

## Dimensional reduction of the master equation for stochastic chemical networks: The reduced-multiplane method

Baruch Barzel,<sup>1</sup> Ofer Biham,<sup>1</sup> Raz Kupferman,<sup>2</sup> Azi Lipshtat,<sup>3</sup> and Amir Zait<sup>1</sup>

<sup>1</sup>*Racah Institute of Physics, The Hebrew University, Jerusalem 91904, Israel*

<sup>2</sup>*Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel*

<sup>3</sup>*The Gonda Brain Research Center, Bar-Ilan University, Ramat-Gan 52900, Israel*

(Received 8 March 2010; published 16 August 2010)

Chemical reaction networks which exhibit strong fluctuations are common in microscopic systems in which reactants appear in low copy numbers. The analysis of these networks requires stochastic methods, which come in two forms: direct integration of the master equation and Monte Carlo simulations. The master equation becomes infeasible for large networks because the number of equations increases exponentially with the number of reactive species. Monte Carlo methods, which are more efficient in integrating over the exponentially large phase space, also become impractical due to the large amounts of noisy data that need to be stored and analyzed. The recently introduced multiplane method [A. Lipshtat and O. Biham, *Phys. Rev. Lett.* **93**, 170601 (2004)] is an efficient framework for the stochastic analysis of large reaction networks. It is a dimensional reduction method, based on the master equation, which provides a dramatic reduction in the number of equations without compromising the accuracy of the results. The reduction is achieved by breaking the network into a set of maximal fully connected subnetworks (maximal cliques). A separate master equation is written for the reduced probability distribution associated with each clique, with suitable coupling terms between them. This method is highly efficient in the case of sparse networks, in which the maximal cliques tend to be small. However, in dense networks some of the cliques may be rather large and the dimensional reduction is not as effective. Furthermore, the derivation of the multiplane equations from the master equation is tedious and difficult. Here we present the reduced-multiplane method in which the maximal cliques are broken down to the fundamental two-vertex cliques. The number of equations is further reduced, making the method highly efficient even for dense networks. Moreover, the equations take a simpler form, which can be easily constructed using a diagrammatic procedure, for any desired network architecture. It is shown that the method provides accurate results for the population sizes of the reactive species and their reaction rates.

DOI: [10.1103/PhysRevE.82.021117](https://doi.org/10.1103/PhysRevE.82.021117)

PACS number(s): 02.50.Ga, 05.10.Gg, 82.65.+r

### I. INTRODUCTION

Reaction networks commonly appear in physical, chemical, and biological systems [1]. They describe the interaction between chemical species in a liquid solution or on a surface, that react with each other to form more complex species. The reactions are mediated by the mobility of the reacting particles, which may be due to thermal diffusion or convection. In some cases the system is isolated from the environment with no inflow or outflow of particles. In other cases, there is a flow of particles into the system, which react and leave the system by outflow, evaporation, or desorption. In homogeneous systems it is convenient to assume that the systems are well mixed and that the reaction rates do not depend on the location of the particles. Under this assumption, the reaction networks can be analyzed using mean-field models which account for the average concentrations of the reactive species.

Consider a chemical network which involves  $J$  species  $X_i$ ,  $i=1, \dots, J$  with reactions of the form  $X_i+X_j \rightarrow X_k$ . Such networks may be described by graphs. Each species,  $X_i$ , is represented by a node (vertex) in the graph, while the reaction between a pair of species  $X_i$  and  $X_j$  ( $i \neq j$ ) is represented by an edge that connects the corresponding nodes. The reaction product,  $X_k$ , may be indicated near the edge. Networks may also include self reactions of the form  $X_i+X_i \rightarrow X_k$ , represented by loops attached to the node  $X_i$ . Some of the reaction

products may be reactive and participate in further reactions, while others are nonreactive products. Reaction networks are often sparse, namely, only a small fraction of the  $J(J+1)/2$  pairs of nodes are connected with each other. They may include hubs, namely, highly reactive species that are connected to a large fraction of all the other species in the network.

When the reactions take place in a macroscopic system, the number of reactants of each species is typically large. The law of large numbers applies and fluctuations in the concentrations and in the reaction rates become negligible. As a result, such reaction networks can be analyzed using rate equation models. These models account for the average concentrations of all the species involved in the network and their time dependence. The rate equations are based on the mean-field approximation and as such they ignore stochastic fluctuations.

However, one often encounters systems that are partitioned into small domains (such as biological organisms that consist of cells or the surfaces of polycrystalline solids that consist of microcrystals). In such systems, the diffusion of reactants between domains may be suppressed. When the domains are sufficiently small, the copy numbers of the reactive species in each domain may be low. Under these conditions, the temporal fluctuations in these copy numbers in a single domain as well as the variations between different domains cannot be ignored. As a consequence, the rate equa-

tions fail and the analysis of these reactions requires stochastic methods [2,3]. Such methods can be implemented either by direct integration of the master equation or by Monte Carlo (MC) simulations [4].

An important example of the situation described above appears in interstellar clouds of gas and dust [5]. These clouds exhibit a combination of gas-phase chemistry and reactions taking place on the surfaces of microscopic dust grains [5–7]. The essential reactions taking place on dust grains include the formation of molecular hydrogen [8,9] water ice and simple organic molecules such as formaldehyde and methanol [7]. Due to the low gas density and the microscopic size of the grains, the flux of particles (atoms and molecules) impinging on a single grain is low. After sticking to the surface, these particles diffuse on the surface of the grain and may react with each other or desorb back to the gas phase. The copy numbers of the reactive species on a single grain are determined by the balance between the incoming flux on the one hand and the reaction and desorption rates on the other hand. In a broad range of astrophysically relevant conditions, these numbers are small and their fluctuations are significant. It was shown that under these conditions the rate equations are not suitable for the evaluation of the reaction rates [10–14] and stochastic methods are needed [15–17].

In biology, organisms consist of cells, where each cell exhibits a complex network of metabolic processes, signal transduction cascades, protein networks, and gene regulation [18,19]. Some of the crucial components of these networks, such as certain transcription factors and their binding sites appear in low copy numbers. As a result, stochastic fluctuations play an important role in cellular processes [20–22]. Recent technological advances have made it possible to measure the fluctuations in the copy numbers of proteins in single cells, revealing the importance of stochastic effects [23,24].

Theoretically, the most complete description of stochastic reaction networks is in the form of the master equation [2,3]. This is a set of coupled ordinary differential equations, which includes one equation for each microscopic state of the system. The dynamical variables of these equations are the probabilities of these states. The quantities of interest, namely, the average copy numbers of the reactive species and the reaction rates are obtained by proper summations. For some simple networks, the master equation can be solved using a generating function [25,26]. In this approach, the set of coupled ordinary differential equations is transformed into a single partial differential equation for the generating function. This equation can be solved numerically and in a few cases it can also be solved analytically [15,27]. The master equation can also be approximated using the Fokker-Planck equation [28]. This is a partial differential equation in which the population sizes of the reactive species are represented by continuous variables, making the Fokker-Planck equation unsuitable for reaction networks in the limit of small populations, where the discrete nature of the reactive species is important. For complex reaction networks using a generating function is not feasible. Moreover, the direct numerical integration of the master equation is impractical because the number of equations proliferates exponentially with the number of reactive species [29,30].

Since the late 1970s the Gillespie algorithm [31,32] has emerged as the method of choice for the simulation of stochastic chemical networks. This is a kinetic Monte Carlo approach, namely, an algorithm that generates “paths” of the stochastic process. 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 weight. After each move the elapsed time is properly advanced, the list of available processes is updated and their new rates are evaluated. The Gillespie algorithm is exact in the sense that it converges to the exact results as more statistics is accumulated. It also turned out to be remarkably efficient in integrating over an exponentially large phase space. It has enabled researchers to simulate networks of increasing size and complexity and to analyze them in great detail. Improved versions of the algorithm were developed, extending the applicability of the method. Most notably, the tau-leaping method provides a significant speedup [33].

However, the Gillespie approach suffers from fundamental drawbacks. In order to extract expectation values for desired moments and correlations, one needs to perform statistical analyses over large amounts of noisy data. Another problem is that the stochastic systems are often embedded in a macroscopic reservoir (modeled by rate equations). There is no effective way to couple the Gillespie simulation to the surrounding environment. This problem limits the applicability of the method for systems such as interstellar grain chemistry, where the dust grains interact with the surrounding molecular gas. Furthermore, in systems that combine fast and slow processes, the broad range of time scales sharply reduces the efficiency of the algorithm, where the very fast and repetitive moves consume most of the simulation time.

Recently, the multiplane method for the analysis of stochastic reaction networks was introduced [34]. This is an efficient equation-based method, which is derived from the master equation and provides a dramatic reduction in the number of equations. The method is based on breaking the reaction network into a set of maximal fully connected subnetworks, or maximal cliques (Fig. 1). A lower dimensional master equation is derived for the marginal probability distribution associated with each clique. This is done by tracing over all the species that do not belong to the clique and neglecting the correlations between pairs of species that do not react with each other. The result is a set of lower dimensional master equations with suitable coupling between them. These equations are referred to as the multiplane equations. The multiplane method was tested extensively for a class of simple networks [35]. The results obtained for the copy numbers of reactants and for the reaction rates were found to be in excellent agreement with the master equation.

The multiplane method exhibits significant advantages over Monte Carlo methods such as the Gillespie approach. For typical networks found in empirical systems, which tend to be sparse, the number of equations scales moderately with the number of reactive species. The multiplane method provides the marginal probability distributions for the cliques with great efficiency, from which the full probability distribution can be approximated. The method is suitable for the analysis of systems under steady state as well as time-

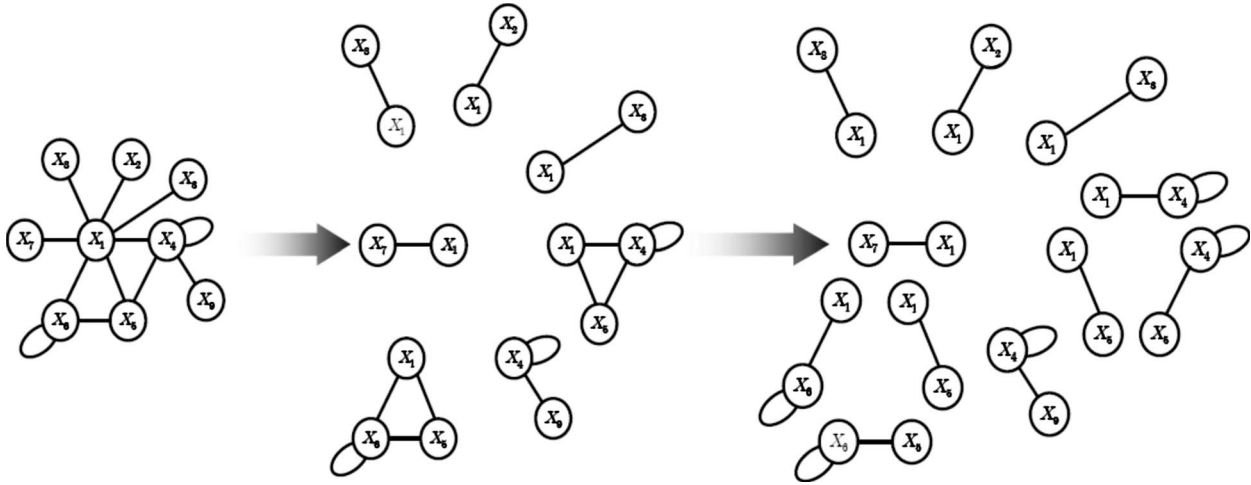


FIG. 1. Graphic illustration of the multiplane and the reduced-multiplane methods for the analysis of stochastic reaction networks. In the multiplane method the network is broken to a set of maximal cliques. A lower dimensional master equation is written for the marginal probability distribution associated with each clique, with suitable couplings between them. In the reduced-multiplane method the maximal cliques are broken into the fundamental two-vertex cliques, each including two molecular species.

dependent conditions. They can be easily coupled to the rate equations that describe the macroscopic dynamics of the environment in which the stochastic system is embedded. Moreover, unlike Monte Carlo methods, the multiplane method deals efficiently with situations involving a large separation of time scales between the fast and slow processes.

The multiplane method is highly efficient in the case of sparse networks, in which a typical node is connected only to a small fraction of the other nodes. However, in dense networks the cliques may be rather large and the dimensional reduction is not as effective. Furthermore, the derivation of the multiplane equations from the master equation (by tracing over the species that are not included in each clique) is tedious and difficult to automate.

In this paper we present the reduced-multiplane method in which the maximal cliques are replaced by the fundamental two-vertex cliques, each including a pair of species that react with each other (Fig. 1). The number of equations is further reduced, making the method suitable not only for sparse networks but for dense networks as well. Moreover, we provide a simple diagrammatic procedure for the construction of the reduced-multiplane equations directly from the network structure. We analyze the method in detail and test its validity and efficiency by applying it to simple network architectures over a range of parameters. The results are compared to those obtained from the master equation. It is shown that the reduced-multiplane method provides accurate results for both the population sizes and reaction rates.

The paper is organized as follows. In Sec. II we review the rate equation approach using a simple example network. In Sec. III we present the master equation formulation. The original multiplane method is described in Sec. IV and its strengths and limitations are discussed. In Sec. V we introduce the reduced-multiplane method. In Sec. VI we present a simple diagrammatic procedure for the construction of the reduced-multiplane equations. In Sec. VII we test the performance of the reduced-multiplane method. The main findings are summarized and discussed in Sec. VIII.

## II. RATE EQUATIONS

Consider a system which consists of a large ensemble of identical domains, with the same reaction network taking place in each domain under the same physical conditions. In the rest of the paper these domains will be referred to as cells. To be specific we discuss below the simple reaction network shown in Fig. 2(a). This network includes five reactive species denoted by  $X_i$ ,  $i=1, \dots, 5$  and six nonreactive product species, denoted by  $X_i$ ,  $i=6, \dots, 11$ . The reactions in this system are  $X_1+X_1 \rightarrow X_6$ ,  $X_1+X_2 \rightarrow X_7$ ,  $X_1+X_3 \rightarrow X_8$ ,  $X_2+X_3 \rightarrow X_9$ ,  $X_2+X_4 \rightarrow X_{10}$ , and  $X_3+X_5 \rightarrow X_{11}$ . Each cell is exposed to an incoming flow  $F_i$  ( $s^{-1}$ ),  $i=1, \dots, 5$  of  $X_i$  particles. Each particle of species  $X_i$ , located in one of the cells, may leave the system at a rate  $W_i$  ( $s^{-1}$ ). The rate constant for the reaction between particles  $X_i$  and  $X_j$  is denoted by  $K_{i,j}$  ( $s^{-1}$ ). For pairs of species that do not react with each other  $K_{i,j}=0$ . We denote the average number of particles of the  $X_i$  species in a cell by  $N_i$ . This is a continuous variable that may take any non-negative value. Below we construct the rate equations for this network. These are coupled ordinary differential equations for the time derivatives of the  $N_i$ 's. The rate equations that describe the network shown in Fig. 2(a) take the form

$$\frac{dN_1}{dt} = F_1 - W_1N_1 - K_{1,2}N_1N_2 - K_{1,3}N_1N_3 - 2K_{1,1}N_1^2,$$

$$\frac{dN_2}{dt} = F_2 - W_2N_2 - K_{1,2}N_1N_2 - K_{2,3}N_2N_3 - K_{2,4}N_2N_4,$$

$$\frac{dN_3}{dt} = F_3 - W_3N_3 - K_{1,3}N_1N_3 - K_{2,3}N_2N_3 - K_{3,5}N_3N_5,$$

$$\frac{dN_4}{dt} = F_4 - W_4N_4 - K_{2,4}N_2N_4,$$

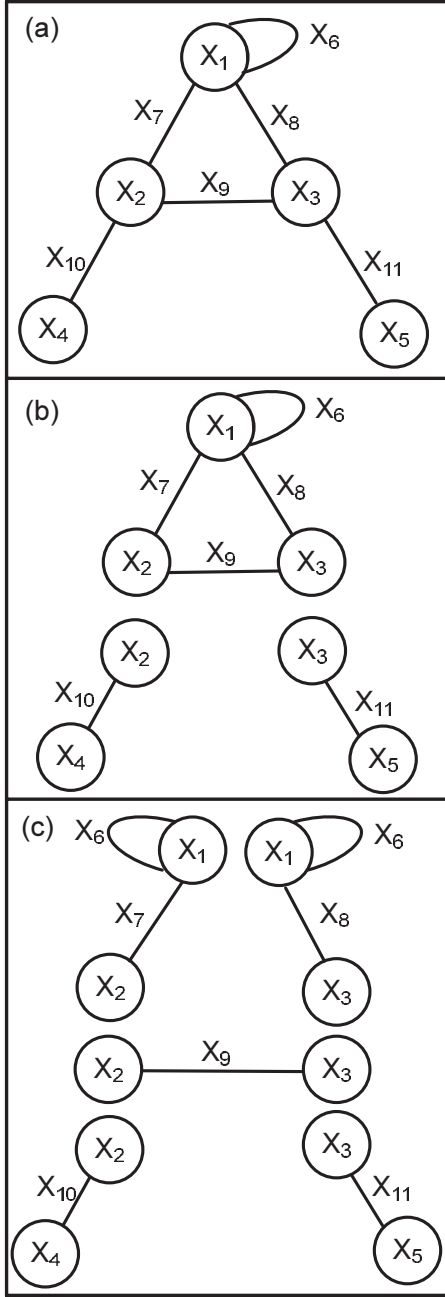


FIG. 2. (a) Graphic representation of the reaction network introduced in Sec. II, which involves five reactive species. The nodes represent reactive species and the edges represent reactions between pairs of species. The reaction products are specified near the edges; (b) the same network broken down into maximal cliques, as described in Sec. IV; (c) the same network broken down into fundamental two-vertex cliques, as described in Sec. V.

$$\frac{dN_5}{dt} = F_5 - W_3N_5 - K_{3,5}N_3N_5. \quad (1)$$

The first terms on the right hand side represent the flow of  $X_i$  particles into the cells. The second terms represent the removal of  $X_i$  particles, which is proportional to their average population size in the cells. The remaining terms account for the reactions between particles. More specifically, terms of

the form  $K_{i,j}N_iN_j$  represent reactions between particles of different species, while terms of the form  $2K_{i,i}N_i^2$  represent reactions between particles of the same species. The average production rate of the  $X_6$  particles per cell is given by  $R_6 = K_{1,1}N_1^2$ , while the production rates of the other product species  $X_k$ ,  $k=7, \dots, 11$  are given by  $R_k = K_{i,j}N_iN_j$ , where  $i$  and  $j$  are the corresponding reactive species. For simplicity, we assume here that nonreactive product species leave the system immediately upon formation. In this case the system approaches a steady state condition in which the average populations of the reactive species in a cell are fixed.

For large cells in which all the reactive species are abundant, Eq. (1) accounts correctly for their copy numbers and for the reaction rates. However, in the limit of small cells, some of the average population sizes,  $N_i$ , may become small. In this case the discrete nature of the particles becomes important and the fluctuations cannot be ignored. As a result, the reaction rates obtained from the rate Eqs. (1) are incorrect. The domain of validity of the rate equations was studied systematically in Ref. [36]. An expression was derived for the cell size below which stochastic fluctuations should be taken into account.

### III. MASTER EQUATION

To account for the discrete nature of the particles and for the stochastic effects, we use the master equation. Here, the population sizes of the reactive species are discrete, namely,  $N_i$  takes only non-negative integer values. In a system that includes  $J$  reactive species, the dynamical variables of the master equation are the joint probabilities  $P(N_1, N_2, \dots, N_J)$  of a cell to contain  $N_i$  particles of species  $X_i$ . The master equation for the network presented in Fig. 2(a) is

$$\begin{aligned} & \frac{dP(N_1, \dots, N_5)}{dt} \\ &= \sum_{i=1}^5 F_i [P(\dots, N_i - 1, \dots) - P(N_1, \dots, N_5)] \\ &+ \sum_{i=1}^5 W_i [(N_i + 1)P(\dots, N_i + 1, \dots) - N_i P(N_1, \dots, N_5)] \\ &+ \sum_{i < j} K_{i,j} [(N_i + 1)(N_j + 1)P(\dots, N_i + 1, N_j + 1, \dots) \\ &- N_i N_j P(N_1, \dots, N_5)] \\ &+ \sum_i K_{i,i} [(N_i + 2)(N_i + 1)P(\dots, N_i + 2, \dots) \\ &- N_i(N_i - 1)P(N_1, \dots, N_5)]. \end{aligned} \quad (2)$$

The first term in Eq. (2) describes the flow of particles into the cell. The second term accounts for the removal of particles from the cell. The third term describes the reactions between pairs of particles of different species. The last term accounts for the reaction between pairs of  $X_1$  particles, resulting in dimers denoted by  $X_6$ .

The average population size of  $X_i$ ,  $i=1, \dots, 5$  particles in a cell is given by

$$\langle N_i \rangle = \sum_{N_1, \dots, N_5} N_i P(N_1, \dots, N_5). \quad (3)$$

Consider a reaction of the form  $X_i + X_j \rightarrow X_k$ . For  $i \neq j$  the production rate of  $X_k$  particles in a cell is given by  $R_k = K_{i,j} \langle N_i N_j \rangle$ , while for  $i = j$  the production rate is  $R_k = K_{i,i} (\langle N_i^2 \rangle - \langle N_i \rangle)$ , where the second moments are given by

$$\langle N_i N_j \rangle = \sum_{N_1, \dots, N_5} N_i N_j P(N_1, \dots, N_5). \quad (4)$$

In the network of Fig. 2(a), the production rate of  $X_6$  particles per cell is given by  $R_6 = K_{1,1} (\langle N_1^2 \rangle - \langle N_1 \rangle)$ . The production rates of the other product species  $X_k$ ,  $k = 7, \dots, 11$  are given by  $R_k = K_{i,j} \langle N_i N_j \rangle$ , where  $X_i$  and  $X_j$  are the corresponding reactants according to Fig. 2(a).

In numerical studies the master equation must be truncated in order to keep the number of equations finite. A convenient way to achieve this is to assign upper cutoffs  $N_i^{\max}$ ,  $i = 1, \dots, J$  on the population sizes. The number of coupled equations is thus  $N_E = \prod_{i=1}^J (N_i^{\max} + 1)$ . The truncated master equation is valid if the probability to have, in a single cell, larger populations beyond the cutoffs is vanishingly small. However, the number of equations,  $N_E$ , grows exponentially as the number of reactive species increases, severely limiting the applicability of the master equation [29,30].

#### IV. MULTIPLANE METHOD

The recently introduced multiplane method [34] provides an efficient computational framework for the analysis of

complex reaction networks. To construct the multiplane equations one first has to break the network down into a set of maximal fully connected subnetworks (maximal cliques). For example, in the network of Fig. 2(a) the maximal cliques are  $C_1^{\text{MP}} = \{X_1, X_2, X_3\}$ ,  $C_2^{\text{MP}} = \{X_2, X_4\}$ , and  $C_3^{\text{MP}} = \{X_3, X_5\}$ , as shown in Fig. 2(b). A lower dimensional master equation for the reduced probability distribution associated with each clique is obtained by tracing over all the species that do not belong to that clique.

As an example we will describe this construction for the clique  $C_1^{\text{MP}}$ . The resulting equation includes conditional moments of the form  $\langle N_i \rangle_{N_1, N_2, N_3}$  where  $i = 4, 5$ . Such conditional moments are defined by

$$\langle N_i \rangle_{N_1, N_2, N_3} = \sum_{N_i} N_i P(N_i | N_1, N_2, N_3). \quad (5)$$

We then introduce an approximation that amounts to the assumption that the correlation between a pair of species that do not react with each other is weak and can be neglected. This allows us to replace the moment  $\langle N_4 \rangle_{N_1, N_2, N_3}$  by  $\langle N_4 \rangle_{N_2}$  and the moment  $\langle N_5 \rangle_{N_1, N_2, N_3}$  by  $\langle N_5 \rangle_{N_3}$ . The equations for different maximal cliques are coupled via these conditional moments. For instance, the conditional moment  $\langle N_4 \rangle_{N_2}$  is evaluated from the marginal probability distribution  $P(N_2, N_4)$  associated with the clique  $C_2^{\text{MP}}$ . The resulting lower dimensional master equations are referred to as the multiplane equations. For the clique  $C_1^{\text{MP}}$ : the multiplane equations take the form

$$\begin{aligned} \frac{dP(N_1, N_2, N_3)}{dt} = & \sum_{i=1}^3 F_i [P(\dots, N_i - 1, \dots) - P(N_1, N_2, N_3)] \\ & + \sum_{i=1}^3 W_i [(N_i + 1)P(\dots, N_i + 1, \dots) - N_i P(N_1, N_2, N_3)] \\ & + \sum_{j=1, i < j}^3 K_{i,j} [(N_i + 1)(N_j + 1)P(\dots, N_i + 1, N_j + 1, \dots) - N_i N_j P(N_1, \dots, N_5)] \\ & + K_{2,4} [(N_2 + 1) \langle N_4 \rangle_{N_2+1} P(N_1, N_2 + 1, N_3) - N_2 \langle N_4 \rangle_{N_2} P(N_1, N_2, N_3)] \\ & + K_{3,5} [(N_3 + 1) \langle N_5 \rangle_{N_3+1} P(N_1, N_2, N_3 + 1) - N_3 \langle N_5 \rangle_{N_3} P(N_1, N_2, N_3)] \\ & + K_{1,1} [(N_1 + 2)(N_1 + 1)P(N_1 + 2, N_2, N_3) - N_1(N_1 - 1)P(N_1, N_2, N_3)]. \end{aligned} \quad (6)$$

Using a similar procedure, one can construct the multiplane equations for the cliques  $C_2^{\text{MP}}$  and  $C_3^{\text{MP}}$ . The resulting system of coupled ordinary differential equations can be integrated numerically using a standard stepper such as a fourth order Runge Kutta. In case that one is interested only in the steady state solution, the equations can also be solved as a set of algebraic equations by setting the time derivatives on the left-hand side to zero. From the resulting (approximated)

marginal probability distributions one can obtain all the physically interesting quantities such as the average population sizes of all the reactive species and the reaction rates. Moreover, from the steady state solution one can obtain single time measures for the fluctuation levels such as the variance in the population size of a single species  $\sigma_i^2 = \langle N_i^2 \rangle - \langle N_i \rangle^2$  and the two-species correlation  $C_{i,j} = (\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle) / (\sigma_i \sigma_j)$ . From the full time-dependent solution one

obtain additional information such as the autocorrelation function for a single species.

For the network of Fig. 2(a) the multiplane method achieves a dimensional reduction from a five-dimensional master equation to three sets of multiplane equation, one of which is three-dimensional and the other two are two-dimensional. In general, for sparse reaction networks most cliques do not include more than two or three species. Under these conditions the multiplane method is very efficient. However, in dense networks the dimensional reduction achieved by the multiplane method is not as effective. Furthermore, the derivation of the multiplane equations from the master equation is tedious and difficult to automate.

To solve these problems we introduce a simplified version of the multiplane method, referred to as the reduced-multiplane method. In this approach the maximal cliques are broken down into the fundamental two-vertex cliques, each containing a pair of species reacting with each other.

### V. REDUCED-MULTIPLANE METHOD

Consider the clique  $\mathcal{C}_1^{\text{MP}}$  discussed above. It consists of three species ( $X_1$ ,  $X_2$  and  $X_3$ ) and three reactions, that produce the species  $X_7$ ,  $X_8$ , and  $X_9$  (in addition to the dimer species  $X_6$ ). The production rates of these species are given in terms of the moments  $\langle N_i N_j \rangle$ . Each of these moments can be derived from the corresponding two-dimensional marginal probability distribution  $P(N_i, N_j)$ . In order to obtain the population sizes of the reactive species and the reaction rates it is thus sufficient to calculate the two-dimensional marginal

probability distributions for all the pairs of species that react with each other.

In the reduced-multiplane method, one divides the network into the fundamental two-vertex cliques, each corresponding to a pair of species that react with each other. For the network shown in Fig. 2(a) the fundamental two-vertex cliques are

$$\begin{aligned} \mathcal{C}_1^{\text{RMP}} &= \{X_1, X_2\}, \\ \mathcal{C}_2^{\text{RMP}} &= \{X_1, X_3\}, \\ \mathcal{C}_3^{\text{RMP}} &= \{X_2, X_3\}, \\ \mathcal{C}_4^{\text{RMP}} &= \{X_2, X_4\}, \\ \mathcal{C}_5^{\text{RMP}} &= \{X_3, X_5\}, \end{aligned} \tag{7}$$

as shown in Fig. 2(c). To derive the reduced-multiplane equations for a certain clique, one follows the procedure used in the multiplane method and traces over all the species that do not belong to that clique. As in the original procedure, one obtains conditional moments, which can be approximated by the rule stated above. According to this rule, only the correlations between species that directly react with each other are maintained. For example, in the term describing the reaction  $X_2 + X_4 \rightarrow X_{10}$  in the reduced-multiplane equations for the clique  $\mathcal{C}_1^{\text{RMP}}$ , the moment  $\langle N_4 \rangle_{N_1, N_2}$  is approximated by  $\langle N_4 \rangle_{N_2}$  (because  $X_1$  does not react with  $X_4$ ). The reduced-multiplane equations for the clique  $\mathcal{C}_1^{\text{RMP}}$  take the form

$$\begin{aligned} \frac{dP(N_1, N_2)}{dt} &= \sum_{i=1}^2 F_{i1} [P(\dots, N_i - 1, \dots) - P(N_1, N_2)] \\ &+ \sum_{i=1}^2 W_{i1} [(N_i + 1)P(\dots, N_i + 1, \dots) - N_i P(N_1, N_2)] \\ &+ K_{1,2} [(N_1 + 1)(N_2 + 1)P(N_1 + 1, N_2 + 1) - N_1 N_2 P(N_1, N_2)] \\ &+ \sum_{i=1}^2 K_{i,3} [(N_i + 1) \langle N_3 \rangle_{N_i+1} P(\dots, N_i + 1, \dots) - N_i \langle N_3 \rangle_{N_i} P(N_1, N_2)] \\ &+ K_{2,4} [(N_2 + 1) \langle N_4 \rangle_{N_2+1} P(N_1, N_2 + 1) - N_2 \langle N_4 \rangle_{N_2} P(N_1, N_2)] \\ &+ K_{1,1} [(N_1 + 2)(N_1 + 1)P(N_1 + 2, N_2) - N_1(N_1 - 1)P(N_1, N_2)]. \end{aligned} \tag{8}$$

When computing the reaction term for  $X_1$  and  $X_3$  the conditional moment  $\langle N_3 \rangle_{N_1}$ , is computed from the clique  $\mathcal{C}_2^{\text{RMP}}$  but when computing the reaction term between  $X_2$  and  $X_3$  we use  $\langle N_3 \rangle_{N_2}$ , which is obtained from the clique  $\mathcal{C}_3^{\text{RMP}}$ . It should be noted that the average population size  $\langle N_3 \rangle$  can be computed independently, using either one of these two cliques. Since the equations associated with the two cliques are consistent,

the results are expected to be similar. The result of this approximation is a set of reduced-multiplane equations for each clique, with suitable couplings between adjacent cliques.

Beyond the specific example network of Fig. 2(a) we now consider the approximations involved in the construction of the reduced-multiplane equations for a general reaction network of  $J$  reactive species. In order to obtain the reduced-multiplane equations  $dP(N_i, N_j)/dt = \dots$ , for the pair of spe-

cies  $X_i$  and  $X_j$  which react with each other, one first traces the master equation over all the other species. The resulting equations contain moments conditioned on the values  $N_i$  and  $N_j$ . The approximation performed in the reduced-multiplane method consists of eliminating the dependence of such moments on either  $N_i$ ,  $N_j$  or both, according to the following scheme. For mixed moments, containing two species that do not belong to the pair, the dependence is dropped completely. The marginal probability distributions  $P(N_k, N_l)$  available in this method, do not include the correlations required in order to evaluate an expression of the form  $\langle N_k N_l \rangle_{N_i, N_j}$ . Therefore, we use the approximation

$$\langle N_k N_l \rangle_{N_i, N_j} \approx \langle N_k N_l \rangle, \quad (9)$$

where the right hand side can be evaluated from the pairwise marginal probability distribution  $P(N_k, N_l)$ . The availability of  $P(N_k, N_l)$  is guaranteed, otherwise the term  $\langle N_k N_l \rangle_{N_i, N_j}$  would not appear in the first place.

For moments containing only one species, which is neither  $X_i$  nor  $X_j$  (expressed in general as  $\langle N_k^m \rangle_{N_i, N_j}$ , where  $n = 1, 2$ ), one first has to identify the term in the master equation that generated this moment. One then examines the reaction associated with that term. If this is a reaction of the form  $X_i + X_k \rightarrow X_l$ , it is clear that the conditional moment  $\langle N_k^m \rangle_{N_i, N_j}$  depends more strongly on  $N_i$  than on  $N_j$ . Therefore, we ignore its dependence on  $N_j$ , leading to the approximation

$$\langle N_k^m \rangle_{N_i, N_j} \approx \langle N_k^m \rangle_{N_i}. \quad (10)$$

Similarly, if the reaction is of the form  $X_j + X_k \rightarrow X_l$ , the approximation we use is

$$\langle N_k^m \rangle_{N_i, N_j} \approx \langle N_k^m \rangle_{N_j}. \quad (11)$$

If the reaction which gives rise to the moment  $\langle N_k^m \rangle_{N_i, N_j}$  is of the form  $X_k + X_k \rightarrow X_i$ , one needs to distinguish between three cases. If in addition to the above reaction,  $X_i$  and  $X_k$  interact through a reaction of the form  $X_i + X_k \rightarrow X_l$ , it is clear that there is a stronger dependence of the aforementioned conditional moment on  $N_i$  than on  $N_j$ . As a result, the following approximation is sensible:

$$\langle N_k^m \rangle_{N_i, N_j} \approx \langle N_k^m \rangle_{N_i}. \quad (12)$$

If no such reaction exists, the dependence of the moment on  $N_i$  is not directly available in any of the existing marginal probability distributions. Therefore, it cannot be evaluated within the reduced-multiplane approach. One still needs to distinguish between two cases: either a reaction of the form  $X_j + X_k \rightarrow X_l$  appears in the network, or there is no such reaction. In case that such a reaction appears, the dependence on  $N_j$  can still be evaluated using the marginal probability distribution  $P(N_j, N_k)$  which is guaranteed to exist. The approximation which is taken in this case is

$$\langle N_k^m \rangle_{N_i, N_j} \approx \langle N_k^m \rangle_{N_j}. \quad (13)$$

If the reaction  $X_j + X_k \rightarrow X_l$  does not appear in the network both dependencies cannot be evaluated, leading to the approximation

$$\langle N_k^m \rangle_{N_i, N_j} \approx \langle N_k^m \rangle. \quad (14)$$

This unconditioned moment can be evaluated from any marginal probability distribution containing  $X_k$ .

The dimensional reduction obtained by the multiplane and by the reduced-multiplane methods for a network of  $J$  reactive species is evaluated below. For simplicity, we use a simple truncation scheme in which the cutoffs for all the reactive species are the same, namely,  $N_i^{\max} = N^{\max}$ . The number of equations in the master equation is  $N_E^{\text{Master}} = (N^{\max} + 1)^J$ . The dimensional reduction obtained from the multiplane method depends on the connectivity of the network. Consider a network that consists of  $m$  maximal cliques,  $C_k^{\text{MP}}$ ,  $k = 1, \dots, m$ , where the number of species in the  $k$ th clique is  $J_k$ . The number of multiplane equations will be  $N_E^{\text{MP}} = \sum_{k=1}^m (N^{\max} + 1)^{J_k}$ . In the reduced-multiplane method, for a network with  $r$  reactions of the form  $X_i + X_j \rightarrow X_k$  ( $i \neq j$ ), one divides the network into  $r$  fundamental two-vertex cliques. The number of equations will then be  $N_E^{\text{RMP}} = r(N^{\max} + 1)^2$ . The exponential proliferation of the master equation as a function of the number of reactive species is lowered, in the reduced-multiplane method, to at most a quadratic dependence.

The reduced-multiplane method has been tested for different network architectures and for a broad range of parameters. The results for the population sizes and for the reaction rates were found to be in excellent agreement with the master equation. Moreover, for a class of simple networks, we performed an asymptotic analysis of the reduced-multiplane equations in both the *large cell* limit and the *small cell* limit [37]. The simplifying assumption made in these networks is that all the product species leave the system and do not participate in any further reactions. In addition, it is assumed that each reaction involves only two particles. In both limits the reduced-multiplane method was found to be asymptotically exact.

In the so-called *large cell* limit, the incoming flows of reactants are large, and the reaction rates constants are small, leading to large population sizes. In this limit the population sizes of the reactive species can be considered as continuous variables. The distribution of population sizes is narrowly concentrated around the average value. The master equation can then be transformed into a partial differential equation, for which an approximate solution is obtained, where the deviation from the average population size is considered as a small parameter. Formally, the analysis of both the reduced-multiplane equations and the master equation was performed using Van Kampen's  $\Omega$ -expansion [2,38]. To leading order, the production rates and population sizes obtained from the reduced-multiplane equations are in perfect agreement with those obtained from the master equation.

For the *small cell* limit, where the reactions rate constants are large and the incoming flows are low, the steady state distributions were obtained using a small-parameter expansion. Here the small parameter is taken to be the system size. The incoming fluxes of particles are proportional to the system size while the reaction rates are inversely proportional to the system size. Both the reduced multiplane and the master equation can be solved order by order in this small param-

eter. Again, the results are found to be in perfect agreement with those obtained from the master equation.

In this paper we study more general networks, in which reaction products can participate in further reactions as well as dissociation reactions in which one particle is transformed to one or more product particles. It turns out to be difficult to extend the asymptotic analysis to the more general networks described above. The main difficulty is that there is an *a priori* assumption about the size of conditional moments appearing in the reduced-multiplane method. This assumption greatly simplifies the analysis and is verified *a posteriori*. This assumption no longer holds in the case of more general reactions such as dissociation.

## VI. CONSTRUCTING THE REDUCED-MULTIPLANE EQUATIONS

Since the construction of the reduced-multiplane equations from the master equation is tedious, we have developed a diagrammatic procedure that enables to obtain the equations directly from the network graph. This procedure applies to networks that include reactions of the form



We first choose a suitable cutoff  $N_i^{\max}$ ,  $i=1, \dots, J$  for each reactive species. Then we proceed to write down the set of equations for each one of the fundamental two-vertex cliques. To demonstrate the construction of the reduced multiplane equations in practice, consider the single clique including the nodes  $X_i$  and  $X_j$ , which react to form  $X_m$  ( $X_i + X_j \rightarrow X_m$ ). The reduced-multiplane equations for the marginal probability distribution  $P(N_i, N_j)$  include two types of terms, internal and external. The internal terms are those associated with these two species alone and would exist even in an isolated network which includes only these two species and the reaction between them. These terms describe the incoming flux of each of the two species, the desorption of the two species, and the reaction between them. The external terms describe all the other reactions in which  $X_i$  or  $X_j$  participate, either as reactants or as products. While the internal terms have a fixed form, the external terms depend on the local structure of the network in the vicinity of the  $X_i$ - $X_j$  clique. The reduced-multiplane equations for this clique take the form

$$\begin{aligned} \frac{dP(N_i, N_j)}{dt} = & F_i [P(N_i - 1, N_j) - P(N_i, N_j)] \\ & + F_j [P(N_i, N_j - 1) - P(N_i, N_j)] \\ & + W_i [(N_i + 1)P(N_i + 1, N_j) - N_i P(N_i, N_j)] \\ & + W_j [(N_j + 1)P(N_i, N_j + 1) - N_j P(N_i, N_j)] \\ & + K_{i,j} [(N_i + 1)(N_j + 1)P(N_i + 1, N_j + 1) \\ & - N_i N_j P(N_i, N_j)] \\ & + \text{External Terms (taken from Table I),} \end{aligned} \quad (16)$$

where the internal terms are written explicitly and the exter-

nal terms should be assembled from Table I according to the network components adjacent to the clique. These components include reactions between  $X_i$  (or  $X_j$ ) and species that belong to other cliques, self reactions, as well as reactions in which  $X_i$  (or  $X_j$ ) are products. Whenever the network includes a reaction element listed in the Table, the corresponding term should be added to the reduced-multiplane equations for  $P(N_i, N_j)$ .











## VII. SIMULATIONS AND RESULTS

The formulation of chemical networks used in this paper is very general and can be applied to many different systems. For example, the cells in which the reactions take place can be three-dimensional or two-dimensional. Examples of three dimensional cells are liquid droplets (such as aerosols) and biological cells, in which the reactions take place in the bulk. Examples of two-dimensional cells include dust grains in interstellar clouds, where the reactions take place on the surface. In order to apply the reaction network formulations to these systems one has to calculate the rate constants  $F_i$ ,  $W_i$ , and  $K_{i,j}$  as a function of the cell size and the physical conditions.

To be specific, we will consider the case in which the cells consist of spherical liquid droplets or grains of radius  $r$ , surrounded by a gas phase. In this case the influx  $F_i$  of  $X_i$  particles into each cell is given by  $F_i = \rho_i v_i \sigma$  where  $\rho_i$  is the density of  $X_i$  particles in the gas phase,  $v_i$  is their average velocity and  $\sigma = \pi r^2$  is the cross section of the spheres. Here the rate constant  $W_i$  describes the evaporation rate of a single particle from a liquid droplet or the desorption rate from a grain surface. In the case of grains  $W_i$  does not depend on the radius. In the case of liquid droplets  $W_i$  is inversely proportional to  $r$  because only those particles adjacent to the surface may evaporate. In the case of liquid droplets the reactions are mediated by diffusion in the liquid solution while in the case of solid grains they are mediated by diffusion on the surface. On grain surfaces, the adsorbed particles reside in adsorption sites. The density of these adsorption sites on the surface is denoted by  $s$  ( $\text{cm}^{-1}$ ). The total number of sites on the surface of a grain of radius  $r$  is  $S = 4\pi r^2 s$ . While residing on a grain, the particles diffuse on its surface via hopping between adjacent sites. The hopping rate of  $X_i$  particles is given by  $a_i$  (hops  $\text{s}^{-1}$ ). A more complete description of the diffusion and desorption mechanisms and their temperature dependence is given in Refs. [16,39]. It is convenient to define the sweeping rate  $A_i = a_i/S$ , which is approximately the inverse of the time it takes an adsorbed  $X_i$  atom to visit nearly all the adsorption sites on the grain surface. It has been argued that this approximation neglects the nature of two-dimensional diffusion (back diffusion) as well as the fundamental first-passage problem expressed by the competition between diffusion-mediated encounters and desorption. A more precise expression for the sweeping rate, taking these effects into account, is given in Ref. [40]. Since the purpose of the present paper is to demonstrate the method and not to provide specific quantitative results, we chose to stick to the simpler expression. The analog of the sweeping rate in the case of liquid droplets would take the form  $A_i = a_i/V$ , where



TABLE I. Diagrammatic approach for the construction of the reduced-multiplane equations. To construct the equations for the fundamental two-vertex clique of species  $X_i$  and  $X_j$ , each element in the graph is substituted by the corresponding equation term, which is added to Eq. (16). Note that for each diagram in the table, one should also consider the analogous diagram in which  $i$  and  $j$  are exchanged and add the corresponding term as well. Also, note that  $k$  and  $l$  represent species that do not belong to this particular clique, namely,  $k, l \notin \{i, j\}$ . Regarding the last three diagrams, which represent the reaction  $X_k + X_k \rightarrow X_i$ , only one of them should be included, depending on the appearance of the network element shown in the condition column. The upper one should be chosen in case that  $X_i$  and  $X_k$  react with each other. If not, the second term should be included if  $X_j$  reacts with  $X_k$ . Otherwise, the last term should be included.

Element	Condition	Equation term
		$K_{i,k}[(N_i+1)\langle N_k \rangle_{N_i+1} P(N_i+1, N_j-1) - N_i \langle N_k \rangle_{N_i} P(N_i, N_j)]$
		$K_{i,k}[(N_i+1)\langle N_k \rangle_{N_i+1} P(N_i+1, N_j) - N_i \langle N_k \rangle_{N_i} P(N_i, N_j)]$
		$K_{k,l} \langle N_k N_l \rangle [P(N_i-1, N_j) - P(N_i, N_j)]$
		$K_{i,i}[(N_i+2)(N_i+1)P(N_i+2, N_j-1) - N_i(N_i-1)P(N_i, N_j)]$
		$K_{i,i}[(N_i+2)(N_i+1)P(N_i+2, N_j) - N_i(N_i-1)P(N_i, N_j)]$
		$K_{k,k}[\langle N_k^2 - N_k \rangle_{N_i-1} P(N_i-1, N_j) - \langle N_k^2 - N_k \rangle_{N_i} P(N_i, N_j)]$
		$K_{k,k} \langle N_k^2 - N_k \rangle_{N_j} [P(N_i-1, N_j) - P(N_i, N_j)]$
		$K_{k,k}(\langle N_k^2 \rangle - \langle N_k \rangle) [P(N_i-1, N_j) - P(N_i, N_j)]$

$V$  is the volume of the droplet and  $a_i$  is a parameter, which is linearly proportional to the diffusion coefficient of  $X_i$  particles.

Having defined the sweeping rates  $A_i$ ,  $i=1, \dots, J$  we can now relate them to the reaction rate coefficients  $K_{i,j}$ . Consider the reaction between particles of species  $X_i$  and  $X_j$  ( $i \neq j$ ), that takes place when two such particles encounter each other. Since both of them diffuse simultaneously, the rate constant for their reaction is given by  $K_{i,j}=A_i+A_j$ . For pairs of species  $X_i$  and  $X_j$  that do not react with each other  $K_{i,j}=0$ . The case of reactions between particles of the same species ( $i=j$ ) should be considered separately. In this case, the sum of the sweeping rates for each pair is  $2A_i$ . However, the number of pairs of  $X_i$  particles is  $N_i(N_i-1)/2 \approx N_i^2/2$ . Absorbing this factor of  $1/2$  into the rate coefficient we obtain  $K_{i,i}=A_i$  for species  $X_i$  which form dimers and  $K_{i,i}=0$  for all other species. From this construction we obtain the full  $J \times J$  matrix of reaction rate coefficients.

In the simulations presented below we use the parametrization of reactions on the surfaces of spherical grains. This example is particularly convenient because the incoming fluxes  $F_i$  are linearly proportional to  $S$  while the reaction rates  $K_{i,j}$  are inversely proportional to  $S$ . To be more specific, we consider the chemical network presented in Fig. 2, taking place on the surfaces of small spherical grains of radius  $r$ . In order to test the reduced-multiplane method, we solve the reduced-multiplane equations for this network by direct numerical integration. The steady state results are then compared to those of the master equation and to the results obtained from the rate equations.

In the network of Fig. 2 the grains are exposed to fluxes of the reactive species  $X_i$ ,  $i=1, \dots, 5$ . These fluxes are given by  $F_1=10^{-7}S$  ( $s^{-1}$ ),  $F_2=0.01F_1$ ,  $F_3=0.1F_1$ ,  $F_4=0.01F_1$ , and  $F_5=0.05F_1$ . These parameters are chosen so that  $X_1$  is the dominant species. The desorption rates are  $W_1=10^{-4}$  ( $s^{-1}$ ),  $W_2=W_1$ ,  $W_3=W_4=0.1W_1$  and  $W_5=0.01W_1$ . The sweeping rates are:  $A_1=1/S$  ( $s^{-1}$ ),  $A_2=A_5=0.1A_1$ ,  $A_3=A_1$  and  $A_4=0.01A_1$ .

The results obtained under steady state conditions are presented in Fig. 3. In Fig. 3(a) we show the population sizes of the  $X_1$ ,  $X_2$ ,  $X_3$ , and  $X_5$  species on a grain versus  $S$ , obtained from the reduced-multiplane equations (symbols). The results are in excellent agreement with those obtained from the master equation (solid lines). The results of the rate equations (dashed lines) coincide with the multiplane and master equations for large grains, but exhibit significant deviations for small grains. In Fig. 3(b) we show the production rates of  $X_6$ ,  $X_8$ , and  $X_{10}$  particles on a single grain, versus  $S$ , obtained from the reduced-multiplane method (symbols). Again, the results are in excellent agreement with those obtained from the master equation (solid lines), while the rate equations (dashed lines) exhibit large deviations for small grains.

We next consider a more complex network described by the graph shown in Fig. 4. The topology of this network is taken from the methanol-producing network on ice-coated grains in the interstellar medium [29,34]. In this network, some of the products of the various reactions take part as reactants in other reactions. This gives rise to additional correlations between the population sizes of different species. It was already shown that the multiplane method is suitable for

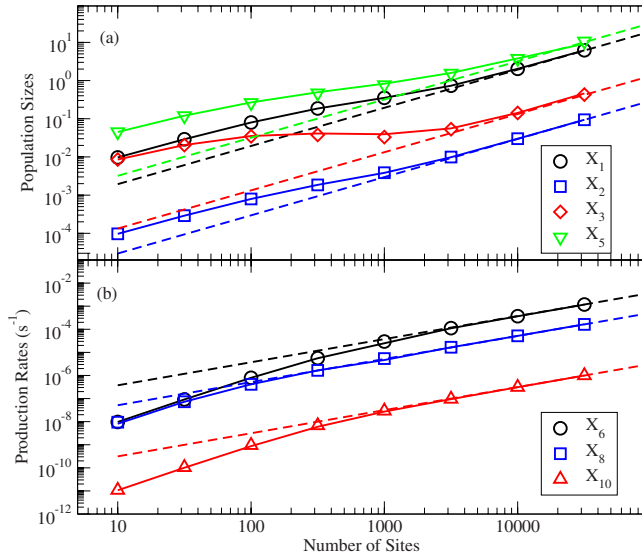


FIG. 3. (Color online) The average population sizes of the reactive species (a) and the production rates of product species (b), obtained from the reduced-multiplane equations (symbols) for the network of Fig. 2, under steady state conditions. The results are in perfect agreement with those obtained from the master equation (solid line). The rate equations (dashed lines) exhibit significant deviations in the small system limit.

the analysis of such systems [34,35]. Below we demonstrate that the reduced-multiplane method also deals successfully with the additional correlations that appear in these more complex networks.

The parameters we used in the analysis of the network of Fig. 4 are given below. The fluxes for this network are  $F_1 = 10^{-8}S$  (s<sup>-1</sup>),  $F_2 = 0.01F_1$ , and  $F_7 = 0.001F_1$ . All other species have no incoming flux. The desorption rates are  $W_1 = 10^{-4}$  (s<sup>-1</sup>),  $W_2 = W_1$ ,  $W_3 = W_{10} = 0.05W_1$ ,  $W_7 = 0.01W_1$ ,  $W_8 = 0.001W_1$ , and  $W_9 = 0.05$ . The sweeping rates are  $A_1 = 1/S$  (s<sup>-1</sup>),  $A_2 = 0.1A_1$ ,  $A_3 = 0.01A_1$ ,  $A_7 = A_8 = 0.001A_1$ ,  $A_9 = 0.0005A_1$ , and  $A_{10} = 0.00005A_1$ .

In Fig. 5 we present the steady state solutions for the rate equations (dashed lines), master equation (solid lines) and reduced-multiplane equations (symbols) describing this net-

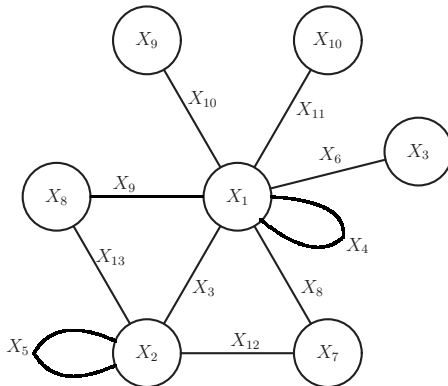


FIG. 4. Graphic representation of a more complex reaction network, which is analyzed in Sec. VII. This network includes seven reactive species, some of them products of reactions between other species.

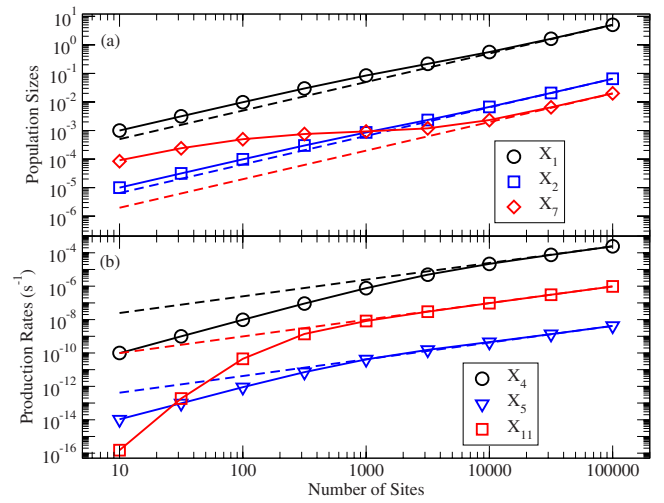


FIG. 5. (Color online) The average population sizes of the reactive species (a) and the production rates of product species (b), obtained from the reduced-multiplane equations (symbols) for the network of Fig. 4, under steady state conditions. The results are in perfect agreement with those obtained from the master equation (solid line). The rate equations (dashed lines) exhibit significant deviations in the small system limit.

work. The results computed using the reduced-multiplane method (symbols) are found to be in excellent agreement with those of the master equation. The rate equations, however, exhibits significant deviations in the limit of small grains.

So far we have considered only the steady state solutions obtained from the reduced-multiplane method. In Fig. 6 we present the time-dependent population sizes [Fig. 6(a)] and production rates [Fig. 6(b)] of some of the species in the reaction network shown in Fig. 4 for  $S=1000$ . The initial condition for this plot was the state in which the system is empty, namely,  $N_i=0$  for  $i=1, \dots, 7$ . The results of the reduced-multiplane method (symbols) are found to be in excellent agreement with those obtained from the master equation (solid lines). Our numerical tests show that the reduced-multiplane method is applicable away from steady state conditions and provides accurate results for the first and second moments.

### VIII. SUMMARY AND DISCUSSION

We have presented an efficient method for the analysis of stochastic reaction networks. The method is based on breaking the network into a set of fundamental two-vertex cliques, each consisting of two species that react with each other. This is a simplification of the multiplane method, in which the network is decomposed into a set of maximal cliques. The advantages of the reduced-multiplane method are particularly significant in the case of large networks in which the connectivity is dense. In this case the numerical integration of the master equation is impractical. The multiplane method is of limited efficiency due to the appearance of large cliques. A further advantage is that the new method leads to simpler equations, which can be constructed by an automated

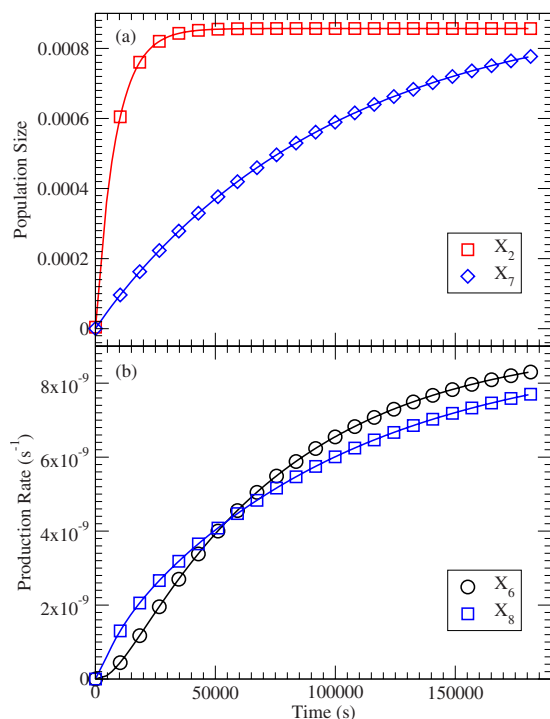


FIG. 6. (Color online) The time-dependent average population sizes of two of the reactive species (a) and the production rates of two product species (b), obtained from the reduced-multiplane equations (symbols) for the network of Fig. 4. The results are in perfect agreement with those obtained from the master equation (solid line).

procedure. To be concrete, we applied the method to reaction networks taking place on microscopic surfaces. However, it is applicable in a large variety of other situations such as reactions in liquid solutions and in living systems. The method can be further generalized to encompass additional reactions which are relevant empirically, such as dissociation reactions, as well as reactions involving multiple products and branching ratios.

In this paper we focused on systems that exhibit a single steady state. One should be more careful when attempting to model systems that exhibit multiple steady states or stochastic oscillations. When running Monte Carlo simulations, such features are easily seen when examining the time-dependent population sizes. However, in the master equation formalism and similarly in the reduced-multiplane method, such features are more difficult to analyze. Let us examine a system with two distinct steady states. Such a system displays two

distinct peaks in the probability distribution. Assuming the approximation performed by the reduced-multiplane method is still valid, and there is no reason to think that is not the case, the original question reduces to the following: can one observe bistability, or any other similar effects, using only the two-species marginal probability distributions instead of the full probability distribution. Generically, the answer appears to be positive. Since at least one species must have a different population size in each of the two quasistationary states of a bistable system, any marginal probability distribution involving this species will display a peak around each of the values. This can be used as an indication of the existence of bistability. Stochastic oscillations can be identified using similar considerations.

While most chemical and biological reactions in nature are two-species reactions, some models do consist of three-body or generally  $n$ -body reactions. The method proposed in this paper has been suited specifically for the common cases. Generalizing the reduced-multiplane method to networks including  $n$ -body reactions requires some adjustments to the original formulation. The notion of a reaction graph is no longer applicable in such cases. It should be replaced by a hypergraph in which each hyperedge includes all the species involved in the corresponding reaction. In this case, the fundamental two-vertex cliques are replaced by fundamental  $n$ -vertex cliques. The equations can then be written down using a similar procedure.

In addition to the reduced-multiplane method, a related method based on moment equations, provides efficient stochastic simulations of reaction networks [41]. In this approach, one constructs differential equations for the first and second moments of the probability distribution. These equations are brought into a closed form using a suitable truncation scheme. As a result, the number of equations is reduced to one equation for each reactive species and one equation for each reaction. In applications such as interstellar chemistry, in which the main objective is to calculate the reaction rates, the moment equations are advantageous. However, in many other applications in which one is interested in marginal probability distributions the multiplane method and the reduced-multiplane method are more suitable.

#### ACKNOWLEDGMENTS

Baruch Barzel is appreciative of support through the ‘‘Hoffman Leadership and Responsibility’’ program, at the Hebrew University. This work was supported by the US-Israel Binational Science Foundation.

- [1] A. R. Peacock, *An Introduction to the Physical Chemistry of Biological Organization* (Oxford University Press, New York, 1989).
- [2] C. W. Gardiner, *Handbook of Stochastic Methods* (Springer, Berlin, 2004).
- [3] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [4] M. E. J. Newman and G. T. Barkema, *Monte Carlo Methods in Statistical Physics* (Clarendon Press, Oxford, 1999).
- [5] A. G. G. M. Tielens, *The Physics and Chemistry of the Interstellar Medium* (Cambridge University Press, Cambridge, England, 2005).
- [6] T. I. Hasegawa, E. Herbst, and C. M. Leung, *Astrophys. J.* **82**, 167 (1992).

- [7] E. Herbst, *Annu. Rev. Phys. Chem.* **46**, 27 (1995).
- [8] D. Hollenbach and E. E. Salpeter, *J. Chem. Phys.* **53**, 79 (1970).
- [9] R. J. Gould and E. E. Salpeter, *Astrophys. J.* **138**, 393 (1963).
- [10] A. G. G. M. Tielens and W. Hagen, *Astron. Astrophys.* **114**, 245 (1982).
- [11] S. B. Charnley, A. G. G. M. Tielens, and S. D. Rodgers, *Astrophys. J.* **482**, L203 (1997).
- [12] P. Caselli, T. I. Hasegawa, and E. Herbst, *Astrophys. J.* **495**, 309 (1998).
- [13] O. M. Shalabiea, P. Caselli, and E. Herbst, *Astrophys. J.* **502**, 652 (1998).
- [14] T. Stantcheva, P. Caselli, and E. Herbst, *Astron. Astrophys.* **375**, 673 (2001).
- [15] N. J. B. Green, T. Toniazzo, M. J. Pilling, D. P. Ruffle, N. Bell, and T. W. Hartquist, *Astron. Astrophys.* **375**, 1111 (2001).
- [16] O. Biham, I. Furman, V. Pirronello, and G. Vidali, *Astrophys. J.* **553**, 595 (2001).
- [17] S. B. Charnley, *Astrophys. J.* **562**, L99 (2001).
- [18] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Chapman & Hall; CRC, London, 2006).
- [19] B. Ø. Palsson, *Systems Biology: Properties of Reconstructed Networks* (Cambridge University Press, Cambridge, England, 2006).
- [20] H. H. McAdams and A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).
- [21] T. B. Kepler and T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
- [22] J. Paulsson, *Nature (London)* **427**, 415 (2004).
- [23] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
- [24] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, *Nat. Genet.* **31**, 69 (2002).
- [25] M. Assaf and B. Meerson, *Phys. Rev. E* **74**, 041115 (2006).
- [26] D. A. McQuarrie, C. J. Jachimowski, and M. E. Russell, *J. Chem. Phys.* **40**, 2914 (1964).
- [27] O. Biham and A. Lipshtat, *Phys. Rev. E* **66**, 056103 (2002).
- [28] H. Risken, *The Fokker-Planck Equation* (Springer, Berlin, 1989).
- [29] T. Stantcheva, V. I. Shematovich, and E. Herbst, *Astron. Astrophys.* **391**, 1069 (2002).
- [30] T. Stantcheva and E. Herbst, *Mon. Not. R. Astron. Soc.* **340**, 983 (2003).
- [31] D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976).
- [32] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [33] Y. Cao, D. T. Gillespie, and L. R. Petzold, *J. Chem. Phys.* **124**, 044109 (2006).
- [34] A. Lipshtat and O. Biham, *Phys. Rev. Lett.* **93**, 170601 (2004).
- [35] B. Barzel, O. Biham, and R. Kupferman, *Phys. Rev. E* **76**, 026703 (2007).
- [36] A. Lederhendler and O. Biham, *Phys. Rev. E* **78**, 041105 (2008).
- [37] B. Barzel, O. Biham, and R. Kupferman, *Multiscale Model. Simul.* **6**, 963 (2007).
- [38] N. G. van Kampen, *Phys. Rep.* **124**, 69 (1985).
- [39] N. Katz, I. Furman, O. Biham, V. Pirronello, and G. Vidali, *Astrophys. J.* **522**, 305 (1999).
- [40] I. Lohmar, J. Krug, and O. Biham, *Astron. Astrophys.* **504**, L5 (2009).
- [41] B. Barzel and O. Biham, *J. Chem. Phys.* **127**, 144703 (2007).