

## Minimal moment equations for stochastic models of biochemical reaction networks with partially finite state space

Jakob Ruess

Citation: *The Journal of Chemical Physics* **143**, 244103 (2015); doi: 10.1063/1.4937937

View online: <http://dx.doi.org/10.1063/1.4937937>

View Table of Contents: <http://scitation.aip.org/content/aip/journal/jcp/143/24?ver=pdfcov>

Published by the [AIP Publishing](#)

---

### Articles you may be interested in

[Modeling delay in genetic networks: From delay birth-death processes to delay stochastic differential equations](#)

*J. Chem. Phys.* **140**, 204108 (2014); 10.1063/1.4878662

[An effective method for computing the noise in biochemical networks](#)

*J. Chem. Phys.* **138**, 084106 (2013); 10.1063/1.4792444

[Steady-state fluctuations of a genetic feedback loop: An exact solution](#)

*J. Chem. Phys.* **137**, 035104 (2012); 10.1063/1.4736721

[A partial-propensity formulation of the stochastic simulation algorithm for chemical reaction networks with delays](#)

*J. Chem. Phys.* **134**, 014106 (2011); 10.1063/1.3521496

[An equation-free probabilistic steady-state approximation: Dynamic application to the stochastic simulation of biochemical reaction networks](#)

*J. Chem. Phys.* **123**, 214106 (2005); 10.1063/1.2131050

---



**AIP** | APL Photonics

*APL Photonics* is pleased to announce  
**Benjamin Eggleton** as its Editor-in-Chief



# Minimal moment equations for stochastic models of biochemical reaction networks with partially finite state space

Jakob Ruess<sup>a)</sup>

*Institute of Science and Technology Austria, Am Campus 1, 3400 Klosterneuburg, Austria*

(Received 24 July 2015; accepted 27 November 2015; published online 22 December 2015)

Many stochastic models of biochemical reaction networks contain some chemical species for which the number of molecules that are present in the system can only be finite (for instance due to conservation laws), but also other species that can be present in arbitrarily large amounts. The prime example of such networks are models of gene expression, which typically contain a small and finite number of possible states for the promoter but an infinite number of possible states for the amount of mRNA and protein. One of the main approaches to analyze such models is through the use of equations for the time evolution of moments of the chemical species. Recently, a new approach based on conditional moments of the species with infinite state space given all the different possible states of the finite species has been proposed. It was argued that this approach allows one to capture more details about the full underlying probability distribution with a smaller number of equations. Here, I show that the result that less moments provide more information can only stem from an unnecessarily complicated description of the system in the classical formulation. The foundation of this argument will be the derivation of moment equations that describe the complete probability distribution over the finite state space but only low-order moments over the infinite state space. I will show that the number of equations that is needed is always less than what was previously claimed and always less than the number of conditional moment equations up to the same order. To support these arguments, a symbolic algorithm is provided that can be used to derive minimal systems of unconditional moment equations for models with partially finite state space. © 2015 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4937937>]

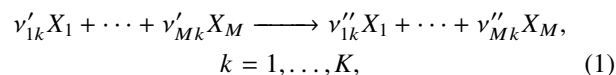
## INTRODUCTION

Stochastic kinetic models governed by the chemical master equation (CME) are widely used to describe biochemical reaction networks in which the inherent randomness of molecular interactions cannot be ignored.<sup>1</sup> However, computing the time evolution of the probability distribution for such models is typically challenging because the CME cannot be solved exactly,<sup>2,3</sup> except in some special cases.<sup>4–6</sup> An approach that has recently gained popularity is based on deriving systems of differential equations from the CME that only describe the time evolution of moments—up to some desired order  $L$ —of the underlying probability distribution.<sup>7–9</sup> The solution of such moment equations can often be computed or at least approximated in cases where the whole CME is intractable.<sup>10</sup> However, if more than just moments of very low order are desired (i.e.,  $L \gg 1$ ), the size of the system of moment equations increases quickly in the number of chemical species that play a role for the system. A common feature of reaction networks is that some of the reacting chemical species can only be present in finite amounts of molecules (e.g., gene copies, conserved enzymes). Recently, it has been argued that in such cases, the stochastic dynamics of the network can be described more accurately with fewer equations through the use of conditional moments.<sup>11,12</sup> Here, I show that the result that less

equations are needed is only true if the specific structure of the reaction network is not taken into account in the derivation of the classical unconditional moment equations. I provide a formula to determine how many classical moment equations are minimally needed to compute all moments up to any desired order  $L$ , and show that, for all  $L$ , this number is smaller than the number of conditional moment equations—in line with the mathematical intuition that more accurate descriptions of the whole underlying probability distribution cannot be obtained with less equations. Furthermore, I provide a simple MATLAB script that can be used to symbolically derive the minimal system of moment equations and show that these equations can be solved numerically much faster than previously used systems of moment equations in which the specific structure of models with partially finite state space is not taken into account.

## BINARY VARIABLE REPRESENTATION OF REACTION NETWORKS WITH PARTIALLY FINITE STATE SPACE

Consider a reaction network of  $M$  chemical species  $X_1, \dots, X_M$  that interact stochastically according to  $K$  reactions,



where the coefficients  $v'_{ik}$  and  $v''_{ik}$  determine how many molecules of the  $i$ th species are consumed and produced

<sup>a)</sup>Electronic mail: [jruess@ist.ac.at](mailto:jruess@ist.ac.at)

in the  $k$ th reaction, respectively. Under the assumption that the system is well-stirred and in thermal equilibrium, the probability distribution describing the time evolution of the number of molecules of the different species is governed by the chemical master equation.<sup>13</sup> Furthermore, ordinary differential equations (ODEs) that describe the time evolution of moments of this probability distribution, i.e.,  $\mathbb{E}[X_i(t)]$ ,  $i = 1, \dots, M$ ;  $\mathbb{E}[X_i(t)X_j(t)]$ ,  $i, j = 1, \dots, M$ ,  $i \leq j$ ;  $\dots$ , can be derived from the CME. For details of these results, the reader is referred to the provided references.<sup>7-9</sup> In this paper, I will consider a subclass of general reaction networks. Specifically, I will focus on systems in which the structure of the reaction network ensures that some of the chemical species can only be present in finite amounts of molecules. Without loss of generality let us assume that

$$X_1 \in \{0, \dots, m_1\}, \dots, X_j \in \{0, \dots, m_j\},$$

with  $m_1, \dots, m_j \in \mathbb{N}$  and  $1 \leq j \leq M$ . (2)

In fact, we can go one step further and assume that only one species has a finite state space. This is because an equivalent description of the system can be derived where the  $j$  species in (2) are replaced by one species that can take  $m = \prod_{i=1}^j (m_i + 1)$  different values (by using a mapping that uniquely relates each state in  $\{0, \dots, m_1\} \times \dots \times \{0, \dots, m_j\}$  to a state in  $\{0, \dots, m\}$ ). It should be noted that this equivalent system cannot be written in standard form (1) anymore and that specifying the correct transition rates on the finite state space is somewhat more involved in such a description. From the theoretical perspective, however, it is sufficient to know that such an equivalent system exists. Partially finite state spaces arise naturally in many applications, often due to conservation laws in the system. For instance, models of enzymatic reactions typically contain an influx of substrate to the system but a finite and conserved number of enzymes that catalyze the reaction.<sup>14</sup> This implies that the enzyme and any complex that involves the enzyme can only be present in finite numbers of molecules, whereas substrate and product species can theoretically be present in arbitrarily large amounts.

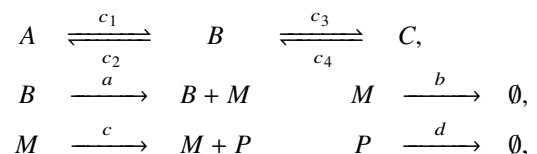
Another important example is provided by models of gene expression.<sup>15,16</sup> Such models typically consist of one or more species that describe the state of the promoter of the gene,<sup>17,18</sup> one species for mRNA, one for protein and possibly one each for some further quantities like protein dimers or fluorescent protein.<sup>19</sup> Thereby, the different possible states of the promoter may, for instance, correspond to bound or unbound transcription factors,<sup>20</sup> different states of the chromatin structure,<sup>21</sup> or the presence of DNA loops.<sup>22</sup> In any case, there is usually only one copy of the gene present, which leads to a finite state space for the promoter species, whereas the other chemical species can in principle be present in arbitrarily large amounts. At the level of the chemical master equation, it is very natural to represent the promoter with a single chemical species  $X_1 \in \{0, \dots, m\}$  that can switch between  $m + 1$  values where each of these values corresponds to one possible state of the promoter. The corresponding reaction network, however, cannot be directly written in the form given in (1), because the different values of  $X_1$  are only abstract representations of the promoter states and do not correspond to actual molecule counts. From the mathematical

perspective, this is not necessarily a problem. It may, however, become a problem when moment equations need to be derived for the system. Specifically, the derivation of moment equations<sup>23</sup> typically operates on a structure like the one in (1) and it is not immediately clear how it can be performed if such a structure is not available. This seems to be the reason why a different formulation of models of gene expression is more prevalent in the field.<sup>11,21</sup> In this formulation, the promoter is represented by  $m + 1$  binary variables  $X_1, \dots, X_{m+1} \in \{0, 1\}$  where each variable represents a possible state of the promoter. The understanding here is that if a variable takes the value of one at a certain time point, the promoter is in the state that is represented by this variable and all other variables must necessarily be zero. This implies that we can fully characterize the promoter with  $m$  chemical species that follow the additional rules given by

$$X_i(t) \in \{0, 1\}, i = 1, \dots, m$$

with  $X_i(t) \cdot X_j(t) = 0$  for  $i \neq j$ . (3)

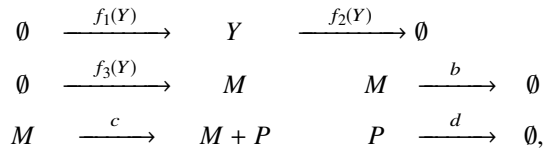
Note that the species  $X_{m+1}$  has been dropped in (3), because its value is always uniquely determined by the values of  $X_1, \dots, X_m$ . The advantage of this formulation is that the values of all these variables can be interpreted as amounts of molecules and that it is straightforward to formulate the reaction propensities as functions of the state of the system. While researchers have been intuitively using this formulation, its implications for the moment equations do not seem to be entirely clear in the field. More specifically, the promoter species  $X_i$ ,  $i = 1, \dots, m$  have so far only been treated in the same way as any other species and the specific properties in (3) were neglected.<sup>11</sup> This does not lead to incorrect results, but the derived moment equations are unnecessarily complicated. In the following, I will explain how smaller systems of moment equations can be obtained by using the structure in (3) and what the implications of these equations are for general systems with partially finite state space. The first thing to realize is that for any formulation of the model with  $m + 1$  binary variables, there exists an equivalent formulation with one species that can take  $m + 1$  values. Importantly, the reverse also holds true, which means that any species that can take only finitely many values can be represented by a set of binary variables with the structure in (3). To illustrate this equivalence, let us consider a simple model of gene expression in which the promoter can be in one of three possible states that are represented in the following with the binary variables  $A$ ,  $B$ , and  $C$ :



where at time zero it holds that  $A(0), B(0), C(0) \in \{0, 1\}$  and  $A(0) + B(0) + C(0) = 1$ . The conservation law  $A + B + C = 1$  implies that one of the binary variables can be dropped from the network, but this is not necessary for the discussion here and will only become important later.

This network is equivalent to the following network in which the promoter is represented by one species  $Y \in \{0, 1, 2\}$

in the understanding that  $Y = 0$  corresponds to  $A = 1$ ,  $Y = 1$  to  $B = 1$  and  $Y = 2$  to  $C = 1$ :



where  $f_1(Y) = c_1 \cdot \mathbb{1}_{\{Y=0\}} + c_3 \cdot \mathbb{1}_{\{Y=1\}}$ ,  $f_2(Y) = c_2 \cdot \mathbb{1}_{\{Y=1\}} + c_4 \cdot \mathbb{1}_{\{Y=2\}}$  and  $f_3(Y) = a \cdot \mathbb{1}_{\{Y=1\}}$ .

In this paper, we will explore what this equivalence implies for the moment equations and how comparisons between classical and conditional moment equations can be made in the light of these results.

## THE NUMBER OF CLASSICAL AND CONDITIONAL MOMENT EQUATIONS

Let us consider a reaction network with  $m$  binary variables  $X_1, \dots, X_m$  with the properties in (3) that describe, for instance,  $m + 1$  possible promoter states or more generally any reaction network with partially finite state space where the cardinality of the finite part of the state space is  $m + 1$ . Furthermore, assume that  $n = M - m$  species  $X_{m+1}, \dots, X_M$  with an infinite state space are present in the network. If we do not take into account the specific structure in (3) in the derivation of moment equations and treat the system just like a standard reaction network with  $M$  species, one can deduce that the number of moment equations  $N_{M,L}$  up to order  $L$  is given by

$$N_{M,L} = \sum_{i=1}^L \binom{M+i-1}{i}. \quad (4)$$

As already mentioned in the Introduction, it has been argued that a description with less equations for the same truncation order  $L$  can be obtained through the use of conditional moment equations. The idea of such approaches is to use equations for the probabilities of the binary variables, i.e., for

$$\begin{aligned} \mathbb{P}(X_i(t) = 1, X_1(t) = 0, \dots, X_{i-1}(t) = 0, X_{i+1}(t) \\ = 0, \dots, X_m(t) = 0), \quad i = 1, \dots, m, \end{aligned} \quad (5)$$

and equations for conditional moments, given each possible state of the binary variables, for the remaining species, i.e., for the conditional means

$$\mathbb{E}[X_j(t)|X_i(t) = 1, X_1(t) = 0, \dots, X_{i-1}(t) = 0, X_{i+1}(t) \\ = 0, \dots, X_m(t) = 0], \quad i = 1, \dots, m, \quad j = m + 1, \dots, M \quad (6)$$

and

$$\mathbb{E}[X_j(t)|X_1(t) = 0, \dots, X_m(t) = 0], \quad j = m + 1, \dots, M. \quad (7)$$

The conditional moments in Eq. (7) are needed because the conservation law cannot be used to express these conditional moments as functions of the conditional moments in Eq. (6). This is essentially a consequence of the fact that for a Bernoulli random variable  $B$  the conditional mean  $\mathbb{E}[X|B]$  of another random variable  $X$  does not uniquely determine the conditional mean  $\mathbb{E}[X|1-B]$ , unless  $X$  and  $B$  are independent.

For truncation orders  $L > 1$ , equivalent conditional higher order moments are required. Using this approach, the number

of equations  $N_{m,n,L}^c$  for a truncation order  $L$  is given by

$$N_{m,n,L}^c = m + (m + 1) \sum_{i=1}^L \binom{n+i-1}{i}, \quad (8)$$

where the first term stems from the  $m$  equations for the probabilities and the second term arises because for each of the  $m + 1$  possible promoter states a system of moment equations up to order  $L$  for the  $n$  species  $X_{m+1}, \dots, X_M$  is needed. A consequence of Eq. (8) is that when the order up to which the moments have to be computed is increased from  $L - 1$  to  $L$ , the number of equations increases by

$$N_{m,n,L}^{c,+} = (m + 1) \cdot \binom{n+L-1}{L}. \quad (9)$$

To compare this to the number of classical moment equations, we can deduce from Eq. (4) that the number of classical moment equations increases by

$$N_{M,L}^+ = \binom{M+L-1}{L} \quad (10)$$

when the order is increased from  $L - 1$  to  $L$ . Writing out the binomial coefficients in Eqs. (9) and (10), it is easy to see that for any  $n$  and  $m$  there exists a truncation order  $l$  such that  $N_{m,n,L}^{c,+} < N_{M,L}^+ \forall L > l$ . In other words, the number of equations in the conditional moments framework grows slower in the truncation order than the number of classical moment equations, and thus, there always exists a truncation order for which the system can be described with less equations by using conditional moments. This observation was originally stated in Ref. 11 where the authors considered a specific example with  $m = 1$  and  $n = 2$  and deduced that  $N_{m,n,L}^c < N_{M,L}$  for  $L > 3$  (see Table II in Ref. 11). Together with the fact that the classical moments can be recovered from the conditional moments, one reaches the mathematically quite unintuitive conclusion that more properties of the underlying probability distribution can be captured with a smaller number of equations.<sup>11,12</sup> In this paper, I will show that this result is only obtained because the  $N_{M,L}$  classical moment equations contain many equations that are redundant. Specifically, while Eq. (4) provides the number of moment equations that would be generated by a naive use of current toolboxes,<sup>23,24</sup> it cannot be taken as a measure of the number of classical moments that are actually necessary. In other words, the result that  $N_{m,n,L}^c < N_{M,L}$  is not very meaningful from a mathematical perspective because the conditional moments framework automatically incorporates the specific structure of the network given in Eq. (3) and any comparison to classical moment equations must take this into account.

## THE MINIMAL NUMBER OF CLASSICAL MOMENT EQUATIONS FOR REACTION NETWORKS WITH PARTIALLY FINITE STATE SPACE

Essentially, there is only one simple aspect that needs to be taken into account to deduce the minimally needed number of moment equations. One has to realize that the structure in Eq. (3) implies that all moments that contain products  $X_i(t) \cdot X_j(t)$  with  $i, j \leq m$  and  $i \neq j$  are zero and can be dropped, whereas all moments that contain powers  $X_i(t)^s$ ,  $i \leq m, s > 1$  can be replaced by lower order moments, because



$X_i(t) \in \{0, 1\}$  implies that  $X_i(t)^2 = X_i(t)$ . To understand why such simplifications must necessarily be possible, it is helpful to remember the discussion in the first section in this paper where we saw that an equivalent representation of the system can be obtained by replacing the  $m$  binary variables with one species  $Y \in \{0, 1, \dots, m\}$ . As discussed earlier, this equivalence can be understood in the sense that  $X_i(t) = 1$  corresponds to  $Y(t) = i - 1$  for  $i = 1, \dots, m$  and  $X_i(t) = 0 \forall i = 1, \dots, m$  to  $Y(t) = m$ . Consequently, the first order moments  $\mathbb{E}[X_i(t)]$ ,  $i = 1, \dots, m$  can be understood as the probabilities  $\mathbb{P}(Y(t) = i - 1)$ ,  $i = 1, \dots, m$ , which completely characterize the full probability distribution of  $Y$ . In this light, it is not surprising that higher order moments of  $X_i(t)$ ,  $i = 1, \dots, m$  cannot contain any additional information. These arguments only explain why no additional moments for the marginal distribution of the finite species are needed, but equivalent comparisons can also be invoked for the joint distribution of finite and infinite species to explain why all of the cross moments between finite and infinite species that contain terms of the form  $X_i(t) \cdot X_j(t)$  with  $i, j \leq m$  must necessarily be redundant. In essence, the use of the binary variable representation together with moment equations can be understood as an approach that leads to moments for the infinite species but automatically retains the complete probability distribution of the finite species.

While these considerations provide simple and significant reductions in the number of moment equation, they are currently not used in the literature; the most likely reason being that the currently available toolboxes have not been developed to incorporate the specific structure of reaction networks with partially finite state space. The provided supplementary material,<sup>25</sup> however, shows that it is straightforward to incorporate these simplifications in symbolic derivations of the moment equations. Using these simplifications, the number of minimally needed moment equations can be obtained as

$$N_{m,n,L}^{min,+} = \sum_{i=1}^L \binom{n+i-1}{i} + m + m \sum_{i=1}^{L-1} \binom{n+i-1}{i}. \quad (11)$$

This number consists of the number of moments that do not contain any binary variables (first term), the means of the binary variables that uniquely determine their full marginal distribution (second term), and the cross moments, such as  $\mathbb{E}[X_1(t)X_{m+1}(t)^{L-1}]$ , that contain exactly one power of a binary variable and no more than  $L - 1$  powers of other species (third term). An aspect to point out in Eq. (11) is that, contrary to Eq. (8), the last sum is multiplied by  $m$  and not by  $m + 1$ . This is a consequence of the conservation law, which can be used here to reduce the number of equations because for two random variables  $B$  and  $X$  it holds that  $\mathbb{E}[X(1 - B)] = \mathbb{E}[X] - \mathbb{E}[XB]$  and this implies that the  $(m + 1)$ th binary variable does not lead to any additional moments. From Eq. (11), it follows that when the order is increased from  $L - 1$  to  $L$ , the number of moments increases by

$$N_{m,n,L}^{min,+} = \binom{n+L-1}{L} + m \cdot \binom{n+L-2}{L-1}. \quad (12)$$

We are now at the point where we can state the result that provides the main backbone of the discussion in this paper.

*Corollary 1.* The number of minimally needed classical moment equations is never larger than the number of conditional moment equations and always smaller if  $n > 1$ .

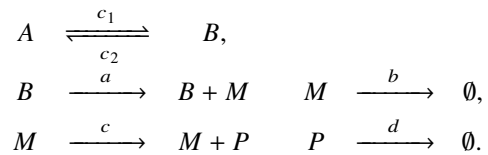
*Proof.* It suffices to show that for any  $n, m, L \in \mathbb{N}$  it holds that  $N_{m,n,L}^{min,+} \leq N_{m,n,L}^{c,+}$  with equality holding if and only if  $n = 1$ . To see this, note that

$$\begin{aligned} N_{m,n,L}^{min,+} &\leq N_{m,n,L}^{c,+} \\ &\Leftrightarrow \binom{n+L-1}{L} + m \cdot \binom{n+L-2}{L-1} \leq (m+1) \cdot \binom{n+L-1}{L} \\ &\Leftrightarrow m \cdot \binom{n+L-2}{L-1} \leq m \cdot \binom{n+L-1}{L} \\ &\Leftrightarrow L \leq n+L-1 \\ &\Leftrightarrow n \geq 1, \end{aligned}$$

which concludes the proof.

## COMPARISON OF THE NUMBER OF MOMENT EQUATIONS FOR BENCHMARK REACTION NETWORKS

*Example 1.* As a first benchmark example, let us consider a reaction network that consists of two possible states for the promoter, denoted here  $A$  and  $B$ , and of messenger RNA  $M$  and protein  $P$ . The purpose of this example is only to discuss the number of moment equations and not to determine their solution. Since this number does not depend on the reaction propensities, it is not of importance here which reactions take place. For instance, the system could be the standard random telegraph model,



Alternatively, the reactions could also be as in Ref. 11, which means that a direct comparison of the number of moment equations obtained here and in the reference is possible. Noting that for this example we have that  $n = 2$  and  $m = 1$ , we can directly compare the number of moment equations using (4), (8), and (11). The results are listed in Table I. The first two rows in this table are exactly the same as those listed in Table II in Ref. 11 (except that we obtain one conditional moment less for each closure order because of the conservation law). Based on the comparison of these two rows, it was concluded that less conditional moments are needed for large closure orders. We can now see, however,

TABLE I. Comparison of the number of moment equations for a reaction network with  $n = 2$  and  $m = 1$ . The first row shows the number of classical moments obtained when the structure of the system is not used, the second row provides the number of conditional moment equations, and the third row the number of minimally needed classical moments.

Truncation order $L$	1	2	3	4	5	6
Classical	3	9	19	34	55	83
Conditional	5	11	19	29	41	55
Minimal classical	3	8	15	24	35	48

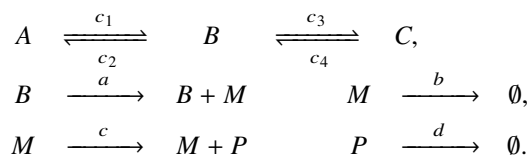
TABLE II. List of required moments up to order  $L = 3$  for Example 1 ( $n = 2$  and  $m = 1$ ). The moments listed in each row are only the moments of the corresponding order, i.e., for the moment equations up to order three the moments of all the rows together are needed.

Order $L$	Minimal classical	Conditional
1	$\mathbb{E}[A], \mathbb{E}[M], \mathbb{E}[P]$	$P(A = 1), \mathbb{E}[M A = 1], \mathbb{E}[P A = 1],$ $\mathbb{E}[M A = 0], \mathbb{E}[P A = 0]$
2	$\mathbb{E}[AM], \mathbb{E}[AP], \mathbb{E}[M^2],$ $\mathbb{E}[MP], \mathbb{E}[P^2]$	$\mathbb{E}[M^2 A = 1], \mathbb{E}[MP A = 1], \mathbb{E}[P^2 A = 1],$ $\mathbb{E}[M^2 A = 0], \mathbb{E}[MP A = 0], \mathbb{E}[P^2 A = 0]$
3	$\mathbb{E}[AM^2], \mathbb{E}[AMP], \mathbb{E}[AP^2],$ $\mathbb{E}[M^3], \mathbb{E}[M^2P], \mathbb{E}[MP^2],$ $\mathbb{E}[P^3]$	$\mathbb{E}[M^3 A = 1], \mathbb{E}[M^2P A = 1], \mathbb{E}[MP^2 A = 1],$ $\mathbb{E}[P^3 A = 1], \mathbb{E}[M^3 A = 0], \mathbb{E}[M^2P A = 0],$ $\mathbb{E}[MP^2 A = 0], \mathbb{E}[P^3 A = 0]$

that this result was only obtained because the structure of the system was not used in the derivation of the classical moment equations, and accordingly, many redundant moments were counted. The third row in Table I shows that the number of non-redundant classical moments is always less and grows slower than the number of conditional moments.

To make this comparison more intuitive for the reader, Table II provides a list of the conditional and minimal classical moments up to order  $L = 3$ . Note that  $A(t), B(t) \in \{0, 1\}$  and  $A(t) = 1 - B(t)$ , which leads to  $\mathbb{E}[A(t)] = P(A(t) = 1) = 1 - P(B(t) = 1) = 1 - \mathbb{E}[B(t)]$  and  $\mathbb{E}[M(t)|A(t) = 0] = \mathbb{E}[M(t)|B(t) = 1]$ , but this does not imply any relation between  $\mathbb{E}[M(t)|A(t) = 0]$  and  $\mathbb{E}[M(t)|A(t) = 1]$ .

*Example 2.* As a second example, consider a reaction network with a third binary variable,



Here, we have that  $m = 2$  and  $n = 2$  and obtain the number of moment equations listed in Table III. Comparing the first and the third row, it can be seen that at a truncation order of  $L = 6$  more than two third of the naively derived classical moment equations are actually redundant.

### NUMERICAL EVALUATION OF THE COMPUTATIONAL COST OF SOLVING THE CLASSICAL AND MINIMAL CLASSICAL MOMENT EQUATIONS

In this paper, it was established that the number of classical moment equations that are really needed for models with partially finite state space may be significantly less than the equations that are obtained in a naive derivation of the moment equations. This suggests that computational time can

TABLE III. Comparison of the number of moment equations for a network with  $n = 2$  and  $m = 2$ .

Truncation order $L$	1	2	3	4	5	6
Classical	4	14	34	69	125	209
Conditional	8	17	29	44	62	83
Minimal classical	4	11	21	34	50	69

be saved if the moments are computed using the results of this paper. To test whether this is really the case, and to evaluate quantitatively how large the computational savings are, I performed a study where the classical and minimal classical moment equations were solved numerically for the two examples considered in this paper with many different parameter values. More specifically, for both examples, all parameters were sampled 10 000 times (uniformly from the interval  $[0, 1)$ ) and the moment equations were solved for  $L = 2, 3, 4, 5$  up to a final time of  $t = 100$  using the CVODE solver of the SUNDIALS toolbox.<sup>26</sup> Thereby, I assumed that  $M(0) = P(0) = 0$  for both examples, for Example 1 that  $A(0) = 1$  and  $B(0) = 0$ , and for Example 2 that  $A(0) = 0$ ,  $B(0) = 0$ , and  $C(0) = 1$ . The implementation of this comparison is provided as supporting MATLAB files and the precise settings of the numerical solver can be found in the code. The results are listed in Tables IV (for Example 1) and V (for Example 2). It can be seen that the saved computational time increases quickly with the order of the moment equations. For higher orders, the solution of the minimal moment equations was on average up to 80%–90% faster. Specifically for applications where moments have to be computed iteratively, such as for parameter inference<sup>21</sup> or experiment design,<sup>19,27</sup> this is extremely important and may determine whether algorithms that are typically used in such applications require hours or days of running time.

TABLE IV. Comparison of the computational cost (in seconds) of numerically solving classical and minimal classical moment equations for Example 1.

Truncation order $L$	2	3	4	5
Classical	0.0095	0.0134	0.0209	0.1488
Minimal classical	0.0095	0.0114	0.0151	0.0211
Saved time (%)	0.16	14.76	27.94	85.85

TABLE V. Comparison of the computational cost (in seconds) of numerically solving classical and minimal classical moment equations for Example 2.

Truncation order $L$	2	3	4	5
Classical	0.0117	0.0211	0.2206	0.5886
Minimal classical	0.0116	0.0133	0.0175	0.1059
Saved time (%)	1.48	36.81	92.09	82.01

## SYMBOLIC DERIVATION OF THE MINIMAL MOMENT EQUATIONS

The MATLAB code for deriving minimal systems of moment equations and a description of the code are provided in the supplementary material.<sup>25</sup> I would like to specifically point out that this code is a working tool and not meant to be a full toolbox for solving moment equations. The tool only derives moment equations and is currently neither directly interfaced with moment closure methods nor with ODE solvers. In addition, the reaction network has to be specified in the binary variable representation. An automatic derivation of this description for any reaction network with partially finite state space is currently not implemented. In the future, I plan to work on extensions for these tasks.

## DISCUSSION

As stochastic kinetic models gain more and more importance for applications in molecular biology,<sup>17,19,21,28,29</sup> also the question of how to best work with these models is being debated increasingly often.<sup>30-32</sup> It is clear that there will never exist one approach that is universally best for any reaction network,<sup>8</sup> but it may very well be that for networks with specific structure, some approaches are better suited than others. In this paper, I focused on networks with partially finite state space, which typically arise due to conservation laws in the chemical reactions. The main focus of the examples considered here was on models of gene expression where DNA is a conserved quantity. Equally important applications can, however, also be found in (single) enzyme kinetics with small and conserved numbers of enzymes.<sup>33,34</sup> I showed how the minimal number of moment equations of any order can be determined for such models and provided a symbolic algorithm for deriving these equations. The results revealed that the implications of a partially finite state space had not been completely understood in earlier studies and that the number of classical moment equations that are really needed is smaller than what was previously claimed.<sup>11,12,35</sup>

The results of this paper can also be seen as the basis for a hybrid approach for the analysis of stochastic models of biochemical reaction networks. The binary variable representation of the network, and the derivation of moment equations from it, automatically keeps the full probability distribution over the finite part of the state space while resorting to low-order moments over the infinite part. Essentially the same is done when the conditional moment equations are used<sup>11</sup> and one may question how it is possible that the number of minimally needed classical moment equations that was derived in this paper turned out to be smaller than the number of conditional moment equations. The reason for this is that the conditional moment equations up to order  $L$  provide information about the full probability distribution that is “partially of order  $L + 1$ ” from the perspective of the minimal classical moment equations. For instance, in Example 1, the five conditional moments of order  $L = 1$  listed in Table II uniquely determine the three minimal classical moments of order  $L = 1$ , but they also determine  $\mathbb{E}[AM]$  and  $\mathbb{E}[AP]$ , which are of second

order. If we would interpret all cross moments between binary variables and infinite variables (i.e., moments of the type  $\mathbb{E}[X_1(t)X_{m+1}(t)^{L-1}]$ ) as belonging to order  $L - 1$  instead of  $L$  in the formulation of the minimal classical moment equations, then exactly the same number of minimal classical and conditional moment equations would be obtained. In fact, in this interpretation, minimal classical and conditional moment equations capture exactly the same properties of the full probability distribution and can be derived from each other. In this light, one could argue that the conditional moment equations do not provide a new way of representing stochastic models of biochemical reaction networks. They do, however, suggest new approximation techniques<sup>36,37</sup> and this may help future efforts in modeling stochastic reaction networks.

## ACKNOWLEDGMENTS

This work has received funding from the German Research Foundation (DFG) as part of the Transregional Collaborative Research Center “Automatic Verification and Analysis of Complex Systems” (SFB/TR 14 AVACS, <http://www.avacs.org/>), by the European Research Council (ERC) under Grant No. 267989 (QUAREM), by the Austrian Science Fund (FWF) under Grant Nos. S11402-N23 (RiSE) and Z211-N23 (Wittgenstein Award) and by the People Programme (Marie Curie Actions) of the European Union’s Seventh Framework Programme (FP7/2007-2013) under REA Grant Agreement No. 291734.

<sup>1</sup>J. Paulsson, *Phys. Life Rev.* **2**, 157 (2005).

<sup>2</sup>T. Jahnke, *Multiscale Model. Simul.* **9**, 1646 (2011).

<sup>3</sup>B. Munsky and M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).

<sup>4</sup>T. Jahnke and W. Huisinga, *J. Math. Biol.* **54**, 1 (2007).

<sup>5</sup>P. Bokes, J. King, A. Wood, and M. Loose, *J. Math. Biol.* **64**, 829 (2012).

<sup>6</sup>R. Grima, D. Schmidt, and T. Newman, *J. Chem. Phys.* **137**, 035104 (2012).

<sup>7</sup>S. Engblom, *Appl. Math. Comput.* **180**, 498 (2006).

<sup>8</sup>J. Goutsias and G. Jenkinson, *Phys. Rep.* **529**, 199 (2013).

<sup>9</sup>J. Hespanha, *IEEE Proc.: Control Theory Appl.* **153**, 520 (2006).

<sup>10</sup>A. Singh and J. Hespanha, *IEEE Trans. Autom. Control* **56**, 414 (2011).

<sup>11</sup>J. Hasenauer, V. Wolf, A. Kazerooni, and F. Theis, *J. Math. Biol.* **69**, 687 (2014).

<sup>12</sup>A. Andreychenko, L. Mikeev, and V. Wolf, *J. Coupled Sys. Multiscale Dyn.* **3**(2), 156 (2015).

<sup>13</sup>D. Gillespie, *Physica A* **188**, 404 (1992).

<sup>14</sup>A. Miliadis-Argeitis, S. Engblom, P. Bauer, and M. Khammash, *J. R. Soc. Interface* **12**(113), 20150831 (2015).

<sup>15</sup>H. El Samad, M. Khammash, L. Petzold, and D. Gillespie, *Int. J. Robust Nonlinear Control* **15**, 691 (2005).

<sup>16</sup>B. Munsky, B. Trinh, and M. Khammash, *Mol. Syst. Biol.* **5**, 318 (2009).

<sup>17</sup>G. Neuert, B. Munsky, R. Tan, L. Teytelman, M. Khammash, and A. van Oudenaarden, *Science* **339**, 584 (2013).

<sup>18</sup>G. Rieckh and G. Tkacik, *Biophys. J.* **106**, 1194 (2014).

<sup>19</sup>J. Ruess, F. Parise, A. Miliadis-Argeitis, M. Khammash, and J. Lygeros, *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8148 (2015).

<sup>20</sup>C. Zechner, M. Unger, S. Pelet, M. Peter, and H. Koepl, *Nat. Methods* **11**, 197 (2014).

<sup>21</sup>C. Zechner, J. Ruess, P. Krenn, S. Pelet, M. Peter, J. Lygeros, and H. Koepl, *Proc. Natl. Acad. Sci. U. S. A.* **109**, 8340 (2012).

<sup>22</sup>J. Vilar and S. Leibler, *J. Mol. Biol.* **331**, 981 (2003).

<sup>23</sup>J. Hespanha, available at <http://www.ece.ucsb.edu/~hespanha/software> (2007).

<sup>24</sup>D. Schnoerr, G. Sanguinetti, and R. Grima, *J. Chem. Phys.* **143**, 185101 (2015).

- <sup>25</sup>See supplementary material at <http://dx.doi.org/10.1063/1.4937937> for the symbolic algorithm for the derivation of minimal systems of moment equations.
- <sup>26</sup>A. Hindmarsh, P. Brown, K. Grant, S. Lee, R. Serban, D. Shumaker, and C. Woodward, *ACM Trans. Math. Software (TOMS)* **31**, 363 (2005).
- <sup>27</sup>J. Ruess, A. Miliadis-Argeitis, and J. Lygeros, *J. R. Soc., Interface* **10**, 20130588 (2013).
- <sup>28</sup>A. Singh, B. Razooky, R. Dar, and L. Weinberger, *Mol. Syst. Biol.* **8**, 607 (2012).
- <sup>29</sup>F. Bertaux, S. Stoma, D. Drasdo, and G. Batt, *PLoS Comput. Biol.* **10**, e1003893 (2014).
- <sup>30</sup>R. Grima, *J. Chem. Phys.* **136**, 154105 (2012).
- <sup>31</sup>J. Ruess and J. Lygeros, *ACM Trans. Model. Comput. Simul. (TOMACS)* **25**, 8 (2015).
- <sup>32</sup>B. Munsy, Z. Fox, and G. Neuert, *Methods* **85**, 12 (2015).
- <sup>33</sup>H. Qian and E. Elson, *Biophys. Chem.* **101**, 565 (2002).
- <sup>34</sup>S. Kou, B. Cherayil, W. Min, B. English, and X. Sunney Xie, *J. Phys. Chem. B* **109**, 19068 (2005).
- <sup>35</sup>A. Andreychenko, L. Bortolussi, R. Grima, P. Thomas, and V. Wolf, e-print [arXiv:1509.09104](https://arxiv.org/abs/1509.09104) [q-bio.QM] (2015).
- <sup>36</sup>M. Soltani, C. Vargas-Garcia, and A. Singh, *IEEE Trans. Biomed. Circuits Syst.* **9**, 518 (2015).
- <sup>37</sup>P. Thomas and R. Grima, *Phys. Rev. E* **92**, 012120 (2015).