_i}{v_i}$ and $\binom{N_i}{v_i}$ is the binomial coefficient. To illustrate the motivation for this definition, consider the reaction $\vec{n} \rightarrow \vec{m}$. 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 $\mathbf{W}(\vec{N}, \vec{n})$.

Let $P(\vec{N})$ represent the time dependent probability for the system to be in the state \vec{N} . The master equation for $P(\vec{N})$ takes the form

$$\frac{dP(\vec{N})}{dt} = \sum_{\vec{n}, \vec{m}} T_{\vec{n}} P_{\vec{n}}^{\vec{m}} \left[\mathbf{W}(\vec{N} + \vec{n} - \vec{m}, \vec{n}) P(\vec{N} + \vec{n} - \vec{m}) - \mathbf{W}(\vec{N}, \vec{n}) P(\vec{N}) \right]. \quad (3)$$

In this equation one sums over all the reactions $\vec{n} \rightarrow \vec{m}$ in the system. These reactions yield a positive contribution to $P(\vec{N})$ when the state of the system is $\vec{N} + \vec{n} - \vec{m}$, and a negative contribution to $P(\vec{N})$ when the system is in the state \vec{N} . In numerical simulations the master equation must be truncated in order to maintain a finite number of equations. This is achieved by setting upper cutoffs C_i , $i = 1, \dots, J$, such that $P(\vec{N}) = 0$ if $N_i > C_i$ for any value of i . The number of coupled equations, $N_E = \prod_{i=1}^J (C_i + 1)$, grows exponentially with the number of reactive species, J . This severely limits the applicability of the master equation to complex reaction networks [9, 10].

A more compact description of stochastic reaction networks can be obtained using moment equations. These equations, derived by tracing over the master equation, consist of

ordinary differential equations for the time derivatives of the moments $\langle N_1^{a_1} \cdots N_J^{a_J} \rangle$, where a_i are integers. The order of a specific moment is given by $k = \sum_{i=1}^J a_i$. The difficulty with moment equations is that higher order moments appear on the right hand side of each equation. 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 [11–14]. In practice, the truncation is typically done at the level of third order moments, namely only the first and second order moments are taken into account. Making the truncation at higher orders turns out to be prohibitively complicated even for relatively simple networks. Below we introduce a different formulation of the moment equations, which enables to extend the truncation up to any desired order.

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

$$\langle \mathbf{W}_{\vec{v}} \rangle = \sum_{\vec{N}} \mathbf{W}(\vec{N}, \vec{v}) P(\vec{N}). \quad (4)$$

To understand the meaning of the binomial moments, consider the case where $\vec{v} = \vec{e}_i$. Here the corresponding binomial moment, $\langle \mathbf{W}_{\vec{e}_i} \rangle$, is given by the average number of combinations 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 $\langle \mathbf{W}_{\vec{v}} \rangle = \langle N_i N_j \rangle$, 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 $\langle \mathbf{W}_{\vec{v}} \rangle = (\langle N_i^2 \rangle - \langle N_i \rangle^2)/2$. Similarly, the number of X_i triplets is given by $\langle \mathbf{W}_{\vec{v}} \rangle = (\langle N_i^3 \rangle - 3\langle N_i^2 \rangle + 2\langle N_i \rangle^2)/6$, where $\vec{v} = 3\vec{e}_i$. The order the binomial moment $\langle \mathbf{W}_{\vec{v}} \rangle$ is $k = \sum_{i=1}^J v_i$.

To obtain the binomial moment equations we express the time derivative of $\langle \mathbf{W}_{\vec{v}} \rangle$ using Eq. (4), where the time derivative of $P(\vec{N})$ is taken from the master equation [Eq. (3)], and a summation is taken over \vec{N} [15]. The resulting binomial moment equations take the form

$$\frac{d\langle \mathbf{W}_{\vec{v}} \rangle}{dt} = \sum_{\vec{n}, \vec{m}} \left[B_{\vec{n}}^{\vec{m}} \binom{\vec{v} + \vec{n} - \vec{m}}{\vec{n}} \langle \mathbf{W}_{\vec{v} + \vec{n} - \vec{m}} \rangle - T_{\vec{n}} \binom{\vec{n}}{\vec{m}} \binom{\vec{v} + \vec{m}}{\vec{n}} \langle \mathbf{W}_{\vec{v} + \vec{m}} \rangle \right], \quad (5)$$

where

$$B_{\vec{n}}^{\vec{m}} = \sum_{\vec{w}} T_{\vec{n}} P_{\vec{n}}^{\vec{w}} \binom{\vec{w}}{\vec{m}}. \quad (6)$$

The first term in Eq. (5) accounts for positive contributions of the reactions to $\langle \mathbf{W}_{\vec{v}} \rangle$, while the second term accounts for the negative contributions. We first consider the positive contributions. Consider a single combination $\mathcal{Q}_{\vec{v}+\vec{n}-\vec{m}}$. There are on average $\langle \mathbf{W}_{\vec{v}+\vec{n}-\vec{m}} \rangle$ such combinations in the system. For each one of these combinations, the reaction $\vec{n} \rightarrow \vec{m}$, produces a new $\mathcal{Q}_{\vec{v}}$ combination, so $\langle \mathbf{W}_{\vec{v}} \rangle$ increases. The rate constant $B_{\vec{n}}^{\vec{m}}$ accounts for the rate of formation of $\mathcal{Q}_{\vec{m}}$ combinations by the reactions $\vec{n} \rightarrow \vec{w}$ for all possible choices of \vec{w} . The reaction rate is also proportional to the binomial coefficient $\binom{\vec{v}+\vec{n}-\vec{m}}{\vec{n}}$, that accounts for the number of combinations $\mathcal{Q}_{\vec{n}}$ in $\mathcal{Q}_{\vec{v}+\vec{n}-\vec{m}}$, each of which may undergo the reaction.

We now refer to the negative contributions. Consider the combination $\mathcal{Q}_{\vec{v}+\vec{m}}$ which undergoes a reaction of the form $\vec{n} \rightarrow \vec{w}$, 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 $\mathcal{Q}_{\vec{n}}$ is removed from the system. The removed combination $\mathcal{Q}_{\vec{n}}$ can be decomposed into $\mathcal{Q}_{\vec{m}}$ and $\mathcal{Q}_{\vec{n}-\vec{m}}$. Note that there are $\binom{\vec{n}}{\vec{m}}$ different possibilities to perform this decomposition of $\mathcal{Q}_{\vec{n}}$. When $\mathcal{Q}_{\vec{m}}$ is removed, the combination $\mathcal{Q}_{\vec{v}+\vec{m}}$ is replaced by $\mathcal{Q}_{\vec{v}}$. This $\mathcal{Q}_{\vec{v}}$ combination is then eliminated by the subsequent removal of $\mathcal{Q}_{\vec{n}-\vec{m}}$.

Eq. (5) 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 $\langle \mathbf{W}_{\vec{v}} \rangle$ of order $k = \sum_{i=1}^J v_i$. It represents the average number of appearances of a certain combination $\mathcal{Q}_{\vec{v}}$ which consists of k molecules in the system. In a small system, where the average copy numbers are small, $\langle \mathbf{W}_{\vec{v}} \rangle$ tends to decrease as k increases. This enables us to use the following truncation scheme. We choose a cutoff C such that $\langle \mathbf{W}_{\vec{v}} \rangle$ 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 carried out, is $N_E = \sum_{k=1}^C \binom{k+J-1}{k}$. 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 fact, in an earlier version of this formulation, it was shown that for a broad range of conditions the cutoff $C = 2$ is sufficient [13, 14]. In this case, the equations include first order moments that

account for the average population sizes and second order moments that account for the number of pairs of species X_i and X_j in the system, from which the reaction rates are evaluated. 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.

To demonstrate the method, we apply it to the reaction network shown in Fig. 1. 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 species, and the dissociation of X_6 and X_7 . The four first order reactions involving X_6 are given by $\vec{e}_6 \rightarrow 0$ (degradation), $\vec{e}_6 \rightarrow 2\vec{e}_1 + \vec{e}_4$, $\vec{e}_6 \rightarrow \vec{e}_4 + \vec{e}_7$ and $\vec{e}_6 \rightarrow \vec{e}_1 + \vec{e}_5$, where $P_{\vec{e}_6}^0 = 0.999$ and $P_{\vec{e}_6}^{2\vec{e}_1 + \vec{e}_4} = P_{\vec{e}_6}^{\vec{e}_4 + \vec{e}_7} = P_{\vec{e}_6}^{\vec{e}_1 + \vec{e}_5} / 2 = 0.00025$. The two first order reactions involving X_7 are $\vec{e}_7 \rightarrow 0$ and $\vec{e}_7 \rightarrow 2\vec{e}_1$, where $P_{\vec{e}_7}^0 = 0.999$, $P_{\vec{e}_7}^{2\vec{e}_1} = 0.001$. 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, $\vec{e}_1 + \vec{e}_4 \rightarrow \vec{e}_5$, $\vec{e}_1 + \vec{e}_4 \rightarrow \vec{e}_3 + \vec{e}_7$ and $\vec{e}_1 + \vec{e}_4 \rightarrow 2\vec{e}_1 + \vec{e}_3$, where $P_{\vec{e}_1 + \vec{e}_4}^{\vec{e}_5} = 0.25$, $P_{\vec{e}_1 + \vec{e}_4}^{\vec{e}_3 + \vec{e}_7} = 0.5$ and $P_{\vec{e}_1 + \vec{e}_4}^{2\vec{e}_1 + \vec{e}_3} = 0.25$. The paths for the reaction of X_5 and X_6 are $\vec{e}_5 + \vec{e}_6 \rightarrow 5\vec{e}_1 + 2\vec{e}_3$, $\vec{e}_5 + \vec{e}_6 \rightarrow \vec{e}_5 + \vec{e}_6$ and $\vec{e}_5 + \vec{e}_6 \rightarrow \vec{e}_9$, with the probabilities $1/4$, $1/4$ and $1/2$, respectively. This means that upon encounter of a pair of X_5 and X_6 molecules, 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 [16]. We thus set $T_0 = gS$, $T_{\vec{e}_i} = d_i$ and $T_{\vec{e}_i + \vec{e}_j} = a_{ij}/S$. The parameters we use are $g = 1 \text{ s}^{-1}$ for the zero order reactions, $d_i = 0.3 \text{ s}^{-1}$, $i = 1, \dots, 3$, for the first order reactions and $a_{ij} = 1 \text{ s}^{-1}$ for the second order reactions that are included in the network.

For this reaction network, setting a cutoff of $C = 3$ 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. A lower cutoff of $C = 2$ results in 65 equations, compared with approximately 6×10^4 equations in the master equation. After

solving the binomial moment equations, one can extract the average population sizes, which are given by the $\langle \mathbf{W}_{\vec{e}_i} \rangle$ moments.

In Fig. 2 we show the population sizes of several selected species vs. the system size, S . These results were obtained from the binomial moment equations with $C = 2$ (circles), under steady state conditions. For small systems, the results are in perfect agreement with those of the master equation (solid lines). The master equation results were obtained using the Gillespie algorithm [7], since direct integration of the master equation is already infeasible in this case [17]. For large systems, the stochastic results converge to those obtained from the rate equations (dashed lines). For most species, the results of the binomial moment equations with such low cutoff, coincide with those of the master equation not only in the small system limit but also for $S \gg 1$. 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. This surprising feature was discussed in an earlier formulation of the moment equations, presented in Refs. [13, 14]. In case that the the cutoff of $C = 2$ is insufficient one may raise it to 3 or 4, until a smooth convergence to the rate equation results is obtained in the large system limit. Results obtained for a cutoff of $C = 3$ (+) are shown for $\langle N_1 \rangle$ and $\langle N_2 \rangle$. Consider the population of the species X_9 in the system. This species is produced by the reaction $\vec{e}_5 + \vec{e}_6 \rightarrow \vec{e}_9$. Thus, in order for X_9 to be produced, a pair of X_5 and X_6 must be simultaneously present in the system. However, such pairs form only when there are triplets in the system. For instance, when the combination $\mathcal{Q}_{\vec{e}_1 + \vec{e}_4 + \vec{e}_6}$ transforms into $\mathcal{Q}_{\vec{e}_5 + \vec{e}_6}$ through the second order reaction $\vec{e}_1 + \vec{e}_4 \rightarrow \vec{e}_5$. Thus, a cutoff lower than $C = 3$ in the binomial moment equations would terminate the production of X_9 even in the stochastic limit. In such cases, the ability to extend the equations to higher cutoffs is crucial. In Fig. 3 we show the convergence of the binomial moment equations to the rate equation results (dashed lines) for the average population of X_2 . Results are shown for cutoffs of $C = 2$ (circles), $C = 3$ (+), $C = 4$ (triangles), $C = 5$ (\times) and $C = 6$ (squares). Already for $C = 5$ or 6 the convergence to the deterministic results is smooth. When a higher cutoff is required, one can safely use the rate equations.

In conclusion, the binomial moment equations provide a highly efficient equation-based methodology for the simulation of stochastic reaction networks. The binomial moments capture the essence of the combinatorics that governs the reaction rates, reflecting the stoichiometric structure of the reactions. The 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. Unlike the ordinary moment equations, the binomial moment equations can be constructed to any order using an automated procedure, taking the network structure as input. The number of equations is dramatically reduced. This method opens 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.

Baruch Barzel thanks the Hoffman Leadership and Responsibility program at the Hebrew University for support. This work was supported by the US-Israel Binational Science Foundation.

-
- [1] S.B. Charnley, *Astrophys. J.* **509**, L121 (1998).
- [2] H.H. McAdams and A. Arkin, *Proc. Natl. Acad. Sci. US* **94**, 814 (1997).
- [3] M. Kaern, T.C. Elston, W.J. Blake and J.J. Collins, *Nature Review Genetics* **6**, 451 (2005).
- [4] R. Lande, S. Engen, B.E. Sæther, *Stochastic Population Dynamics in Ecology and Conservation* (Oxford University Press, New York, 2003).
- [5] N.G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [6] C.W. Gardiner, *Handbook of Stochastic Methods* (Springer-Verlag, Berlin, 2004).
- [7] D.T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [8] D.T. Gillespie, *Annu. Rev. Phys. Chem.* **58**, 35 (2007).
- [9] T. Stantcheva and E. Herbst, *Mon. Not. R. Astron. Soc.* **340**, 983 (2003).
- [10] A. Lipshtat, O. Biham, *Phys. Rev. Lett.* **93**, 170601 (2004).
- [11] D.A. McQuarrie, *J. Appl. Prob.* **4**, 413 (1967).
- [12] C.A. Gómez-Uribe and G.C. Verghese, *J. Chem. Phys.* **126**, 024109 (2007).
- [13] B. Barzel and O. Biham, *Astrophys. J.* **L115**, 20941 (2007).
- [14] B. Barzel and O. Biham, *J. Chem. Phys.* **127**, 144703 (2007).
- [15] The details of this derivation will appear in B. Barzel and O. Biham, to be submitted to *Phys. Rev. E*.
- [16] I. Lohmar, J. Krug and O. Biham, *Astron. Astrophys.* **504**, L5 (2009).
- [17] Note that for some species the solid line in Fig. 2 does not extend all the way to the left. This is due to the fact that the relevant reactions become rare in the small system limit and cannot be evaluated by the Gillespie algorithm.

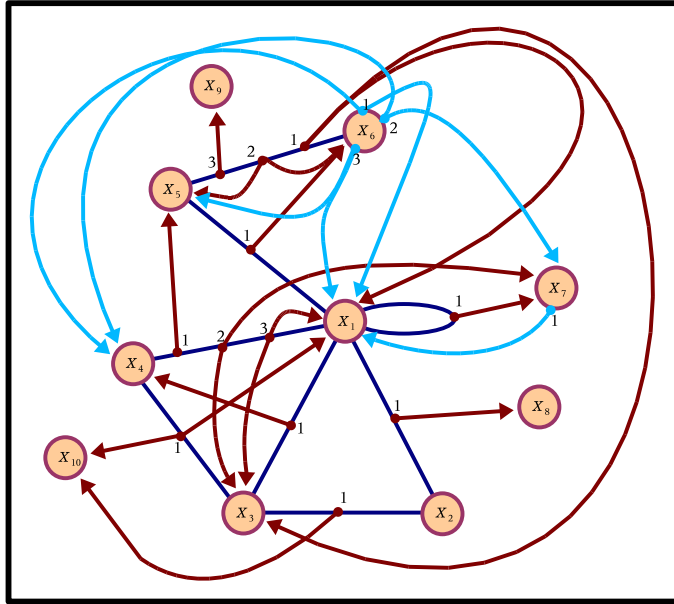


FIG. 1: (Color online) A reaction network consisting of 10 reactive species. First order reactions appear as arrows connecting the reacting species to its products. Second order reactions are denoted by edges connecting the two reacting species. Arrows are drawn from the edge to the reaction products. Different reaction paths are marked by the indices appearing by the arrows.

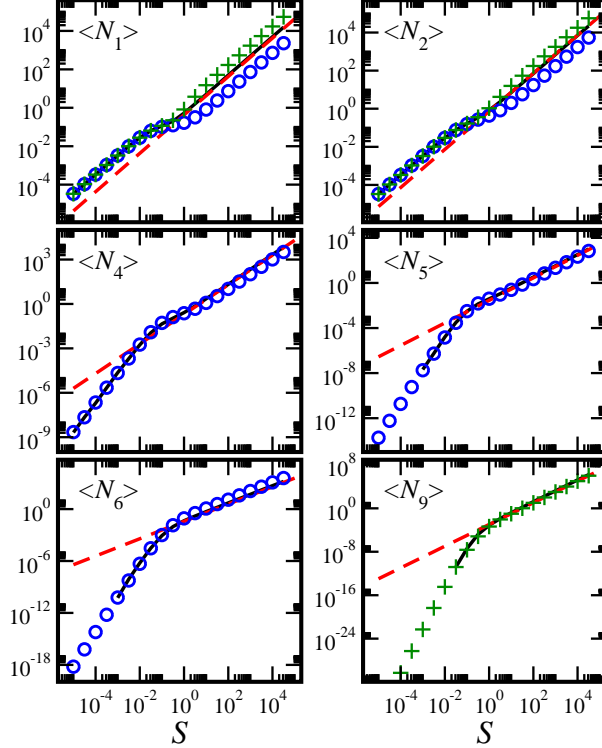


FIG. 2: (Color online) The average populations sizes, $\langle N_i \rangle$, of several species from the reaction network appearing in Fig. 1, vs. the system size, S , as obtained from the moment equations under steady state conditions (circles). These results were obtained for a cutoff of $C = 2$. In the limit of small system sizes the results are in perfect agreement with the results obtained from the master equation (solid lines). The rate equation results (dashed lines) show significant deviations for small sizes. For the species X_1 and X_2 the moment equations deviate for large systems, and in order to get accurate results one has to use a higher cutoff. For these species we display results obtained from the moment equations with a cutoff of $C = 3$ (+). The species X_9 cannot be produced without the presence of triplets in the system (bottom right display). A cutoff of 2 is thus insufficient for obtaining its population size. However using the moment equations with a cutoff of 3 (+) provides accurate results for the entire range of system sizes.

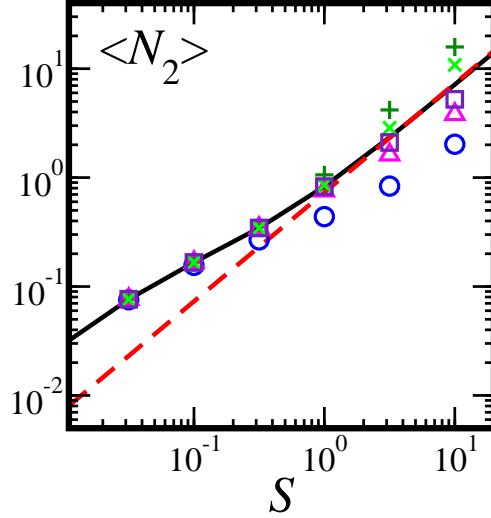


FIG. 3: (Color online) Here we focus on the transition between the stochastic regime $S < 1$ and the deterministic regime $S > 1$. Results are shown for $\langle N_2 \rangle$ obtained from the moment equations with a cutoff of 2 (circles), 3 (+), 4 (triangles), 5 (\times) and 6 (squares). As the cutoff used in the moment equations is raised, the convergence to the rate equation results (dashed line) becomes smoother. When a cutoff higher than 6 is required, one reliably enters the range of validity of the rate equations.