

Fast stochastic simulation of biochemical reaction systems by alternative formulations of the chemical Langevin equation

Bence Mélykúti,^{1,a)} Kevin Burrage,^{2,b)} and Konstantinos C. Zygalakis³

¹*Department of Statistics, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, United Kingdom*

²*Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD, United Kingdom*

³*Oxford Centre for Collaborative Applied Mathematics, Mathematical Institute, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, United Kingdom*

(Received 13 November 2009; accepted 12 March 2010; published online 26 April 2010)

The Chemical Langevin Equation (CLE), which is a stochastic differential equation driven by a multidimensional Wiener process, acts as a bridge between the discrete stochastic simulation algorithm and the deterministic reaction rate equation when simulating (bio)chemical kinetics. The CLE model is valid in the regime where molecular populations are abundant enough to assume their concentrations change continuously, but stochastic fluctuations still play a major role. The contribution of this work is that we observe and explore that the CLE is not a single equation, but a parametric family of equations, all of which give the same finite-dimensional distribution of the variables. On the theoretical side, we prove that as many Wiener processes are sufficient to formulate the CLE as there are independent variables in the equation, which is just the rank of the stoichiometric matrix. On the practical side, we show that in the case where there are m_1 pairs of reversible reactions and m_2 irreversible reactions there is another, simple formulation of the CLE with only $m_1 + m_2$ Wiener processes, whereas the standard approach uses $2m_1 + m_2$. We demonstrate that there are considerable computational savings when using this latter formulation. Such transformations of the CLE do not cause a loss of accuracy and are therefore distinct from model reduction techniques. We illustrate our findings by considering alternative formulations of the CLE for a human ether a-go-go related gene ion channel model and the Goldbeter–Koshland switch.

© 2010 American Institute of Physics. [doi:10.1063/1.3380661]

I. INTRODUCTION

It is now well known that intrinsic noise can drive important biochemical processes.^{1,2} Intrinsic noise is associated with the uncertainty of knowing when a reaction occurs and what that reaction is. These effects are particularly accentuated when there are small numbers of molecules in the system. Two important scientific fields where intrinsic noise effects are significant are genetic regulation^{3–7} and ion channel dynamics.⁸ When there are small numbers of molecules in the system, the kinetics between the species is best described by discrete Markov processes. The *stochastic simulation algorithm* (SSA) describes the time evolution of a set of chemical species chemically reacting in a well-stirred environment.⁹ As a stochastic process, it has a probability mass function that is the solution of a discrete parabolic differential equation described by the so-called *chemical master equation* (CME).^{10,11}

The basic idea of the SSA is that, at each time point, a

waiting time to the next reaction must be sampled from an exponential distribution, then the occurring reaction is sampled based on how large the propensities of the different reaction channels are relative to one another. In the limit that the number of molecules of all species becomes large, the waiting time becomes, on average, very small and under the law of mass action the time evolution of the kinetics is described by a system of ordinary differential equations (ODEs). This system is just the *reaction rate equation* and describes, approximately, the time evolution of the mean of the evolving Markov process described by the SSA.

There is also a representation of the SSA with Poisson processes.^{12–14} On one hand, this facilitates an analytical examination of the asymptotics of the kinetic equations.¹³ On the other hand, it naturally leads to a way of accelerating the SSA by allowing all the reactions to fire within the one step.¹⁴ If the number of reactions in this step are sampled from a Poisson distribution the resulting method is called a Poisson τ -leap method,¹⁵ while if they are sampled from a binomial distribution the method is called a binomial τ -leap method.^{16,17}

What is less well known is that there is an intermediate regime between the discrete stochastic and continuous deterministic regimes. In this regime, the intrinsic noise is still important, but there are sufficient number of molecules to describe the evolving kinetics by continuous models. This regime is called the chemical Langevin equation (CLE)

^{a)}Also at the Life Sciences Interface Doctoral Training Centre, University of Oxford, Wolfson Building, Parks Road, Oxford, OX1 3QD, UK and at the Control Group, Department of Engineering Science, University of Oxford, Parks Road, Oxford, OX1 3PJ, UK. Electronic mail: melykuti@stats.ox.ac.uk.

^{b)}Also at the Oxford Centre for Integrative Systems Biology, Department of Biochemistry, University of Oxford, South Parks Road, Oxford, OX1 3QU, UK and at the Institute for Molecular Bioscience, University of Queensland, St Lucia, QLD 4072, Australia.

regime^{11,18} and the mathematical framework is that of systems of Itô stochastic differential equations (SDEs) driven by a multidimensional Wiener process.¹⁹ This system of SDEs is called the CLE.

Gillespie¹⁸ derived a particular form of the CLE via two simplifying assumptions; first, that the propensities of reaction channels are independent of each other in short time intervals, and second, that even in these short time intervals the expected number of firings of each reaction channel is sufficiently large. A number of authors have realized that the CLE (and SDEs in general) can be written in alternative, equivalent forms using arguments from stochastic analysis and multivariate statistics.^{20–22} These changes to the equation do not change the finite-dimensional distribution of its solution. As such, the alternative formulations do not arise from model reduction. These authors used this insight solely to reformulate the CLE with the minimum number of Wiener processes.

Our contribution in this article is a detailed exploration of this insight. We investigate the minimum number of Wiener processes in the CLE *in conjunction* with a state space reduction by removing linearly dependent variables through conservation laws exhibited by the chemical system. This will shed light on the structure of the CLE from a linear algebraic viewpoint. Second, we show how from the standard form CLE one Wiener process can be omitted for each pair of reversible reactions. If there are m_1 pairs of reversible reactions and m_2 irreversible reactions then only m_1+m_2 Wiener processes are needed, rather than $2m_1+m_2$. We demonstrate that this simplification, in addition to its appealing explanatory power, can significantly speed up numerical simulations. Adalsteinsson *et al.*²³ also use this formulation but they failed to point out that this formulation is different to the original one and did not compare the computational costs. Singer *et al.*²⁴ derived this new form for a certain application from the Euler–Maruyama discretization²⁵ of the CLE which is not a rigorous method.

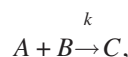
We will initially set out to construct the CLE in such a way that its mean and variance match the mean and the variance of the CME at any time instance t . It will become clear that with our current knowledge this is an insurmountable task in the general case because the moment equations are not closed. Still, we will arrive at the standard, well-known form of the CLE (Ref. 18) and will use that as the best approximation available. All our alternative formulations will be equivalent to this standard form. The drift (deterministic) component of the CLE is just the right-hand side of the ODE describing the deterministic regime (that is, of the reaction rate equation), while the diffusion (stochastic) component takes a very special form as noted by Van Kampen¹¹ by using the linear noise approximation theory. Gillespie¹⁸ captured this stochastic component by using one Wiener process per chemical reaction, whereas we will explore other possibilities.

Thus in Sec. II we give the necessary background on discrete and continuous stochastic models for biochemical reaction kinetics. In Sec. III we give the mathematical framework that allows us to manipulate the form of the CLE in terms of the number of Wiener processes in it, and also the

dimension of the state space. In Sec. IV we illustrate our approach on three instructive cases. These are a ring of reactions, a system of linear reversible kinetics describing different states of a human ether a-go-go related gene (HERG) ion channel,²⁶ and a nonlinear system known as the Goldbeter–Koshland switch.²⁷ In Sec. V we present numerical simulations of the corresponding CLEs for each of these three problems which we use for computational benchmarking and Sec. VI gives conclusions.

II. STOCHASTIC BIOCHEMICAL KINETICS

Let us suppose that there are N chemical species S_1, \dots, S_N , undergoing m chemical reactions in an environment that is well stirred and of constant temperature and fixed volume. Let $x_i(t)$ ($i \in \{1, \dots, N\}$) denote the number of molecules of species S_i at time t and let $x(t) = (x_1(t), \dots, x_N(t))^T$. Now any set of chemical reactions is uniquely characterized by two sets of quantities. The first are the *update (stoichiometric) vectors* ν_1, \dots, ν_m for each of the m reactions. These column vectors form the *stoichiometric matrix* $\nu = [\nu_1 \dots \nu_m] \in \mathbb{Z}^{N \times m}$. The second set of quantities are the *propensity functions* $a(x) = (a_1(x), \dots, a_m(x))^T$ that reflect the probabilities of each of the reactions to occur: if the chemical system is in state x , then the probability of a single firing of reaction channel j in an infinitesimal time interval of length h is $a_j(x)h$. Two very common assumptions are that in any reaction at most two molecules interact (the reactions are at most bimolecular, or second order) and that the kinetics follow the law of mass action. Under these assumptions, the propensity functions are polynomials of degree at most 2, and specifically are of the form $k_r, k_s x_i, k_u x_i x_k$ or $k_v x_i (x_i - 1)$ (with $k_r, k_s, k_u, k_v > 0$ constants). For example, given the reaction



then with $x(t) = (x_A(t), x_B(t), x_C(t))^T$, $\nu_1 = (-1, -1, 1)^T$, and $a_1(x(t)) = kx_A(t)x_B(t)$.

Given $x(t)$ at time t , the SSA draws a random waiting time τ to the next reaction from an exponential distribution with parameter $a_0(x(t)) = \sum_{j=1}^m a_j(x(t))$. Then it randomly selects which reaction to occur based on the relative sizes of $a_1(x(t)), \dots, a_m(x(t))$: reaction j is chosen with probability $a_j(x(t))/a_0(x(t))$. If reaction j is selected indeed, then the state vector is updated as

$$x(t + \tau) = x(t) + \nu_j,$$

and the algorithm repeats.

Given an initial condition $x(t_0) = x_0$, the probability mass function $p(x, t)$ ($t \geq t_0$) of the SSA is the solution of the CME given by

$$\frac{\partial p(x, t)}{\partial t} = \sum_{j=1}^m a_j(x - \nu_j) p(x - \nu_j, t) - \sum_{j=1}^m a_j(x) p(x, t). \quad (1)$$

In this representation, there is one equation for each configuration x of the state space and if $p(t)$ denotes the vector of probabilities for each of these configurations then Eq. (1) becomes

$$\dot{p}(t) = Ap(t), \quad p(0) = p_0,$$

with an appropriate matrix A . This is simply a linear ODE, so

$$p(t) = e^{At}p(0).$$

Although the size of this system becomes very large very quickly with the increase of N and of the upper bounds on each molecular population (very often such upper bounds are impossible to establish, in which case this system of ODEs is infinite), there are a number of techniques that allow the computation of the probability mass function directly in even quite complicated biochemical reaction systems. These include truncating the state space,²⁸ the use of Krylov subspace techniques^{29,30} and quasi-steady-state approximations.^{31,32}

Given the CME (1), it is possible to write down equations that describe the evolution of, for example, the mean and the covariance matrix. In particular, by multiplying Eq. (1) by x_i and summing, we can show

$$\frac{d}{dt}E(x(t)) = \sum_{j=1}^m \nu_j E(a_j(x(t))) = \nu E(a(x(t))). \quad (2)$$

Similarly, the time evolution of the second moment is given^{11,33} by

$$\begin{aligned} \frac{d}{dt}E(x(t)x(t)^T) &= E(h(x(t))x(t)^T) + E(x(t)h(x(t))^T) \\ &\quad + E(B(x(t))), \end{aligned} \quad (3)$$

where $h(x) = \nu a(x)$, and the *diffusion matrix* $B(x)$ is defined by

$$B(x) = \nu \operatorname{diag}(a(x))\nu^T.$$

The equation for the covariance matrix follows easily. Under the law of mass action kinetics, both Eqs. (2) and (3) are closed only if all reactions are at most first order (that is, all propensity functions a_j have degree at most one). If there is at least one bimolecular reaction, then the corresponding propensity function is a quadratic polynomial which will render the right-hand side of Eq. (2) dependent on the appropriate second moment. At the same time, Eq. (3) will become dependent on third moments through the $E(h(x(t))x(t)^T)$ and $E(x(t)h(x(t))^T)$ terms. Similarly, the equation for any higher moment will be dependent on even higher moments. In such a case, we say that the moment equations are not closed.

There are a few proposals on how to relax this problem through approximation.^{34,35} We sketch another simple method. With $\mu(t) = E(x(t))$, the simplest Taylor expansion of the propensity functions (truncation after the first-order term) leads to the approximate equality

$$\frac{d\mu(t)}{dt} = \sum_{j=1}^m \nu_j a_j(\mu(t)) = \nu a(\mu(t)). \quad (4)$$

In the deterministic regime, under the law of mass action, the ODE describing the evolution of the kinetics is just given by Eq. (4).

For the covariance matrix

$$\sigma(t) = E((x(t) - \mu(t))(x(t) - \mu(t))^T)$$

by the same-order truncation of the Taylor expansion of the propensity functions, we can show that

$$\frac{d\sigma(t)}{dt} = H(\mu(t))\sigma(t) + \sigma(t)H(\mu(t))^T + B(\mu(t)) \quad (5)$$

holds approximately, where $H(\mu(t))$ is the Jacobian matrix of $h(x)$ evaluated at $x = \mu(t)$. Importantly, Eqs. (4) and (5) are closed, they can be solved numerically. These two approximations are in fact exact when the law of mass action kinetics is assumed and all reactions are at most first order. (That is, there is at most one molecule on the reactant side of any reaction channel.) More accurate expressions for μ and σ can be developed by taking more terms in the Taylor series expansion of $f(x)$ and $B(x)$.²²

III. THEORETICAL RESULTS

A. The CLE as a parametric family of SDEs

In the Langevin regime, one hopes to be able to construct an SDE such that the evolution equations for its mean and the second moments are exactly Eqs. (2) and (3) in order to get the same mean and covariance as in the CME at any time instance t . The approach of matching the first two moments is implicit in Gillespie's derivation,¹⁸ and is more explicitly expressed by, for instance, Wilkinson²⁰ but has not been explored in depth. Formally, we are seeking $f: \mathbb{R}^N \rightarrow \mathbb{R}^N$ and $g: \mathbb{R}^N \rightarrow \mathbb{R}^{N \times d}$ such that the solution to

$$dx(t) = f(x(t))dt + g(x(t))dW(t), \quad (6)$$

with a d -dimensional standard Wiener process W , has its first two moments given in Eqs. (2) and (3). Recall that $W(t) = (W_1(t), \dots, W_d(t))^T$ is a vector of d independent one-dimensional Wiener processes. A Wiener process $W(t)$ starting at zero has the properties

$$E(W(t)) = 0, \quad \text{for all } t \geq 0,$$

$$\operatorname{Var}(W(t) - W(s)) = |t - s|, \quad \text{for all } t, s \geq 0,$$

and its increments on nonoverlapping intervals are independent normal random variables.

We can derive ODEs that describe the evolution of the first two moments of Eq. (6). By taking the expectation on both sides of Eq. (6), it is easily seen that

$$\frac{d}{dt}E(x(t)) = E(f(x(t))). \quad (7)$$

To calculate the second moment for $x(t)$ it is enough to calculate $E(x_i(t)x_k(t))$ for all $i, k \in \{1, \dots, N\}$, $i \leq k$.

Proposition III.1. For Eq. (6),

$$\begin{aligned} \frac{d}{dt}E(x_i(t)x_k(t)) &= E(f_i(x(t))x_k(t)) + E(x_i(t)f_k(x(t))) \\ &\quad + \sum_{j=1}^d E(g_{ij}(x(t))g_{kj}(x(t))). \end{aligned} \quad (8)$$

Proof. The proof of this proposition is found in the Appendix. \square

Comparing Eq. (2) with Eq. (7) suggests we should choose

$$f(x) = \nu a(x). \quad (9)$$

Once this is fixed, a comparison between Eqs. (3) and (8) suggests that

$$\sum_{j=1}^d g_{ij}(x)g_{kj}(x) = B_{ik}(x)$$

for all i and k is the most natural choice for g , that is,

$$g(x)g(x)^T = \nu \operatorname{diag}(a(x))\nu^T. \quad (10)$$

It is important to point out here that such a choice of f and g can only guarantee that the first two moments for the CME and the CLE are the same when all reactions are at most first order (under the law of mass action). In the case when there is at least one bimolecular reaction (when the moment equations are not closed), the mere formal matching of the two pairs of equations is not sufficient to match the first and second moments because Eqs. (3) and (8) are dependent on higher moments which we do not attempt to match.

Gillespie's argument¹⁸ gave the CLE in the following form:

$$dx(t) = \nu a(x(t))dt + \sum_{j=1}^m \nu_j \sqrt{a_j(x(t))} dW_j(t). \quad (11)$$

This is the CLE as the scientific community knows and uses it. Clearly, it satisfies both Eqs. (9) and (10) with $d=m$.

Corollary III.2. *Under the law of mass action kinetics, if all reactions are at most first order, then the CLE (11) gives the same first and second moments (means and covariances) as the CME.*

This result has already appeared in Ref. 36 (Theorem 4.1) but we think our notations are more succinct and therefore this derivation is more transparent. As we have just noted, in the general case, this does not imply that the first two moments of the CLE are the same as those of the CME. Instead of further studying the relationship of the CME and the CLE, we accept Eq. (11) as the standard reference SDE model for biochemical reaction systems and derive alternative formulations thereof based on the following insight.

Proposition III.3. *Different solutions g to the factorization problem (10) all give CLEs that have the same finite-dimensional distributions (in different terminology, which coincide in law).*

Proof. This can be easily derived by applying Theorem 8.4.3 of Ref. 19, but there is a direct proof. We will, however, assume previous knowledge of a standard tool, the Kolmogorov forward equation (also known as the Fokker-Planck equation).¹⁹ For the solution g of Eq. (10), the probability distribution function $p_t(x_0, x)$ of a transition from x_0 to x in a time interval of length t evolves according to the partial differential equation

$$\begin{aligned} \frac{\partial p_t(x_0, x)}{\partial t} = & - \sum_{i=1}^N \frac{\partial(p_t(x_0, x)(\nu a(x))_i)}{\partial x_i} \\ & + \frac{1}{2} \sum_{i,k=1}^N \frac{\partial^2(p_t(x_0, x)(g(x)g(x)^T)_{ik})}{\partial x_i \partial x_k}. \end{aligned}$$

$p_t(x_0, x)$ evolves identically for all solutions g to Eq. (10) because $p_0(x_0, x) = \delta_{x_0}(x)$ (the Dirac delta function at x_0) does not depend on g , and the parameters in the Kolmogorov forward equation $\nu a(x)$ and $g(x)g(x)^T = B(x)$ are identical for any g . \square

This proposition implies that the SDE

$$dx(t) = \nu a(x(t))dt + \sum_{j=1}^d g_j(x(t))dW_j(t) \quad (12)$$

defined with different g , which each satisfy Eq. (10) are equivalent in the sense that the distributions of their solutions are the same at any time instance t . It follows that *all* their moments will be identical.

The main goal of this paper is to explore the different possibilities of how the SDE (12) can be parameterized with different g such that the multidimensional Itô diffusion processes given by these different parameterizations all preserve the distribution of the standard CLE (11). Note that although the number of rows of g is fixed from the beginning to be the number of chemical species N , the number of columns, d , is not *a priori* specified. Our first line of inquiry will examine the minimum d for which the factorization (10) is possible—this is also the minimum number of Wiener processes necessary to describe the Itô diffusion process. Then, using the mathematical framework we have developed, we will be able to construct and prove the validity of a reduced formulation of the original CLE where there is only one Wiener process associated with each pair of reversible reactions. This formulation may be considered as a more natural model of chemical reaction systems than the original model (11). We will also demonstrate that this reduced formulation can speed up numerical simulations considerably without compromising accuracy.

B. Gillespie's original formulation

Construction 1. *Assuming that the number of firings of different reaction channels are independent in short time intervals and that the expected number of firings is large enough for each reaction channel, in his seminal paper, Gillespie¹⁸ derived that g is of the form*

$$g(x) = \nu \operatorname{diag}(\sqrt{a_1(x)}, \dots, \sqrt{a_m(x)}).$$

As we said, this will be our reference model and is also a special case of Eq. (10) with $d=m$. Here every independent Wiener process corresponds to one reaction channel. Hence the physical interpretation of this model is quite clear. Every variable is forced by as many Wiener processes as there are reaction channels which change its count.

Gillespie mentioned¹⁸ that this is not the only possible formulation and that other formulations with differing numbers of Wiener processes are possible. He referred to his

former work,³⁷ where equations were laid down which if satisfied by both a g_1 and a g_2 then the two Langevin equations with either g_1 or g_2 would have increments with identical distributions. This is analogous to our Proposition III.3.

C. The minimal formulation

In what follows, we will often refer to $\dim(\text{Ker } \nu)$, the dimension of the (right) nullspace of ν , and to $\dim(\text{Ker } \nu^T)$, the dimension of the left nullspace of ν . Left nullvectors correspond to *conservation laws* in the reaction system, that is, to preserved linear combinations of different species' counts. Each right nullvector corresponds to a sequence of firings of reaction channels such that if starting in state x all reactions occur the number of times given by the right nullvector, then the chemical system will eventually return to the original molecular counts x .

In this communication, we are exploring different natural choices for the formulation of the CLE (12) and hence for the choice of d . A natural question to ask is what is the minimum number of Wiener processes in the CLE, or equivalently, what is the minimum d for which the factorization of

$$B(x) = \nu \text{diag}(a(x)) \nu^T$$

in Eq. (10) is possible?

As $B(x)$ is a symmetric, positive semidefinite square matrix for all x , it can be diagonalized by a change of basis with an orthonormal matrix $U(x)$ whose columns are eigenvectors of $B(x)$,

$$B(x) = U(x)D(x)U(x)^T.$$

We partition the eigenvectors based on whether they belong to zero eigenvalue ($U_0(x)$) or some positive eigenvalue ($U_1(x)$) and arrange them such that $U(x)=[U_1(x) U_0(x)]$. Then there are $N-\dim(\text{Ker } B(x))$ nonzero eigenvalues, so $D(x)$ is of the form

$$D(x) = \begin{bmatrix} D_1(x) & 0 \\ 0 & 0 \end{bmatrix}$$

with a diagonal

$$D_1(x) \in \mathbb{R}^{(N-\dim(\text{Ker } B(x))) \times (N-\dim(\text{Ker } B(x)))}.$$

The construction for $g(x)$ is then

$$g(x) = U(x)D(x)^{1/2} = [U_1(x)D_1(x)^{1/2} \quad 0],$$

or simply

$$g(x) = U_1(x)D_1(x)^{1/2} \in \mathbb{R}^{N \times (N-\dim(\text{Ker } B(x)))}.$$

Indeed, $g(x)g(x)^T = U(x)D(x)^{1/2}D(x)^{1/2}U(x)^T = B(x)$. This formulation shows that $d=N-\dim(\text{Ker } B(x))$ independent Wiener processes are enough to define Eq. (12).

This factorization is minimal indeed since the rank of $g(x)$ cannot be less than the rank of $B(x)=g(x)g(x)^T$, that is, $N-\dim(\text{Ker } B(x))$.

The next proposition shows that the number of columns of $g(x)$ is independent of the state x . In order to avoid digression the proofs of the following two propositions are found in the Appendix.

Proposition III.4. *For every strictly positive x (it is enough that for all x and each reaction channel j , $a_j(x) > 0$ holds), $\dim(\text{Ker } B(x))$ is equal to the number of linearly independent conservation laws of the reaction network, $\dim(\text{Ker } \nu^T)$. In fact, a vector $y \in \mathbb{R}^N \setminus \{0\}$ is a (right) nullvector of $B(x)$ if and only if it is a left nullvector of the stoichiometric matrix ν .*

The following proposition states that this construction reduces the number of Wiener processes compared to the m Wiener processes of Gillespie's construction by the dimension of the right nullspace of ν .

Proposition III.5. $N-\dim(\text{Ker } \nu^T) = m - \dim(\text{Ker } \nu)$.

We summarize the results of this section.

Construction 2. *The previously described*

$$g(x) = U_1(x)D_1(x)^{1/2}$$

gives a CLE (12) with $N-\dim(\text{Ker } \nu^T) = m - \dim(\text{Ker } \nu)$ independent Wiener processes. Any CLE requires at least this many independent Wiener processes.

Note that Ref. 18 (Appendix B) and Ref. 20 (p. 189) are both inaccurate when claiming that generally the number of Wiener processes d must be no less than N . We will return to the problem of state space reduction where we prove that there is an equivalent formulation of the CLE with $N-\dim(\text{Ker } \nu^T)$ states, and as we see here, $N-\dim(\text{Ker } \nu^T)$ Wiener processes (Sec. III E).

The minimum number of Wiener processes needed is interesting for efficient numerical simulation.²⁵ Notice that the solution in Construction 2 is not satisfactory since U_1 is dependent on x . Hence, in a numerical simulation scheme at each time step a new diagonalization of $B(x)$ is required, which is computationally expensive.

As a first improvement, we propose another approach that results in a g of the same size, but potentially decreases the requirement for repeated computation at the cost of increased initial, one-off computation. A substantially different construction will be presented in Construction 4. Let $W = [W_1 W_0] \in \mathbb{R}^{N \times N}$ be an orthogonal matrix such that the columns of $W_0 \in \mathbb{R}^{N \times \dim(\text{Ker } \nu^T)}$ form an orthonormal basis in the left nullspace of ν , $\text{Ker } \nu^T$, and the columns of $W_1 \in \mathbb{R}^{N \times (N-\dim(\text{Ker } \nu^T))}$ are an orthonormal basis in the orthogonal complement, the image space of ν , $\text{Im } \nu$. Let us define the square root $\bar{M} = \sqrt{M}$ of a square matrix $M \in \mathbb{R}^{k \times k}$ as any square matrix $\bar{M} \in \mathbb{R}^{k \times k}$ such that $\bar{M}\bar{M}^T = M$, if such an \bar{M} exists.

Construction 3. *For notational brevity let $A(x) = \text{diag}(a(x))$. Then*

$$g(x) = W_1 \sqrt{W_1^T \nu A(x) \nu^T W_1}$$

gives a CLE (12) with $N-\dim(\text{Ker } \nu^T) = m - \dim(\text{Ker } \nu)$ independent Wiener processes.

Proof. We verify that $\hat{g}(x) = W \sqrt{W^T \nu A(x) \nu^T W}$ is an equally valid diffusion term [it satisfies Eq. (10)] and that the stated g is equivalent to \hat{g} . Note that $W^T \nu A(x) \nu^T W$ and $W_1^T \nu A(x) \nu^T W_1$ are symmetric positive semidefinite matrices, therefore their square root can be evaluated as for $B(x)$ earlier. Thus

$$\hat{g}(x)\hat{g}(x)^T = WW^T\nu A(x)\nu^T WW^T = \nu A(x)\nu^T$$

since W is orthogonal, so Eq. (10) is satisfied. Also,

$$\begin{aligned} W^T\nu A(x)\nu^T W &= \begin{bmatrix} W_1^T\nu \\ 0 \end{bmatrix} A(x) \begin{bmatrix} \nu^T W_1 & 0 \end{bmatrix} \\ &= \begin{bmatrix} W_1^T\nu A(x)\nu^T W_1 & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

shows that it is enough to use the top left block with $W_1^T\nu A(x)\nu^T W_1$. This is because when constructing $g(x)$ from this, on the left the columns of W_0 would be multiplied by zeros, and when constructing the CLE (12) on the right the last $\dim(\text{Ker } \nu^T)$ Wiener processes would be multiplied by zeros. Hence we can omit those. \square

This is an improvement over Construction 2 in that here the square root of a state-dependent $(N - \dim(\text{Ker } \nu^T)) \times (N - \dim(\text{Ker } \nu^T))$ matrix is used instead of an $N \times N$ matrix.

D. A general, state-independent reduction technique

In the previous subsection, a practical constraint for numerical simulations was discussed. Constructions that require in each time step an eigendecomposition of a state-dependent matrix are computationally too costly. In the following, we develop a construction in which to compute $g(x)$ only matrix products and taking the square root of a state-dependent diagonal matrix are required. This construction will give a CLE that generally may need more than $N - \dim(\text{Ker } \nu^T) = m - \dim(\text{Ker } \nu)$ independent Wiener processes, but certainly not more than m .

For a positive integer k , let I_k denote the $k \times k$ identity matrix. We say two nonzero vectors $y_1, y_2 \in \mathbb{R}^N \setminus \{0\}$ represent the same *direction*, if there is a $\lambda \in \mathbb{R} \setminus \{0\}$ such that $y_1 = \lambda y_2$.

Construction 4. Let s be the number of different directions given by the columns of ν . There exist matrices $J \in \mathbb{R}^{m \times s}$ and $V \in \mathbb{R}^{s \times m}$ such that $VA(x)V^T \in \mathbb{R}^{s \times s}$ is diagonal with only nonnegative entries and

$$g(x) = \nu J \sqrt{VA(x)V^T} \quad (13)$$

gives a CLE (12) with s independent Wiener processes, $m - \dim(\text{Ker } \nu) \leq s \leq m$.

Proof. Permute the columns of $\nu \in \mathbb{R}^{N \times m}$ such that $\nu = [\nu_1 \nu_2]$, where $\nu_1 \in \mathbb{R}^{N \times s}$ has one representative column vector for each direction given by the columns of ν . Then the columns that are left (ν_2) are each a constant multiple of one column in ν_1 . We permute the entries of $A(x)$ accordingly.

Let

$$\nu_2 = [\nu_1 \nu^{(1)} \dots \nu_1 \nu^{(m-s)}],$$

where for all i , $\nu^{(i)} \in \mathbb{R}^s$ has one nonzero entry.

Introducing $M = [\nu^{(1)} \dots \nu^{(m-s)}] \in \mathbb{R}^{s \times (m-s)}$, the definitions are

$$J = \begin{bmatrix} I_s \\ 0 \end{bmatrix} \in \mathbb{R}^{m \times s},$$

$$V = [I_s \ M] \in \mathbb{R}^{s \times m}.$$

First, partitioning $A(x)$ according to the sizes of blocks of V ,

$$\begin{aligned} VA(x)V^T &= [I_s \ M] \begin{bmatrix} A_1(x) & 0 \\ 0 & A_2(x) \end{bmatrix} \begin{bmatrix} I_s \\ M^T \end{bmatrix} \\ &= A_1(x) + MA_2(x)M^T \\ &= A_1(x) + \sum_{j=1}^{m-s} (A_2(x))_{jj} \nu^{(j)} \nu^{(j)T}, \end{aligned}$$

where the last step follows from

$$(MA_2(x)M^T)_{ik} = \sum_{j=1}^{m-s} (\nu^{(j)})_i (A_2(x))_{jj} (\nu^{(j)})_k^T.$$

Since $\nu^{(j)}$ has only one nonzero entry for all j , $\sum_{j=1}^{m-s} (A_2(x))_{jj} \nu^{(j)} \nu^{(j)T}$ is diagonal with only nonnegative entries, and consequently $VA(x)V^T$ is too.

Second,

$$\nu J V = [\nu_1 \ \nu_2] \begin{bmatrix} I_s & M \\ 0 & 0 \end{bmatrix} = [\nu_1 \ \nu_1 M] = [\nu_1 \ \nu_2] = \nu.$$

Hence $\sqrt{VA(x)V^T}$ exists trivially, and

$$\begin{aligned} g(x)g(x)^T &= \nu J \sqrt{VA(x)V^T} (\nu J \sqrt{VA(x)V^T})^T \\ &= \nu J V A(x) V^T J^T \nu^T \\ &= \nu A(x) \nu^T \end{aligned}$$

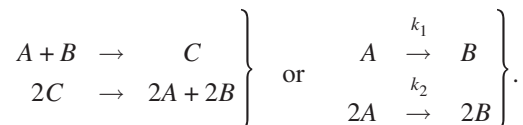
so Eq. (10) is satisfied. The actual form of g is

$$\begin{aligned} g(x) &= [\nu_1 \ \nu_2] \begin{bmatrix} \sqrt{A_1(x) + MA_2(x)M^T} \\ 0 \end{bmatrix} \\ &= \nu_1 \sqrt{A_1(x) + MA_2(x)M^T} \\ &= \nu_1 \sqrt{A_1(x) + \sum_{j=1}^{m-s} (A_2(x))_{jj} \nu^{(j)} \nu^{(j)T}}. \end{aligned}$$

\square

Corollary III.6. There is a formulation of the CLE (12) that is constructed from Gillespie's original CLE by omitting one independent Wiener process for each pair of reversible reactions and assigning to the retained Wiener process either respective stoichiometric vector multiplied by the square root of the sum of the two propensities. This is computationally inexpensive to numerically simulate. If m_1 is the number of pairs of reversible reactions, then in Gillespie's formulation there would be $2m_1$ Wiener processes for the reversible reactions, while in this formulation there would only be m_1 .

In fact, the result is slightly more general than this. Consider chemical systems with reactions



In both cases, one independent Wiener process can be spared. Note that the reactions in these examples are at most bimolecular.

E. State space reduction

Another form of model reduction we have not discussed yet is the reduction of the number of variables. The conservation laws describe linear dependencies between the counts of molecular species. This can be used to express certain variables as functions of others. With $\dim(\text{Ker } \nu^T)$ linearly independent conservation laws it is possible to reduce the number of variables from N to $N - \dim(\text{Ker } \nu^T)$ without loss of accuracy.

To this end, we specify an invertible matrix $T \in \mathbb{R}^{N \times N}$ such that $T\nu$ will take over the role of ν . (For esthetic reasons one may prefer $T \in \mathbb{Z}^{N \times N}$.) T is just a change of basis of the state space. To see this, multiply the CLE (12) with T from the left to get an equation in a new variable $z = Tx$:

$$d(Tx) = T\nu a(T^{-1}Tx)dt + Tg(T^{-1}Tx)dW(t),$$

or, by letting \circ denote the composition of functions, and \cdot multiplication (a special composition), we have

$$dz = (T \cdot \nu \cdot a \circ T^{-1})(z)dt + (T \cdot g \circ T^{-1})(z)dW(t).$$

We define T such that the last $\dim(\text{Ker } \nu^T)$ coordinates of the new state variable z are the conservation laws, which do not change at all.

We give T for Construction 1 first. Order the columns of $\nu \in \mathbb{R}^{N \times m}$ such that $\nu = [\nu_b \ \nu_c]$, where the columns of $\nu_b \in \mathbb{R}^{N \times (m - \dim(\text{Ker } \nu))}$ form a basis for $\text{Im } \nu$, and $\nu_c \in \mathbb{R}^{N \times \dim(\text{Ker } \nu)}$ is the collection of the rest of the column vectors. These are linearly dependent on columns of ν_b . Then, similarly to Construction 4, there are vectors

$$w^{(1)}, \dots, w^{(\dim(\text{Ker } \nu))} \in \mathbb{R}^{m - \dim(\text{Ker } \nu)},$$

and a matrix

$$R = [w^{(1)} \ \dots \ w^{(\dim(\text{Ker } \nu))}] \in \mathbb{R}^{(m - \dim(\text{Ker } \nu)) \times \dim(\text{Ker } \nu)}$$

such that $\nu_c = \nu_b R$.

Define $\nu_b^\perp \in \mathbb{R}^{N \times \dim(\text{Ker } \nu^T)}$ such that its columns form a basis of the orthogonal complement space of $\text{Im } \nu$, and let

$$T = \begin{bmatrix} (\nu_b^T \nu_b)^{-1} \nu_b^T \\ (\nu_b^\perp)^T \end{bmatrix}.$$

(To get an integer-valued T , we may put an appropriate diagonal matrix $D_0 \in \mathbb{Z}^{(N - \dim(\text{Ker } \nu^T)) \times (N - \dim(\text{Ker } \nu^T))}$ in front of $(\nu_b^T \nu_b)^{-1} \nu_b^T$, and choose $\nu_b^\perp \in \mathbb{Z}^{N \times \dim(\text{Ker } \nu^T)}$.)

Hence

$$\begin{aligned} T\nu &= \begin{bmatrix} (\nu_b^T \nu_b)^{-1} \nu_b^T \\ (\nu_b^\perp)^T \end{bmatrix} [\nu_b \ \nu_c] \\ &= \begin{bmatrix} I_{m - \dim(\text{Ker } \nu)} & (\nu_b^T \nu_b)^{-1} \nu_b^T \nu_c \\ 0 & 0 \end{bmatrix} \\ &= \begin{bmatrix} I_{m - \dim(\text{Ker } \nu)} & R \\ 0 & 0 \end{bmatrix}. \end{aligned}$$

Therefore, in no CLE formulation will the last $\dim(\text{Ker } \nu^T)$ variables be affected by the drift term $T\nu a(x)$. Since in Constructions 1 and 4 the first factor in $g(x)$ is ν , the last $\dim(\text{Ker } \nu^T)$ rows of the diffusion term $Tg(x)$ will vanish too.

Consequently, the last $\dim(\text{Ker } \nu^T)$ variables of z are constant and can be omitted from a numerical simulation.

The same argument holds for Construction 3, using W_1 and W_0 instead of ν_b and ν_b^\perp , respectively, in T . In the case of Construction 2, the state space reduction must precede the reduction of the number of Wiener processes. This method is very similar to Construction 3. For Construction 4, a finer partitioning of matrices ν, J, V is proposed. The detailed calculations are in the Appendix. These considerations prove the following result.

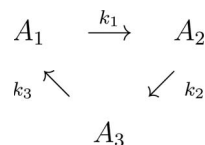
Theorem III.7. *For Constructions 1–4, a state space transformation is possible which reduces the number of variables from N to $N - \dim(\text{Ker } \nu^T) = m - \dim(\text{Ker } \nu)$ without changing the number of independent Wiener processes.*

IV. EXAMPLES

We demonstrate the reduction of the number of independent Wiener processes in the CLE by three examples. The first example is a simple one merely to illustrate our ideas, while the other two systems are models of real interest to biologists, namely, a Markov model for a HERG encoded K^+ channel,²⁶ and the Goldbeter–Koshland switch,²⁷ which plays a vital role in many cellular pathways.³⁸ In order to focus on the application of our main results, we will not carry out the fairly well-known state space reduction in any example.

A. A cyclical reaction system

Consider the following ring of $m=3$ reactions with $N=3$ species, $(A_1, A_2, A_3)^T$:



The indexing of reactions follows that of rate constants k_j . This specifies the order of columns in the stoichiometric matrix

$$\nu = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix},$$

which has rank 2. The propensity vector function is just

$$a(x) = (k_1 x_1, k_2 x_2, k_3 x_3)^T.$$

Gillespie's diffusion term (Construction 1) is

$$g^1(x) = \begin{bmatrix} -\sqrt{k_1 x_1} & 0 & \sqrt{k_3 x_3} \\ \sqrt{k_1 x_1} & -\sqrt{k_2 x_2} & 0 \\ 0 & \sqrt{k_2 x_2} & -\sqrt{k_3 x_3} \end{bmatrix}.$$

As there are no parallel stoichiometric vectors, Construction 4 cannot reduce the number of Wiener processes.

Constructions 2 and 3 can be computed analytically for such a small example. In Construction 2, finding the eigenvalues of the rank 2, 3×3 matrix requires the solution of a cubic equation (roots of the characteristic polynomial). But

we know that one eigenvalue is zero and this reduces the problem to a quadratic. This gives $D(x)$. Finding the eigenvectors is done by solving a linear equation for each nonzero eigenvalue, and then the vectors need to be normalized to create $U_1(x)$.

The calculations giving Construction 3 can be coded in step-by-step instructions. The orthogonal matrix W can be chosen as

$$W = [W_1 W_0] = \begin{bmatrix} -1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{6} & 1/\sqrt{3} \\ 0 & 2/\sqrt{6} & 1/\sqrt{3} \end{bmatrix}.$$

This is computed only once, therefore, its computational cost is almost irrelevant. Then

$$W_1^T \nu A(x) \nu^T W_1 = \begin{bmatrix} 2a_1(x) + \frac{1}{2}a_2(x) + \frac{1}{2}a_3(x) & -\frac{\sqrt{3}}{2}a_2(x) + \frac{\sqrt{3}}{2}a_3(x) \\ -\frac{\sqrt{3}}{2}a_2(x) + \frac{\sqrt{3}}{2}a_3(x) & \frac{3}{2}a_2(x) + \frac{3}{2}a_3(x) \end{bmatrix}.$$

To take the square root of this or, in general, of a matrix

$$\begin{bmatrix} M_{11} & M_{12} \\ M_{12} & M_{22} \end{bmatrix},$$

we can compute the two eigenvalues as the roots of the quadratic characteristic polynomial. These are

$$\lambda_{1,2} = \frac{M_{11} + M_{22} \pm \sqrt{(M_{11} - M_{22})^2 + 4M_{12}^2}}{2}.$$

The corresponding normalized eigenvectors are

$$v_1 = \frac{1}{\sqrt{(\lambda_1 - M_{22})^2 M_{12}^{-2} + 1}} \begin{pmatrix} (\lambda_1 - M_{22})M_{12}^{-1} \\ 1 \end{pmatrix},$$

$$v_2 = \frac{1}{\sqrt{(\lambda_2 - M_{22})^2 M_{12}^{-2} + 1}} \begin{pmatrix} (\lambda_2 - M_{22})M_{12}^{-1} \\ 1 \end{pmatrix}.$$

Thus

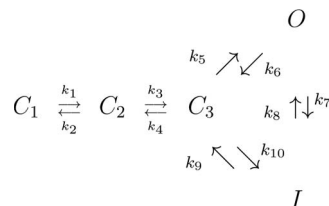
$$g^3(x) = W_1 [\sqrt{\lambda_1(x)} v_1(x) \quad \sqrt{\lambda_2(x)} v_2(x)]$$

is the product of a 3×2 and a 2×2 matrix, and the CLE requires two Wiener processes.

The construction that requires the least computation time hinges on how the cost of these computations compares to the cost of generating independent Wiener increments (that is, normal random variables).

B. A K^+ channel

We model the transformations of HERG encoded K^+ channels between three closed states (C_1, C_2, C_3), one open state (O) and one inactivation state (I) as $N=5$ chemical species (C_1, C_2, C_3, O, I)^T reacting through $m=10$ reactions



(For details, see Ref. 26 and references therein.) Thus the stoichiometric matrix is

$$\nu = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -1 & -1 & 1 \end{bmatrix},$$

and the propensity vector function is

$$a(x) = (k_1 x_1, k_2 x_2, k_3 x_2, k_4 x_3, k_5 x_3, k_6 x_4, k_7 x_4, k_8 x_5, k_9 x_5, k_{10} x_3)^T.$$

Gillespie's formulation (Construction 1) needs ten Wiener processes with

$$g^1(x) = \nu \sqrt{\text{diag}(a(x))}.$$

The rank of the stoichiometric matrix ν is 4, which allows for a CLE specification with four Wiener processes. Thus the minimal solutions g^2 and g^3 from Constructions 2 and 3, respectively, are of the form

$$g^2(x) = U_1(x) D_1(x)^{1/2},$$

$$g^3(x) = W_1 \sqrt{W_1^T \nu A(x) \nu^T W_1},$$

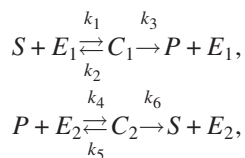
where $U_1(x)$, W_1 are 5×4 , $D_1(x)$ and $\sqrt{W_1^T \nu A(x) \nu^T W_1}$ are 4×4 matrices, respectively. With the exception of W_1 , we could only compute either of these matrices analytically if we solved a quartic equation. To avoid this laborious task, we can use standard numerical computations that we do not present here.

On the other hand, Construction 4 gives a simple closed form diffusion term. Indeed, this is a straightforward example where the number of Wiener processes can be decreased by half, to five, with

$$g^4(x) = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix} \text{diag} \left(\begin{array}{c} \sqrt{a_1(x) + a_2(x)} \\ \sqrt{a_3(x) + a_4(x)} \\ \sqrt{a_5(x) + a_6(x)} \\ \sqrt{a_7(x) + a_8(x)} \\ \sqrt{a_9(x) + a_{10}(x)} \end{array} \right).$$

C. The Goldbeter–Koshland switch

This example studied by Goldbeter and Koshland²⁷ is a system of covalent modifications facilitated by two converter enzymes, E_1 and E_2 . A typical example is a phosphorylation-dephosphorylation system. It consists of the following $m=6$ reactions:



with $N=6$ chemical species, $(S, E_1, C_1, P, E_2, C_2)^T$. The corresponding stoichiometric matrix is

$$\nu = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 1 \\ -1 & 1 & 1 & 0 & 0 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 1 & 1 \\ 0 & 0 & 0 & 1 & -1 & -1 \end{bmatrix},$$

while the propensity vector function is given by

$$a(x) = (k_1 x_1 x_2, k_2 x_3, k_3 x_3, k_4 x_4 x_5, k_5 x_6, k_6 x_6)^T.$$

Gillespie's formulation (Construction 1) with six Wiener processes is

$$g^1(x) = \nu \sqrt{\text{diag}(a(x))}.$$

However, the rank of the stoichiometric matrix ν is 3, which implies that only three Wiener processes are needed in the CLE. As with the K^+ channel, this can only be practically computed through numerical computation.

The closed form diffusion term from Construction 4 requires four Wiener processes. Removing the stoichiometric vectors corresponding to reactions 2 and 5, and introducing $u_1(x) = \sqrt{a_1(x) + a_2(x)}$, $u_2(x) = \sqrt{a_3(x)}$, $u_3(x) = \sqrt{a_4(x) + a_5(x)}$, and $u_4(x) = \sqrt{a_6(x)}$, we have

$$g^4(x) = \begin{bmatrix} -u_1(x) & 0 & 0 & u_4(x) \\ -u_1(x) & u_2(x) & 0 & 0 \\ u_1(x) & -u_2(x) & 0 & 0 \\ 0 & u_2(x) & -u_3(x) & 0 \\ 0 & 0 & -u_3(x) & u_4(x) \\ 0 & 0 & u_3(x) & -u_4(x) \end{bmatrix}.$$

These examples demonstrate cases in which the stoichiometric matrix is rank deficient and a reduction in the number of Wiener processes is possible. In Example 1 there were no parallel stoichiometric vectors, thus Construction 4 could not be deployed. In Examples 2 and 3 some Wiener processes could be spared for reversible reactions. These were also cases in which Constructions 2 and 3 could reduce the system size even further.

V. SIMULATIONS

In this section, we present computational benchmarking of numerical simulations of the examples described in Sec. IV. In addition to this, in order to demonstrate the theory we developed, we compare the numerically computed empirical means and variances from simulations that use different constructions for g in the CLE (12). As we have already stated in Proposition III.3, all these different CLEs have the same

TABLE I. Comparison of the empirical means and variances in the cyclical reaction system at time 5 for CLE Constructions 1 and 3.

	Construction 1	Construction 3
$E(x_1(5))$	98.41	98.52
$E(x_2(5))$	87.76	87.62
$E(x_3(5))$	93.83	93.86
$\text{Var}(x_1(5))$	50.17	51.19
$\text{Var}(x_2(5))$	47.37	47.49
$\text{Var}(x_3(5))$	47.80	48.46

finite-dimensional distributions and thus we expect all moments calculated with different g to agree (up to Monte Carlo sampling error).

A. A cyclical reaction system

For this example we chose rate constants to be $k_1 = k_2 = k_3 = 0.1$ and set the initial state to be $(100, 80, 100)^T$. Our numerical computations were carried out in MATLAB. We integrated using the Euler–Maruyama method²⁵ up to time 5 with a time step size 0.005 and generated 10^4 realizations for each simulated construction.

We simulated two different CLE formulations: the standard formulation (Construction 1) and Construction 3. The first needs three Wiener processes, while the latter only needs two Wiener processes. In our simulations of Construction 3, we used the explicit formula from Sec. IV A for the square root of the 2×2 matrix.

For this simple example, the running time required to generate the sample with g given by Construction 1 was 255 s, while with Construction 3 it was 256 s. This lack of computational improvement ought not to be surprising since the time saved by using one less Wiener process could be expected to be comparable to the time spent evaluating the complicated exact formula for the matrix square root.

The results of the comparison of the simulated means and variances using the two different constructions are presented in Table I. As one can see they agree very accurately, as expected.

B. A K^+ channel

We chose all rate constants to be $k_j = 0.1$ ($j \in \{1, \dots, 10\}$) and set the initial state to be $(100, 50, 100, 50, 100)^T$. We used the Euler–Maruyama integration scheme with a time step size 0.005 to generate 10^4 realizations up to time 5.

We simulated two different CLE constructions: the standard formulation (Construction 1) and Construction 4. The first needs ten Wiener processes, while the second needs five Wiener processes.

For this example, the running time required to generate the sample with g given by Construction 1 was 455 s, while with Construction 4 it was 261 s. This is a saving of approximately 42%. This saving compares to a 50% decrease in the number of Wiener processes and arose mainly from the decrease in the computational cost of matrix multiplications to compute the diffusion term. This observation accentuates the considerable benefit our reduction method can provide. As in

Sec. V A the means and the variances calculated using the two different constructions agreed (data not shown).

C. The Goldbeter–Koshland switch

We chose rate constants $k_1=0.05$, $k_2=0.1$, $k_3=0.1$, $k_4=0.01$, $k_5=0.1$, and $k_6=0.1$ and set the initial state to be $(110, 100, 30, 30, 100, 30)^T$. We generated 10^4 realizations up to time 5 with the Euler–Maruyama method, with a time step size 0.005.

In our simulations, we compared the standard formulation (Construction 1) with the reduced one, Construction 4. Whereas the first requires six Wiener processes, the latter only needs four.

The running time required to generate the sample for Construction 1 was 349 s, while for Construction 4 it was 254 s. Therefore the saving in computational time was approximately 27%, which is the result of a 33% reduction in the number of Wiener processes. Just as in the previous two cases, the means and the variances calculated using the two different constructions agreed (data not shown).

VI. CONCLUSION

In this paper, we have given a thorough analysis of the form of the CLE. We proved that the finite-dimensional distribution of the Itô diffusion process described by the CLE does not change if the diffusion term $g(x)=\nu\sqrt{\text{diag}(a(x))}$ of the standard form is replaced by another one as long as it satisfies Eq. (10), $g(x)g(x)^T=\nu\text{diag}(a(x))\nu^T$. We explored different ways how the factorization of the right-hand side can be carried out.

Via the combination of the factorization with the minimum number of columns in g (Sec. III C) with a state space reduction by the removal of dependent variables (Sec. III E), we showed that the CLE can be given in a form where there are as many independent Wiener processes as there are linearly independent variables. This is also the number of independent stoichiometric vectors, that is, the dimension of the linear space spanned by the update vectors, rank ν . Neither the number of variables nor the number of Wiener processes can be further reduced without loss of accuracy of the system description. The number of Wiener processes can be reduced independently from the state space reduction: one can carry out just one or the other or both. Indeed, the state space reduction was achieved by multiplying the state x and the functions f and g by a matrix T from the left and by compensating for this by “fitting” both f and g with an “input converter” T^{-1} , which are “external” changes. The reduction in the number of Wiener processes relies on what happens with the entirety of $g(x)g(x)^T$, that is, on the internal structure of $g(x)$.

Due to the stoichiometric constraints, after its release from the initial state a chemical reaction model can only move within a (rank ν)-dimensional affine subspace of the full N -dimensional state space. The state space reduction reflects this geometric constraint. One of the contributions of this paper is that we found that this many, rank ν independent Wiener processes are sufficient (and also necessary) to describe the distribution of the process given by the CLE at

any time instance t . The eigendecomposition used for Construction 2 is apparently just an algorithm to disentangle the directions (locally) in which the m Wiener processes of the standard CLE fluctuate to rank ν orthogonal directions.

Another contribution of this paper is that we showed in the same mathematical framework that in the case where there are m_1 pairs of reversible chemical reactions and m_2 irreversible reactions there is another, transparent formulation of the CLE with only m_1+m_2 Wiener processes, whereas the standard form uses $2m_1+m_2$. This new form can be seen to be more intuitive than the standard approach. At the heart of this construction is a transformation in which the two Gaussian noise processes that correspond to the two directions of a reversible reaction are replaced by a single one with variance equal to the sum of the two variances.

It is important to distinguish our system-size reduction methods from model reduction techniques, such as time scale separation applied to multiscale systems.³⁹ Ours are not approximations but transformations of the CLE which avoid information loss about the statistical properties of the kinetics of the chemical system. In computer jargon, ours are *lossless compression* methods [of the information content of the diffusion matrix $B(x)$] as opposed to model reduction approaches that are *lossy*.

We illustrated these ideas by considering alternative forms of the CLE for a HERG ion channel model and the Goldbeter–Koshland switch. We showed that considerable savings in running time can be achieved when using the reduced form of the CLE for numerical simulation. We believe that all software implementations of the CLE should include this reduced form. This would only require a small change in code and would accelerate simulation without changing the statistical properties of the generated stochastic process. The CLE is an important tool for the analysis and simulation of multiscale chemical reaction systems and it is vital to choose its most appropriate or most efficient formulation according to the requirements of the application.

ACKNOWLEDGMENTS

The authors wish to thank Thomas G. Kurtz for observing a mistake in an earlier version of this material which has been corrected subsequently. B.M.’s work was funded by the Engineering and Physical Sciences Research Council through a Doctoral Training Centre grant for the Life Sciences Interface Doctoral Training Centre, University of Oxford (Grant No. EP/E501605/1). K. C. Z. was supported by Award No. KUK-C1-013-04 made by the King Abdullah University of Science and Technology (KAUST).

APPENDIX: PROOFS DEFERRED FROM THE MAIN TEXT

1. Proof of Proposition III.1

Proof. For the ease of notation, we will drop the time variable t from $x(t)$. We apply the multidimensional Itô’s formula. This claims that when substituting time t and a diffusion process $x(t)$ into a function $u(t, y): \mathbb{R} \times \mathbb{R}^N \rightarrow \mathbb{R}$,

$$du(t,x) = \frac{\partial u(t,x)}{\partial t} dt + \sum_{i=1}^N \frac{\partial u(t,x)}{\partial y_i} dx_i + \frac{1}{2} \sum_{i,j=1}^N \frac{\partial^2 u(t,x)}{\partial y_i \partial y_j} dx_i dx_j$$

holds, where the rules for computing $dx_i dx_j$ are $dt dt = dt dW_{j,t} = dW_{j,t} dt = 0$, $dW_{j,t} dW_{j',t} = \delta_{jj'} dt$ (Kronecker delta),¹⁹ where these are approximations that hold in an $o(dt)$ sense. Applying the formula with $u(t,y) = y_i y_k$ gives

$$\begin{aligned} d(x_i x_k) &= 0 + (x_k dx_i + x_i dx_k) + \frac{1}{2} (dx_i dx_k + dx_k dx_i) \\ &= (x_k dx_i + x_i dx_k) \\ &\quad + \frac{1}{2} 2 \left(f_i(x) dt + \sum_{j=1}^d g_{ij}(x) dW_j(t) \right) \\ &\quad \times \left(f_k(x) dt + \sum_{j'=1}^d g_{kj'}(x) dW_{j'}(t) \right) \\ &= (x_k dx_i + x_i dx_k) \\ &\quad + \sum_{j=1}^d \sum_{j'=1}^d g_{ij}(x) g_{kj'}(x) dW_j(t) dW_{j'}(t) \\ &= \left(x_k f_i(x) dt + x_k \sum_{j=1}^d g_{ij}(x) dW_j(t) \right. \\ &\quad \left. + x_i f_k(x) dt + x_i \sum_{j=1}^d g_{kj}(x) dW_j(t) \right) \\ &\quad + \sum_{j=1}^d g_{ij}(x) g_{kj}(x) dt. \end{aligned}$$

Taking the expectation on both sides yields

$$dE(x_i x_k) = E(x_k f_i(x)) dt + E(x_i f_k(x)) dt + \sum_{j=1}^d E(g_{ij}(x) g_{kj}(x)) dt,$$

which is just another form of Eq. (8). \square

2. Proof of Proposition III.4

Proof. If $y \in \mathbb{R}^N \setminus \{0\}$ is a left nullvector of ν , $y^T \nu = 0$, then it is trivially a nullvector of $B(x)$,

$$B(x)y = \nu \operatorname{diag}(a(x)) \nu^T y = 0.$$

In the other direction, if $B(x)y = 0$, we first prove that y is a left nullvector of the factor in Gillespie's factorization,

$$\sigma(x) = \nu \operatorname{diag}(\sqrt{a_1(x)}, \dots, \sqrt{a_m(x)}).$$

Indeed, $0 = B(x)y = \sigma(x)\sigma(x)^T y$, hence $0 = y^T \sigma(x)\sigma(x)^T y = \|\sigma(x)^T y\|^2 = \|y^T \sigma(x)\|^2$, so $y^T \sigma(x) = 0$ and y is a left nullvector to $\sigma(x)$.

The left nullvectors of $\sigma(x)$ and ν are the same, since $\sqrt{a_1(x)}, \dots, \sqrt{a_m(x)}$ are all positive by assumption. Therefore y is a left nullvector of ν as claimed. \square

3. Proof of Proposition III.5

Proof. The column rank of $\nu \in \mathbb{R}^{N \times m}$ is just $\dim(\operatorname{Im} \nu)$. It is well known that

$$\dim(\operatorname{Im} \nu) + \dim(\operatorname{Ker} \nu) = m.$$

The row rank of ν is the column rank of ν^T , or $\dim(\operatorname{Im} \nu^T)$. Similarly,

$$\dim(\operatorname{Im} \nu^T) + \dim(\operatorname{Ker} \nu^T) = N.$$

It is also well known that the column and row ranks are always equal. Therefore

$$m - \dim(\operatorname{Ker} \nu) = \operatorname{rank} \nu = N - \dim(\operatorname{Ker} \nu^T).$$

\square

4. State space reduction for Construction 4

For Construction 4, a finer partitioning of matrices ν, J, V is proposed. Let us order the columns of $\nu \in \mathbb{R}^{N \times m}$ such that $\nu = [\nu_1 \nu_2 \nu_3 \nu_4]$, where the columns of $\nu_1 \in \mathbb{R}^{N \times (m - \dim(\operatorname{Ker} \nu))}$ form a basis for $\operatorname{Im} \nu$; ν_3 is the collection of the column vectors that are constant multiples of any single column of ν_1 ; the columns of ν_2 represent all the directions specified by columns of ν that are distinct to directions of the columns of ν_1 (columns of ν_2 are linearly dependent on columns of ν_1 , they are a linear combination of more than one); and finally ν_4 is the collection of the column vectors that are constant multiples of any single column of ν_2 . Let the sizes of these matrices define r_2, r_3 and r_4 such that $\nu_2 \in \mathbb{R}^{N \times r_2}$, $\nu_3 \in \mathbb{R}^{N \times r_3}$, $\nu_4 \in \mathbb{R}^{N \times r_4}$. Obviously, $m - \dim(\operatorname{Ker} \nu) + r_2 = s$, and $r_2 + r_3 + r_4 = \dim(\operatorname{Ker} \nu)$. The entries of $A(x)$ are permuted accordingly, and then $A(x)$ is partitioned into blocks.

This uniquely specifies the matrices

$$R \in \mathbb{R}^{(m - \dim(\operatorname{Ker} \nu)) \times r_2},$$

$$M_3 = [v^{(1)} \dots v^{(r_3)}] \in \mathbb{R}^{(m - \dim(\operatorname{Ker} \nu)) \times r_3},$$

$$M_4 = [w^{(1)} \dots w^{(r_4)}] \in \mathbb{R}^{r_2 \times r_4},$$

such that $\nu_2 = \nu_1 R$, $\nu_3 = \nu_1 M_3$, and $\nu_4 = \nu_2 M_4$, and all $v^{(i)}$ and $w^{(k)}$ have only one nonzero entry each. Then let

$$J = \begin{bmatrix} I_{m - \dim(\operatorname{Ker} \nu)} & 0 \\ 0 & I_{r_2} \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \in \mathbb{R}^{m \times s},$$

$$V = \begin{bmatrix} I_{m - \dim(\operatorname{Ker} \nu)} & 0 & M_3 & 0 \\ 0 & I_{r_2} & 0 & M_4 \end{bmatrix} \in \mathbb{R}^{s \times m},$$

J having first r_3 then r_4 rows of zeros.

The construction is again as in Eq. (13). For the sake of notational clarity, let

$$\begin{aligned} C_1(x) &= A_1(x) + \sum_{j=1}^{r_3} (A_3(x))_{jj} v^{(j)} v^{(j)T} \\ &\in \mathbb{R}^{(m - \dim(\operatorname{Ker} \nu)) \times (m - \dim(\operatorname{Ker} \nu))}, \end{aligned}$$

$$C_2(x) = A_2(x) + \sum_{j=1}^{r_4} (A_4(x))_{jj} w^{(j)} w^{(j)T} \in \mathbb{R}^{r_2 \times r_2}.$$

Then

$$VA(x)V^T = \begin{bmatrix} C_1(x) & 0 \\ 0 & C_2(x) \end{bmatrix}$$

is diagonal, $\nu JV = \nu$ and (10) hold. Defining T with ν_1 in the role of ν_b ,

$$g(x) = \nu_1 [\sqrt{C_1(x)} \quad R\sqrt{C_2(x)}],$$

$$Tg(x) = \begin{bmatrix} \sqrt{C_1(x)} & R\sqrt{C_2(x)} \\ 0 & 0 \end{bmatrix},$$

whose nonzero blocks together are in $\mathbb{R}^{(m-\dim(\text{Ker } \nu)) \times s}$, as required.

¹T. A. Turner, S. Schnell, and K. Burrage, *Comput. Biol. Chem.* **28**, 165 (2004).

²E. Klipp, R. Herwig, A. Kowald, C. Wierling, and H. Lehrach, *Systems Biology in Practice: Concepts, Implementation and Application* (Wiley-VCH, Weinheim, 2005).

³H. H. McAdams and A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).

⁴J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 2075 (2000).

⁵M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).

⁶F. J. Isaacs, J. Hasty, C. R. Cantor, and J. J. Collins, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 7714 (2003).

⁷T. Tian and K. Burrage, *J. Theor. Biol.* **227**, 229 (2004).

⁸J. G. Restrepo, J. N. Weiss, and A. Karma, *Biophys. J.* **95**, 3767 (2008).

⁹D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).

¹⁰D. T. Gillespie, *Physica A* **188**, 404 (1992).

¹¹N. G. Van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed., (Elsevier, Amsterdam, 2007).

¹²T. G. Kurtz, *J. Chem. Phys.* **57**, 2976 (1972).

¹³K. Ball, T. G. Kurtz, L. Popovic, and G. Rempala, *Ann. Appl. Probab.* **16**, 1925 (2006).

¹⁴D. F. Anderson, *J. Chem. Phys.* **127**, 214107 (2007).

¹⁵D. T. Gillespie, *J. Chem. Phys.* **115**, 1716 (2001).

¹⁶T. Tian and K. Burrage, *J. Chem. Phys.* **121**, 10356 (2004).

¹⁷A. Auger, P. Chatelain, and P. Koumoutsakos, *J. Chem. Phys.* **125**, 084103 (2006).

¹⁸D. T. Gillespie, *J. Chem. Phys.* **113**, 297 (2000).

¹⁹B. Øksendal, *Stochastic Differential Equations: An Introduction with Applications*, 6th ed. (Springer, New York, 2007).

²⁰D. J. Wilkinson, *Stochastic Modelling for Systems Biology, Mathematical and Computational Biology Series* (CRC Press, Boca Raton, 2006).

²¹E. J. Allen, L. J. S. Allen, A. Arciniega, and P. E. Greenwood, *Stochastic Anal. Appl.* **26**, 274 (2008).

²²M. Ullah and O. Wolkenhauer, *J. Theor. Biol.* **260**, 340 (2009).

²³D. Adalsteinsson, D. McMillen, and T. Elston, *BMC Bioinf.* **5**, 24 (2004), <http://www.biomedcentral.com/1471-2105/5/24>.

²⁴A. Singer, R. Erban, I. G. Kevrekidis, and R. R. Coifman, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 16090 (2009), <http://www.pnas.org/content/106/38/16090> (abstract).

²⁵P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations* (Springer, Berlin, 1992).

²⁶T. Brennan, M. Fink, and B. Rodriguez, *Eur. J. Pharm. Sci.* **36**, 62 (2009).

²⁷A. Goldbeter and D. E. Koshland, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 6840 (1981).

²⁸B. Munsky and M. Khammash, *J. Chem. Phys.* **124**, 044104 (2006).

²⁹S. MacNamara, K. Burrage, and R. B. Sidje, *Multiscale Model. Simul.* **6**, 1146 (2008), <http://dx.doi.org/10.1137/060678154>.

³⁰S. MacNamara, A. M. Bersani, K. Burrage, and R. B. Sidje, *J. Chem. Phys.* **129**, 095105 (2008).

³¹W. E. D. Liu, and E. Vanden-Eijnden, *J. Chem. Phys.* **123**, 194107 (2005).

³²E. L. Haseltine and J. B. Rawlings, *J. Chem. Phys.* **117**, 6959 (2002).

³³R. Tomioka, H. Kimura, T. J. Kobayashi, and K. Aihara, *J. Theor. Biol.* **229**, 501 (2004).

³⁴C. A. Gómez-Urbe and G. C. Verghese, *J. Chem. Phys.* **126**, 024109 (2007).

³⁵A. Singh and J. P. Hespanha, in Decision and Control, 2006 45th IEEE Conference, pp. 2063–2068, available online at <http://dx.doi.org/10.1109/CDC.2006.376994>.

³⁶D. J. Higham and R. Khanin, *The Open Applied Mathematics Journal* **2**, 59 (2008).

³⁷D. T. Gillespie, *Am. J. Phys.* **64**, 1246 (1996).

³⁸C.-Y. F. Huang and J. E. Ferrell, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10078 (1996), <http://www.pnas.org/content/93/19/10078> (abstract).

³⁹V. Sotiropoulos, M.-N. Contou-Carrere, P. Daoutidis, and Y. N. Kaznesis, *IEEE/ACM Trans. Comput. Biol. Bioinf.* **6**, 470 (2009).