Stochastic dynamics and non-equilibrium thermodynamics of a bistable chemical system: the Schlögl model revisited

Melissa Vellela* and Hong Qian*

Department of Applied Mathematics, University of Washington, Seattle, WA 98195, USA

Schlögl’s model is the canonical example of a chemical reaction system that exhibits bistability. Because the biological examples of bistability and switching behaviour are increasingly numerous, this paper presents an integrated deterministic, stochastic and thermodynamic analysis of the model. After a brief review of the deterministic and stochastic modelling frameworks, the concepts of chemical and mathematical detailed balances are discussed and non-equilibrium conditions are shown to be necessary for bistability. Thermodynamic quantities such as the flux, chemical potential and entropy production rate are defined and compared across the two models. In the bistable region, the stochastic model exhibits an exchange of the global stability between the two stable states under changes in the pump parameters and volume size. The stochastic entropy production rate shows a sharp transition that mirrors this exchange. A new hybrid model that includes continuous diffusion and discrete jumps is suggested to deal with the multiscale dynamics of the bistable system. Accurate approximations of the exponentially small eigenvalue associated with the time scale of this switching and the full time-dependent solution are calculated using MATLAB. A breakdown of previously known asymptotic approximations on small volume scales is observed through comparison with these and Monte Carlo results. Finally, in the appendix section is an illustration of how the diffusion approximation of the chemical master equation can fail to represent correctly the mesoscopically interesting steady-state behaviour of the system.

Keywords: bistable systems; chemical master equation; stochastic processes; multiscale dynamics; entropy production rate; non-equilibrium steady state

1. INTRODUCTION

Molecular reactions are the chemical basis of cellular biological functions. The complex biochemical reaction networks in terms of proteins and macromolecular complexes play a wide range of important roles in cellular signalling, control and regulation (Zhu et al. 2004). In recent years, a great amount of attention has been given to the formulation and analysis of models describing reaction networks. We are still looking for new concepts and ideas to illustrate and quantify what is important in these systems.

In this paper, we wish to use the example of a simple model first studied by Schlögl (1972) to illustrate and interlace some of these recent ideas. The major issues are as follows.

1.1. Stochastic modelling

Chemical reactions inside a cell (especially those of signalling networks involving transcription regulation, protein phosphorylation and GTPases) often involve a small number of molecules of one or more of the reactants. Thus, the traditional method of modelling systems by describing concentration changes with ordinary differential equations (ODEs) and the law of mass action (Zheng & Ross 1991; Murray 2002) is inappropriate. A Markov chain (or master equation) model accounts for the discrete, probabilistic nature of the chemical reactions at the molecular level, but can be difficult to analyse. Diffusion (Fokker–Planck) approximations to the master equation were first developed by Van Kampen (1981) and shown by Kurtz (1976, 1978) to be valid in the thermodynamic limit for some finite time. However, unless the steady state is unique in the macroscopic description (Mansour et al. 1981), the two models can disagree in the infinite time limit (see appendix A). Microscopic simulations have validated the master equation as the most accurate description of a reactive process; see Baras & Mansour (1997) for an up-to-date review. In this paper, we illustrate the inability of a deterministic model to capture the behaviour of Schlögl’s model. We suggest a new method of analysing the master equation through its spectrum and discuss which model formulation best represents the complete dynamics of the system.
1.2. Functional cellular attractors and multiple time scales

In terms of the master chemical equation (CME) formulation (Delbrück 1940; McQuarrie 1968; Arkin et al. 1998; Gillespie 2007; Beard & Qian 2008), each stable steady state of the deterministic model corresponds to a peak (while an unstable steady state corresponds to a trough or a saddle) in the stationary probability distribution (the ‘steady-state’ solution to the master equation, which is generally unique Schnakenberg 1976). We refer to these states as functional cellular attractors (FCA) because from the standpoint of cell biology, which is the most relevant field for our theory, they are the states in which the system is most likely to be found and in which the system performs its function(s). In fact, Schlögl’s model exhibits multiple time scales: a fast scale where the system relaxes to one of the FCA, and a slow scale over which the system transitions from one FCA to another (Matheson et al. 1974). A key question is how long the system remains in each of the FCA and, as we will show, the CME and Fokker–Planck descriptions can yield conflicting answers to this question. Although we restrict ourselves to the case of well-mixed reactions, spatial considerations are a key component of intracellular modelling (Antoine & Lemarchand 2007). See Elf & Ehrenberg (2004) for an analysis of biological bistability including spatial domains.

1.3. Thermodynamics of open chemical systems

In the environment of a living cell, biochemical reaction systems operate in an ‘open’ setting (Nicolis & Prigogine 1977; Qian 2006, 2007; Ross 2008). That is, there is a flux of material and/or energy acting on the system. We can no longer rely on the traditional theory of equilibrium thermodynamics to create an accurate model of these systems. The non-equilibrium theory allows the possibility of multiple steady states and non-zero steady-state flux (Kjelstrup & Bedeaux 2008), whereas a closed molecular system tends to a chemical equilibrium, which is unique and in which the fluxes are zero (Lewis’ 1925 principle of detailed balance). There is also a non-zero entropy production rate, which characterizes the non-equilibrium steady state (NESS). Recent developments in the area of fluctuation theorem (Andrieux & Gaspard 2007; Sevick et al. 2008) have highlighted the importance of entropy production and its relationship with the irreversible rate, which characterizes the non-equilibrium steady state (NESS). The CME may also be referred to as a birth and death process in the theory of Markov processes (Gardiner 1985), where \( k_n \) and \( \mu_n \) are the birth and death rates, respectively.

\[
\frac{dx}{dt} = k_1 a x^2 - k_2 x^3 - k_4 x + k_3 b.
\]

Depending on the parameters, there can be one, two or three non-negative steady states. We are particularly interested in the set of parameters for which there are two stable steady states separated by an unstable steady state, the bistable case.

The stochastic model is given in the form of the CME, an infinite system of coupled ODEs (Beard & Qian 2008). Let integer random variable \( n_X(t) \) be the number of X molecules at time \( t \), \( n_A \) and \( n_B \) be the number of A and B molecules, which are fixed for a fixed \( V \), and \( p_n(t) \) be the probability of having \( n \) X molecules at time \( t \): \( p_n(t) = \Pr\{n_X(t) = n\} \). The equations are

\[
\frac{dp_0(t)}{dt} = \mu_1 p_1 - \lambda_0 p_0,
\]

and

\[
\frac{dp_n(t)}{dt} = \lambda_{n-1} p_{n-1} + \mu_{n+1} p_{n+1} - (\lambda_n + \mu_n) p_n,
\]

for \( n = 1 \ldots \infty \),

where \( \lambda_n = k_3 n \) and \( \mu_n = k_4 n + k_2 n(n-1)(n-2) \). The CME may also be referred to as a birth and death process in the theory of Markov processes (Gardiner 1985), where \( \lambda_n \) and \( \mu_n \) are the birth and death rates, respectively.
The rates in the stochastic model are related to the rates in the deterministic model by a factor of $V$ which depends on the number of reactants involved in the $i$th reaction. For a reaction involving $m$ reactants, the relation will be $k_i = k_i/V^{m-1}$. In the following analysis, when the volume varies, we consider the ODE reaction rates $k_i$ and the parameters $a$ and $b$ to be constant (as opposed to fixing the $k_i$, $n_A$, and $n_B$). This corresponds to the thought experiment in which the concentrations are fixed while the system volume and numbers of molecules vary correspondingly. Thus, the birth and death rates are written in terms of the ODE parameters as

$$\lambda_n = \frac{ak_1(n-1)}{V} + bk_3 V,$$  

and

$$\mu_n = nk_1 + \frac{k_2(n-1)(n-2)}{V^2}.$$  

The steady-state probability distribution of the stochastic model can be found by noting that $\lambda_{n-1}p_{n-1} = \mu_n p_n$ when the system is stationary. We shall call this equation mathematical detailed balance following the theory of Markov processes (Schnakenberg 1976; Gardiner 1985; Jiang et al. 2004), in which the probability flux in the forward direction is equal to the probability flux in the backward direction at each state in the CME. This leads to the well-known stationary probability,

$$\frac{p_{\text{ss}}^n}{p^0_n} = \prod_{i=0}^{n-1} \frac{\lambda_i}{\mu_{i+1}}, \quad p^0_n = 1 - \sum_{n=1}^{\infty} p_{\text{ss}}^n.$$  

Note that the steady state of the stochastic model (i.e. mesoscopic view) is unique for all values of the parameters. The stable and unstable steady states of the ODE are represented as minima and maxima, respectively, in the steady-state probability distribution of the CME.

3. ANALYSIS

In this section, we analyse and compare the deterministic (ODE) and the stochastic (CME) models. Under equilibrium conditions, the two models are in complete agreement. However, for the parameters that yield a bistable system, there are stark differences in the predicted behaviour, and in the definition of thermodynamic quantities which describe the system.

3.1. Chemical flux and steady-state behaviour

There are four chemical reactions, each of which has a chemical flux, which describes how that reaction contributes to the change of $x$:

$$J_1^x(x) = k_1 a x^2, \quad J_2^x(x) = k_2 x^3 \quad \text{and} \quad J_3^x(x) = k_3 b, \quad J_4^x(x) = k_4 x.$$  

(3.1)

The ODE model can be expressed as the sum of the forward fluxes minus the sum of the backward fluxes,

$$\frac{dx}{dt} = (J_1^+ + J_2^+) - (J_1^- + J_2^-).$$  

(3.2)

In the macroscopic setting, the steady state is reached when $dx/dt = 0$. Because the ODE form of Schrögl’s model is a cubic, there can be one, two or three steady states for a given set of parameters. Note that the cubic term comes from the trimolecular portion of the reaction: bistability is not possible in a one-dimensional model involving only uni- and bimolecular reactions.

We will begin with the simplest case. Recall that, in general, in order to keep the concentrations of A and B at constants $a$ and $b$, there will be a net flux of A going into the system and B coming out of the system, or vice versa. However, there is a particular ratio of $a$ and $b$ at which there will be no need for an external agent to add or remove A and B molecules from the system to maintain their concentrations on average (i.e. ignoring microscopic fluctuations). The values of $a$ and $b$ that make this possible are called equilibrium concentrations. These concentrations lead the system to a unique and completely describable fate: the equilibrium steady state.

Chemical detailed balance occurs when $J_1^+ = J_1^-$, so that the forward and backward fluxes are balanced in each and every reaction in the system. Note that this is a stronger condition than mathematical detailed balance, which simply assumes that the total forward rate is equal to the total backward rate. Using the definitions in equation (3.1), the chemical detailed balance (or equilibrium steady state) condition for Schrögl’s model is

$$\frac{k_1 k_4 a}{k_2 k_3 b} = 1.$$  

(3.3)

This condition is derived later in equation (3.10) by assuming that there is no chemical potential difference between A and B and no free energy lost when exchanging an A for a B or vice versa.

The value of $x$ for which chemical detailed balance occurs is denoted as $x_{\text{ess}}$. This is the equilibrium steady state. When a system is closed to outside influences, or in contact with material baths with a constant chemical potential, this is the unique concentration value of $x$ to which the system will tend. In the CME setting, the steady-state probability distribution of the equilibrium steady state is a Poisson (or multi-Poisson) distribution, as predicted by Gibbs’ theory for grand canonical systems (Gardiner 1985; Heuett & Qian 2006). The number of molecules at which the distribution peaks corresponds (when divided by volume size) to the steady-state concentration in the ODE model. For our example, the steady-state concentration is

$$x_{\text{ess}} = \frac{k_1 a}{k_2} = \frac{k_3 b}{k_4},$$  

which is a function of the rate constants and the volume size.

(3.4)

and the steady-state probability distribution is

$$p^\theta_n = p^0_n \theta^n / n! \theta^\theta, \quad \theta = \frac{ak_1 V}{k_2}.$$  

(3.5)

When the parameter set does not satisfy equation (3.3), the system will reach a NESS. Unlike equilibrium conditions, multiple NESS may exist for a given set of parameters. Since it is not necessary for $J_1^+ = J_1^-$, an
NESS is characterized by the reactions forming a cycle: the reversible reaction in equation (2.1) makes an X from A, and the reaction in equation (2.2) transforms an X to B. Then, an external agent has to constantly make A from B to keep the concentrations of A and B constant. Hence, there is a chemical reaction cycle \( A \rightarrow X \rightarrow B \rightarrow A \). In the stochastic model in §3.3, we shall quantify this reaction cycle by a cycle flux.

We now categorize the types of behaviours that may occur under non-equilibrium conditions through bifurcation analysis, including a ‘bifurcation’ which can only be observed through the CME. Figure 1 shows plots of the ODE, \( \frac{dx}{dt} \) versus \( x \), for an example of chemical equilibrium and for a bistable set of the parameters. When there is only one real root, the steady-state behaviour of the stochastic system will match that of the deterministic system (Mansour et al. 1981). However, as the parameters move away from the chemical detailed balance condition, the shape of the probability steady-state function will deform from its Poissonian shape.

In a biological setting, it is more likely that the concentrations of the pump parameters (A and B) could differ in varying situations, while the rate constants, which are inherent to the types of molecules involved, would stay the same. Thus, we assume that only the rate constants, \( a \) and \( b \), differ in varying situations, while the rate constants, \( k \), are fixed at the values \( k_1=0.5, k_2=0.25 \) for chemical equilibrium and \( a=1, b=1 \) for the bistable case.

Figure 1. Plots of \( \frac{dx}{dt} \) versus \( x \) for the cases of chemical equilibrium (dashed curve) and bistability (solid curve) in the ODE model. The reaction rates are \( k_1=3, k_2=0.6, k_3=0.25 \) and \( k_4=2.95 \) for both plots. The pump parameters are \( a=0.5, b=29.5 \) for chemical equilibrium and \( a=1, b=1 \) for the bistable case.

Figure 2. Regions of monostability and bistability in the \( ab \)-plane with reaction rates fixed at the values \( k_1=3, k_2=0.6, k_3=0.25 \) and \( k_4=2.95 \). in the case of Schrödinger's model, the variance diverges at the bifurcation point, \( \Delta = 0 \) (Nicolis & Turner 1977).

Under bistable conditions, the stochastic and deterministic models yield different predictions (an example of Keizer’s paradox, see Vellela & Qian 2007). In the deterministic model, the system will tend towards one of the two stable fixed points, depending on the initial condition. The stochastic model also predicts that the system will quickly relax towards one stable point (FCA), but randomly switch between the two FCAs on a time scale related to the system size.

Plotting the steady-state solution to the master equation (equation (2.8)), we observe that the two peaks can exchange their relative heights under changes in \( a, b \) or \( V \) individually (i.e. changing one parameter while keeping the other two constant). Figures 3 and 4 illustrate this switch under a change in the pump parameter \( b \), and the volume size \( V \), respectively. The volume size is a critical part of this bifurcation: the value of \( a \) or \( b \) at which the peaks have equal height depends on the volume size. Thus, this change cannot be observed through studying the ODE model.1,2

It has been shown that under multistable conditions, it is impossible to derive a Fokker–Planck equation that correctly describes the behaviour predicted by the master equation (Horsthemke et al. 1977). The relative heights of the two peaks in the steady state are incorrectly approximated in the continuous limit by diffusion approximation. Since these heights dictate the relative stability of the FCA, and the height differences are amplified in the limit of large system size, accuracy of the probability steady state is critical.

None of the continuous models are able to accurately describe the switching behaviour for small volume systems, as we illustrate in §4.3. As of yet, there is no experimental data to confirm the predictions from any

---

1For example, this bifurcation does not correspond to the point at which the potential function \( \int_0^x \varphi \, dx \) has two minima of equal depth.

2In a more strict sense, the relative stability of the two peaks depends not only on the heights of the peak, but also on the curvature. In thermodynamic terms, the former is entropic and the latter is entropic contribution to the free energy. For large system size, the dominant effect is from the height difference, which is exponentially related to the volume \( V \). The curvature converges when \( V \to \infty \).

---

J. R. Soc. Interface (2009)
of these models. However, simulations through molecular dynamics (Baras & Mansour 1997) have shown agreement with the master equation. Thus, when studying non-equilibrium processes occurring on the microscopic and mesoscopic scales, the CME is the most valid mathematical model.

3.2. Deterministic chemical potential and entropy production

In this section, we attempt to characterize the steady-state behaviour using concepts from physical chemistry. In each of the reactions there is a chemical potential difference, between the Gibbs free energy stored in the system before the reaction occurs versus after it has occurred. The first reaction in the Schlögl model changes an A molecule into an X molecule. By definition (Beard & Qian 2008), the amount of chemical potential energy it takes to change one A into one X is

\[ \Delta \mu_{AX} = k_B T \ln \left( \frac{J^+_1}{J^-_1} \right) = k_B T \ln \left( \frac{k_1 a}{k_2 x} \right), \]

where \( k_B \) is Boltzmann’s constant and \( T \) is the temperature (constant throughout our theory). Likewise, the Gibbs free energy used to change one B into one X is

\[ \Delta \mu_{BX} = k_B T \ln \left( \frac{J^+_2}{J^-_2} \right) = k_B T \ln \left( \frac{k_3 b}{k_4 x} \right). \]

The overall result of this reaction system is creating B molecules out of A molecules (or vice versa). Combining the two chemical potentials above, we see that the amount of potential energy lost in exchanging one A for one B is

\[ \Delta \mu_{AB} = \Delta \mu_{AX} + \Delta \mu_{XB} = k_B T \ln \left( \frac{k_1 k_4 a}{k_2 k_3 b} \right) = \Delta G, \]

where we have used \( \Delta \mu_{XB} = -\Delta \mu_{BX} \). We rename this expression \( \Delta G \) because it represents the change in the Gibbs free energy which occurs through one ‘forward’ cycle of the reaction (i.e. producing one B from one A).

The chemical potential difference depends on the values of the rate constants (enthropic) and the amounts of substrate and product (entropic). For a special value of the parameters, namely

\[ \frac{k_1 k_4 a}{k_2 k_3 b} = 1, \]

the chemical potential difference between A and B is zero. This is the same condition as in equation (3.3) and the system will tend towards the equilibrium steady state discussed in §3.1. In this state, there is no preference for producing B from A or vice versa because there is no energy difference between them.

Another important physical quantity is the entropy production rate. Entropy is created when an amount of useful energy is converted to heat in an isothermal process. In order to keep the concentrations of A and B constant, an external agent has to constantly remove the B, convert B back to A, and put A back in the system. All this work becomes heat in the steady-state reaction.

If chemical energy is constantly pumped into the system, a NESS will be reached in which the entropy is produced at a constant rate (related to the amount of work being done on the system) converted to heat and dissipated. The amount of entropy produced in turning one A into one B is equal to the free energy (differing by a constant factor of \( T \), the temperature in Kelvin; equation (3.9)). Thus, the entropy production rate will be this constant multiplied by the rate of producing B from A. This rate is called the steady-state flux

\[ J^+_i = (J^+_1 - J^-_1)|_{x_i} = (J^-_2 - J^+_2)|_{x_i}, \]

where \( x_i \) is the left, middle or right fixed point of equation (2.3) for \( i = -, 0, + \), respectively. The two differences in equation (3.11) represent the net amount of A being used, and the amount of B being produced, respectively, per unit time. These two rates will be equal in the steady state. Thus, we have the entropy production rate in the \( i \)th steady state as

\[ \xi_i = \Delta G \times J^+_i. \]

Note that when the system is in the equilibrium steady state, there is only one entropy production rate and it is zero.

For all other cases, there may be multiple entropy production rates, and they will always be positive. To show this, we rewrite the entropy production rate in terms of each reaction, and use both definitions of the steady-state flux in equation (3.11): 

\[ \xi_i = (\Delta \mu_{AX} + \Delta \mu_{XB}) J^+_i, \]

\[ = k_B T \left[ \ln \left( \frac{J^+_1}{J^-_1} \right) (J^-_1 - J^+_1)|_{x_i} \right. \]

\[ + \ln \left( \frac{J^-_2}{J^+_2} \right) (J^-_2 - J^+_2)|_{x_i} \right]. \]

Note that a term such as \( \ln(x/y)(x-y) \) is always greater than or equal to zero, since \( (x-y) > 0 \) implies \( \ln(x/y) > 0 \) and \( (x-y) < 0 \) implies \( \ln(x/y) < 0 \). Thus, we know that \( \xi_i \geq 0 \) in all cases.
3.3. Stochastic entropy production and cycle flux

The entropy production rate is an important characteristic of the NESS. The entropy production rates can be different for each steady state in the deterministic model of a bistable system. In the stochastic model, there is a unique steady-state distribution, even in the case of bistability. Likewise, there is a unique entropy production rate, even when the system is in a NESS. In this section, we describe how the stochastic entropy production rate is defined and how it relates to the deterministic rate from above.

Let us ‘visualize’ the stochastic dynamics of the chemical reaction system, one reaction at a time. Every time a reaction occurs, there is an associated amount of chemical energy loss (to the solvent via heat) or gain (from the solvent via heat). Since the occurrences of the reactions are stochastic, the system’s net heat output fluctuates. When the system reaches its stationary state, this heat output is not constant, but continues to increase on average (i.e. it increases at a constant rate). Thus, there is a mean stationary entropy production rate.

The entropy production rate for the stochastic model is calculated through a double sum: first over each reaction. We define from equation (2.6) into the transition rates corresponding to each reaction. We define

\[
\begin{align*}
\lambda_n^1 &= k_1 an(n-1)/V, \\
\lambda_n^2 &= k_3 b V, \\
\mu_n^1 &= k_2 n(n-1)(n-2)/V^2, \\
\mu_n^2 &= k_4 n,
\end{align*}
\]

so that \(\lambda_n = \lambda_n^1 + \lambda_n^2\) and \(\mu_n = \mu_n^1 + \mu_n^2\). We may then define the stochastic forward and backward steady-state fluxes analogous to those in equation (3.1):

\[
\begin{align*}
\dot{J}_1^n &= \lambda_n^1 p_n^s, \\
\dot{J}_2^n &= \lambda_n^2 p_n^s,
\end{align*}
\]

where \(p_n^s\) is the distribution given in equation (2.8).

The stochastic entropy production rate (as defined in Luo et al. (2002)) is

\[
\dot{\xi}_{\text{CME}} = \frac{1}{2} \sum_{j=1}^{2} \sum_{n=0}^{\infty} \left( J_j^n(n) - J_j^n(n+1) \right) \ln \left( \frac{J_j^n(n)}{J_j^n(n+1)} \right),
\]

which represents summing the product of flux and chemical potential over all the states in the probability steady-state distribution for each reaction separately.

This definition can be rewritten in terms of the product of the Gibbs free energy change, \(\Delta G\), and the cycle flux, a quantity analogous to the steady-state flux from equation (3.11). Figure 5 illustrates how the rates from equation (3.15) affect the movement on the Markov chain. The net forward movement is equal to \(\dot{J}_1^n(n) - \dot{J}_1^n(n+1)\). This is defined as the forward cycle flux. Likewise, the net movement backward is equal to \(\dot{J}_2^n(n) - \dot{J}_2^n(n+1)\), and is the backward cycle flux. These fluxes represent the rate at which the cycle turns over in the forward (i.e. A→B) and backward (i.e. B→A) directions in a given state \(n\).

The forward and backward cycle fluxes must be equal when the system is in a steady state. Thus, the double sum in equation (3.20) can be rewritten as a single sum using either cycle flux, and the log terms can be combined,

\[
\dot{\xi}_{\text{CME}} = \sum_{n=0}^{\infty} \left( \dot{J}_1^n(n) - \dot{J}_1^n(n+1) \right) \ln \left( \frac{\dot{J}_1^n(n)\dot{J}_2^n(n+1)}{\dot{J}_1^n(n+1)\dot{J}_2^n(n)} \right).
\]

Plugging in the definitions of the single reaction fluxes, the log term becomes the Gibbs free energy from equation (3.9),

\[
\ln \left( \frac{\dot{J}_1^n(n)\dot{J}_2^n(n+1)}{\dot{J}_1^n(n+1)\dot{J}_2^n(n)} \right) = \ln \left( \frac{k_1 k_4 a}{k_2 k_3 b} \right) = \Delta G.
\]

This term represents the entropy production of one forward completion of cycle \(n\) (which is actually independent of \(n\), since the Gibbs free energy is converted into entropy. The stochastic entropy production rate then becomes.
The stochastic entropy production rate can then be written in terms of the deterministic fluxes as

$$\dot{\xi}_{\text{CME}} \approx V \Delta G \sum_{n=0}^{\infty} J_1^{+}(n/V) - J_1^{-}(n/V) p_n^\text{ss} - \mu_{n+1}^1 p_n^\text{ss} + \mu_n^1 p_{n+1}^\text{ss}.$$  

Assuming that the probability steady state is dominated by the values at its two peaks, we can approximate the entropy production rate by the sum of the two terms corresponding to those states:

$$\dot{\xi}_{\text{CME}} \approx V \Delta G \sum_{i=\pm} \left( J_1^+(n/V) - J_1^-(n/V) \right) p_n^\text{ss},$$  

where the $n_\pm$ represent the integers which approximate the $x_\pm$ closest and we have used the approximation $p_{n+1}^\text{ss} \approx p_n^\text{ss}$ around these states. Using the definition of the deterministic entropy production rate from equation (3.12), we have

$$\dot{\xi}_{\text{CME}} \approx V(\xi_{++} + \xi_{--}),$$  

which states that the stochastic mean entropy production rate is approximately the sum of the deterministic entropy production rates of the stable equilibria, weighted by their probabilities, times the system volume. Thus, when all other parameters are held constant, the stochastic entropy production rate will grow linearly with volume size.

As seen in the plots of $p_n^\text{ss}$ in figures 3 and 4, even when there are two peaks in the probability steady state, one of the peaks is usually much more dominant (note the log scale). Owing to this, the stochastic entropy production rate will follow the deterministic entropy production rate of whichever equilibrium is more stable. Figure 6 shows a plot of the possible entropy productions rates (the $\xi_{-\pm}$ and $\dot{\xi}_{\text{CME}}$) over a range of volumes. Only for a small range of the parameters will $\dot{\xi}_{\text{CME}}$ not match one of $\xi_{-\pm}$; this is the range over which the peaks are comparable in height.

A topic of current interest is whether a general principle involving the entropy production rate can be stated (Callen 1957; Glansdorff et al. 1974; Keizer & Fox 1974; Jaynes 1986; Qian 2002; Dewar 2005; Martyusheva & Seleznev 2006): is the rate of entropy production always maximized, minimized or neither? Schlögl’s model illustrates how the entropy production always maximized, minimized or neither?
parameters constant, we can observe how increasingly robust.
production decreases as the more stable FCA becomes
of these dark lines do suggest that the entropy
production rates are indicative of less dissipation cost in
for the more stable FCA only. The negative slopes
stochastic entropy production and barrier height
this. The dark lines show the relationship between the
Figure 7. The stochastic entropy production rate versus the
barrier height of the bold and dashed lines. Grey dashed line,
0.2
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
0.02
0.04
0.06
0.08
0.10
0.12
0.14
0.16
0.18

3.4. Bistable system as a two-state Markov process
When the system is bistable, there exists a non-zero
possibility of jumping from one stable state to another
in the stochastic model. When the system starts near
one of the functional reaction system states (FCA), it
will fluctuate around it for a long time. However, since
stochastic fluctuations are always present, eventually
they will push the system far enough away from one
FCA so that it reaches the point corresponding to the
unstable fixed point of the ODE model. Then the
system will relax towards the other FCA and remain
there for a long time.
This behaviour motivates a new method for viewing
bistable systems. The infinite-state Markov chain
can be compressed into a model with only two states:
the two stable FCAs. Movement near the FCA can
be modelled with simple Gaussian processes (Keizer
1979), and jumping between the FCA can be predicted
using the two-state Markov chain. This simplified
view of the system will capture the behaviours on
both time scales, and allow for a new, simpler method
of simulation.
The key behaviour of a bistable system is the
transition between the FCA. The transition rates
from \( x_- \) to \( x_+ \) have been approximated asymptotically
by Hänggi. He showed that the rates depend
exponentially on the volume size, so that as volume increases,
the likelihood of a transition between the FCA
decreases exponentially. These rates were also derived
through a different method in (Hinch & Chapman
2005). The expression for the transition rate from \( x_- \) to
\( x_+ \) is (Hänggi et al. 1984; Hinch & Chapman
2005)
\[
r_- = \frac{\exp[u^g/x] J^-(x_-) \sqrt{-u^n u^{-}}}{2\pi},
\]
and the transition rate from \( x_+ \) to \( x_- \) is
\[
r_+ = \frac{\exp[u^g/x] J^+(x_+) \sqrt{-u^n u^{+}}}{2\pi}.
\]
In these expressions, \( J_{\pm} = u(0) - u(x_{\pm}) \) where \( u(x) = \int_0^x \ln(J^+ - J^-) dy \) and \( u' \) is the second derivative of \( u(x) \)
evaluated at \( x_{\pm} \). The functions \( J^+ = J^+_1 + J^+_2 \) and
\( J^- = J^-_1 + J^-_2 \) from equation (3.1) are the total continuous
forward and backward fluxes and the volume is
represented as \( 1/\epsilon \).
We may think of the long-term behaviour as a two-
state Markov chain whose infinitesimal matrix contains
the transition rates
\[
\frac{d}{dt} \begin{pmatrix} P_- \\ P_+ \end{pmatrix} = AP = \begin{pmatrix} -r_- & r_- \\ r_+ & -r_+ \end{pmatrix} \begin{pmatrix} P_- \\ P_+ \end{pmatrix}.
\]
The vector \( P \) has two components: \( P_- \) and \( P_+ \) are
the probabilities of being near the FCA \( n_- \) and \( n_+ \),
respectively. The process jumps between the two states,
The right eigenvectors of $Q$, $v_1$, represent contributions to the solution vector (see equation (4.3) in §4.1). The two right eigenvectors of $A$, $v_1$, contain a shortened version of the main contributors, $v_0$ and $v_1$. The right eigenvector $v_0$ represents the probability steady-state distribution. The two values in the eigenvector $v_0$ represent the proportion of probability around each FCA in the steady state. Thus, if we sum the first $n_0$ entries of $v_0$, they will be equal to $r_+/r_+(r_++r_-)$ (the first entry of $v_0$) and the sum of the rest of the entries will equal $r_-/(r_++r_-)$ (the second entry of $v_0$). Likewise, if we sum the first $n_0$ entries of $v_1$, they will be equal to $1/2$ and the sum of the rest will be $-1/2$, the two entries of $v_1$.

The vectors $w_0$ and $w_1$ are both entirely constant, since the equations $wQ=0$ and $wA=0$ are both satisfied by vectors of one. The eigenvector $w_0$ comprises two constant parts and changes values quickly near the unstable point $n_0$. The ratio of these two constants is the same as the ratio of the two entries of $w_1$. Thus, we see that the two-state process based on the transition rates between the FCA shares key characteristics of the original stochastic model.

Assuming that the system acts as a two-state process with the Gaussian noise around each FCA, we can model the probability distribution as the sum of two unnormalized Gaussian distributions. From Hinch & Chapman (2005), the zeroth-order asymptotic approximation to the steady-state probability distribution is

$$p_n^s \approx \frac{A}{\sqrt{J^+J^-}} \exp \left( V \int_0^n \log \left( \frac{J_+}{J_-} \right) dy \right) + O \left( \frac{1}{V} \right),$$

where the constant $A$ can change to match the function properly near each FCA. The unnormalized Gaussian that best matches this function near $n_+$ is

$$p_n^s \approx e^{V u(x_+)} \frac{J^+(x)}{J^+(x_+)} \exp \left( -\frac{u''(x-x_+)^2}{2} \right),$$

and near $n_-$, it is

$$p_n^s \approx e^{V u(x_-)} \frac{J^-(x)}{J^-(x_-)} \exp \left( -\frac{u''(x-x_-)^2}{2} \right).$$

We might consider the ratio of finding the two-state process in one state versus another. By integrating the approximations in equations (3.35) and (3.36) on the left- and right-hand sides of $v_0$ (and using the identity $\int e^{-a(x-b)^2} = \sqrt{2\pi/a}$), we can approximate this ratio by

$$\left[ \frac{p_n^s}{p_0^s} \right] \approx e^{V u(x_+)-u(x_-)} \frac{J^+(x_+)}{J^-(x_-)} \sqrt{\frac{u''_+}{u''_-}}.$$ 

If we instead use the asymptotic equations for the transition rates themselves (equations (3.29) and (3.30)), we find

$$\frac{r_+}{r_-} = e^{u(x_+)-u(x_-)} \frac{J^+(x_+)}{J^-(x_-)} \sqrt{\frac{u''_+}{u''_-}},$$

which is the same result. Thus, we conclude that, asymptotically, the steady-state behaviour is well approximated by the Gaussian noise around both FCA, with the process in equation (3.31) dictating the jumps between the two.

\[\Box\]

This approximation was found by expanding the logarithm of the function in equation (3.34) about $n_0$, and matching the first three terms to fit a Gaussian, noting that $J^+=J^-$ at the fixed points.
When $V$ is large, a realistic system with any external noise will almost surely be found in only one of two FCA. Which state it ends up in depends solely on the parameter values (i.e. whether $u(x_+)>u(x_-)$ or vice versa) and not on the initial condition! Deterministically, which state the system ends up in depends solely on the initial condition. This is a significant contribution from the stochastic model and another example of the Keizer’s paradox (Vellela & Qian 2007). Although there is a result from Kurtz (1976, 1978) which states that the solutions of the two types of models must match to a negligible probability in the thermodynamic limit for all finite time, there is no such statement when $\lim_{V \to \infty}$ and $\lim_{N \to \infty}$ are switched. Here, we observe again the disagreement when we assume the system has reached a steady state before taking the thermodynamic limit.

When working with a small volume system, jumping from one FCA to another is an event that may happen on a reasonable time scale. The system will reach its steady state (i.e. infinite time limit) without reaching the infinite volume limit. Thus, an understanding of the correct steady-state behaviour is crucial. In the §4, we derive a numerical method to examine small volume behaviours, and compare the asymptotic results with direct results from the master equation.

4. NUMERICAL METHODS

In this section, we investigate a numerical method for analysing the Schlögl model. We show that MATLAB is capable of accurately computing the eigenvalues and eigenvectors of a truncated linear system that represents the master equation. The exponentially small eigenvalue is well approximated for appropriately large truncations. This method not only produces the probability steady-state distribution; it also yields the full, time-dependent solution of the truncated system.

4.1. Calculating the solution

The stochastic model is represented by the infinite linear system

$$\frac{dp}{dt} = Qp,$$

where $p = (p_0(t), p_1(t), p_2(t), \ldots)^T$, the vector of probabilities for each state, the matrix,

$$Q = \begin{pmatrix}
-\lambda_0 & \mu_1 & 0 & \ldots \\
\lambda_0 & -\lambda_1 - \mu_1 & \mu_2 & \ldots \\
0 & \lambda_1 & -\lambda_2 - \mu_2 & \ldots \\
& \vdots & \vdots & \ddots & \ddots 
\end{pmatrix},$$

is a stochastic matrix, which has all real eigenvalues, the largest being $\lambda_0=0$.

Although all eigenvalues $\lambda_i$ with $i \geq 1$ are real and negative, the second smallest eigenvalue (or the eigenvalue with the second smallest magnitude), $\lambda_1$, differs from the others. This eigenvalue is exponentially small, i.e. it decays exponentially as the volume size increases. The other eigenvalues are relatively stationary with the volume change, and are much larger in magnitude for large volume. This behaviour is related to the exponential decay seen in the transition rates (see equations (3.29) and (3.30)). In fact, the exponentially small eigenvalue represents the sum of the two transition rates and therefore dictates the slow time scale of the system.

The solution in terms of the eigenvalues $\lambda_i$ and eigenvectors $v_i$ of $Q$ will be

$$p(t) = c_0v_0 + c_1e^{\lambda_1 t}v_1 + \cdots.$$  \hspace{1cm} (4.3)

Because all of the $\lambda_i$ are real and negative, all the terms in this solution will decay to zero as time increases except the first term, corresponding to $\lambda_0=0$. The eigenvector for the zero eigenvalue $v_0$ therefore represents the steady-state probability distribution. The second eigenvalue $\lambda_1$ is related to the time scale of the transitions between the FCA, and the contribution of its eigenvector $v_1$ to the solution will decay on a much slower time scale than the eigenvectors $v_i$ for $i > 1$.

Although this system is infinite in size, the most probable states are those near the two FCA. The states in which $n$ is much larger than $x_+ V$ are very unlikely, and the stationary probability distribution tends to zero as $n$ tends to infinity. Therefore, it is reasonable to cut the system off at some finite value $N$ and analyse the solution under these conditions (see Vellela & Qian (2007) for verification of this method). Thus, we consider the finite system formed by setting $\lambda_N=0$, so that the system cannot move past $N$. Along with setting $\mu_{N+1}=0$, the corresponding transition matrix will remain a stochastic matrix with the expected properties.

We formed the matrix $Q$ using the spdiags command in MATLAB, with the following parameter values (which lie in the bistable region of parameter space):

$$k_1 = 3, k_2 = 0.6, k_3 = 0.25, k_4 = 2.95, a = 1, b = 1.$$ \hspace{1cm} (4.4)

The function $dx/dt = f(x)$ is shown in figure 1 in §3.1. There are two stable steady states: $x_0=0.0935$ and $x_+ = 3.7024i$ and one unstable steady state: $x_0=1.2041$.

The matrix was truncated at a value $N$, past the right fixed point $x_+$. The choice of $N$ was based on a study of the convergence rate of the exponentially small eigenvalue for different choices of $N$. As in Vellela & Qian (2007), the eigenvalue quickly converges for truncation values past $N=x_0 V$. The truncation value of $N=2x_+ V$ is sufficient, and was used for the rest of the calculations below. Eigenvalues and eigenvectors were calculated using the command eig(full(Q)).

4.2. Accuracy of the eigenvalues

The time-dependent solution can be expressed as in equation (4.3) as long as the calculated eigenvalues and eigenvectors are accurate. To test the accuracy of MATLAB’s results, we calculate the absolute residual vector,

$$r_i = Qv_i - \lambda_i v_i,$$ \hspace{1cm} (4.5)

and look at the $L_2$ norm. Table 1 gives a list of the residuals for the first five eigenvalue/eigenvector pairs.
Table 1. The first five eigenvalues and residuals (from equation (4.5)) for the system with parameter values from equation (4.4) and $V=100$.

<table>
<thead>
<tr>
<th>$i$</th>
<th>$\lambda_i$</th>
<th>$|r_i|_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$2.15 \times 10^{-12}$</td>
<td>$7.87 \times 10^{-11}$</td>
</tr>
<tr>
<td>1</td>
<td>$-1.755 \times 10^{-10}$</td>
<td>$7.9064 \times 10^{-11}$</td>
</tr>
<tr>
<td>2</td>
<td>$-1.643$</td>
<td>$8.0805 \times 10^{-11}$</td>
</tr>
<tr>
<td>3</td>
<td>$-2.372$</td>
<td>$7.9897 \times 10^{-11}$</td>
</tr>
<tr>
<td>4</td>
<td>$-3.1778$</td>
<td>$7.8154 \times 10^{-11}$</td>
</tr>
</tbody>
</table>

using a volume size of $V=100$ and the parameter values in equation (4.4). The absolute residual is of the order of $10^{-11}$ for all eigenvalues when $V=100$.

Other values of $V$ were also examined, and it was found that the absolute residual grows slowly and approximately linearly for all eigenvalues as $V$ increases. Although this residual is of the same order for all eigenvalues, $\lambda_1$ grows exponentially with volume size, while the others remain relatively constant. A study of the relative residual, i.e. the absolute residual divided by the magnitude of the eigenvalue, may serve as a better measure of the accuracy of the MATLAB's results. Figure 9 shows a plot of the relative residual, $\|r_i\|_2/|\lambda_i|$ for increasing values of $V$. Similar to the eigenvalues themselves, this residual grows exponentially with $V$ for $\lambda_1$ and is constant for the other eigenvalues. Since the magnitude of $\lambda_0$ is so small, its relative residual is expectedly high. However, this residual remains constant with increasing $V$. For large volumes, such as $V \geq 100$, the absolute residuals are still small, but the relative residual for $\lambda_1$ becomes greater than one and the exponential growth is lost. Thus, we cannot trust the MATLAB’s results for $V$ in this region.

To check the accuracy of the MATLAB’s eigenvector calculations, we compare the eigenvector $v_0$, with $p_{ss}^n$, which can be calculated exactly using equation (2.8). Figure 10 is a plot of the actual probability steady state versus the eigenvector $v_0$, for $V=100$. The eigenvector is a good approximation for the values above machine epsilon ($10^{-16}$). However, as $V$ increases, the vector $v_0$ begins to diverge from the true probability steady state.

Figure 10. Plot of the probability steady state (solid curve), calculated using detailed balance and as the eigenvector $v_0$ in MATLAB (dashed curve). Note that the drastic errors begin to occur in the MATLAB’s solution for values of $p_{ss}$ less than machine epsilon, $10^{-16}$, for values near the FCA (as is seen around $n_+$ in figure 10). A study of the difference between these two vectors, $\|v_0 - p_{ss}^n\|_2$, shows that the error in $v_0$ grows exponentially with volume size as well. As in figure 9, for volumes greater than 100, the error fluctuates wildly and even the exponential growth is no longer observable.

The first two right and left eigenvectors, $w_0$, $w_1$, $w_0$, and $w_1$, respectively, are shown in figure 8. Another check of the MATLAB’s accuracy is to verify the orthogonality condition,

$$ v_i \cdot w_j = 0 \quad \text{for} \quad i \neq j. \quad (4.6) $$

The left eigenvectors were calculated by calculating the right eigenvectors of the transpose matrix with $\text{eig}('full(Q)',)$, and taking the transpose of the eigenvector output. Both $w_0$, $w_1$, and $w_0$, $w_1$ were observed to grow exponentially with increasing $V$. For $V=40$, the dot products are

$$ v_0 \cdot w_1 = -3.5277 \times 10^{-10}, \quad (4.7) $$

and

$$ v_1 \cdot w_0 = -8.7035 \times 10^{-09}, \quad (4.8) $$

and for $V=100$, they are

$$ v_0 \cdot w_1 = -0.00057286, \quad (4.9) $$

and

$$ v_1 \cdot w_0 = -0.006249. \quad (4.10) $$

Again, for volume sizes greater than 100, the error becomes too large to display the exponential growth with $V$.

These tests confirm that the MATLAB’s calculations are accurate for a range of small $V$, but as the volume and the matrix size grow, so do errors in the eigenvalues and the eigenvectors. The first two eigenvalues and their corresponding left and right eigenvalues are the most difficult to calculate, and the error grows exponentially as $V$ increases. In the §4.3, we compare the results for $\lambda_1$ from MATLAB with the asymptotic predictions from Hänggi et al. (1984) and Hinch & Chapman (2005). As we will see, asymptotic...
approximations have the opposite problem: they are accurate for large $V$ and break down for small $V$.

### 4.3. Comparison with simulations and asymptotics

Another way to check the accuracy of the exponentially small eigenvalue $\lambda_1$ is to compare the MATLAB’s $\lambda_1$ with a value obtained from the Monte Carlo simulations and with the sum of the asymptotic transition rates from equations (3.29) and (3.30). From Hänggi et al. (1984), the relationships between $\lambda_1$, the transition rates $dP_\pm/dt$ and the waiting times $T_\pm$ is

$$\lambda_1 = r_+ + r_- = -\frac{1}{2T_+} - \frac{1}{2T_-}. \quad (4.11)$$

The waiting times were calculated from the simulations by keeping track of which FCA the system was near (i.e. which ‘well’ the system was in), and the amount of time it spent while near that FCA. By dividing the total time spent in each well by the number of times that well was visited, we obtained the average waiting time near each stable FCA. The two waiting times were used in equation (4.11) as $T_+$ and $T_-$ to calculate the exponentially small eigenvalue.

The definition of one ‘visit’ to a well must be explicitly stated. The state $n_0 = \text{round}(x_0 V)$ was first used as the dividing line: if the system was greater than $x_0 V$, it was defined to be in the right-hand well, and if it was less than $n_0$, the left-hand well. However, it was observed during simulations that the system does not necessarily fall immediately into one well or another after passing over this point. Occasionally, the simulation would return to $n_0$ several times before heading towards one of the FCA. This behaviour caused the calculation of average waiting time to be underestimated (and $\lambda_1$ to be overestimated), because the number of visits was increased by one each time the state $n_0$ was reached.

Thus, when calculating the average waiting times, we assumed that the simulation had visited a well only when it had reached a particular point, closer to the FCA than just $n_0 \pm 1$. A study was done of the change in $\lambda_1$ with respect to the definition of this point. The eigenvalue displayed a fast, exponential-like convergence as the boundary moved away from $n_0$. For the rest of the simulations, we defined the point at which the system had officially entered a well to correspond to the inflection point in the potential function generated by integrating the cubic ODE.

In figure 11, we see agreement between the simulation predictions, the asymptotic approximations and the MATLAB’s calculations, and the exponential decay of $\lambda_1$ as volume increases. The values obtained through the Monte Carlo simulations and the MATLAB calculations are fairly similar. Although there is a small gap between these two lines, the slopes are very close, which indicates the same rate of predicted exponential decay with volume size.

However, the asymptotic predictions are much smaller than the other two, and the slope of the line is noticeably different. This is because the asymptotic equations are only valid for large volume systems. The simulations and the MATLAB calculations were carried out for relatively small volumes, because of computing restrictions in both cases.

An attempt to unify the asymptotic results with those of the MATLAB and the Monte Carlo was made. The slopes of the lines seen in figure 11 are related to the $\Delta u_\pm$ from equations (3.29) and (3.30). The constants $\Delta u_\pm$ are the asymptotic (continuous space) approximations to the barrier heights, $\Delta p_\pm^{as}$ from equation (3.28). Although it is too computationally intense to compute $\lambda_1$ for large $V$ using the eig command or a simulation, it is possible to compute $p_\pm^{as}$ from the detailed balance equation for larger $V$ using MATLAB. Thus, we can study the change in $\Delta p_\pm^{as}$ with volume. Figure 12 shows the convergence of the barrier heights to the asymptotic values $\Delta u_\pm$ as the system volume increases. We see the asymptotic prediction for the height differences in the probability steady state become accurate for volumes of the order of 100 and higher. Although the asymptotic line for $\lambda_1$ in figure 11 is much lower than the other two lines, figure 12 tells
us that eventually these lines should all converge as $V$ increases further.

We can obtain the MATLAB and the Monte Carlo data for values of $V$ in the approximate range $5 \leq V \leq 40$, and have shown that these two methods yield similar predictions for $\lambda_1$. We have also shown the range of $V$ on which the asymptotic prediction becomes realistic. Unfortunately, these two ranges of volume size do not overlap. Computational limitations do not allow us to confirm the MATLAB’s predictions with those made in Hänggi et al. (1984) and Hinch & Chapman (2005) or vice versa. However, a study of this kind can suggest which method is the best when working with a model of a real biochemical system, in which the volume size is known.

5. CONCLUSION AND FUTURE DIRECTIONS

Schlögl’s model is a simple example of a chemical reaction system that exhibits bistability. Bistable behaviour can be found in many biological networks, including the heart models (Hinch & Chapman 2005), the visual perception (Moreno-Bote et al. 2007), the gene networks (Paulsson 2005), etc. Owing to the ubiquity of switching behaviour, it is important to have comprehensive mathematical models of the bistable chemical reaction systems. In this paper, we used Schlögl’s model as an example to study the differences between the deterministic and the stochastic modelling techniques applied to bistability, and examine this behaviour through a thermodynamic perspective.

For a finite volume system, the steady-state behaviours predicted by the deterministic and the stochastic models can be drastically different. When the system is bistable, the behaviour in the deterministic model depends entirely on the initial condition, i.e. on which side of the unstable steady state the system started. On the contrary, the steady-state behaviour in the stochastic model is completely independent of the initial condition.

In the long term, the stochastic model predicts that the system spends almost all its time in the two FCA (states that correspond to the stable states of the ODE model), and the proportion of time spent in each is dictated by the ratio of the transition rates between them. Because this ratio increases exponentially with volume, as the volume size increases, the proportion of time the system spends in the more stable FCA increases as well. Thus, for a large but finite volume size, the stochastic models predicts that, independent of the starting point, the system will end up in the more stable FCA and spend most of its time there.

The parameters determine which of the FCA is more stable. Asymptotic analysis, which assumes a large range of $V$, we have shown that for small $V$, the FCA may exchange relative stability with the changes solely in $V$. This observation highlights the importance of using a CME model when addressing the intracellular reaction networks. The stability exchange between the FCA has been observed with changes in the pump parameter $b$ and the system volume. In the future, a full bifurcation diagram could be drawn so that the precise effects of $a$, $b$, and $V$ on the system’s stability are known.

When the system is bistable, there are multiple entropy production rates that come from the ODE model, one for each steady state. After calculating this rate in the master equation model, we see that the stochastic entropy production rate is a combination of the rates in each possible state, weighted by the probability of being in each of those states. Thus, the stochastic rate ‘follows’ that of the deterministic system for the more stable FCA (which is the most likely state of the system). This shows that the entropy production rate is neither maximized nor minimized for all NESSs. We conclude, as in (Landauer 1975), that the entropy production rate alone is unable to fully describe stability in the steady state. However, this example does not disprove the conjecture that the entropy production rate will take on some critical value (i.e. a minimum or a maximum) in the NESS. A similar study of a more complicated model, such as a tristable model, may also debunk this idea.

The long-term behaviour of the bistable system is determined by the transition rates between the FCA. For large volumes, the system can be simplified by a two-state Markov chain in which all states to the left of the unstable fixed point are lumped into one state, and all states to the right are considered as another. The transition rates are the entries of the four-by-four stochastic matrix modelling the jumping between the FCA, and while near one FCA, we can model the short time-scale behaviour with Gaussian noise.

The sum of these two transition rates is equal to the second largest (or the second smallest in magnitude) eigenvalue in the stochastic matrix that represents the full Markov chain. This eigenvalue can be calculated with some accuracy using the eig command in MATLAB. For small volume systems, the MATLAB result is more accurate than the asymptotic prediction, and it is faster than calculating the eigenvalue through the Monte Carlo simulations. However, a reliable method has not yet been found to compute this eigenvalue for all ranges of volume. Such a method will be crucial in predicting the time scales of a bistable chemical reaction system for mesoscopic volume sizes.

We thank Profs Ping Ao, Elliot Elson and Annie Lemarchand for their helpful discussions and Joshua Goldwyn and Pei-Zhe Shi for reading the manuscripts. The work is partially supported by the NIH grant no. GM068610.

APPENDIX A: FOKKER–PLANCK APPROXIMATION AND KEIZER’S PARADOX

In appendix A, we illustrate how the steady-state solution of the Fokker–Planck approximation to Schlögl’s model does not agree with the solution obtained through mathematical detailed balance. Let $P(k; t) = \Pr\{X(t) = k\}$, then we have

\[
\frac{dP(k, t)}{dt} = v_{k-1}P(k-1, t) + w_kP(k+1, t) - (v_k + w_{k-1})P(k, t),
\]

\[\text{(A1)}\]
where
\[ v_j = \frac{k_j a k (k - 1)}{V^2} + k_3 b, \]
\[ w_k = \frac{k_2 (k + 1) k (k - 1)}{V^3} + \frac{k_4 (k + 1)}{V}, \]
and \( V \) is the volume of the reaction system. The exact stationary solution to equation (A 1) can be found (Gardiner 1985) as
\[ P(k) = C_0 \prod_{j=0}^{k-1} \frac{v_j}{w_j}, \] (A 2)
where \( C_0 \) is a normalization constant satisfying \( \sum_{k=0}^{\infty} P(k) = 1 \). For large \( V \),
\[ \ln P(k) = \sum_{j=0}^{k-1} \ln \frac{v_j}{w_j} + C_1 \]
\[ = \sum_{j=0}^{k-1} \ln \frac{v(j/V)}{w(j/V)} + \Theta \left( \frac{1}{V} \right) + C_1 \]
\[ = V \int_0^{k/V} \ln \frac{v(z)}{w(z)} \, dz + C_1, \]
in which we introduce the dimensionless functions,
\[ v(z) = z^2 + \sigma, \quad w(z) = z^3 + \mu z, \]
and \( C_1 = \ln C_0 \). Therefore, in terms of the non-dimensionalized concentration \( u = k/V \), we have the probability distribution \( f(u) = VP(Vu) \):
\[ \frac{1}{2V} \ln \hat{f}(u) = \frac{1}{2V} \ln P(Vu) + C_2 \]
\[ = \frac{1}{2V} \int_0^u \ln \frac{v(z)}{w(z)} \, dz + \hat{C}, \] (A 3)
\[ = \frac{1}{2V} \int_0^u \ln \frac{z^2 + \sigma}{z^3 + \mu z} \, dz + \hat{C}. \] (A 4)

When \( V \) is large, the CME (A 1) is often approximated by a Fokker–Planck equation (Gardiner 1985). Let us introduce the finite-difference scheme
\[ f(u + \Delta u) - f(u - \Delta u) \approx 2 f'(u) \Delta u \]
\[ f(u + \Delta u) - 2 f(u) + f(u - \Delta u) \approx f''(u) (\Delta u)^2, \]
equation (A 1) can be approximated by
\[ \frac{\partial f(u, \tau)}{\partial \tau} = \frac{\partial}{\partial u} \left( \frac{u v(u) + w(u)}{2V} \frac{\partial f}{\partial u} - (v(u) - w(u)) f(u) \right), \] (A 5)
where \( f(u, t) du = VP(Vu, t) \). One notes that when \( V \to \infty \), the diffusion term vanishes. Thus, equation (A 5) is a first-order partial differential equation whose characteristics, the macroscopic deterministic dynamics for the Schlögl model, are similar to equation (2.3). This relation provides the law of mass action for the biochemical reactions with a rigorous Markovian stochastic basis: ‘The stochastic model is not an alternative to the deterministic kinetics, it is a more complete kinetic description which is capable of modelling reactions with and without fluctuations.’ (Qian & Elson 2002).

Equation (A 5) is an approximation for equation (A 1) in the limit of large \( V \). What has been neglected is the term of the order of \( V^{-3/2} \). The stationary distributions also can be analytically obtained for equation (A 5):
\[ \frac{1}{2V} \ln f(u) = -\int w(u) - v(u) \, du + C, \] (A 6)
where the constant \( C \) is determined by the normalization of \( f(u) : \int f(u) du = 1 \).
Are stationary distributions \( f(u) \) and \( \hat{f}(u) \) the same asymptotically? Both can be written in compact forms
\[ f(u) = C \exp \left\{ -2V \int_0^u \frac{q(z) - 1}{q(z) + 1} \, dz \right\}, \] (A 7)
\[ \hat{f}(u) = C' \exp \left\{ -V \int_0^u \ln q(z) dz \right\}, \] (A 8)
where \( C \) and \( C' \) are the normalization factors and \( q(u) = \frac{w(u)}{v(u)} \). Both functions have identical local extrema since
\[ \frac{d}{du} f(u) = 0 \quad \text{and} \quad \frac{d}{du} \hat{f}(u) = 0, \]
lead to the same equation for steady states \( q(u) = 1 \). Furthermore, the corresponding local maxima of \( f(u) \) and \( \hat{f}(u) \) are asymptotically equivalent in the limit of \( V \to \infty \). To show this, let \( u^* \) be a root of \( q(u) = 1 \). Then, we have
\[ f(u) = C \exp \left\{ -2V \int_0^{u^*} \frac{q(z) - 1}{q(z) + 1} \, dz \right\} \]
\[ - \frac{V}{2} q'(u^*)(u - u^*)^2, \] (A 9)
and
\[ \hat{f}(u) = C' \exp \left\{ -V \int_0^{u^*} \ln q(z) dz - \frac{V}{2} q'(u^*)(u - u^*)^2 \right\}, \] (A 10)
The quadratic terms in equations (A 9) and (A 10) are the same.

Now let \( u^* \) and \( u^{**} \) be the two roots of \( q(u) = 1 \) with \( q'(u^*) > 0 \), each represent an local maximum for the functions \( f(u) \) and \( \hat{f}(u) \). The important question, then, is which maximum is the larger one. This is determined by the signs of
\[ \int_{u^*}^{u^{**}} \ln q(z) \, dz \quad \text{and} \quad 2 \int_{u^*}^{u^{**}} \frac{q(z) - 1}{q(z) + 1} \, dz. \] (A 11)
A simple counter example can be constructed to show that the signs of expressions in (A 11) can be different.\(^5\) Therefore, in the limit of \( V \to \infty \), \( f(u) \) and \( \hat{f}(u) \) could converge to different local maxima.
Van Kampen repeatedly emphasized that the Fokker–Planck approximation can only be obtained for master equations with small individual jumps.\(^5\)

\(^5\)For example, using the parameter values \( k_1 = 3, k_2 = 0.6, k_3 = 0.25, k_4 = 2.95, a = 1, \) and \( b = 1.305 \), the two integrals have different signs.
A more sophisticated treatment of the diffusion approximation for master equation was given in terms of the $Q$-expansion (ch. 10, Van Kampen 1981). This theory provides a satisfying approximation for the stochastic relaxation in the limit of large $Q$ (our $V$). However, it does not address how to obtain a stationary distribution with multiple equilibria.

Both Hänggi et al. (1984) and Baras et al. (1996) pointed out the delicate issue of $V\to\infty$ and $t\to\infty$ and their non-commutative relation. If one takes the $V\to\infty$ first, one arrives at equation (2.3). If there are multiple stable steady states, the relative population is determined by the initial condition for the deterministic differential equation. How can one obtain a reasonable distribution of initial conditions for the various basins of attraction? Hence, it is clear that for deterministic differential equations, the initial distribution should be chosen according to the region of asymptotic stability.

The one-dimensional example indicates that even though a chemical reaction system is open, the mathematical model (CME) can still satisfy detailed balance, and its steady-state probability thus is completely solvable: $P_k/P_{k+1} = w_k/v_k$. For such a system, one immediately has a physical interpretation: $j_k = (v_k - w_k)$ as the ‘drift’ and $D_k = (v_k + w_k)/2$ as the ‘diffusion coefficient’ (Feller 1971). On the other hand, $A_k = k_bT\ln(w_k/v_k)$ can be viewed as ‘chemical affinity’ (De Donder & Rysellbergh 1936). Thus, there is a relationship between these three quantities $A_k$, $D_k$ and $j_k$:

$$j_k = 2D_k\tanh\left(\frac{-A_k}{2k_bT}\right) = 2D_k\tanh\left(\frac{-\phi_k - \phi_{k+1}}{2k_bT}\right), \quad (A12)$$

if we introduce $\phi_k = k_bT\ln P_k$. Equation (A12), which relates thermodynamic flux to force non-linearly, was introduced by J. Ross and his co-workers (Ross & Vlad 1999). This result is in complete agreement with T. L. Hill’s theory (Hill 1974, 1989) in terms of the forward and backward fluxes (Beard & Qian 2007). Note however, owing to the one-dimensional nature of the problem, that the $v$ and $w$ are not the forward and backward fluxes of a single chemical reaction.

REFERENCES


