THE UNIVERSITY *of* EDINBURGH

Concentration Oscillations in Single Cells: The Roles of Intracellular Noise and Intercellular Coupling

David L K Toner

Doctor of Philosophy
University of Edinburgh
2013

Declaration

This thesis is composed of original work based on the following published papers:

- **D L K Toner** and R Grima. Molecular noise induces concentration oscillations in chemical systems with stable node steady states. The Journal of Chemical Physics, 138(5):055101 2013

- **D L K Toner** and R Grima. Effects of bursty protein production on the noisy oscillatory properties of downstream pathways. Scientific Reports, 3:2438 2013

- Benedicte Wenden*, **David L K Toner*** Sarah K Hodge, Ramon Grima, and Andrew J Millar. Spontaneous spatiotemporal waves of gene expression from biological clocks in the leaf. Proceedings of the National Academy of Sciences of the United States of America, 109(17):675762, 2012 (*denotes equal contribution),

I declare that this thesis was composed by myself and that the work contained therein is my own, except where explicitly stated otherwise in the text. The work has not been submitted for any other degree or professional qualification.

David L K Toner, May 2013

Acknowledgements

I would first like to thank my principal supervisor, Dr. Ramon Grima, as well as my second supervisor, Prof. Andrew J. Millar, for all of their guidance and support throughout my time at Edinburgh University. Thank you to Ramon for the endless hours of scientific discussion on the content of this thesis, for helping me develop my abilities in scientific writing, and for encouraging and helping me in the process of submitting the work for publication. Most of all, thanks for always making time for me. Many thanks too to Benedicte Wenden, for an enjoyable and fruitful collaboration on 'the leaf stuff'. I would also like to thank my colleagues for sharing their time and knowledge with me over the last few years, particularly (and in no particular order) Philipp Thomas, Raffaele Basile, Dr. Natasha Savage, Prof. Peter Swain, Dr. Gerben van Ooijen, Dr. Laura Dixon, Dr Maria Luisa Guerriero, Rob Smith, Dr. Andrea Weisse and Dr. Anne Moore.

I would like to thank all of my new University friends at Edinburgh University for their support and for making the work environment such an enjoyable place (e.g., with cake). It has certainly been far from the typical 'basement isolation' picture of the lonely PhD student, and for that I am extremely grateful! Particular thanks to Dr. Natasha Savage (a pillar of wisdom and friendship at the good and less good times), Dr. Gerben van Ooijen (for being smiley, for probably hundreds of coffee breaks, for darts

and pints and for telling it like it is), Dr. Laura Dixon (aka Norah Bixon - especially for giving me my first warm welcome to the department...and Dolly Parton...) and Dr. Ralph Hector (for being a font of all musical knowledge). Indeed on that musical note, I would like to thank Brian Eno, Susumu Yokota and Stars of the Lid for making the most wonderful music for focusing the brain!

Outside of the University I have been fortunate to have been surrounded by good eggs who have helped to make these last few years a really enjoyable time in my life. Thank you especially to Nancy Burns and Vanessa Salambassi for creating the dream student flat, as well as my friends Marcello Leucci, Luke Humphry and Tom Ridgway for all of their encouragement along the way. A huge thank you to my Mum, Dad, Claire and Judy for being the best family in the world and for your ridiculous amount of love and support - I am so lucky to have you guys. Finally, thank you to lovely 'little Jenny Mullin' for giving me the *real* highlight of my PhD and agreeing to marry me! I can never thank you enough for your love, kindness and words of wisdom that have kept me going and made this PhD possible.

Abstract

Concentration oscillations are a ubiquitous characteristic of intracellular dynamics. The period of these oscillations can vary from few seconds to many hours, well known examples being calcium oscillations (seconds to minutes), glycolytic oscillations (minutes) and circadian rhythms (1 day). Considerable advances into understanding the mechanisms and functionality of concentration oscillations have been made since glycolytic oscillations were observed in the late 1950s, and mathematical methods have been an essential part of this process. With increasing ability to experimentally measure oscillations in single cells as well as in cell ensembles, the gold standard of modelling is to provide tools that can elucidate how the system-wide dynamics in complex organisms emerge from a system of single cells. Both abstract and detailed mechanistic models are complementary in the insight they can bring, and for networks of coupled cells considerations such as intrinsic intracellular noise, cellular heterogeneity and coupling strength are all expected to play a part.

Here, we investigate separately the potential roles played by intrinsic noise arising from finite numbers of interacting molecules and by coupling among cellular oscillators. Regarding the former, it is well known that internal or molecular noise induces concentration oscillations in chemical systems whose deterministic models exhibit damped oscillations. We show, using the linear-noise approximation of the chemical master

equation, that noise can also induce oscillations in biochemical systems whose deterministic descriptions admit no damped oscillations, i.e., systems with a stable node. This non-intuitive phenomenon is remarkable since, unlike noise-induced oscillations in systems with damped deterministic oscillations, it cannot be explained by noise excitation of the deterministic resonant frequency of the system. We here prove the following general properties of stable-node noise-induced oscillations for systems with two species: (i) the upper bound of their frequency is given by the geometric mean of the real eigenvalues of the Jacobian of the system, (ii) the upper bound of the Q-factor of the oscillations is inversely proportional to the distance between the real eigenvalues of the Jacobian, and (iii) these oscillations are not necessarily exhibited by all interacting chemical species in the system. The existence and properties of stable-node oscillations are verified by stochastic simulations of the Brusselator, a cascade Brusselator reaction system, and two other simple chemical systems involving autocatalysis and trimerization. We also show that external noise induces stable node oscillations with different properties than those stimulated by internal noise.

Having demonstrated and explored this non-intuitive effect of noise, we extend the work to investigate the phenomenon of noise induced oscillations in cellular reaction systems characterised by the 'bursty' production of one or more species. Experiments have shown that proteins are typically translated in sharp bursts and similar bursty phenomena have been observed for protein import into subcellular compartments. We investigate the effect of such burstiness on the stochastic properties of downstream pathways by considering two identical systems with equal mean input rates, except in one system molecules (e.g., proteins) are input one at a time and in the other molecules are input in bursts according to some probability distribution. We find that the stochastic behaviour falls in three categories: (i) both systems display or do

not display noise-induced oscillations; (ii) the non-bursty input system displays noise-induced oscillations whereas the bursty input system does not; (iii) the reverse of (ii). We derive necessary conditions for these three cases to classify pathways involving autocatalysis, trimerization and genetic feedback loops. Our results suggest that single cell rhythms can be controlled by regulation of burstiness in protein production.

We go on to investigate roles played by intercellular coupling in whole organ-level oscillations with an experimental analysis of circadian rhythms in *Arabidopsis thaliana*[†]. Circadian clocks in animals are known to be tightly coupled among the cells of some tissues, and this coupling profoundly affects cellular rhythmicity. However, research on the clock in plant cells has largely ignored intercellular coupling. Our research group used luciferase reporter gene imaging to monitor circadian rhythms in leaves of *Arabidopsis thaliana* plants, with both a lower resolution, high throughput method and a high-resolution (cellular level), lower throughput method. Leaves were grown and imaged in a variety of light conditions to test the relative importance of intercellular coupling and cellular coupling to the environmental signal. We analysed the high throughput data and described the circadian phase by the timing of peak expression. Leaves grown for three weeks without entrainment reproducibly showed spatio-temporal waves of gene expression, consistent with intercellular coupling. A range of patterns was observed among the leaves, rather than a unique spatio-temporal pattern, although some patterns were more often observed. All of the measured leaves exposed to light-dark entrainment cycles had fully synchronised rhythms, which could desynchronise rather quickly when placed in a non-entraining environment (i.e., constant light conditions). After four days in constant light some of these leaves were as desynchronised as leaves grown without any rhythmic input, as described by the phase coherence across the leaf. The same fast transition was observed in the reverse experimental scenario,

i.e., applying light-dark cycles to leaves grown in constant light resulted in full synchronisation within two to four days. From these results we conclude that single-cell circadian oscillators were coupled far more strongly to external light-dark cycles than to the other cellular oscillators. Leaves did not spontaneously completely desynchronise, which is consistent with a presence of intercellular coupling among heterogeneous clocks. We also developed a methodology, based on the notion of two functional spatial scales of expression across the leaf, to analyse the high-resolution microscope data and determine whether there is a difference in the phase of circadian expression between vein cells and mesophyll cells in the leaf. The result from a single leaf suggests that the global phase wave dominates the phase behaviour but that there are small delays in the veins compared to their nearby mesophyll cells. This result can be validated by applying the methodology developed here to new microscope leaf data which is currently being collected in the research group.

† *This work was performed as a collaboration between David Toner (DT) and Benedicte Wenden (BW). BW designed and carried out the experiments, DT performed the data analysis and led on data visualisation.*

List of Figures

List of Tables

Contents

Chapter 1

Introduction

1.1 Biochemical Oscillations

Concentration oscillations are a ubiquitous characteristic of intracellular dynamics [1]. The period of these oscillations can vary from few seconds to many hours, well known examples being calcium oscillations (seconds to minutes) [2], glycolytic oscillations (minutes) [3] and circadian rhythms (1 day) [4]. The molecular basis and functionality of the oscillations have fascinated both cell biologists and mathematicians for decades, and since the 1960s a close interplay between the two disciplines has driven forward understanding of the common and unique features of different oscillators [1]. The need for interdisciplinary research into biological rhythms is now well appreciated; researchers of systems and synthetic biology continue to make fast progress in unravelling the mysterious of rhythmicity across taxa, with an increasing focus on the ways in which organism-level oscillations arise from the interaction of noisy single cell mechanisms.

In this introductory chapter we consider the progress made in understanding single cell oscillations in the context of two very different types of oscillations; periodic

behaviour of glycolysis intermediates in yeast and the circadian expression of genes in mammals. Along the way we explain some of the different modelling methodologies which have been used and which feature in this thesis. To reflect the interdisciplinary nature of research into cellular oscillations we describe experimental observations and modelling approaches alongside one another.

1.2 Rhythmic Yeast - Glycolytic Oscillations

Glycolysis is the process which converts glucose to pyruvate by the concerted behaviour of a number of allosterically regulated enzymes, and has been studied extensively in yeast [5]. Concentration oscillations in this process were first observed in 1957, when Duysens and Amesz reported quickly damped oscillations of reduced nicotinamide dinucleotides in a suspension of intact yeast cells [6]. Later studies showed that the oscillations could be sustained over a greater number of cycles, with sustained oscillations being observed for more than 100 cycles when using cell-free extracts (Fig. 1.1, [7, 8]). Quickly, mathematical models were devised to reflect this phenomenon, one of the first being Sel'kov's simplistic model based on autocatalytic regulation [9]. Based on the observation that the enzyme phosphofructokinase (PFK) plays a crucial role in the oscillation, Sel'kov's approach was to use an ordinary differential equation (ODE) model for the concentrations of three biochemical species; the enzyme PFK, its substrate adenosine triphosphate (ATP) and its product adenosine diphosphate (ADP). Autocatalysis in the model is in the form of product activation of the enzyme. Goldbeter recognised a failing of the model to capture a key feature of the experiments, namely that the sustained oscillations should only occur when substrate input rate is within a certain critical range [1], which led to that author's more detailed model for glycolysis

based on allosteric regulation. While more complicated, this model was able to capture the experimentally observed dependency on substrate input rate [10, 11].

Even in this short sequence of events there are four key points to note regarding methodologies in biochemical oscillation research. Firstly, the construction of a mathematical model based on experimental data [9], followed by later critique and refining of the model [10], is central to the way in which biochemical oscillations are studied today. Secondly, we note the type of model used by Sel'kov - an ODE model for concentrations of the interacting species. This approach has been a standard approach for the many years of modelling biochemical oscillations since; we refer to this here and throughout as the rate equation approach. The construction of rate equations, and the requirements for oscillations are introduced below. Thirdly, the recognised presence of autocatalysis in glycolysis is also an important mechanism in other types of biochemical oscillations, and the mathematical reason underpinning this for two-species systems is well understood, as we shall show. Fourthly, the early glycolysis models were constructed to describe experimental results from continuously stirred yeast extracts, i.e., a biochemical scenario in which a vast number of interacting molecules interact in a spatially homogeneous environment. The validity of extrapolating modelling conclusions from such an experimental setup to the normal intracellular operation in single cells is a key point of this thesis, and we discuss its implications for understanding glycolysis below.

1.2.1 Construction of Rate Equation Models

Sel'kov's work in the late 1960's was certainly not the first example of quantitative approaches for investigating chemical reactions; rather these have been used from as early as the mid 19th century [12]. Wilhelmy [13] discovered in 1850 that the rate of

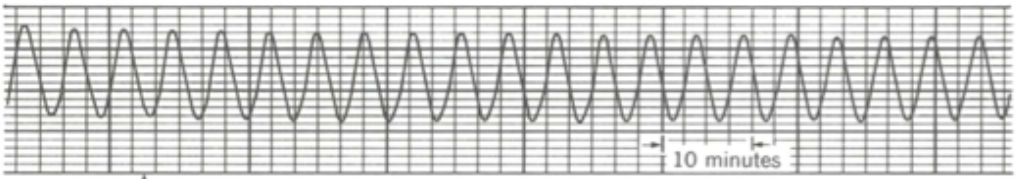


Figure 1.1: Reproduced from [8], the clearly sustained oscillations of diphosphopyridine nucleotide (DPNH) in a yeast cell-free extract.

change of sucrose concentration in the conversion of sucrose to glucose and fructose in an aqueous acidic solution was well described by an ODE, or rate equation. Constructing a rate equation model is straightforward; consider a general reaction system containing N distinct chemical species X_i which react according to R distinct reactions in a vessel of volume Ω. A particular reaction j can be written in the general form [14]:

$$\sum_{i=1}^{N} s_{ij} X_i \xrightarrow{f_j} \sum_{i=1}^{N} r_{ij} X_i, \tag{1.1}$$

where the s_{ij} and r_{ij} are the stoichiometric coefficients. (According to this formulation, if species X_i is not involved in reaction j as a reactant (product) then s_{ij} (r_{ij}) is simply set to zero.) By considering the combined effect of all R reactions of the above type on a particular species, X_i, one can construct the rate equation describing the rate of change of that species' concentration, ϕ_i [14]:

$$\frac{d\phi_i}{dt} = \sum_{j}^{J} S_{ij}\, f_j(\vec{\phi}) \ , \tag{1.2}$$

where we have introduced the net stoichiometric coefficient, $S_{ij} = (r_{ij} - s_{ij})$, i.e., the net gain of species i due to reaction j, and the macroscopic rate function $f_j(\vec{\phi})$. To describe multivariate systems one can write the set of rate equations for the vector of

concentrations, $\vec{\phi}$, in the matrix form;

$$\frac{d}{dt}\vec{\phi} = \mathbf{S}\,\vec{f}(\vec{\phi}), \tag{1.3}$$

where $\vec{\phi}$ is the concentration vector, $\mathbf{S}$ is the stoichiometric matrix with elements S_{ij} and $\vec{f}$ is the vector of macroscopic rate functions.

For 'elementary' reaction mechanisms such as unimolecular degradation or a bimolecular binding reaction, the forms of the rate functions $f_j(\vec{\phi})$ were empirically found, by Wilhelmy and in a series of works from 1864 by Guldberg and Waage [15], to be of a particularly simple form, where the rate of the reaction is proportional to the product of reactant concentrations, each raised to an integer power denoting its 'molecularity' or stoichiometry in the reaction. This type of rate law is commonly referred to as mass action kinetics, and for the general reaction scheme above is written:

$$\frac{d\phi_i}{dt} = \vec{g}(\vec{\phi}) = \sum_j^J S_{ij} \underbrace{k_j \prod_l^N \phi_l^{r_{lj}}}_{f_j(\vec{\phi})}\ , \tag{1.4}$$

where k_j is the reaction rate constant. Sel'kov's model was of such a description; a rate equation model with simple mass action kinetics. Goldbeter's full allosteric model [10] is a a partial differential equation model describing both glycolytic reactions as well as diffusion of metabolites, allowing for spatial as well as temporal description of the behaviour [16] but reduces to an ODE model [10, 11] when the authors eliminate diffusion by arguing that the equations become unstable with respect to spatially homogeneous perturbations before they become unstable with respect to diffusion.

1.2.2 Oscillations from Rate Equations - Limit Cycles and Hopf Bifurcations

Given the constructed rate equation models, the question of interest is 'under what intracellular conditions are oscillations observed?'. Answering this question requires a definition of what is and what is not an oscillation and various possibilities for defining what constitutes an oscillation have been put forward [17]. However, for the ODE approach it is generally the case that existence of oscillations are synonymous with existence of a stable limit cycle solution of the rate equations [1, 18, 19]. Unlike a stable 'fixed point' solution of the rate equations, where all trajectories in the phase space converge in time to constant steady-state values, the stable limit cycle describes a closed orbit in phase space to which all trajectories converge in the limit of long times. This reflects the phenomenon of a perfectly undamped oscillation, and the regular period of oscillation corresponds to the time taken to traverse the orbit. For the system of rate equations in Eq. 1.3, it can easily be shown that systems with two species or more are required for autonomous oscillations [18, 20]. The proof is that for the one-variable system $\frac{d\phi}{dt} = y(\phi)$ to have any sort of oscillatory behaviour it must have a solution curve $\phi(t)$ with an increasing part followed by a decreasing part (or vice versa), separated by a turning point. However, at any prospective turning point $\frac{d\phi}{dt} = y(\phi) = 0$; at that point the fixed point is reached and so $\phi(t)$ does not in fact alter trajectory but rather remains constant in time [20]. For two or more species the system dynamics are far more complex, and determining the existence of limit cycle solutions in a general high-dimensional system is typically only possible through numerical methods. The standard strategy is to perform linear stability analysis, by which one derives an equation for the time evolution of a small perturbation about the

steady state of the system. The Jacobian matrix, **J**, describes this linear stability, and can be constructed directly from the rate equations, with its elements given by [21]:

$$J_{ij} = \frac{\partial}{\partial \phi_j} \sum_{r=1}^{R} S_{ir} f_r. \tag{1.5}$$

The eigenvalues, α_i, of **J** evaluated at the steady state, contain vital information into the behaviour of the system. For a stable system all eigenvalues must have negative real parts, and a damped oscillatory return to the steady state is characterized by a pair of complex conjugate eigenvalues $\alpha_{1,2} = -a \pm ib$, where a and b are positive valued. One of the main techniques for determining the presence of limit cycles is provided by the Hopf bifurcation theorem [18]. This says that as a pair of complex conjugate eigenvalues $\alpha_{1,2} = a \pm ib$ cross the imaginary axis, i.e., a goes positive due to the alteration of some system parameter, λ, the steady state loses stability in such a way that the system passes through a supercritical Hopf bifurcation and a limit cycle is created (Fig. 1.2). This notion pervades the literature of differential equation modelling of glycolysis (e.g., [22]).

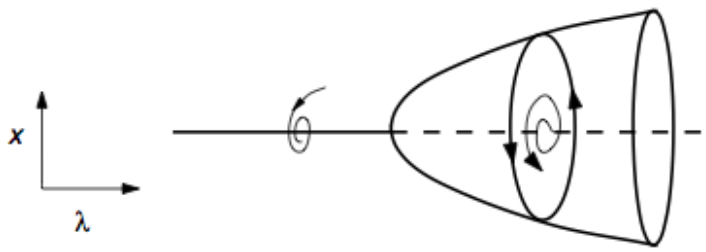


Figure 1.2: Reproduced from [18]. As the control parameter λ increases, the dynamical system transitions from monostable behaviour to limit cycle behaviour through the supercritical Hopf bifurcation.

For a two species system, Bendixson's negative criterion for a limit cycle is particularly useful. The leading diagonal entries $\frac{dg_1(\vec{\phi})}{d\phi_1}$, $\frac{dg_2(\vec{\phi})}{d\phi_2}$ of the 2×2 Jacobian matrix are typically negative in biochemical reactions, simply because reactions generally tend to consume the substrate. However, Bendixson's criterion states that if $\frac{dg_1(\vec{\phi})}{d\phi_1} + \frac{dg_2(\vec{\phi})}{d\phi_2}$ is of a constant sign in some region of the phase space then there can be no periodic solution in that region; what is required is that one of the terms $\frac{dg(\vec{\phi})}{d\phi_i}$ be positive. This can only be brought about by an autocatalytic reaction mechanism. Goldbeter [1] and Tyson [18] point out that such mechanisms are generally rare in biology, but that their presence in different biochemical oscillations is notable.

In larger systems without autocatalysis, a key component of oscillations is the presence of negative feedback [18], and this formed the basis of the famous Goodwin oscillator [23]. More recently, the general principles underlying limit cycle behaviour in biochemical oscillations were collated [19], pointing at conditions which are achieved when the system possesses a negative feedback mechanism with sufficient memory, sufficient nonlinearity in the reaction rates, and a proper balancing of the timescales of species involved in the feedback loop. It has been shown that the minimal truly realistic chemical system, i.e., that composed of a set of unimolecular and bimolecular reactions, which satisfies these properties involves the interaction of three species [24].

1.2.3 The Validity of Glycolysis Rate Equation Models for Single Cell Behaviour

The models described above [11, 9], based on the oscillatory properties of cell-free yeast extracts, describe a spatially homogeneous macroscopic 'soup' of the important biochemical species in glycolysis. This quantitative modelling was important for understanding the autocatalytic process in glycolysis (autocatalytic because two ATP

molecules are 'invested' in the reaction and four ATP molecules produced [5]). However, the experimental setup is not the same as either a population of intact cells or the dynamics of a single yeast cell in isolation (essentially the natural scenario for yeast) [22]. For a suspension of intact yeast cells, obtaining a sustained macroscopic oscillation was found to be highly dependent on the growth phase [25] as well as on the density of cells in the suspension [26]. In experiments on single cells, it was found that after macroscopic oscillations had died out, the oscillation remained in some cells [26]. The picture of macroscopic oscillations in yeast glycolysis was then concluded to be a result of synchrony between single cell oscillators, and the coupling species concluded to be acetaldehyde [27]. It has been suggested that such behaviour could have been the result of evolutionary tendency from unicellular to multicellular behaviour of yeast [22].

Despite the progress made in understanding single cell and macroscopic oscillatory behaviour, the extent to which oscillatory behaviour is determined by intracellular and extracellular cell mechanisms continues to be questioned [5, 28, 29]. It was recently highlighted that an open research question remains as to how common sustained oscillations are in single cells [5]. Very recently, Gustavsson and coworkers used microfluidic approaches to probe the question of whether single cell behaviour was captured well by the population's mean behaviour [29]. They found that sustained oscillations can be observed in single cells in isolation, but notably that the oscillations are heterogeneous, (Fig. 1.3). Intriguingly, the mathematical model built to describe the experimental result was again an ODE model, with heterogeneity between cells introduced by way of varying the activity of one of the enzymes (the authors concede that the variability could come from other sources). Since a deterministic model was used, the experimentally observed oscillations in some single cells (Fig. 1.3) was described as 'limit cycle'

type oscillations. The notion that clear, sustained, heterogeneous oscillations among isolated cells requires a limit cycle in the mathematical model is, however, known not to be true [30]. To understand this point first requires a discussion on the circumstances in which we can expect an ODE model to faithfully describe the reaction mechanisms inside a single cell, which has been an area of increasing research in recent years. A key point from our perspective is that ODE models are deterministic, i.e., given the same initial conditions, the model always gives the same solution. The implications for modelling of single cells is that two uncoupled cells starting in the same conditions would be predicted by an ODE model to behave in exactly the same way over time. Since this is not the case in the experiments of Gustavsson et al [29], they circumvent this problem by an ad hoc addition of variability, or noise, into the differential equations. However, there is an alternative to such an approach for modelling cellular heterogeneity, founded on rigorous physical principles, namely the approach of stochastic chemical kinetics.

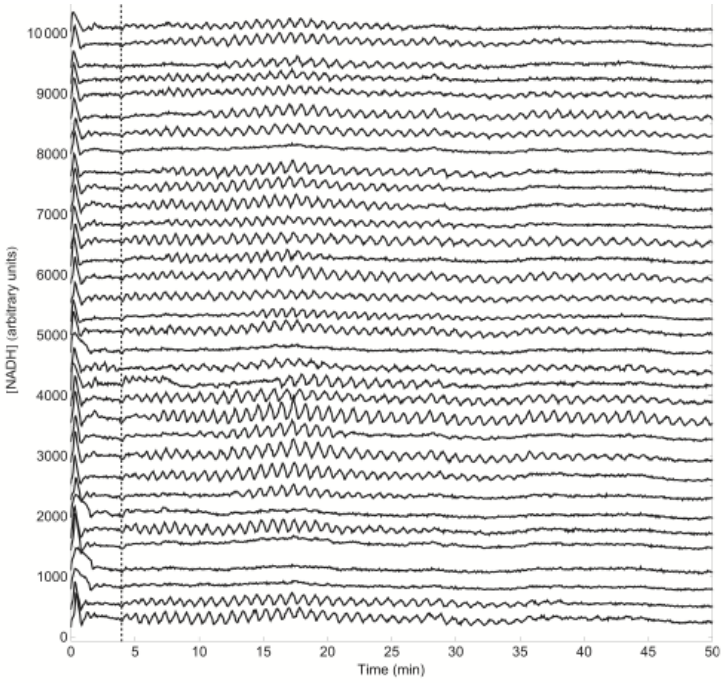


Figure 1.3: Reproduced from [29]. Heterogeneous single cell oscillations of Nicotinamide adenine dinucleotide (NADH) in yeast cells (*Saccharomyces cerevisiae* X2180), as measured by fluorescence in a microfluidic device.

1.3 Stochastic Chemical Kinetics

Whether trying to understand the motion of a cannonball, the dynamics of a weather system, the flow of electricity in a circuit or the motion of pollen grains suspended in water, physicists have sought to find an appropriate scale, or level of abstraction, within which to frame their thinking. In the early 20th century Einstein [31] and, separately, Smoluchowski [32], were only able to make progress on understanding the erratic Brownian motion by working at a new and ingeniously pragmatic scale. A deterministic, macroscopic equation of motion for a pollen grain was obviously inappropriate, given Brown's 'failure' after great efforts to prevent the pollen grains moving erratically. A

deterministic microscopic model, comprising equations of velocity and momentum for each solvent molecule, along with a vast array of initial condition combinations, would have firstly required unavailable computational power and secondly given an output so complex that a useful interpretation might be lost. Instead, Einstein reduced the enormous complexity of the motion of all of the solvent molecules to a probabilistic system of extremely frequent, independent collisions [33], and thus began an intermediate scale of physical modelling, the *mesoscopic* scale, which describes macroscopic variables as stochastic (random) functions of time by the careful elimination of microscopic variables [14].

Arguably the first significant use of stochastic modelling concepts from Brownian motion in the field of chemical kinetics was Kramers' escape problem [34]. Another important early work is Delbrück's investigation of a stochastic model of autocatalysis [35], where he found that the molecule number fluctuations in his simplified first order reaction were of the order of the square root of the number of reactant molecules. Just a few years later, in his classic essay 'What is Life?', Schrodinger was to refer to this as the 'so-called $\sqrt{n}$ law' which describes 'the degree of inaccuracy to be expected in any physical law' [36]. For a Poisson process, the $\sqrt{n}$ scaling can be shown to be exact [14]. Apart from these early studies, the stochastic approach in chemical kinetics really only gained significant application in the 1950s and 1960s [37, 38, 39] when researchers realized that reactions which had exhibited large fluctuations and had been seemingly 'irreproducible' - a word which hints at the same experimental frustrations experienced by Brown in the 1800s - could be quantitatively modelled with a stochastic approach [38]. Indeed by the early 1960s Bartholomay was able to confidently summarize the importance of the stochastic approach, when he stated that the 'basic tenet' of stochastic modelling of chemical kinetics is that 'all reactions are,

to some extent, irreproducible' [40].

Essentially for us, among the observations made in this period was the prediction that one of the major areas of application for the stochastic approach was for modelling biochemical reactions in intracellular regions [39]. Understanding this deduction, i.e., that chemical reactions in test tubes and in single cells should be treated with a fundamentally different approach, all comes down to a simple matter of scale. In a test tube volume containing reactants with molar concentrations, the number of discrete molecules are of the order of 10^{23}, and as such may be very well approximated as a continuum. In this case fluctuations in concentrations caused by individual molecular events are negligible, and the mass action rate equation formalism is appropriate. By this basis, ODE modelling of macroscopic oscillations in yeast cell free extracts is well justified. However, for typical concentrations of biochemical reactants within femtolitre volumes of single cells, the number of copies of reacting species can be of the order of tens to thousands [41, 42], or even down to single molecules in gene expression models. Here, the continuum assumption is invalid - the unpredictable nature of collisions between discrete molecules has a large effect on the concentrations and can thus not be ignored. In such instances deterministic approaches may agree well with the stochastic approach [43], but this cannot be assumed a priori and must generally be investigated on a model-by-model basis. It is important to note that the stochastic chemical kinetics approach is not an alternative to deterministic modelling, but rather that it 'contains' the macroscopic deterministic approach, in the sense that in the thermodynamic limit the two agree [44]. For this reason, stochastic modelling can be seen to be the more rigorous and detailed methodology. One of the most concrete arguments for considering stochasticity is that in some cases the 'details' do not only contribute small, predictable effects to deterministic conclusions, but rather that the qualitative

result from the stochastic model is completely different.

1.3.1 Assumptions of Spatial Homogeneity and 'Well-Mixedness'

Both of the approaches - deterministic and stochastic - assume that the reaction volume is 'well-stirred' and spatially homogeneous. The stirring could either be a direct, convective mechanism of stirring, such as in the experiments on yeast cell free extracts, or rather by natural diffusive stirring of the reaction volume inside cells. In the latter case, the requirement for well stirred reactions can be made more formal by considering a length scale of diffusion for a molecule in its lifetime, sometimes called the Kuramoto length [42, 14], which must be much larger than the length scale of the compartment volume. Framed differently, the diffusive stirring requires that the number of non reactive collisions between particles must be much greater than the number of collisions which result in reactions [45]. It has been noted that when the homogeneity assumption does not hold, one could take the approach of subdividing the reaction volume into smaller volume elements of length δL until it does. Interestingly, by making such a modification the applicability of continuity within the small volume element is even more questionable [42] or, put differently, ensuring homogeneity 'comes at the expense of the first assumption [continuity]' [46].

A further complication for spatial homogeneity is due to the fact that cells are highly organised spaces, with different cellular compartments performing different roles. Moreover, cells are thought to be crowded environments which limits the free diffusion of molecules [47, 48, 49]. Some ODE models have attempted to consider compartmentation, for example in models of circadian rhythms transport between nucleus and cytosol often features to introduce a necessary delay for the rhythm [50, 51, 19], but here the compartmentation amounts to simply labelling species differently in different

compartments, and does not consider the sizes of compartment volumes. Stochastic compartment models do require to account for volumes of compartments, and some work has recently been done on this [52]. Far less has been done regarding the combined effects of macromolecular crowding and reaction stochasticity. The work in this thesis does not consider either compartmentation nor macromolecular crowding, and yet even without these contributions to the system complexity we find non-intuitive roles of stochasticity in oscillations in well-mixed cells.

1.4 Oscillations in Stochastic Systems

1.4.1 Describing Stochastic Processes

The solution of an ODE model describes the deterministic (certain) variation in time of concentration of one or more biochemical species, say $X^d(t)$, where the d denotes deterministic. In contrast, the solution of a stochastic model represents the stochastic (uncertain) variation in time of the species, $X(t)$. Particularly, the full representation of the stochastic process $X(t)$ requires the joint probability distribution of $\{X(t_1), X(t_2), X(t_3) \dots X(t_m)\}$ for all choices of m and all possible values of $t_1, t_2, \dots, t_m$ [53]. It is immediately apparent that determining the properties of stochastic processes in practice could be a formidable challenge, but the problem is greatly simplified by considering a special class of stochastic processes which are stationary, ergodic and Markovian. Stationarity means that the statistical properties of the process do not vary in time; for example the distribution of the random variable $X(t)$ is equal to the distribution of $X(t + t')$, moreover the joint distribution of $\{X(t), X(t + \tau)\}$ is equal to the joint distribution of $\{X(t + t'), X(t + t' + \tau)\}$. What this means on a practical level is that the mean $\langle X \rangle$, the variance $\langle\langle X^2 \rangle\rangle$ and higher moments do not vary

in time, but more than this that the autocorrelation $\langle X(t+t')X(t+t'+\tau)\rangle$ is time invariant. Markov processes are often described as being 'memoryless' in that transition probabilities from one state to another do not depend on the full history of the process but only on the current state. Formally, the conditional probability of being in the state n_m at time t_m given the history $\{n_0, t_0; n_1, t_1; n_2, t_2, ...n_{m-1}, t_{m-1}\}$, i.e. $P(n_m, t_m|\{n_0, t_0; n_1, t_1; n_2, t_2, ...n_{m-1}, t_{m-1}\})$, can be expressed as simply $P(n_m, t_m|n_{m-1}, t_{m-1})$ [14]. Ergodicity is concerned with the extent to which one can equate ensemble averages of a stochastic process with time averages, and the ergodic theorem is a powerful result enabling inference of the properties of stochastic process from real experimental data. Consider that we are interested in finding the mean, $\langle X(t)\rangle$, of a stationary process. We could collect L samples or 'realizations' of this process, x_l, and estimate the average from $\langle X(t)\rangle = \frac{1}{L}\sum x_l(t)$. For this to yield a good estimate L must be sufficiently large, but sometimes only a single sample of a process is available. For stationary processes, the ergodic theorem states that as long as the autocorrelation $\langle X(0), X(\tau)\rangle$ tends to zero as $\tau \to \infty$ then ensemble averages and time averages coincide; this means that one can instead calculate moments using time averages, e.g., $\langle X(t)\rangle = \lim_{T\to\infty} \frac{1}{T}\int_0^T X(t)dt$ [33, 54].

1.4.2 The Fluctuation Power Spectrum

The natural starting point of identifying presence of oscillatory behaviour in a deterministic process is by way of the Fourier transform. This is also the foundation of identifying oscillations in stationary stochastic processes, as the use of Fourier transforms allows a description of how the variance $\langle\langle X^2\rangle\rangle$ is distributed with frequency [55].

This is the so called fluctuation power spectrum, and is defined by Gardiner [56] as

$$S(\omega) = \lim_{T\to\infty} \frac{1}{2\pi T} \left|\tilde{X}(\omega)\right|^2, \tag{1.6}$$

where

$$\tilde{X}(\omega) = \int_0^T X(t) e^{-i\omega t}\, dt. \tag{1.7}$$

Although succinct, these equations gloss over important subtleties, and raise interesting questions. Why is the truncated Fourier transform used? How does one integrate the stochastic process? We believe that it is enlightening and helpful at this point to present Priestley's simple development of the spectrum [53]. As well as providing intuition as to the meaning of the power spectrum, the methodology also naturally connects with our method of generating numerical power spectra using the stochastic simulation algorithm later.

Priestley considers a single sample path, or 'realization', $x(t)$ of the process. Fourier transforming this sample path over all time is not possible because $x(t)$, being a sample from a stationary process, does not decay at infinite times and is therefore not absolutely integrable. A new sample path $x_T(t)$ is defined, which is identical to $x(t)$ over the interval $[-\frac{T}{2}, \frac{T}{2}]$ and zero everywhere else; this $x_T(t)$ is absolutely integrable, and one can define the Fourier transform pair

$$x_T(t) = \frac{1}{\sqrt{2\pi}} \int_\infty^\infty \tilde{x}_T(\omega)\, e^{i\omega t}\, d\omega, \tag{1.8}$$

$$\tilde{x}_T(\omega) = \frac{1}{\sqrt{2\pi}} \int_\infty^\infty x_T(t) e^{-i\omega t}\, dt. \tag{1.9}$$

With this definition, $|\tilde{x}_T(\omega)|^2\, d\omega$ describes the proportion of the total (finite) energy of the signal $x_T(t)$ in the frequency interval $(\omega, \omega + d\omega)$. However, we are interested in

describing the frequency content of the full signal $x(t)$. It is not sensible at this point to take a limit of $T \to \infty$, because the limit $\lim_{T\to\infty} |\tilde{x}_T(\omega)|^2 \to \infty$, i.e., the energy of $x(t)$, is not finite. Rather, the power of the signal in the time window, i.e., $\frac{|\tilde{x}_T(\omega)|^2}{T}$ is the natural candidate for describing the process in the limit $T \to \infty$. In this case, $\lim_{T\to\infty} \frac{|\tilde{x}_T(\omega)|^2}{T}\, d\omega$ describes the proportion of the total power of $x(t)$ in the frequency interval $(\omega, \omega+d\omega)$. Just as $x(t)$ is a sample realisation in time of the stochastic process, this new quantity is a sample spectrum of the process in frequency [55], and so Priestley finally defines the spectrum, or power spectral density, as

$$S(\omega) = \lim_{T\to\infty} \left[\frac{1}{T} \left\langle |\tilde{x}_T(\omega)|^2 \right\rangle \right], \tag{1.10}$$

which is interpreted such that $S(\omega)d\omega$ is the average, over all possible realisations of $X(t)$, of the proportion of total power of $x(t)$ within $(\omega, \omega + d\omega)$.

Priestley's and Gardiner's spectrum equation have the same meaning, and are similar in their presentation. Notable differences between the two final results are the difference by a factor of 2π in the denominator, due to the customary difference among authors of factors of 2π, and the explicit averaging shown in Priestley's equation, emphasising that the fluctuation power spectrum is a statistical quantity which describes the expected spectrum of an individual realization of the process.

The Meaning of Different Shapes of Spectra

Examining the fluctuation power spectra of stochastic processes is a field known widely as spectral analysis, and has been for decades in disciplines as diverse as astronomy and meteorology [57], oceanography [58] and economics [59]. To consider the meaning which can be extracted from spectral analysis, we consider four qualitatively different

types of spectrum.

- The spectrum which is perfectly flat in frequency represents white noise, a stochastic process which contains all frequencies with equal power. White noise is a mathematically useful tool but a physically unrealisable concept, as it requires that a value taken by a measurable fluctuating quantity at time t has no similarity whatsoever with its value at $t + \delta t$ even when δt is vanishingly small. Put differently, the autocorrelation time of the process is precisely zero.

- A spectrum which decays monotonically with ω can describe real processes, and the speed with which the spectrum decays in frequency is directly relatable to the autocorrelation time of the process. Very fast decay of $S(\omega)$, i.e., a spectrum with most of its power at very low frequencies, corresponds to a process with very long autocorrelation time, whereas slow decay of $S(\omega)$ corresponds to short autocorrelation times.

- Most relevant for discussion of oscillations in stochastic processes is the presence of one or more peaks in the power spectrum at a positive value of ω. A spectrum with a single delta spike at ω_1 represents a perfect (deterministic) oscillation at that frequency, with no amplitude or phase modulation.

- More probable than a delta spike for a stochastic process is a spectrum with a smooth peak centred at ω_1. This is taken to represent quasi-periodic behaviour [55, 57] with a characteristic frequency of ω_1; a sharp peak corresponds to a situation in which there is an oscillation with well-defined period and little variability in the phase [60] and amplitude.

1.4.3 The Connection between the Fluctuation Power Spectrum and Single Cell Dynamics

When noise is non-negligible each realization of the stochastic process provides the behaviour of a particular cell in the ensemble [61] and the power spectrum thus provides information of the rhythmicity present at the single-cell level. Individual realizations of the process are, of course, independent and so the power spectrum represents the expected rhythmicity in a single cell operating either in isolation or within an ensemble of dynamically uncoupled cells.

1.5 Chemical Master Equations

Assuming well-mixed conditions, the state of the mesoscopic system at any time t is completely described by the vector of the number of molecules of each species, $\vec{n} = (n_1, n_2, ..., n_N)$. At random times, one of the R reactions (say reaction j) occurs somewhere in the volume Ω and the state vector changes from some $(n_1, n_2, ..., n_N)$ to $(n_1 + S_{1j}, n_2 + S_{2j}, ..., n_N + S_{Nj})$ where $S_{ij} = r_{ij} - s_{ij}$. Whereas rate equations fully describe the macroscopic behaviour in the thermodynamic limit, the statistical properties of the stochastic process are completely described by chemical master equations [14, 62], which are differential-difference equations for the probability of being in state $\vec{n}$ at time t, conditioned on the probability $P(\vec{n_0}, t_0)$.

$$\frac{\partial P(\vec{n}, t|\vec{n_0}, t_0)}{\partial t} = \sum_{j=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ij}} - 1 \right) a_j(\vec{n})\ P(\vec{n}, t|\vec{n_0}, t_0). \tag{1.11}$$

$P(\vec{n}, t|\vec{n_0}, t_0)$ is the probability of the system being in state $\vec{n}$ at time t given that the system was in state $\vec{n_0}$ at time t_0, $dt\, a_j(\vec{n}, \Omega)$ is the probability that a single j^{th} reaction

occurs in the next time interval $[t+dt)$ and E_i^x is a step operator whose action on some function of n_i is to change it to a function of $n_i + x$ [14]. The term a_j is a transition probability per unit time [14], and is often referred to as a propensity function [62].

An example of a chemical master equation for the simple reaction $\emptyset \rightarrow A, A \rightarrow \emptyset$ is given in Appendix A.1. Master equations describing the simplest biochemical reactions schemes, i.e., those involving only linear (unimolecular) reactions, can be solved exactly using moment generating functions and the method of characteristics [14]. For more realistic reaction schemes it is important to consider bimolecular reactions, and in these cases exact solutions are not possible for all but the simplest of schemes (e.g. [63]). The reactions investigated by researchers such as McQuarrie were dictated to a large extent by those that were solvable. Indeed, Gillespie criticised the work in that period for its focus on solving the master equations rather than on the physical meaning of the constructed master equations [64], principally, what is the correct form for the propensity functions? A critical development for the application of stochastic chemical kinetics was Gillespie's seminal work in 1976 [45], which both elegantly described how the propensities a_j of bimolecular reactions can be motivated from physical arguments and provided a method of simulating exact realizations from the stochastic process [45, 65] - exact in the sense that at each time t one obtains a 'stochastically unbiased state of the chemical system' [45]. (Indeed, this exactness is in contrast to numerical methods for ODEs.)

Gillespie used arguments based on thermodynamics and Newtonian physics to motivate a form of the propensity function which he writes as $a_j = h_j\, c_j$, where the reaction parameter c_j (also referred to as the stochastic rate constant [65]) is defined such that $c_j\, dt$ is the 'average' probability (to first order in dt) that a particular combination of the reactant molecules in reaction j will react in the next time interval. h_j is simply

the number of different possible combinations of reactant molecules. Gillespie claims that the requirement that we can find a physically meaningful c_j is the fundamental hypothesis of a valid formulation of stochastic chemical kinetics, and uses simple physical and thermodynamical arguments to show how c_j for a bimolecular reaction depends on the radii of the interacting molecules and their average relative velocity (which is a function of temperature and molecular masses). Because these quantities are physically measurable, the propensity functions can in principle be accurately determined.

1.5.1 Dealing with the Intractability of the Chemical Master Equation

These master equations are typically unsolvable except in special cases (see for example [66]) or for the case where all processing reactions are first-order [67], a very restrictive assumption given that most interactions inside a cell involve the binding of two molecules. The most common approach is to use Gillespie's Stochastic Simulation Algorithm [45], a method which draws exact sample paths from the underlying stochastic process. Monte Carlo sampling generates the random numbers τ and j described by the equation:

$$p(\tau, j|\vec{n}, t) = a_j(\vec{n})\, e^{-\sum_j^J a_j(\vec{n})\,\tau}, \tag{1.12}$$

which is the probability, given the system is in state $\vec{n}$ at time t, that reaction j will occur (immediately and instantaneously) after a waiting time of τ [62].

This method is incredibly useful, due to the ability to draw unbiased trajectories from the underlying distribution, but suffers the following drawbacks:

1. A simulation study can be very computationally expensive, because many realizations are usually required to appreciate the nature of the underlying probability

distribution. This becomes especially relevant when conducting parameter scans of models.

2. It can be difficult to formally relate the model structure and parameters to the simulation output - analysis is more useful than simulation for this.

Alternative to stochastic simulation are approximate analytical methods, which have the principal advantages of both affording greater insight into the stochastic process and not requiring high computational expense. Some methods of approximation including replacing the CME by a Fokker-Planck equation [14], others use moment closure techniques such as derivative matching [68] or the so-called $2MA$ equations [69]. However, Van Kampen [14] cited the "unreliable and contradictory" results obtained by *ad hoc* moment closure methods and formulated his famous and widely used [30, 70, 71, 72] System Size Expansion.

1.5.2 The System Size Expansion - a Systematic Approximation Method for Analysis of CMEs

The principle problem with the determination of the power spectrum from the CME lies in the fact that one cannot generally derive a closed form equation for the autocorrelation (and more generally for the moments) using the chemical master equation and hence an approximation method is needed. One means to circumvent this inherent analytical intractability is a technique popularly referred to as the system-size expansion [14]. We summarize below the calculation of the fluctuation power spectrum to the lowest order of approximation, i.e., the linear noise approximation. This approximation is valid for an arbitrarily complex monostable reaction system provided the fluctuations about the mean concentrations are quite small.

$$\frac{dp(\vec{n},t)}{dt} = \Omega \sum_{j=1}^{J} \left(\prod_{l=1}^{L} \mathbb{E}_l^{-S_{lj}} - 1 \right) a_j\left(\vec{n},\Omega\right) p(\vec{n},t). \tag{1.13}$$

This System Size Expansion is a systematic and consistent expansion of the chemical master equation (Eq. (1.13)) as a series in powers of the inverse square root of the system size, Ω. For chemical master equations, Ω is taken to represent the volume of the reaction compartment. This approximation method uses the rationale that as the number of reactant species increases, the fluctuations in the system can be expected to become relatively small.

The fundamental ansatz is that we can make the transformation of variables:

$$\vec{n} = \Omega\vec{\phi}(t) + \Omega^{\frac{1}{2}}\vec{\xi}, \tag{1.14}$$

which says that we *expect* the number of molecules in the system to be approximately equal to the macroscopic value $\Omega\phi(t)$ as obtained from rate equations plus a fluctuation term which scales approximately with the root of the system size. This rule can be rationalised by considering a Poisson process, for which the standard deviation of the fluctuations exactly matches $\sqrt{n}$ [14]. Note that while Ω represents the reaction volume, it is used a control parameter for the system size in the sense that an increase in Ω is made with the concentration held constant, so that the number of molecules n increases. The fluctuation variable $\vec{\xi}(t) = \Omega^{1/2}(\frac{\vec{n}(t)}{\Omega} - \vec{\phi})$ describes the stochasticity in the system.

Our full explanation of the System Size Expansion is given in A.2. There, we show that to the first order of approximation, known as the Linear Noise Approximation, the System Size Expansion converts the mathematically intractable CME (Eq. (1.13)) in

discrete space to the approximating linear Fokker-Planck equation in continuous space:

$$\frac{\partial \rho(\vec{\xi}, t)}{\partial t} = \sum_{i,j} J_{ij} \frac{\partial (\xi_j \rho)}{\partial \xi_i} + \frac{1}{2} \sum_{i,j} D_{ij} \frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j}, \tag{1.15}$$

where $\rho(\vec{\xi}, t)$ is the probability of the system being in state $\vec{\xi}$ at time t (again, conditioned on the system being in some state state $\vec{\xi_0}$ at time t_0), J_{ij} are the elements of the Jacobian matrix of the macroscopic system of equations, $\mathbf{J}$, and D_{ij} are the elements of the diffusion matrix $\mathbf{D}$. The linear Fokker-Planck equation is equivalently represented as the Langevin stochastic differential equation:

$$d\vec{\xi}(t) = \mathbf{J}\vec{\xi}(t)\, dt + \mathbf{B}\, d\vec{W}(t), \tag{1.16}$$

where $d\vec{W}(t)$ is an N-dimensional Wiener process. The matrices $\mathbf{J}$ and $\mathbf{D} = \mathbf{B}\mathbf{B}^T$ are the same Jacobian and diffusion matrices as in Eq. (1.15). These matrices can be constructed directly from the rate equations and from the stoichiometric matrix; their elements are given by [21]:

$$J_{ij} = \frac{\partial}{\partial \phi_j} \sum_{r=1}^{R} S_{ir} f_r, \tag{1.17}$$

$$D_{ij} = \sum_{r=1}^{R} S_{ir} S_{jr} f_r. \tag{1.18}$$

1.5.3 Information from the Linear Noise Approximation

The solution of the linear Fokker-Planck equation is a multidimensional Gaussian distribution, so it suffices to determine the time evolution of the first and second moments. The procedure for producing these equations is straightforward: for example, to find

the rate of change of the mean of ξ_i we use:

$$\frac{d\langle\xi_i\rangle}{dt} = \int \xi_i \frac{\partial\rho(\vec{\xi},t)}{\partial t} d\vec{\xi} \tag{1.19}$$

and for the covariance between variables ξ_i and ξ_j we use:

$$\frac{d\langle\xi_i\xi_j\rangle}{dt} = \int \xi_i\xi_j \frac{\partial\rho(\vec{\xi},t)}{\partial t} d\vec{\xi}. \tag{1.20}$$

By substituting the Fokker-Planck equation (1.15) into Eqs. (1.19) and (1.20) it is found that the means $\langle\xi_i\rangle$ follow the same dynamics as the linearized system of macroscopic variables, and tend to zero at the stationary state. The implication is that to the linear noise approximation order, i.e., in the limit of large volumes, the time-evolution equations for the mean concentrations predicted by the chemical master equation are the same as the conventional rate equations:

$$\frac{d}{dt}\vec{\phi} = \mathbf{S}\,\vec{f}(\vec{\phi}), \tag{1.21}$$

where $\vec{\phi} = \langle\vec{n}\rangle/\Omega$ is the mean concentration vector, $\mathbf{S}$ is the stoichiometric matrix with elements S_{ij} and $\vec{f} = \lim_{\Omega\to\infty}\vec{\hat{f}}(\langle\vec{n}\rangle,\Omega)$. The second moment equations are slightly more complicated, and tend towards the stationary values of variance and covariance. The fact that the means follow linear equations of motion means that the form of the time correlation matrix is particularly simple [56]; we provide Gardiner's derivation of this time correlation matrix in Appendix A.3.

As highlighted above, the fluctuation power spectrum is the most suitable description for oscillations in stochastic systems. Fortunately, there is a simple result which connects the autocorrelation function in the time domain to the fluctuation power

spectrum in the frequency domain, namely the Wiener-Khinchin theorem. By the Wiener-Khinchin theorem [56], the one-sided power spectrum of the fluctuations in the number of molecules of species i at stationary state is given by the Fourier Transform of the autocorrelation of the number fluctuations in this species:

$$S_i(\omega) = \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-i\omega\tau}\langle[n_i(t) - \langle n_i\rangle_{ss}][n_i(t+\tau) - \langle n_i\rangle_{ss}]\rangle d\tau, \tag{1.22}$$

where the angled brackets denote the statistical average, t is any time for which steady-state conditions have been achieved and $\langle n_i\rangle_{ss}$ is the steady-state number of molecules of species i . A power spectrum peak at some frequency, $\omega = \hat{\omega}$ (the peak frequency), indicates the presence of oscillations in the number of molecules at this particular frequency.

The utility of the linear noise approximation lies in the fact that the linearity of the Fokker-Planck equation (or equivalent Langevin equation) enables one to write down a closed form solution for the autocorrelation function, $\langle\xi_i(t)\xi_i(t+\tau)\rangle$ in terms of the elements of the Jacobian and diffusion matrices. The quantity $\Omega\langle\xi_i(t)\xi_i(t+\tau)\rangle$ is the linear-noise approximation estimate of the correlator $\langle[n_i(t) - \langle n_i\rangle_{ss}][n_i(t+\tau) - \langle n_i\rangle_{ss}]\rangle$ which appears in the spectrum definition, Eq. (1.22). The result is an approximate closed-form equation for the power spectrum of the number fluctuations [56]:

$$S_i(\omega) = \frac{\Omega}{\pi}\left[\left[-\mathbf{J} + \mathbf{I}i\omega\right]^{-1}\mathbf{D}\left[(-\mathbf{J})^T - \mathbf{I}i\omega\right]^{-1}\right]_{ii}, \tag{1.23}$$

where $\mathbf{J}$ and $\mathbf{D}$ are the Jacobian and diffusion matrices shown in Eq. (1.17) and Eq. (1.18).

This methodology for deriving an approximation to the fluctuation power spectrum

is used in Chapters 2 and 3.

1.6 Stochastic Oscillations in Biology

Stochastic models of biochemical oscillators have become more common in the past decade [73, 43, 74, 75], their dynamics being explored principally by means of the stochastic simulation algorithm. To a lesser extent, the problem has also received theoretical attention by analysis of the chemical master equation [76, 30, 77]. Noise has often considered to be an inconvenient disturbance for the timing and regularity of oscillations, and techniques have been developed to quantify this robustness [78, 79]. However, noise can also play a constructive role for oscillations, in that when deterministic models predict a stable steady state, the noise serves to sustain the oscillation [30]. We refer to this phenomenon as a noise-induced oscillation.

There has been much research in the physics literature into the phenomenon of stochastic resonance and coherence resonance [80, 81]. The late 1990*s* saw a drive to apply the methods to real chemical and biochemical models, although the introduction of stochasticity to the models was at first rather ad hoc. Later, researchers showed that intrinsic fluctuations in a circadian model cause a smoothed transition to oscillatory behaviour near the Hopf bifurcation by using a Chemical Langevin Equation description of the stochastic kinetics [76] (Fig. 1.4). Because the signal to noise ratio of the oscillations exhibited a maximum at an intermediate system size, these authors identified the behaviour of the model with that of coherence resonance (alternatively known as autonomous stochastic resonance [82] or stochastic resonance without periodic forcing [80]). The phenomenon of 'noisy precursors' to bifurcations was known before Hou's work in the literature on stochastic and coherence resonance in physics [83].

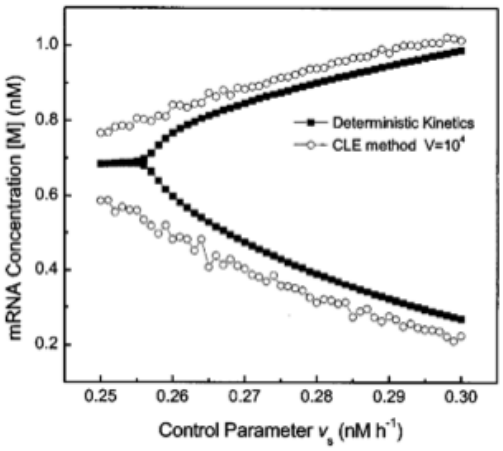


Figure 1.4: Reproduced from [76]. Using the Chemical Langevin Equation approach stochastic oscillations in a circadian model were observed outside of the deterministic limit cycle regime. In this sense the noise is seen to blur the transition from non-oscillatory to oscillatory behaviour.

The presence of a Hopf bifurcation is, in fact, not necessary for inducing oscillatory behaviour in stochastic biochemical systems. Davis and Roussel explored a simple enzyme kinetic model featuring competitive inhibition [84], reporting peaks in the fluctuation power spectra and pronounced stochastic oscillations, yet no Hopf bifurcation is possible for any choice of parameters. Rather, the noise is deemed to induce oscillations because the linear stability of the deterministic model features one or more pairs of complex conjugate eigenvalues [84, 77]. Instead of a numerical simulation approach, McKane et al [30] used van Kampen's linear noise approximation to generate the fluctuation power spectra for a genetic regulation model and Sel'kov's glycolysis model; the peaks at positive frequency which they observed were also explained in terms of the complex eigenvalues. The explanation which they provide is that the resonant frequency in the system is given approximately by the imaginary part of the eigenvalue and intrinsic noise in the system, being white (in time), unselectively amplifies oscillations which would otherwise die out in time [30]. The analysis was extended

by researchers in the same group to an autocatalytic reaction system with more than one pair of complex eigenvalues, and they demonstrated presence of two peaks in the spectrum. The peak frequencies were expected to be close to the imaginary parts of the different pairs of eigenvalues when the real parts of the eigenvalues are small.

This led to the important insight that stochasticity is not necessarily detrimental to the production of sustained oscillations but rather that it can promote oscillations. In particular, using the linear-noise approximation [14] of the chemical master equation, it was shown that internal noise induces sustained oscillations in biochemical networks whose deterministic rate equations predict stable foci, i.e., those whose perturbations decay in an oscillatory manner [30, 77]. The intuitive reason for this phenomenon is that the "underlying stochasticity has a flat spectrum in frequency space (i.e., white noise), and this automatically excites the resonant frequencies of the system" [30], i.e., the frequency of the damped oscillations in the deterministic model. Similar conclusions were reached by means of multiple-scale analysis in the context of oscillating numbers of infected individuals in a population [85] and by analysis of the Q-matrix in simple models of biochemical reactions [86].

1.7 Stochasticity and Coupling in Single Cell Circadian Rhythms

As most habitats are characterized by 24-h day-night cycles and seasonal changes, endogenous circadian rhythms are important for organisms to anticipate and adapt to their environment. Like models of glycolysis, circadian rhythms in unicellular and multicellular organisms have principally been investigated within the rate equation formalism [50, 51, 87], and these models have identified transcription-translation feedback

loops (TTFLs) as a central clock mechanism. Other recent research has showed that circadian rhythms can arise from non TTFL mechanisms [88]. However, in recent years there has been considerable interest in the mammalian clock literature [89, 90, 91] concerning the relationship between observed experimental behaviour in an ensemble of cells and the circadian mechanism within single cells. There is growing interest, too, in the plant systems biology community [92]. Imaging assays of rhythmic LUCIFERASE (LUC) reporter genes allowed noninvasive measurements with spatial resolution [93], showing that individual cells support autonomous circadian oscillators [94, 95]. Further studies such as that by Welsh and coworkers [90] changed perceptions of how the clock functions in peripheral mammalian cells by demonstrating that the damped oscillations in an ensemble of cells reflected not single-cell damping, but a loss of synchrony amongst oscillating cells (see Figure 1.5). This was an exemplar experimental observation that the implicit assumption of behaviour inside each individual cell approximately matching that of the population average may be incorrect.

This heterogeneity is similar to that observed by Gustavsson in glycolysis (Fig. 1.3), and there the modellers described such heterogeneity between cells by allowing variation in a parameter. However, it is not unreasonable to think the heterogeneity is due to internal noise, since some key genes and proteins in circadian clock pathways are present in small copy numbers [96, 74]. More recent investigations of the data from single fibroblasts and SCN neuron cells have probed the question of whether clocks can be thought to operate essentially as limit cycle oscillators, to which noise plays a destructive, desynchronising role, or essentially as damped oscillators with noise sustaining the oscillation [97, 98]. By comparing with simple stochastic models, these papers are in agreement that the result is inconclusive as cell dynamics are best described as operating close to the Hopf bifurcation, but there is some evidence that

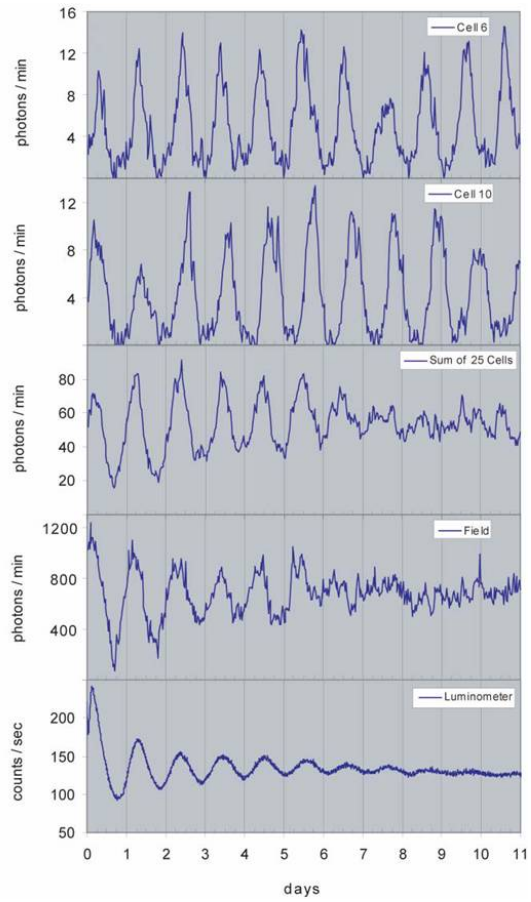


Figure 1.5: Damped population rhythms resultant from single cell sustained oscillations (reproduced from [90]). Bioluminescence from two individual fibroblasts (*mPer2::LUC-SV40* knockin mice) are shown in the top two subplots; we begin to see the damping effect through loss of synchrony in the sum of luminescence of 25 cells in the third subplot.

noise plays a constructive role [98].

In a follow-on study from [90], Liu et al [91] showed that understanding of single-cell behaviour versus tissue-level behaviour yielded profound insights into the nature of intercellular coupling. They found that the rhythmic gene expression of the clock gene *Per2* in *Per1*- or *Cry1*-deficient mice was severely disrupted in individual suprachiasmatic nuclei (SCN) neurons and individual fibroblasts. Remarkably, however, rhythms were still observed in an SCN slice while peripheral tissue explants exhibited the severe rhythmic disruption expected from an ensemble of rhythmically compromised cells. This implies that the nature of intercellular coupling in terms of its ability to maintain correct clock functioning is quite different in the SCN and in peripheral tissues. The coupling of clocks among cells is now a topic of intense interest, because dynamical systems theory shows that such coupling can profoundly alter the period and entrainment behaviour of multioscillator systems [99, 100]. The mammalian suprachiasmatic nucleus (SCN) has been most studied and tight coupling by synaptic transmission among SCN neurons is crucial to sustain rhythmicity [100, 101]. Heterogeneity among individual neurons leads to spatiotemporal waves of rhythmic gene expression in SCN slice cultures [102, 103, 104].

1.8 Outline of Thesis

In this thesis we study separately the effects of intracellular noise and intercellular coupling on oscillatory dynamics in biochemical systems. Regarding noise, in Chapter 2 we use the linear noise approximation analysis to investigate the minimal conditions for any type of oscillation in the presence of intracellular noise (in contrast to the limit cycle requirement of deterministic approach) and explore an intriguing new regime in

which noise can play a constructive role for oscillations. We extend the results in Chapter 3 to investigate the effect of burstiness in the production of proteins on the oscillatory properties of downstream pathways.

Regarding coupling, by investigating experimental data we find that weak local coupling between Arabidopsis cells in the leaf limits desynchronisation in constant conditions and generates well defined spatiotemporal patterns, but that the patterns are different between different leaves. We also present investigations into high resolution microscope data, and show that using the combination of phase at different spatial scales we can confirm the earlier hypothesis that the phase lags in the veins, and that coupling between mesophyll cells appears to bypass the veins in some instances.

Chapter 2

Mesoscopic Concentration Oscillations in Biochemical Systems with Stable Node Steady States

2.1 Introduction

It is now well known that internal noise induces sustained oscillations in biochemical networks whose deterministic rate equations predict stable foci, i.e., those whose perturbations decay in an oscillatory manner [30, 77]. In this chapter, we show that internal noise induces oscillations even in biochemical systems whose deterministic rate equation models predict a stable node. This is remarkable since it cannot be explained by the intuitive reasoning used for noise-induced oscillations (NIO) associated with stable foci. In particular stable nodes are not characterized by a damped oscillatory return to steady state and hence there does not exist a resonant frequency which white noise can excite. We also show that external noise (noise whose origin lies outside the chemical system under consideration [105]) can as well lead to NIO in stable node systems, albeit these having different properties than NIO produced by internal noise.

The chapter is divided as follows. In Section 2.2 we use the linear noise approximation theory to show that NIO can be observed for systems with a stable node. General properties of these NIO such as their frequency and quality and the fundamental reason for their origin are studied in Section 2.3. The existence of these new types of NIO are verified by means of stochastic simulations of three simple chemical systems involving the interaction of two species in Section 2.4. Therein we also study the relationship between the quality of the NIO and the distance from the node-focus borderline in phase space, as well as the robustness of the linear-noise approximation results for small molecule numbers. In Section 2.5 we show that in some cases, the quality of stable node NIO improves dramatically with the number of interacting species thus lending evidence of their possible importance in large biochemical networks. In Section 2.6 we show that external noise generates stable node NIO with properties differing from

those produced by internal noise. Section 2.7 describes some minimal requirements for a stable node NIO biological motif, as well as offering an interpretation of the physical meaning of oscillations in only one of two interacting species. Finally we conclude with a discussion in Section 2.8.

2.2 A Classification of NIO for Two Species Systems

In this section we restrict ourselves to chemical reaction schemes involving two species and use the linear-noise approximation result of section 1.5.3 to explore the intimate relationship between the existence of NIO and the type of steady state (stable node or stable focus) in the deterministic equations.

For systems with only two interacting species, the resultant functional form of the spectrum equation, Eq. (1.23), for species X_i is:

$$S_i(\omega) = \frac{\Omega}{\pi} \frac{\alpha_i(\mathbf{J}, \mathbf{D}) + \beta_i(\mathbf{D})\,\omega^2}{p(\mathbf{J}) + q(\mathbf{J})\,\omega^2 + \omega^4}, \tag{2.1}$$

where the parameteric dependencies on the Jacobian and diffusion matrices are explicitly shown. In particular, the functions p and q are equal to $\lambda_1^2\lambda_2^2$ and $\lambda_1^2 + \lambda_2^2$ respectively, where the λ's are the eigenvalues of the Jacobian matrix, $\mathbf{J}$ (see Appendix A.4.1). For a stable node, the eigenvalues are real and negative while for a stable focus, the eigenvalues are a complex pair, $\lambda_{1,2} = -\mu \pm i\tilde{\omega}$ where both μ and $\tilde{\omega}$ are positive real numbers. Hence it follows that generally $p \in \mathbb{R}_{>0}$, and $q \in \mathbb{R}$. The parameter $\beta_i = D_{ii}$ and hence by Eq. (1.18) one can deduce that $\beta_i \in \mathbb{R}_{>0}$. Lastly the parameter α_i has to be a positive real-valued number since $\alpha_i = (\pi/\Omega)pS_i(0)$.

We now decompose the spectrum Eq. (2.1) for a single species into two sub-spectra,

$S_i^\alpha(\omega)$ and $S_i^\beta(\omega)$:

$$S_i(\omega) = \frac{\Omega}{\pi}\big(S_i^\alpha(\omega) + S_i^\beta(\omega)\big), \tag{2.2}$$

$$S_i^\alpha(\omega) = \frac{\alpha_i}{p + q\,\omega^2 + \omega^4}\,, S_i^\beta(\omega) = \frac{\beta_i\,\omega^2}{p + q\,\omega^2 + \omega^4}. \tag{2.3}$$

In the large ω limit, $S_i(\omega)$ monotonically decreases as ω^{-2}. In the opposite limit of small ω, we have:

$$S_i(\omega) = \frac{1}{p}\alpha_i + \frac{1}{p}\left(\beta_i - \frac{\alpha_i\,q}{p}\right)\omega^2 + O\left(\omega^4\right).$$

A peak in the power spectrum will then exist if the spectrum increases for small ω, i.e., if the ω^2 term above is positive. As shown above, the parameters p, α_i and β_i are positive so when $q < 0$ the ω^2 term is positive and hence a peak always exists. When $q \geq 0$ a peak only exists if $\beta_i > \alpha_i q/p$. Further understanding of this is gleaned by considering the sub-spectra S_i^α and S_i^β. If both sub-spectra have a peak, so too will the total spectrum. S_i^β always has a single peak ($\frac{dS_i^\beta}{d\omega} = 0$ at $\omega = p^{\frac{1}{4}}$). S_i^α is only peaked when $q < 0$. We summarise this as follows:

- **Stability Dominated NIO**: The parameter $q < 0$ and hence both sub-spectra have a peak, guaranteeing a peak in $S_i(\omega)$.

- **Noise Dependent NIO**: The parameter $q \geq 0$ such that only the sub-spectrum S_i^β has a peak, but the magnitude of β_i relative to α_i is sufficiently large that a peak exists in $S_i(\omega)$. The exact criterion in this case is $\beta_i > \alpha_i q/p$.

For both cases, it can be further shown that the peak power in $S_i(\omega)$ occurs at a frequency lying between the peak frequencies of the two sub-spectra. The proof is

as follows. Let the S_i^α and S_i^β subspectra have peaks at the frequencies $\hat{\omega}_{S^\alpha}$ and $\hat{\omega}_{S^\beta}$ respectively (note that we include the possibility that $\hat{\omega}_{S^\alpha} = 0$) . By Taylor expansion of $S_i(\omega)$ at these two frequencies, one finds that the slope of $S_i(\omega)$ at $\omega = \hat{\omega}_{S^\alpha}$ is > 0, and the slope at $\omega = \hat{\omega}_{S^\beta}$ is < 0. Since there can only be a single peak in the spectrum of a two species system (see Appendix A.4.2), it follows that the peak power in $S_i(\omega)$ is in the range $(\hat{\omega}_{S^\alpha}, \hat{\omega}_{S^\beta})$.

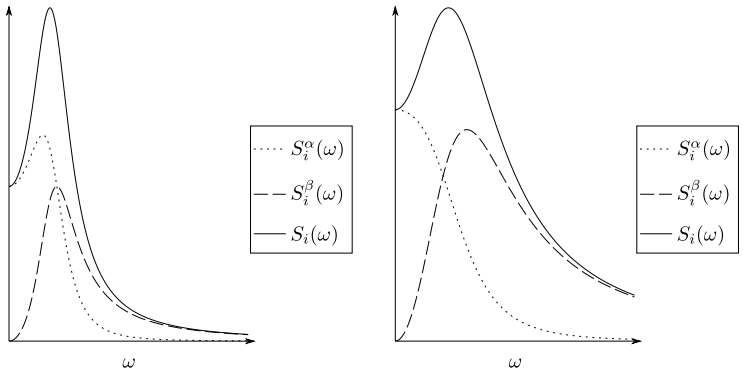
The first case is termed Stability Dominated NIO because the existence of the NIO solely rests on the sign of q which is a function of the eigenvalues of the Jacobian and hence of the type of steady state. The second case is termed Noise Dependent NIO because β_i, being a function of the diffusion matrix, is a measure of the strength of the noise and this needs to be larger than a critical threshold for NIO to exist. Note the important implication that for the Stability Dominated NIO case, NIO present in one species dictates that NIO are present in the other species, however this is not necessarily true for Noise Dependent NIO since the α_i and β_i are species dependent.

Consider the case of Stability Dominated NIO. If we have a stable focus then the eigenvalues can be denoted as $\lambda_{1,2} = -\mu \pm i\tilde{\omega}$ where both μ and $\tilde{\omega}$ are positive real numbers. The condition $q < 0$ is then fulfilled if $\mu < \tilde{\omega}$, in other words if the timescale of decay of the amplitude of the damped oscillations in the deterministic model is larger than the period of the same oscillations.

Now we consider the case of Noise Dependent NIO. The condition $q \geq 0$ is satisfied for stable foci with $\mu \geq \tilde{\omega}$ and also for stable nodes, i.e., those steady states characterized by two real and negative eigenvalues. Hence these type of steady states can also be expected to give rise to NIO, this condition being dictated by the magnitude of the elements of the diffusion matrix.

Hence in summary it is clear that not only stable foci can give rise to NIO but

that stable nodes can do as well. In the rest of this chapter we focus on studying the existence and properties of the latter type of NIO, which we refer to as stable node NIO.



(a) **Stability Dominated NIO** (b) **Noise Dependent NIO**

Figure 2.1: Plots of the spectrum (2.2) along with the sub-spectra (2.3) showing the two possible ways in which a peak in the power spectrum can arise: (a) The parameters p and q are such that both sub-spectra have a peak ($p = 1, q = -1$) and a peak in $S_i(\omega)$ is guaranteed; (b) The parameters p and q are such that only the sub-spectrum S_i^β has a peak ($p = 5, q = 1$), but the magnitude of β_i is sufficiently large that a peak exists in $S_i(\omega)$.

2.3 General Properties of Stable Node NIO

2.3.1 Frequency of the Oscillations

While one would expect that the frequency of stable foci NIO is close to the frequency of damped oscillations in the deterministic model, it is not at all clear what should be the frequency of stable node NIO. A simple expression for this can be deduced by differentiating $S_i(\omega)$ in Eq. (2.1), setting the resulting equation to zero and solving for the roots of the quartic in ω. Using the fact that $q > 0$ for a node, one finds that the only positive (and hence admissible) solution for the peak frequency of NIO in species

i is:

$$\hat{\omega}_i = \left(\frac{[p(p + q^2 x_i(x_i - 1))]^{1/2} - p}{q x_i} \right)^{1/2}, \tag{2.4}$$

where $x_i = \beta_i p / \alpha_i q$. Note that the criterion for a peak for a Noise Dependent NIO (see Section 2.2) implies $x_i > 1$. From the above equation we can see that $\hat{\omega}_i$ varies with x_i between 0 and a maximum value of $p^{1/4} = \sqrt{\lambda_1 \lambda_2}$. Hence the frequency of NIO in the node case is bounded from above by the geometric mean of the two decay timescales of non-oscillatory transients in the deterministic model.

Furthermore, one can deduce that the peak frequency increases monotonically with x_i, meaning that this frequency increases with the fraction of power contributed by the sub-spectrum $S_i^{\beta}(\omega)$ to the total power spectrum $S_i(\omega)$. It is also easy to show that the maximum peak frequency value of $p^{1/4} = \sqrt{\lambda_1 \lambda_2}$ for $S_i(\omega)$ is equal to the peak frequency of the $S_i^{\beta}(\omega)$ sub-spectrum. These observations are intuitively clear since for a Noise Dependent NIO (of which a stable node NIO is a special case) only the $S_i^{\beta}(\omega)$ has a peak and hence its size relative to the peakless $S_i^{\alpha}(\omega)$ sub-spectrum crucially dictates the spectral properties of the NIO.

2.3.2 Quality of the Oscillations

Next we discuss the quality of the stable node NIO. Some researchers have described the strength of noise-induced oscillations using the ratio of heights $S_i(\hat{\omega})/S_i(0)$ [30], while others have multiplied the traditional quality factor with the peak height of the spectrum to form a signal to noise ratio (SNR) [80]. One of the main reasons for using the SNR is to show a resonance of the SNR with noise in the system. In our models, noise magnitude is quantified by the system size parameter Ω. Analysis of the number fluctuation spectrum to linear approximation (1.23) shows that the volume only affects

the magnitude of the spectrum, and further that this is through a linear scaling of power in Ω. The linear noise approximation therefore fails to capture any resonance of SNR with volume. Hence, we focus exclusively on the time properties of the signal and not its magnitude, and therefore on the signal's quality. The well-known classical measure of oscillation quality is the Q-factor, defined as:

$$Q = \frac{\hat{\omega}}{\Delta\omega}, \tag{2.5}$$

where the bandwidth $\Delta\omega$ is the difference of the two frequencies at which the power takes its half-maximum value. To make connection with later developments, we denote this conventional Q-factor as $Q^{50\%}$.

The $Q^{50\%}$ of the spectrum $S_i(\omega)$ can be written as a function of two parameters, $R_\epsilon = \lambda_1/\lambda_2$ which is the ratio of the two real eigenvalues, and $R_{H\alpha\beta}$ which is the relative weighting of the α and β subspectra, as quantified by the ratio of their maximum heights (see Fig. 2.2(a) for a contour plot). This choice of parameters is convenient because R_ϵ is a parameter which is dictated solely by deterministic stability considerations whereas $R_{H\alpha\beta}$ is a parameter which is determined by the properties of the internal noise. Note that our R_ϵ is limited to the interval $(0, 1]$ by specifiying that $|\lambda_1| < |\lambda_2|$. The quality of the NIO increases with increasing R_ϵ and with decreasing $R_{H\alpha\beta}$ and reaches a maximum of $Q^{50\%} = 1/2$ when $R_\epsilon = 1$ and $R_{H\alpha\beta} = 0$.

The first observation can be explained as follows. As R_ϵ approaches its maximum of one, it approaches the case of a stable degenerate node. Such a steady state lies at the node-focus borderline and is characterized in phase space by deterministic trajectories which approach the origin tangent to the single linearly independent eigenvector[1]; the

[1]Note that a degenerate node could be of the star node type whereby the Jacobian has two independent eigenvectors and there is no curvature in the phase space trajectories. This is, however, a special

deterministic trajectories in this case display a degree of curvature, in a sense trying to wind around in a spiral but not quite making it. The stochastic trajectories will follow to some extent the curvature of the deterministic trajectories and for some cases, the noise will cause the trajectory to close on itself hence leading to a noise-induced oscillation (see Fig. 2.2(b) for an illustration). Hence as one approaches the node-focus borderline, i.e., as $R_\epsilon \to 1$, one expects the quality of NIO to increase since the stochastic trajectories can sample increasingly curved deterministic trajectories.

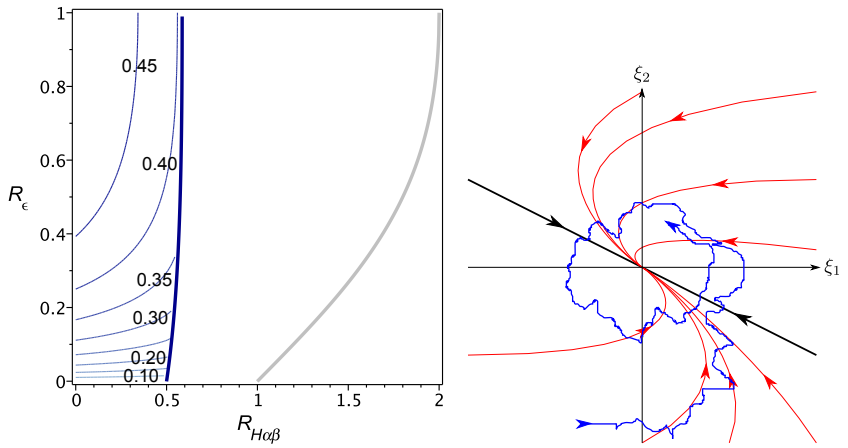
The observation that the quality increases with decreasing $R_{H\alpha\beta}$ is intuitively obvious when one considers that out of the two sub-spectra composing the total spectrum, only the $S_i^\beta(\omega)$ sub-spectrum has a peak, and hence its increased contribution to the total spectrum necessarily improves the quality. The same argument hints that the maximum $Q^{50\%} = 1/2$ for the total spectrum $S_i(\omega)$ equals the $Q^{50\%}$ of the $S_i^\beta(\omega)$ sub-spectrum. This is indeed the case. The latter is given by the remarkably simple expression:

$$Q^{50\%}_{S^\beta} = \frac{\sqrt{R_\epsilon}}{R_\epsilon + 1}, \tag{2.6}$$

which maximizes at a value of $1/2$ when the eigenvalues become equal.

In Fig. 2.2(a), it is observed that there is a large region of parameter space for which NIO exist but the shape of the power spectrum is such that it does not fall to its half-maximum value on the low frequency side of the peak and hence a $Q^{50\%}$ cannot be defined (this is the area between the thick blue and grey lines in the figure). To allow a measure of quality over a greater range of the parameter space, we introduce a more general version of the conventional Q-factor in Eq. (2.5): $Q^{f\%} = \hat{\omega}/\Delta\omega^{f\%}$, where

case appearing when the Jacobian is diagonal and hence when the two species are non-interacting; the power spectrum Eq. (2.1) does not display a maximum in frequency for this case and hence no NIO either. Thus the only degenerate node case where NIO appear is where the Jacobian is not diagonal, i.e., the case of one independent eigenvector which is also characterized by curved trajectories, as discussed in the main text.



(a) (b)

Figure 2.2: (a) The quality of stable node oscillations, as described by the conventional quality factor, $Q^{50\%}$. Contours of $Q^{50\%}$ (blue) are shown as a function of the two parameters R_ϵ and $R_{H\alpha\beta}$ (see text for definitions). The thick grey line demarks the parameter regions in which NIO are/are not observed and the thick dark blue line demarks the regions in which the peak is strong enough to obtain its half-maximum value on the low frequency side of the peak (i.e., $Q^{50\%}$ is defined). It is clear that $Q^{50\%}$ is not a suitable measure for describing the quality of weak stable node NIO as it is only defined for a small region of the full parameter space in which NIO is observed. (b) Illustration of an example noisy evolution of the fluctuation variables ξ_1, ξ_2 in a system with degenerate node stability, i.e., $R_\epsilon = 1$, corresponding to the maximum $Q^{50\%}$ case. Red curves are deterministic trajectories, the blue curve is the stochastic trajectory and the black line is the single linearly independent eigenvector of the Jacobian.

$f \in (0, 100)$ and $\Delta\omega^{f\%}$ is the difference of the frequencies at which the spectrum takes $f\,\%$ of its maximum value. We choose to use the Q-factor $Q^{99\%}$ in order to describe the vast majority of the region which cannot be captured by the $Q^{50\%}$ measure, i.e. the region between the thick blue and thick grey lines in Fig. 2.2(a), as confirmed by Fig. 2.3.

This measure, while more complicated analytically than $Q^{50\%}$, is also a function of only R_ϵ for $S_i^\beta(\omega)$ and a function of R_ϵ and $R_{H\alpha\beta}$ for $S_i(\omega)$, and shows very similar behaviour to the $Q^{50\%}$ measure over the parameter regions in which $Q^{50\%}$ is defined (compare the blue contours in Figs. 2.2(a) and 2.3)). It can be shown that the maximum possible value of $Q^{99\%}$ is $3\sqrt{11}/2 \simeq 5$. The $Q^{99\%}$ measure only uses spectrum

information very close to the peak frequency; it is a very localized measure which directly captures the high curvature at the peak. However, we note that the measure successfully captures the behaviour of the spectrum over wider frequency ranges. As an example, a large amplification value $\frac{S_i(\hat{\omega})}{S_i(0)}$ (ratio of the power at the peak frequency to the power at zero frequency) has been highlighted as being an important parameter in describing pronounced stochastic oscillations [30]. In Fig. 2.3 this amplification factor is also shown (lighter green contours represent higher amplification), and it is observed that it is not possible to obtain a very high $Q^{99\%}$ value without an associated high amplification value. We provide further validation of the $Q^{99\%}$ measure in section 2.4.4, following the investigation of exemplar biochemical models.

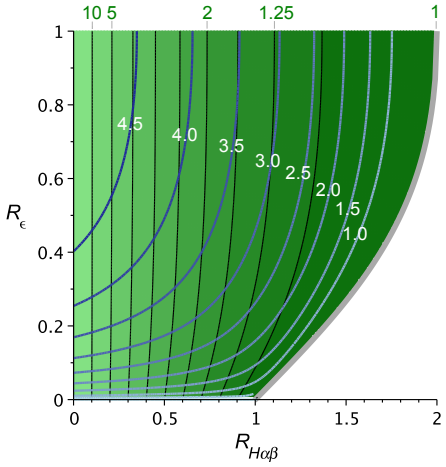


Figure 2.3: The quality of stable node oscillations, as described by $Q^{99\%}$. Contours of $Q^{99\%}$ (blue) are shown as a function of the two parameters R_ϵ and $R_{H\alpha\beta}$. Green filled contours show the variation of the amplification factor $\frac{S_i(\hat{\omega})}{S_i(0)}$, with selected contour values shown. The thick grey line demarks the parameter regions in which NIO are/are not observed.

We note that it is difficult to further develop the general theory of stable node NIO for biochemical systems with internal noise because generally the **J** and **D** matrices are both dependent on the rate constants in the system as well as the steady-state values of concentrations of the rate equations (see Eqs. (1.17) and (1.18)). This places

important constraints on the values that can be taken by the parameters $\alpha_i(\mathbf{J}, \mathbf{D})$ and $\beta_i(\mathbf{D})$ and hence constraints on the region of parameter space where stable node NIO exist. Hence the rest of the chapter is devoted to understanding stable node NIO in the context of specific biochemical systems.

2.4 Two Species Biochemical Systems with Stable Node NIO

In this section we use the linear-noise approximation to study the relationship between the type of steady state and the existence of NIO for three biochemical systems involving the interaction of two species. These numerical classifications are listed in Table 2.1 (pg. 47). For selected points in parameter space where stable node NIO are predicted to exist, we use the periodogram method to numerically estimate the power spectra from stochastic simulations using the stochastic simulation algorithm (see Appendix A.5) and compare these with the theoretical spectra predicted by the linear-noise approximation. In all cases we find good agreement with the two spectra, and hence verify that a peak in the power spectrum of number fluctuations does indeed exist for certain stable node steady states in real chemical systems. We shall also investigate the relationship between the $Q^{99\%}$ quality factor of stable NIO and the distance of the node from the node-focus borderline in phase space, as well as the robustness of the linear-noise approximation predictions for small molecule numbers.

2.4.1 Example 1: The Brusselator

The Brusselator is the only known chemical scheme involving just two interacting species and whose deterministic equations admit limit cycle oscillations [106, 107, 108,

109]:

$$\emptyset \xrightarrow{k_0} A,\ 2A + B \xrightarrow{k_1} 3A,\ A \xrightarrow{k_2} B,\ A \xrightarrow{k_3} \emptyset. \tag{2.7}$$

The autocatalytic step $2A + B \rightarrow 3A$ is not an elementary reaction but rather an effective reaction composed of simpler reaction steps (more on this later). The possible biological relevance of this reaction scheme stems from the fact that the autocatalytic step can be produced by a series of enzyme-catalyzed reactions [108].

We start by defining two dimensionless parameters:

$$\Lambda_1 = \frac{k_0^2\, k_1}{k_3^3},\ \Lambda_2 = \frac{k_2}{k_3}. \tag{2.8}$$

The stoichiometric matrix and the macroscopic rate function vector for this system are:

$$\mathbf{S} = \begin{pmatrix} 1 & 1 & -1 & -1 \\ 0 & -1 & 1 & 0 \end{pmatrix}, \tag{2.9}$$

$$\vec{f} = \{k_0, k_1[A]^2[B], k_2[A], k_3[A]\}^T, \tag{2.10}$$

where $[A]$ and $[B]$ are the macroscopic concentrations of species A and B. The rate equations are then given by $\{\partial_t[A], \partial_t[B]\}^T = \mathbf{S}\vec{f}$. Linear stability analysis of these equations reveals that the regions in Λ_1-Λ_2 space which characterize a stable node and a stable focus respectively are:

$$\Lambda_1 > \Lambda_2 - 1, \quad 1 + (\Lambda_1 - \Lambda_2)^2 - 2(\Lambda_1 + \Lambda_2) \geq 0, \tag{2.11}$$

$$\Lambda_1 > \Lambda_2 - 1, \quad 1 + (\Lambda_1 - \Lambda_2)^2 - 2(\Lambda_1 + \Lambda_2) < 0. \tag{2.12}$$

Now we use the linear-noise approximation to deduce the conditions for the existence

of NIO. Substituting the stoichiometric matrix and rate function vector into Eqs. (1.17) and (1.18) we obtain the Jacobian and diffusion matrices. Finally substituting the latter in Eq. (1.23) we obtain equations for $S_1(\omega)$, the power spectrum of species A, and for $S_2(\omega)$, the power spectrum of species B. The latter equations are of the form given by Eq. (2.1) with global parameters $p = k_3^4\Lambda_1^2$, $q = k_3^2(1 + (\Lambda_1 - \Lambda_2)^2 - 2\Lambda_2)$ and species-specific parameters given by:

$$\alpha_1 = 2k_0k_3^2\Lambda_1^2, \tag{2.13}$$

$$\beta_1 = 2k_0(1 + \Lambda_2), \tag{2.14}$$

$$\alpha_2 = 2k_0k_3^2\Lambda_2(1 + \Lambda_2), \tag{2.15}$$

$$\beta_2 = 2k_0\Lambda_2. \tag{2.16}$$

Differentiating the expressions for $S_1(\omega)$ and $S_2(\omega)$ with respect to ω, one finds the conditions for the peak in the power spectrum and hence for the existence of NIO in the number fluctuations of species A and of species B respectively are:

$$(\Lambda_1 - \Lambda_2)^2 - 3\Lambda_2 < 0, \tag{2.17}$$

$$\Lambda_2(1 + 2\Lambda_1 + \Lambda_2 - (\Lambda_1 - \Lambda_2)^2) - 1 > 0. \tag{2.18}$$

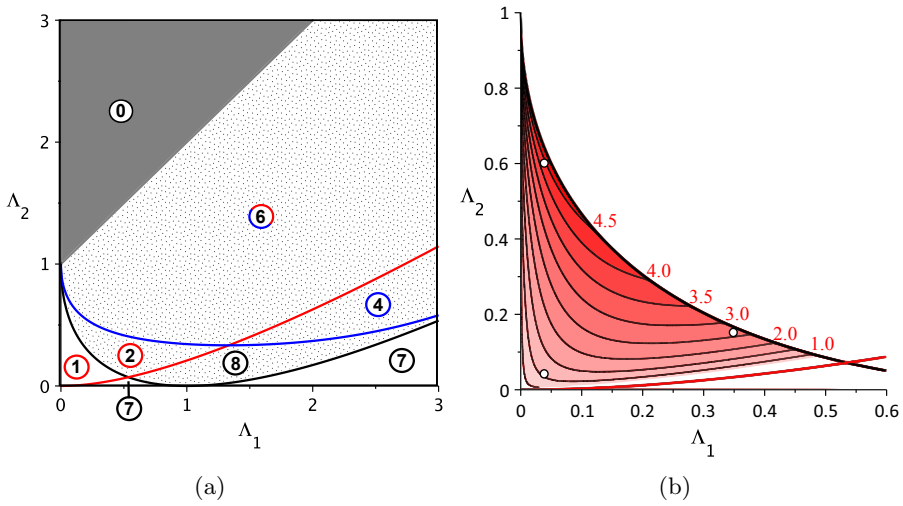
In Fig. 2.4(a) we plot in Λ_1-Λ_2 space the inequalities Eqs. (2.11)-(2.12) which determine the type of steady state (solid black line), together with the inequalities Eqs. (2.17)-(2.18) which determine the existence of NIO (red and blue lines). The intersection of the regions defined by these inequalities are numbered according to the classification set forth in Table 2.1. A large region of the space, i.e., region 6, exhibits NIO in both species and the steady state is a focus. This is the conventionally studied

case. However one can also see that there is a region, namely region 1, wherein the steady state is a node but there are NIO for species A. Interestingly there is no region where there are NIO in species B and the steady state is a node. We shall return to this point in Section 2.6.

Numerical ID	Classification	
	NIO	Stability of Steady State
0	None	Unstable
1	Species A	Node
2	Species A	Focus
3	Species B	Node
4	Species B	Focus
5	A and B	Node
6	A and B	Focus
7	None	Node
8	None	Focus

Table 2.1: Existence of NIO and linear stability classifications as used in Figures 2.4, 2.7 and 2.12.

In Fig. 2.4(b) we plot the $Q^{99\%}$ quality factor for region 1 in Fig. 2.4(a). This shows that, broadly speaking, the quality of the stable node NIO increases with decreasing distance from the node-focus boundary. This relationship is only approximate however, the reason being that the quality is also significantly influenced by the properties of internal noise which are not single handedly captured by the Jacobian of the rate equations. In Fig. 2.5 we show the spectra obtained from the linear-noise approximation and from stochastic simulations for the three points marked by white circles in Fig. 2.4(b). A comparison of cases (a) and (b) in this figure shows the large difference in quality factor even though the location of the nodes in Λ_1-Λ_2 space puts both of them approximately the same distance from the node-focus border (see Fig. 2.4(b)). In all cases we find good agreement between theory and simulations confirming the existence



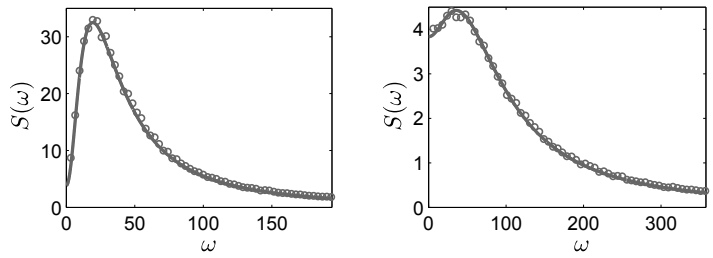
(a) (b)

Figure 2.4: (a) NIO existence and stability classification for the Brusselator in Λ_1-Λ_2 space. Number classifications are as in Table 2.1. Red (blue) circles are used to emphasise the existence of NIO in the number fluctuations of species A (B); the dotted region corresponds to the stable focus regime; white regions correspond to the stable node regimes and the grey region is where the fixed point is unstable (limit cycle). (b) Contour plot of the variation of the Q factor $Q^{99\%}$ with Λ_1, Λ_2 in the stable node regime (① in Figure 2.4(a)). The black contour lines, labelled with red numbers, represent $Q^{99\%}$ values from 0.5 to 4.5 in 0.5 increments. White circles indicated on the $Q^{99\%}(\Lambda_1, \Lambda_2)$ surface correspond to the three power spectra in Figure 2.5.

of stable node NIO.

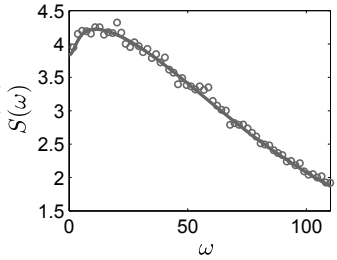
Finite volume effects and elementary reaction versions of the Brusselator

We used stochastic simulations to explore two further questions: (i) given that the linear-noise approximation theory is valid in the limit of large volumes / large molecule numbers, how well do its results for stable node NIO hold when one has small volume / small molecule numbers? (ii) the Brusselator is composed of one trimolecular reaction, a reaction which in practice occurs very rarely due to the unlikely event of three colliding molecules [110]. Thus in many instances such a reaction approximately models a set of underlying (fast) elementary (unimolecular and bimolecular) reactions. Is it the case



(a) $\Lambda_1 = 0.04, \Lambda_2 = 0.6$: $Q^{99\%} = 4.38$.

(b) $\Lambda_1 = 0.35, \Lambda_2 = 0.15$: $Q^{99\%} = 2.63$



(c) $\Lambda_1 = 0.04, \Lambda_2 = 0.04$: $Q^{99\%} = 1.06$

Figure 2.5: Power spectrum plots of the number fluctuations in species A in the Brusselator reaction system for three different sets of dimensionless parameters for which a stable node steady state exists. Solid lines show the analytical spectrum from the linear-noise approximation; open circles show the numerical spectrum calculated by averaging the periodograms of 2500 realizations of the stochastic simulation algorithm. The constants are $\Omega = 1 \times 10^{-15} l$, $k_0 = 1 \times 10^{-4} M\, s^{-1}$, and $k_3 = 100 s^{-1}$. In addition k_1 and k_2 take values of $4 \times 10^{12} M^{-2} s^{-1}$ and $60 s^{-1}$ for case (a), $3.5 \times 10^{13} M^{-2} s^{-1}$ and $15 s^{-1}$ for case (b) and $4 \times 10^{12} M^{-2}\, s^{-1}$ and $4 s^{-1}$ for case (c). Note that the units for concentration, time and frequency ω are Molar (M), second (s) and radians per second ($rad\, s^{-1}$) respectively.

that stable node NIO can also be predicted from elementary reaction models of the Brusselator?

The first question is important since molecule numbers of many species inside cells are quite small, typically in the range of few tens to few thousands [41, 42]. In Fig. 2.6 we show the results of stochastic simulations investigating how the power spectra for the parameter sets (a) and (b) used in Fig. 2.5 change with decreasing volume and a corresponding decrease in the mean molecule numbers of species A. The linear-noise approximation result is shown as a solid line. Note that in both cases the linear-noise

theory is accurate for mean molecule numbers of the order of a thousand molecules. For parameter set (b) the theory remains remarkably accurate for mean molecule numbers less than that of a single molecule while for parameter set (a) considerable deviations from the theory are evident for mean molecule numbers below a hundred molecules. In particular one observes a clear deterioration of quality with decreasing mean molecule numbers, which we quantified as follows. For each of the numerical power spectra S^{SSA} from the stochastic simulation algorithm shown in Fig. 2.6(a), the $Q^{50\%}$ was determined by fitting the numerical power spectra using a function of the form:

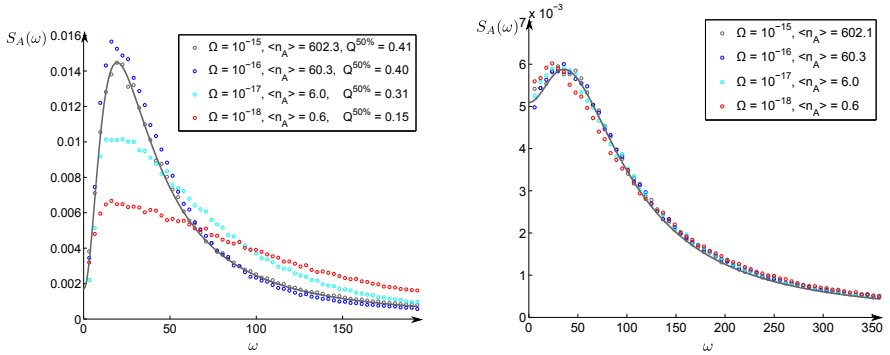
$$S^{fit} = \frac{\alpha + \beta\omega^2}{p + q\omega^2 + \omega^4}.$$

We then used the parameterized function to find $\hat{\omega}$ (the frequency at which the power maximizes) and $\Delta\omega$ (the difference of the two frequencies at which the power is half maximum) analytically, from which we finally obtain $Q^{50\%} = \hat{\omega}/\Delta\omega$. To determine the parameters of S^{fit}, the MATLAB optimization toolbox function fminsearch was used to perform an unconstrained parameter search which minimized the Euclidean norm between the data and the fitted function, i.e., $C = \sqrt{\sum_{k=1}^{L/2}\left(S_k^{fit} - S_k^{SSA}\right)^2}$. The fitted functions are shown in Fig. A.1 in Appendix A.6 for reference.

These results show that stable node NIO can exist for molecule numbers typical of those inside cells and suggest that the quality of the spectra predicted by the linear-noise approximation provides an upper bound for the quality of spectra at finite mean molecule numbers (finite volumes). The reason that the linear noise approximation performs less well for the parameters in (a) than in (b) is not well understood, but the excellent performance of the linear approximation in Fig. 2.6(b) even for a very low mean number of n_A molecules is an example of performance well beyond that which

would be reasonably expected of an approximation best suited to infinite volumes.

The importance of the second question stems from the fact that the linear-noise approximation can generally give different results for effective models and their elementary versions [111, 112] and hence there exists the possibility that the stable node NIO is an artifact of modeling a trimolecular reaction. We simulated an elementary reaction version of the Brusselator put forward by Cook et al [113] where the trimolecular reaction $2A + B \rightarrow 3A$ is broken down into the pair of bimolecular reactions: $A + A \rightleftharpoons X, X + B \rightarrow X + A$ (this is labeled Scheme III in the aforementioned paper). The system displayed stable node NIO with similar quality which hence verifies that such oscillations are not an artifact of the non-elementary reaction in the Brusselator. Further support to this conjecture will be evident in the next subsections where we study two systems composed of purely elementary reactions and in both cases we find stable node NIO.



(a) (b)

Figure 2.6: Normalised power spectrum plots of the number fluctuations in species A in the Brusselator reaction system for parameter sets (a) and (b) in Fig. 2.5, as a function of compartment volume Ω. Solid lines show the analytical spectrum (normalised by the total power) from the linear-noise approximation; open circles show the normalised numerical spectrum for each reaction volume, calculated by averaging the periodograms of 2500 realizations of the stochastic simulation algorithm and then normalizing by the total power (sum of all spectrum values × frequency resolution). In the figure legends, compartment volumes Ω are shown in units of litres, $\langle n_A \rangle$ is the steady-state mean number of molecules of A and $Q^{50\%}$ is the quality factor calculated from the numerical power spectrum (see text for details). The unit for frequency ω is $rad\,s^{-1}$.

2.4.2 Example 2: A Simpler Autocatalytic Reaction

We now consider another autocatalytic reaction scheme in which two distinct species are input to a reaction volume wherein they are involved in a bimolecular autocatalytic reaction, before the product species is exported from the reaction volume:

$$\emptyset \xrightarrow{k_0} A,\ \emptyset \xrightarrow{k_1} B,\ A + B \xrightarrow{k_2} 2B,\ B \xrightarrow{k_3} \emptyset \tag{2.19}$$

This reaction is simpler than the Brusselator and indeed more realistic in the sense that it is composed of at most bimolecular (and hence elementary) reactions. It is also the case that such autocatalytic reactions appear in various biological contexts such as the autocatalytic conversion of normal prion protein to its pathogenic form [114], and the activation of MPF complex in the cell division cycle [115].

Its analysis proceeds as for the previous example. We define two non-dimensional parameters:

$$\Lambda_1 = \frac{k_1}{k_0},\ \Lambda_2 = \frac{k_0 k_2}{k_3^2}. \tag{2.20}$$

The stoichiometric matrix and the macroscopic rate function vector for this system are:

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 1 & -1 \end{pmatrix}, \tag{2.21}$$

$$\vec{f} = \{k_0, k_1, k_2[A][B], k_3[B]\}^T, \tag{2.22}$$

from which we obtain the deterministic rate equations $\{\partial_t[A], \partial_t[B]\}^T = \mathbf{S}\vec{f}$. Note that in contrast with the Brusselator, this system's deterministic rate equations do not exhibit limit cycle behaviour. Linear stability analysis of the rate equations shows that

its steady state is a node if the condition:

$$(\Lambda_1 + (1+\Lambda_1)^2\Lambda_2)^2 - 4(1+\Lambda_1)^3\Lambda_2 \geq 0, \tag{2.23}$$

is satisfied; otherwise the steady state is a focus. The linear-noise analysis proceeds as before, namely one uses Eqs. (1.17) and (1.18) to construct the Jacobian and diffusion matrices from $\mathbf{S}$ and $\vec{f}$ and then substitutes these into Eq. (1.23) to obtain the spectra equations: $S_1(\omega)$ for species A, and $S_2(\omega)$, for species B. The latter equations are of the form given by Eq. (2.1) with global parameters $p = k_3^4(1+\Lambda_1)^2\Lambda_2^2$, $q = k_3^2(\Lambda_1^2 - 2(1+\Lambda_1)^2\Lambda_2 + (1+\Lambda_1)^4\Lambda_2^2)/(1+\Lambda_1)^2$ and species-specific parameters given by:

$$\alpha_1 = 2k_0k_3^2, \tag{2.24}$$

$$\beta_1 = 2k_0, \tag{2.25}$$

$$\alpha_2 = 2k_0k_3^2\Lambda_2^2(1+\Lambda_1)^3, \tag{2.26}$$

$$\beta_2 = 2k_0(1+\Lambda_1). \tag{2.27}$$

Maxima of the power spectra at non-zero frequencies (and hence NIO) for species A and B respectively occur when the conditions:

$$2(1+\Lambda_1)^2\Lambda_2 - \Lambda_1^2 > 0, \tag{2.28}$$

$$1 + 2\Lambda_1 + (1+\Lambda_1)^2\Lambda_2(2 - (1+\Lambda_1)^2\Lambda_2) > 0, \tag{2.29}$$

are satisfied. From the inequalities Eq. (2.23), (2.28) and (2.29) we obtain the complete phase space plot of the system's deterministic and NIO behaviour (see Fig. 2.7(a)). Note that unlike for the Brusselator, stable node NIO are possible for both species, i.e.,

regions 1 and 3. However, note that it is not possible to simultaneously have stable node NIO in both species; this situation is only achievable for stable focus NIO (region 6, Figure 2.7(a)). We postpone discussion of the meaning of an oscillation in only one variable to the conclusions section of this chapter. In Fig. 2.7(b) we show the $Q^{99\%}$ quality measure of stable node NIO in both species (hues of red indicate the strength of NIO in A and hues of blue indicate the strength of NIO in B). Note that as for the case of the Brusselator, there is an inverse relationship between the quality and the distance from the node-focus borderline in phase space (darker hues of red and blue approaching the node-focus borderline). It is also found that the quality of NIO in species B are better than those in species A.

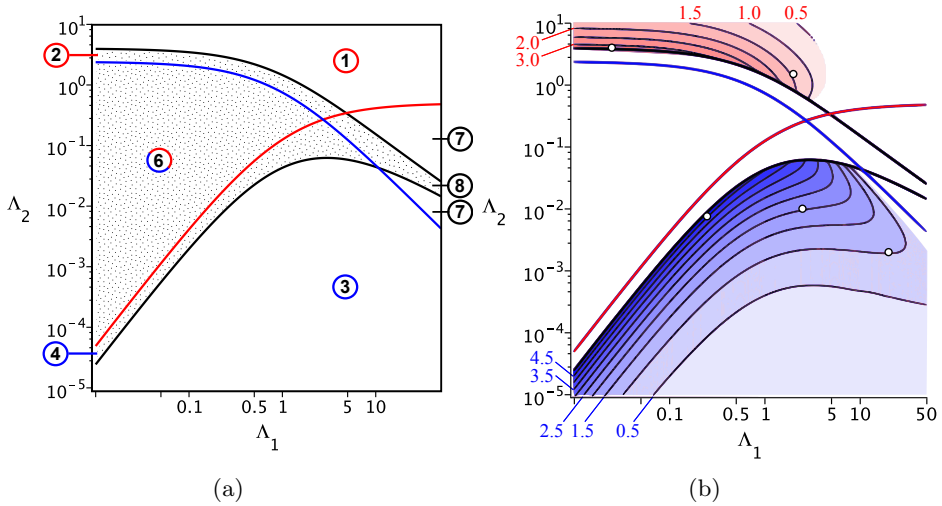
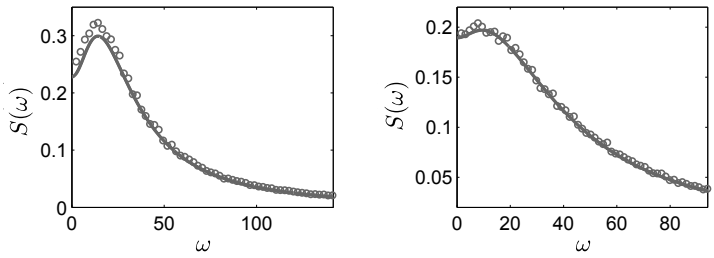


Figure 2.7: (a) NIO existence and stability classification for a simple autocatalytic reaction system in Λ_1-Λ_2 space. Number classifications are as in Table 2.1. Red (blue) circles are used to emphasise the existence of a peak in the power spectrum of species A (B); the dotted region corresponds to the stable focus regime; white regions correspond to the stable node regimes. (b) Contour plot of the variation of the Q factor $Q^{99\%}$ with Λ_1, Λ_2 in the stable node regimes (① and ③ in Figure 2.7(a)). The black contour lines, labelled with red (for species A) and blue numbers (for species B), represent $Q^{99\%}$ from 0.5 to 4.5 in 0.5 increments. Open circles indicated on the $Q^{99\%}$ (Λ_1, Λ_2) surface correspond to the Λ values of the five power spectra in Figs. 2.8 and 2.9.



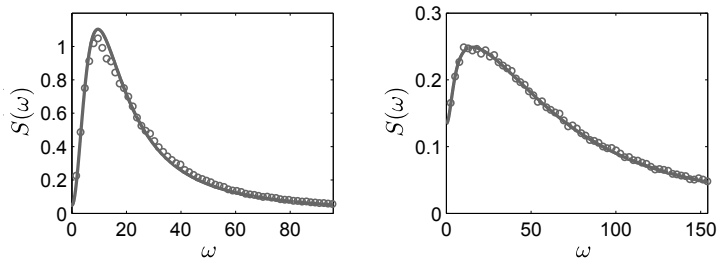
(a) $\Lambda_1 = 2.5 \times 10^{-2}, \Lambda_2 = 4$: $Q^{99\%} = 3.21$ (b) $\Lambda_1 = 2, \Lambda_2 = 1.5$: $Q^{99\%} = 1.43$

Figure 2.8: Power spectrum plots of number fluctuations in species A in the autocatalysis reaction system for two different sets of dimensionless parameters for which a stable node steady state exists. Solid lines show the analytical spectrum from the linear-noise approximation; open circles show the numerical spectrum calculated by averaging the periodograms of 5000 realizations of the stochastic simulation algorithm. The parameters common to both cases are: $\Omega = 1 \times 10^{-15} l, k_0 = 1 \times 10^{-6} M\, s^{-1}, k_3 = 10 s^{-1}$. The case specific rate constants are: (a) $k_1 = 2.5 \times 10^{-8} M\, s^{-1}, k_2 = 4 \times 10^{8} M^{-1}\, s^{-1}$ and (b) $k_1 = 2 \times 10^{-6} M\, s^{-1}, k_2 = 1.5 \times 10^{8} M^{-1}\, s^{-1}$. The unit for frequency ω is $rad\, s^{-1}$.

We again present the power spectra from the linear-noise approximation alongside simulation results for the simple autocatalytic system for selected parameter sets (see Figs. 2.8 and 2.9). The locations of these parameter sets in phase space are shown as white circles in Fig. 2.7(b). In all cases the linear-noise approximation and simulation results agree very well, confirming the existence of stable node NIO in a chemical system which does not admit limit cycle behaviour. It is also noteworthy that the $Q^{99\%}$ measure corresponds well with the more visually pronounced, sharp peaks in the power spectra, which indicates its general usefulness in quantifying the quality of NIO of all types.

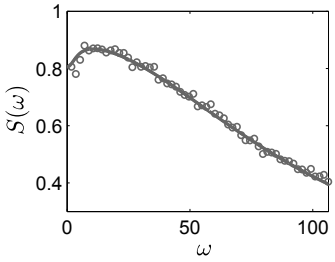
Are oscillations evident in individual time series?

As mentioned in Section 2.3, the maximum $Q^{99\%}$ of a stable node which can be observed in a two species system is ≈ 5. The simple autocatalytic reaction can give rise to NIO with quality approaching this maximum (see Fig. 2.9(a)). However, stochastic



(a) $\Lambda_1 = 0.25, \Lambda_2 = 7.5 \times 10^{-3}$: $Q^{99\%} = 4.55$

(b) $\Lambda_1 = 2.5, \Lambda_2 = 1 \times 10^{-2}$: $Q^{99\%} = 2.04$

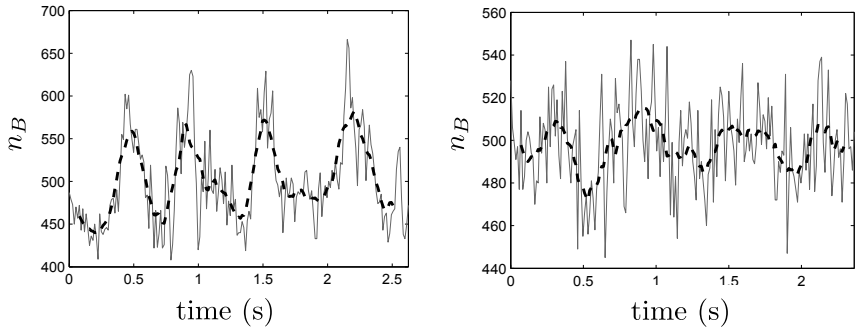


(c) $\Lambda_1 = 20, \Lambda_2 = 2 \times 10^{-3}$: $Q^{99\%} = 1.02$

Figure 2.9: Power spectrum plots of number fluctuations in species B in the autocatalysis reaction system for three different sets of dimensionless parameters for which a stable node steady state exists. Solid lines show the analytical spectrum from the linear-noise approximation; open circles show the numerical spectrum calculated by averaging the periodograms of 5000 realizations of the stochastic simulation algorithm. The parameters common to both cases are: $\Omega = 1 \times 10^{-15} l, k_0 = 1 \times 10^{-6} M\, s^{-1}, k_3 = 100 s^{-1}$. The case specific rate constants are: (a) $k_1 = 2.5 \times 10^{-7} M\, s^{-1}, k_2 = 7.5 \times 10^{7} M^{-1}\, s^{-1}$; (b) $k_1 = 2.5 \times 10^{-6} M\, s^{-1}, k_2 = 1 \times 10^{8} M^{-1}\, s^{-1}$; (c) $k_1 = 2 \times 10^{-5} M\, s^{-1}, k_2 = 2 \times 10^{7} M^{-1}\, s^{-1}$. The unit for frequency ω is $rad\, s^{-1}$.

trajectories of these near-maximum quality NIO, as produced by the stochastic simulation algorithm, give rise to oscillations in the time series data which are generally not easily discernible by the naked eye. An example of the most visually observable NIO that can be expected using the parameters for the high and low $Q^{99\%}$ cases of Fig. 2.9(a) and Fig. 2.9(c), respectively, are presented in Fig. 2.10. An approximation of the visually observable underlying oscillation is represented by a running time average of the data (dotted line). As well as aiding the reader in visualising the shape and regularity of the NIO, we note that this running time average provides a simple measure of the proportion of the variance in the data which is attributable to the visually

discernible oscillatory behaviour. For the high $Q^{99\%}$ and low $Q^{99\%}$ parameters, the oscillation obtained from time averaging accounts for 54% and 22% of the variance of the non-time averaged data respectively. This again supports the quality measure used to describe the weak NIO in this thesis.



(a) (b)

Figure 2.10: Time series plots of the number of B molecules in the autocatalysis reaction system for (a) the kinetic parameters of the high Q case in Fig. 2.9(a); (b) the kinetic parameters of the low Q case in Fig. 2.9(c). For each case 100 realizations of the stochastic simulation algorithm were obtained and the time series with the most visually observable oscillation is here shown. To allow fair comparison, volumes were chosen in each case to give a mean number of B molecules $\langle n_B \rangle = 500$. Based on the characteristic period T found from the peak frequencies of the linear noise approximation spectra in Figs. 2.9(a) and 2.9(c), four period-lengths of data were recorded at time intervals $\Delta t = \frac{1}{50}T$. In each figure, the solid line is the simulated data and the dotted line is a running average of the data over 13 points (approximately one quarter of a period).

It is interesting to compare the time series from these stable-node oscillations with other types of noisy oscillation. Fig. 2.11(a) shows the strong noise-induced oscillations generated by the Autocatalysis model for a particular set of parameters within the stable focus regime. These are virtually as coherent in terms of their regularity in period and amplitude as the 'noise-affected oscillations' generated by the Brusselator in the deterministic limit cycle regime (Fig. 2.11(b)). The similarity of behaviour of these noisy oscillations is notable given that they result from rather different phenomena, i.e., noise induces the oscillations in Fig. 2.11(a) but is deleterious for the regularity of

the wave form in Fig. 2.11(c). Fig. 2.11(c) shows four period-lengths of the time series from the Autocatalysis model when parameters are chosen such that the eigenvalues are complex (stable focus regime) and the resultant $Q^{99\%}$ value for the spectrum of species n_B matches that of the stable node oscillations in Fig. 2.10(a). Again, there is a hint of a visually observable oscillation but it is clearly not easily discernible by the naked eye.

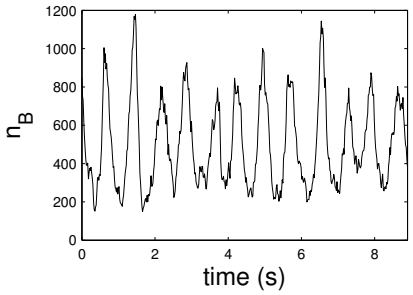
We have identified our weak oscillations by the peaked shape of the fluctuation spectrum at some intrinsic frequency and it is apparent that this definition does not necessarily lead to clearly visible oscillations in the time series. However, we go on in Section 2.5 to show two main points; firstly, that low quality oscillation 'modules' with an intrinsic frequency, such as our stable node oscillators, can work in harmony to create a clearly visible and regular oscillation at that frequency, and secondly that these clearly visible oscillations are indeed possible within the stable node regime.

2.4.3 Example 3: Trimerization Reaction

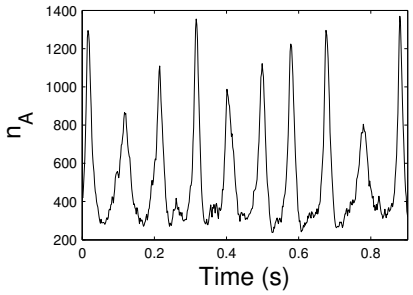
We now study a simple reaction scheme which describes a trimerization process:

$$\emptyset \xrightarrow{k_0} A,\ 2A \xrightarrow{k_1} B,\ A + B \xrightarrow{k_2} \emptyset \tag{2.30}$$

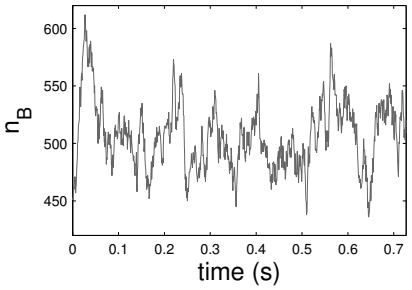
In this scheme the monomer A is produced, binds to another A to form the dimer B, and finally both monomer and dimer bind to form a trimer. The trimer is not explicitly represented in this two-species model; independent of whether it accumulates, decays, or simply exits from the reaction volume, no effect is observed on the behaviour of A or B since the last reaction step is irreversible. This simple reaction is of relevance to various biological situations such as the trimerization of receptor proteins and of



(a) $\Lambda_1 = 7.5 \times 10^{-3}, \Lambda_2 = 7.5 \times 10^{-3}$



(b) $\Lambda_1 = 5 \times 10^{-1}, \Lambda_2 = 1.75$



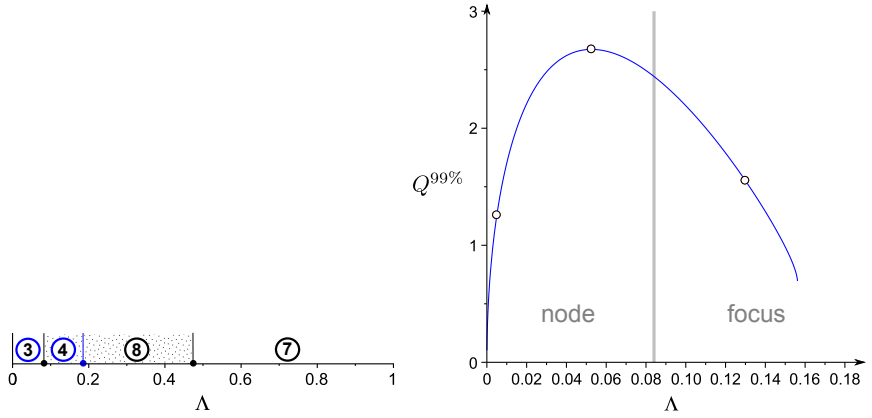
(c) $\Lambda_1 = 5 \times 10^{-1}, \Lambda_2 = 1.823 \times 10^{-1}$: $Q^{99\%} = 4.55$

Figure 2.11: Time series plots showing (a) clearly visible noise-induced oscillations in the number of B molecules for parameters in the stable focus regime of the autocatalysis model (b) clearly visible noise-affected oscillations in the number of A molecules for parameters in the limit cycle regime of the Brusselator model and (c) four period lengths of non-visibly clear oscillations in the number of B molecules in the autocatalysis reaction system for kinetic parameters in the stable focus regime giving $Q^{99\%} = 4.55$, i.e., matching that of the stable node case in Fig. 2.10(a). Volumes were chosen in each case to give a mean number of molecules $\langle n \rangle = 500$.

heat-shock factors [116, 117].

As we shall see the behaviour of this system can be quantified by means of a single non-dimensional parameter:

$$\Lambda = \frac{k_1}{k_2}. \tag{2.31}$$



(a) (b)

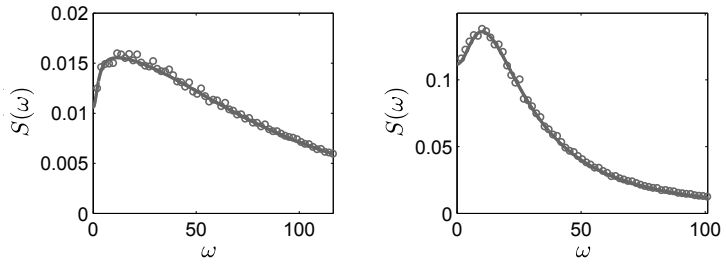
Figure 2.12: (a) NIO existence and stability classification for the trimerization system with the parameter Λ. Number classifications are as in Table 2.1. Blue circles are used to emphasise the existence of a peak in the power spectrum of variable B; the dotted region corresponds to the stable focus regime whereas white regions correspond to the stable node regimes. (b) Variation of the Q factor $Q^{99\%}$ with Λ (regions ③ and ④ in Figure 2.12(a)). Open circles indicated on the $Q^{99\%}(\Lambda)$ curve correspond to the Λ values of the three power spectra in Fig. 2.13.

.

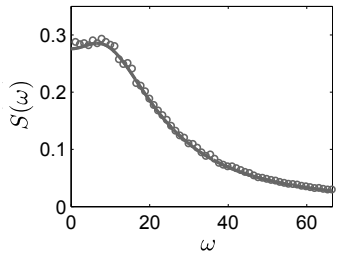
As before from the stoichiometric and macroscopic rate function vector one obtains both the type of steady state and the linear-noise approximation. We here just state the relevant results. The condition for a stable node is:

$$1 + \Lambda(25\Lambda - 14) \geq 0. \tag{2.32}$$

A stable focus is obtained otherwise; similar to the simple autocatalysis scheme previously studied, there is no Hopf bifurcation in the system and hence no deterministic oscillations are possible. The linear-noise analysis gives power spectra of the form



(a) $\Lambda = 0.005$: $Q^{99\%} = 1.26$ (b) $\Lambda = 0.053$: $Q^{99\%} = 2.67$



(c) $\Lambda = 0.13$: $Q^{99\%} = 1.55$

Figure 2.13: Power spectrum plots of number fluctuations in species B in the trimerization reaction system for three different sets of dimensionless parameters for which a stable node steady state exists. Solid lines show the analytical spectrum from the linear-noise approximation; open circles show the numerical spectrum calculated by averaging the periodograms of 2500 realizations of the stochastic simulation algorithm. The parameters common to both cases are: $\Omega = 1 \times 10^{-15} l, k_0 = 1 \times 10^{-6} M\, s^{-1}, k_2 = 1.204 \times 10^{8} M^{-1}\, s^{-1}$. The case specific rate constant is: (a) $k_1 = 6.022 \times 10^{5} M^{-1}\, s^{-1}$; (b) $k_1 = 6.339 \times 10^{6} M^{-1}\, s^{-1}$; (c) $k_1 = 1.566 \times 10^{7} M^{-1}\, s^{-1}$. The unit for frequency ω is $rad\, s^{-1}$.

given by Eq. (2.1) with global parameters $p = 4k_0^2 k_2^2$, $q = k_0 k_2 (1 + \Lambda(25\Lambda - 2))/3\Lambda$ and species-specific parameters given by:

$$\alpha_1 = 4k_0^2 k_2 / 3\Lambda, \tag{2.33}$$

$$\beta_1 = 8k_0/3, \tag{2.34}$$

$$\alpha_2 = 16k_0^2 k_2 \Lambda/3, \tag{2.35}$$

$$\beta_2 = 2k_0/3. \tag{2.36}$$

From these we deduce that the conditions for the existence of NIO in species A and B

are:

$$(\Lambda - 1)^2 < 0, \tag{2.37}$$

$$1 + 4\Lambda - 50\Lambda^2 > 0. \tag{2.38}$$

Note that the first of these two conditions cannot be met and hence there are no NIO in species A; however NIO is possible for species B. Plotting the inequalities Eqs. (2.32) and (2.38) we obtain the phase space diagram for the existence of species B NIO in stable node and focus regions (see Fig. 2.12(a)). The theoretical quality factor of the stable node and stable focus NIO (regions 3 and 4) are shown in Fig. 2.12(b). The maximum quality of this extremely simple system is limited ($Q^{99\%} = 2.67$) in comparison to the simple autocatalysis and Brusselator schemes. A comparison of linear-noise approximation and simulation derived power spectra at three Λ values (whose position in phase space is marked by open circles in Fig. 2.12(b)) are shown in Fig. 2.13; the spectra are in good agreement and as for previous cases confirm the existence of stable node NIO. The simulations also confirm that optimal quality is obtained in the stable node regime, and not in the stable focus regime. This is intriguing given that one would expect noise to generate the largest oscillations by exciting the resonant frequencies of the damped oscillations in the focus regime; furthermore this clearly shows that the quality is not always inversely proportional to the distance from the node-focus borderline. In Fig. 2.12(b) the change in $Q^{99\%}$ from the node-focus transition point (i.e., $\Lambda \approx 0.084$) to the maximum quality point ($\Lambda \approx 0.053$) is approximately 9.4%. Both the peak frequency and the 99% bandwidth increase due to this small move away from the node-focus transition point, but the percentage increase in peak frequency (13.6%) outweighs the percentage increase in the bandwidth (3.8%),

yielding the quality increase. Validation of the increase in quality when moving away from the node-focus transition point is given by inspection of the ratio of maximum power to zero-frequency power, i.e., $\frac{S_B(\widehat{\omega})}{S_B(0)}$; this measure also increases by 8.7% when Λ changes from $\Lambda \approx 0.084$ to $\Lambda \approx 0.053$.

2.4.4 Validation of the $Q^{99\%}$ Measure for Stable Node NIO in Biochemical Systems

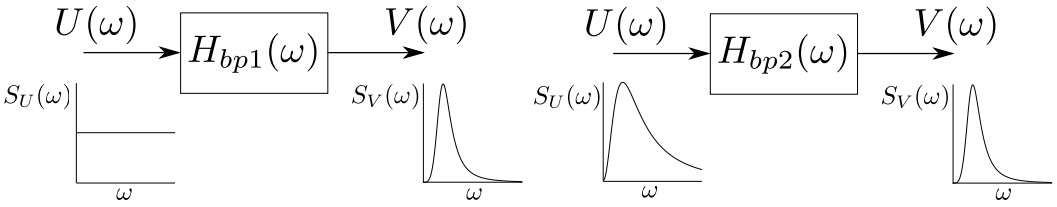


Figure 2.14: From LTI filter theory, the filtering action of the band-pass filter H on an input stationary process with spectrum $S_U(\omega)$ results in an output stationary process with spectrum $S_V(\omega)$. (a) Band-pass filter H_{bp1} with white noise input, (b) A different band-pass filter H_{bp2} whose input process has a spectrum which matches that of a selected variable in one of the parameterized biochemical models. For an output spectrum $S_V(\omega)$ with the same quality in (a) and (b), the required band-pass filter parameter $Q_{bp2} < Q_{bp1}$, i.e., the filter in (b) requires less filtering to achieve an optimal output than the filter in (a) since the input to the filter in (b) is of higher quality than the input to the filter in (a).

We validate the $Q^{99\%}$ measure as follows. Consider a filter whose input signal has the same power spectrum as a selected variable in one of the three biochemical models studied here. Now say that the output of the filter should be such that its power spectrum has some chosen optimal quality factor. It then follows that the lower the quality of the input signal, the larger the degree of filtering needed to be performed by the filter. Hence if the $Q^{99\%}$ measure is reliable then we expect an inverse relationship between it and the degree of filtering. In what follows we now flesh out these ideas using the theory of linear time invariant (LTI) filters.

If a stable LTI system with gain function H takes as its input (U) a stationary process with spectrum $S_U(\omega)$, the output V is also a stationary process with power spectrum given by [118]:

$$S_V(\omega) = [H(\omega)]^2 S_U(\omega).$$

Fig. 2.14(a) shows an illustrative example where the input is stationary white noise and H is the gain function of a simple band-pass filter, given by:

$$H_{bp1}(\omega) = \left| \frac{\hat{\omega}_{bp1}\,\omega}{-Q_{bp1}\,\omega^2 + i\,\hat{\omega}_{bp1}\,\omega + \hat{\omega}_{bp1}^2\,Q_{bp1}} \right|.$$

With this system, the output spectrum $S_V(\omega)$ has a peak at $\omega = \hat{\omega}_{bp1}$ and (conventional) Q-factor given by $Q_V^{50\%} = Q_{bp1}$. Note that this Q-factor is a measure of the degree of filtering that the filter performs.

We validated the $Q^{99\%}$ quality measure of stable node oscillations by considering another LTI band-pass filter system in which the spectrum of the input process is the fluctuation spectrum of a particular species in one of the biochemical models (Fig. 2.14(b)). Specifically, for each of the three biochemical systems evaluated at the parameters in Figs. 2.5, 2.8, 2.9 and 2.13, we considered the problem of obtaining an output quality $Q_V^{50\%} = 1$ by selecting an appropriate value of Q_{bp2} for a band-pass filter described by:

$$H_{bp2}(\omega) = \left| \frac{\hat{\omega}_U\,\omega}{-Q_{bp2}\,\omega^2 + i\,\hat{\omega}_U\,\omega + \hat{\omega}_U^2\,Q_{bp2}} \right|.$$

Note that here we tune the parameter $\hat{\omega}_{bp2}$ to the value of the peak frequency of the particular input fluctuation spectrum, $\hat{\omega}_U$, so that the band-pass filter works constructively with the input to create a higher quality output. If the $Q^{99\%}$ measure is valid then we would expect that for input spectra with high $Q^{99\%}$ values, lower values

of Q_{bp2} would be required to obtain $Q_V^{50\%} = 1$. This was indeed found to be the case, as shown in Fig. 2.15, which validates the use of $Q^{99\%}$.

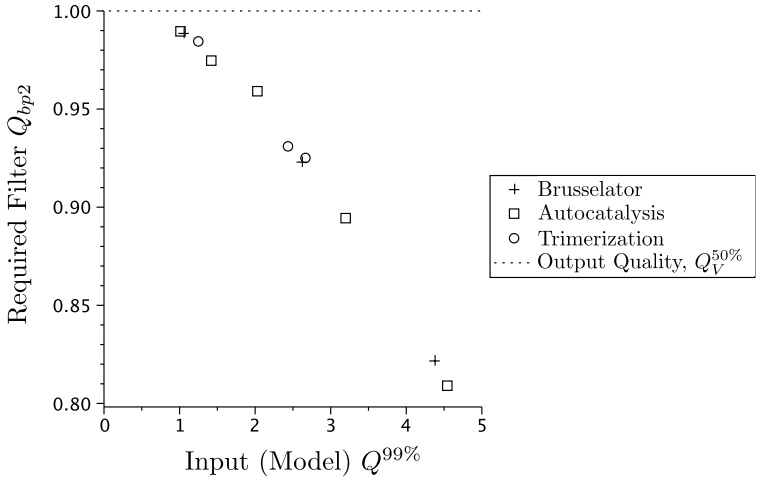


Figure 2.15: With the setup as shown in Fig. 2.14(b) and inputs with spectra from the parameterized biochemical models (Figs. 2.5, 2.8, 2.9, 2.13), the required additional filter quality Q_{bp2} was determined for a 'system' (output) quality of $Q_V^{50\%} = 1$. The figure shows that a biochemical species whose spectrum has a low value of $Q^{99\%}$ (here the input (model) $Q^{99\%}$) requires a greater degree of filtering (larger Q_{bp2} value) for an output $Q_V^{50\%} = 1$. This inverse relationship validates the use of $Q^{99\%}$.

2.5 Stable Node NIO in Cascade Chemical Reaction Systems

As mentioned in Section 2.3, the maximum $Q^{99\%}$ of a stable node which can be observed in a two species system is ≈ 5. In Section 2.4, we saw how both the Brusselator and the simpler autocatalytic reaction can give rise to NIO with quality approaching this maximum (see Fig. 2.5(a) and Fig. 2.9(a)). However stochastic trajectories of these near-maximum quality NIO, as produced by the stochastic simulation algorithm, give rise to oscillations in the time series data which are not easily discernible by the naked eye. In other words, even though the oscillation is present, the noise is so large that it masks the former. In this section we show that for certain classes of chemical systems,

the quality of stable node NIO considerably improves with the number of interacting species, eventually leading to pronounced oscillations in the time series data.

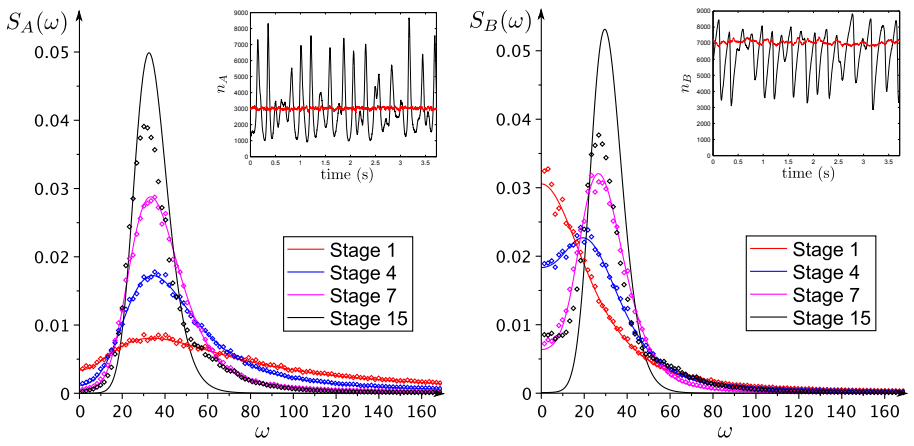
Consider the following reaction scheme:

$$
\begin{aligned}
&\emptyset \xrightarrow{k_0} A_1, \\
&2A_1 + B_1 \xrightarrow{k_1} 3A_1, \quad A_1 \xrightarrow{k_2} B_1, \\
&A_1 \xrightarrow{k_3} A_2, \\
&2A_2 + B_2 \xrightarrow{k_1} 3A_2, \quad A_2 \xrightarrow{k_2} B_2, \\
&A_2 \xrightarrow{k_3} A_3, \\
&..... \\
&2A_N + B_N \xrightarrow{k_1} 3A_N, \quad A_N \xrightarrow{k_2} B_N, \\
&A_N \xrightarrow{k_3} \emptyset.
\end{aligned}
\tag{2.39}
$$

This describes a chain of N downstream-connected Brusselator modules, and was first introduced by Shibata [119] and further investigated by Ramaswamy et al. [120] (a similar model has been studied by means of deterministic rate equations by Tyson and coworkers [108]) .

The cascade is composed of a chain of similar modules connected by an irreversible reaction. Steady-state conditions imply that the influx into a given module equals the outflux from this module to the next. Hence we have the condition $k_3[A_i]^* = k_0$ for $i \in \{1, N\}$, where $[A_i]^*$ is the steady-state concentration of species A_i. At all stages, the steady-state concentrations are the same, i.e., $[A_i]^* = [A_1]^*; [B_i]^* = [B_1]^*$ for all i. However the dynamics of the fluctuations at each stage are very different; for example, Ramaswamy et al. reported a downstream amplification of stable focus NIO [120].

Here we investigate whether this process also leads to a downstream improvement of the quality of stable node NIO.



(a) (b)

Figure 2.16: Normalised power spectrum plots of the number fluctuations in species A (a) and in species B (b) involved in the cascade Brusselator reaction system illustrated in scheme (2.39) with $N = 15$. The parameters are $\Lambda_1 = 0.15$ and $\Lambda_2 = 0.35$, as defined in Eq. (2.8). For selected stages along the cascade reaction, solid lines show the analytical spectrum from the linear-noise approximation (normalised by the total power) and open circles show the normalised numerical spectrum calculated by averaging the periodograms of 750 realizations of the stochastic simulation algorithm and then normalizing by the total power (sum of all spectrum values $\times$ frequency resolution). The inset shows a realization of a time series of the number of A/B molecules at the first stage ($n = 1$, red) and last stage ($n = 15$, black) of the cascade reaction, as obtained from the stochastic simulation algorithm. The unit for frequency ω is $rad\, s^{-1}$.

Fig. 2.16 shows the variation in the normalised power spectra (from the stochastic simulation algorithm and the linear-noise approximation) of A and B molecule fluctuations at different stages along a cascade chain of $N = 15$ Brusselators. At every stage of the cascade, the eigenvalues of the Jacobian of the deterministic rate equations describing this system are -30 and -50 (in units of s^{-1}) and hence the steady state is a stable node. Note how the quality of the stable node NIO improves as it is processed by successive stages, finally leading to remarkably large NIO (insets). Because each stage of the cascade is identical in the rate constants, the dramatic increase in the NIO

quality seen along the cascade can be understood as a resonance resulting from the action of a weak oscillator feeding back its output spectrum to itself, which therefore further excites the natural frequency. The agreement between the spectra obtained from simulations and the linear-noise approximation is generally good, although some deviations can be discerned in the final stage. The drastic improvement in quality is evident for both species (see insets of Fig. 2.16; in Fig. 2.17 we also show a plot of the quality of the NIO in species A versus the cascade stage); this improvement in NIO quality is particularly spectacular for species B since in this case in the first stage there is not even a stable node NIO and yet pronounced stable node NIO ensues at the final stage. Hence it is clear by this example that certain systems with more than 2 interacting species can lead to very high quality stable node NIO.

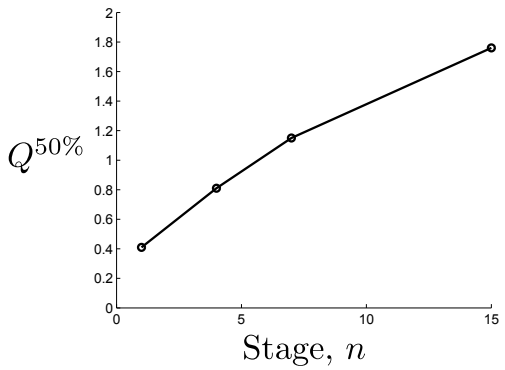
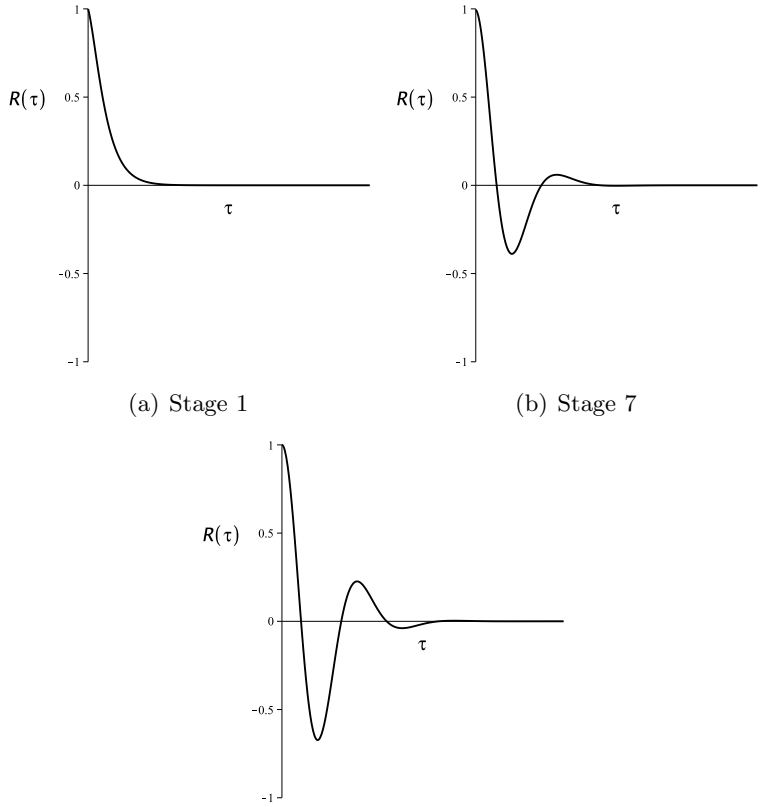


Figure 2.17: Plot of the quality of the spectrum of the number fluctuations of species A versus the stage of the cascade Brusselator reaction system. The quality is calculated using the theoretical linear noise approximation spectra shown in Fig. 2.16(a). Note that the quality of the NIO increases as the noisy oscillatory signal makes its way downstream through the cascade.

Further understanding of the cascade oscillation is obtained by examination of the autocorrelation functions at different stages. In Fig. 2.18 we show that for variable B, which develops a peaked power spectrum along the stages of the cascade, its autocorrelation function has the form of monotonic decay at Stage 1 (expected due to the lack

of peak in the power spectrum) but at further stages (for which we have seen in Fig. 2.16 that the spectrum for variable B becomes peaked) we note that the autocorrelation function also develops a signature akin to damped oscillations. We suggest that this feature of the autocorrelation function may be useful for identifying these stronger types of oscillation in which oscillatory behaviour is visually clear in the time series (Fig. 2.16).



(a) Stage 1

(b) Stage 7

(c) Stage 12

Figure 2.18: Autocorrelation functions $R(\tau)$ for species B at different stages along the cascade Brusselator reaction. The signature akin to damped oscillations for later stages supports the observation of clear oscillations in the time series.

2.6 External versus Internal Noise Induced Stable Node Oscillations

In our investigation of internal noise NIO in exemplary biochemical systems, we observed two counter-intuitive properties: (i) for all of the reaction schemes, it was not possible to have NIO in both species within the stable node region, indicating an unusual type of oscillatory behaviour that we discuss further in Section 2.7.2; (ii) for the trimerization reaction, the maximum quality is attained at a point in parameter space within the stable node region and not the stable focus region. Since in section 2.2 we established that stable node NIO are of the noise dependent type, it is a natural step to question whether the particular constraints placed on the diffusion matrix are fundamental in generating these counter-intuitive properties.

For a two-species system, the diffusion matrix $\mathbf{D}$ (describing either internal or external noise) can be written in the form:

$$\mathbf{D} = \begin{pmatrix} c_1 & c_2 \\ \\ c_2 & c_3 \end{pmatrix} = \sigma_m \begin{pmatrix} \sigma_r^{-1/2} & \sigma_c \\ \\ \sigma_c & \sigma_r^{1/2} \end{pmatrix},$$

where $\sigma_m = \sqrt{c_1 c_3}, \sigma_r = \frac{c_3}{c_1}, \sigma_c = \frac{c_2}{\sqrt{c_1 c_3}}$. The parameter σ_m only scales the fluctuation spectrum in power (y-axis) and has no effect on the frequency composition of the spectrum. The other parameters satisfy the positive semi-definite requirement of $\mathbf{D}$ when $\sigma_r \in (0, \infty)$ and $\sigma_c \in [-1, 1]$.

In biochemical systems with internal noise, as described by the linear noise approximation, the diffusion and Jacobian matrices are in general intimately and non-trivially linked by their dependence on the rate constants (see Eqs. (1.17) and (1.18)). This

coupling means that it is impossible to vary the parameters σ_r and σ_c independently of the underlying form of system stability. However, it is possible to uncouple these matrices and therefore directly study the importance of the precise form of the noise by reframing the biochemical model as one in which internal noise is negligible but external noise is significant.

2.6.1 External Noise Trimerizor Details

Below we provide the details of an approach for introducing external noise for the trimerization reaction. The starting point is to assume that external noise sources introduce stochasticity in the rate constants, i.e., $\widetilde{k_i}(t) = k_i(1 + \epsilon\,\widetilde{\eta_i}(t))$, where $\widetilde{\eta_i}(t)$ is Gaussian white noise and ϵ is a small parameter such that $\epsilon \times \ std(\widetilde{\eta_i}) << 1$. This results in an alternative linear stochastic differential equation for the concentration fluctuations (c.f. Eq. (1.16)) given by:

$$d\vec{\xi^e}(t) = \mathbf{J}\vec{\xi^e}(t)\,dt + \mathbf{B}^e\,d\vec{W}(t), \tag{2.40}$$

where the superscript e denotes external noise. The matrix $\mathbf{J}$ is the same as for internal noise, but in the external noise case the diffusion matrix $\mathbf{D}^e = \mathbf{B}^e.(\mathbf{B}^e)^T$ can be changed independently of $\mathbf{J}$ by changing the external noise parameters.

Consider again the trimerization reaction scheme. When there is no noise of any form, the deterministic rate equations for the concentrations are given by:

$$\partial_t\phi_A(t) = k_0 - 2k_1\phi_A^2(t) - k_2\phi_A(t)\phi_B(t), \tag{2.41}$$

$$\partial_t\phi_B(t) = k_1\phi_A^2(t) - k_2\phi_A(t)\phi_B(t). \tag{2.42}$$

Now consider the case when there is some external noise present in the system. One method of representing external noise is by allowing the rate constants to be stochastic [121]. White Gaussian noise processes $\widetilde{\eta}_i(t)$ are added to the rate constants and scaled by a small parameter, ϵ, so that the new stochastic rate constants read:

$$\widetilde{k}_i(t) = k_i(1 + \epsilon\,\widetilde{\eta}_i(t)), \tag{2.43}$$

where the tilde is used throughout to denote a random variable. It follows that the dynamics in the presence of external noise is described by the stochastic rate equations:

$$\partial_t\widetilde{\phi_A}(t) = \widetilde{k}_0(t) - 2\widetilde{k}_1(t)\widetilde{\phi_A}^2(t) - \widetilde{k}_2(t)\widetilde{\phi_A}(t)\widetilde{\phi_B}(t), \tag{2.44}$$

$$\partial_t\widetilde{\phi_B}(t) = \widetilde{k}_1(t)\widetilde{\phi_A}^2(t) - \widetilde{k}_2(t)\widetilde{\phi_A}(t)\widetilde{\phi_B}(t). \tag{2.45}$$

We solve these equations by making the ansatz:

$$\widetilde{\phi}_i(t) = \phi_i^* + \epsilon\,\widetilde{\xi}_i^e(t), \tag{2.46}$$

where $\phi_A^* = \sqrt{\frac{k_0}{3k_1}}$ and $\phi_B^* = \frac{k_1}{k_2}\sqrt{\frac{k_0}{3k_1}}$ are the steady-state solutions of the deterministic rate equations Eqs. (2.41)-(2.42) in the absence of any noise and $\epsilon\,\widetilde{\xi}_i^e(t)$ represents the stochastic contribution about ϕ_A^*, ϕ_B^*. Collecting terms of order ϵ we obtain:

$$\partial_t \begin{bmatrix} \widetilde{\xi_A^e}(t) \\ \widetilde{\xi_B^e}(t) \end{bmatrix} = \sqrt{\frac{k_0}{3k_1}} \begin{bmatrix} -5k_1 & -k_2 \\ k_1 & -k_2 \end{bmatrix} \begin{bmatrix} \widetilde{\xi_A^e}(t) \\ \widetilde{\xi_B^e}(t) \end{bmatrix} + \frac{k_0}{3} \begin{bmatrix} 3 & -2 & -1 \\ 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \widetilde{\eta_0}(t) \\ \widetilde{\eta_1}(t) \\ \widetilde{\eta_2}(t) \end{bmatrix},$$

or

$$\partial_t \begin{bmatrix} \widetilde{\xi_A^e}(t) \\ \widetilde{\xi_B^e}(t) \end{bmatrix} = \sqrt{\frac{k_0}{3k_1}} \begin{bmatrix} -5k_1 & -k_2 \\ k_1 & -k_2 \end{bmatrix} \begin{bmatrix} \widetilde{\xi_A^e}(t) \\ \widetilde{\xi_B^e}(t) \end{bmatrix} + \begin{bmatrix} \widetilde{w_A}(t) \\ \widetilde{w_B}(t) \end{bmatrix},$$

$$\widetilde{w_A}(t) = \frac{k_0}{3} \left(3\widetilde{\eta_0}(t) - 2\widetilde{\eta_1}(t) - \widetilde{\eta_2}(t)\right),$$
$$\widetilde{w_B}(t) = \frac{k_0}{3} \left(\widetilde{\eta_1}(t) - \widetilde{\eta_2}(t)\right).$$

The symmetric positive semi-definite external noise diffusion matrix $\mathbf{D}^e$ is defined by:

$$\langle \widetilde{w_i}(t)\widetilde{w_j}(t')\rangle = D^e_{i,j}\delta(t-t'),\ i,j = 1,2,$$

where in the above notation $w_1 = w_A$ and $w_2 = w_B$. We obtain this matrix by writing the 3×3 symmetric positive semi-definite covariance matrix $\mathbf{C}$ for the external noise η_0, η_1 and η_2 as:

$$\langle \widetilde{\eta_i}(t)\widetilde{\eta_j}(t')\rangle = C_{i+1,j+1}\delta(t-t'),\ i,j = 0,1,2.$$

As an example, the term $D^e_{1,1} = \langle \widetilde{w_A}(t)\widetilde{w_A}(t')\rangle$ is given by:

$$\begin{aligned} D^e_{1,1} &= \langle \widetilde{w_A}(t)\widetilde{w_A}(t')\rangle, \\ &= \frac{k_0^2}{9}\left(9C_{1,1} - 12C_{1,2} - 6C_{1,3} + 4C_{2,2} + 4C_{2,3} + C_{3,3}\right). \end{aligned}$$

Writing the matrix $\mathbf{D}^e$ in the form:

$$\mathbf{D}^e = \sigma_m \begin{pmatrix} \sigma_r^{-1/2} & \sigma_c \\ \sigma_c & \sigma_r^{1/2} \end{pmatrix},$$

we find that the noise parameters σ_r and σ_c are given as functions only of the $C_{i,i}$ (but in a combination which is specific for the trimerization reaction scheme):

$$\sigma_r = \frac{C_{2,2} - 2\,C_{2,3} + C_{3,3}}{9\,C_{1,1} - 12\,C_{1,2} - 6\,C_{1,3} + 4\,C_{2,2} + 4\,C_{2,3} + C_{3,3}},$$
$$\sigma_c = \frac{3\,C_{1,2} - 3\,C_{1,3} - 2\,C_{2,2} + C_{2,3} + C_{3,3}}{\sqrt{(9\,C_{1,1} - 12\,C_{1,2} - 6\,C_{1,3} + 4\,C_{2,2} + 4\,C_{2,3} + C_{3,3})\,(C_{2,2} - 2\,C_{2,3} + C_{3,3})}}.$$

Varying the external noise parameters, $C_{i,i}$, allows a large range of possible σ_r, σ_c. Given no prior information about the external noise, we might propose that noise in the input reaction parameter k_0 is uncorrelated with the noise in the other parameters, i.e., $C_{1,2} = C_{1,3} = C_{2,1} = C_{3,1} = 0$, but that the other parameters are only constrained so far as to give a positive semi-definite $\mathbf{C}$ matrix. We were able to find noise matrices to satisfy these criteria for all of the σ_r, σ_c values used in the investigation of external noise in Fig. 2.19.

The power spectra of the new stochastic differential equation can be calculated using Eq. (2.1) in the main text. NIO due to external noise are obtained whenever the spectrum displays a maximum at some peak frequency.

2.6.2 Results of External Noise in the Trimerizor

For the simple trimerization reaction, the internal noise forces the constants to take values: $\sigma_r = 1/4$ and $\sigma_c = -1/4$. In contrast, for external noise these constants can take any value provided they satisfy the positive semi-definite constraints mentioned earlier. In Fig. 2.19(a) we show the existence of NIO obtained from external noise for four different values of σ_c, and with σ_r fixed to unity. Note that for internal noise stable node NIO were only possible in one species (see Fig. 2.12(a)), but for external noise it is possible to obtain NIO in both species in the stable node regime (region 5

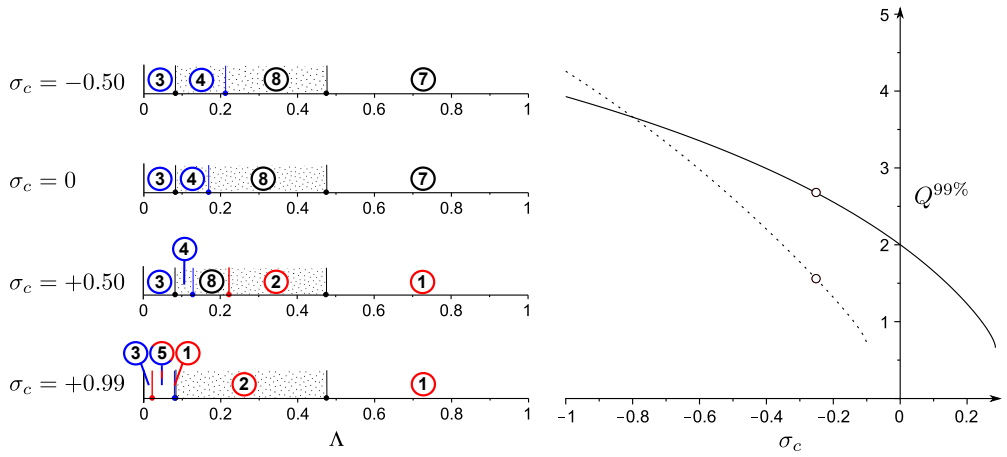


Figure 2.19: Properties of NIO stimulated by external noise in the trimerization reaction. (a) Existence of external noise NIO as a function of the noise coupling strength parameter σ_c with σ_r held constant and equal to unity. When $\sigma_c \leq 0$ there are NIO only in species B, with similar behaviour as for internal noise (see Fig. 2.12(a)). For larger, positive, noise coupling $\sigma_c = +0.50$, NIO in A become possible (region ①); for very large noise coupling $\sigma_c = +0.99$ it is possible to have NIO in both A and B (region ⑤). (b) Quality of external noise NIO as a function of the noise coupling strength parameter σ_c with σ_r held constant and equal to 1/4. The solid line is for the stable node point, $\Lambda \approx 0.053$, and the dotted line is for the stable focus point, $\Lambda = 0.13$. The open circles here correspond directly to the open circles in Fig. 2.12(b) as they indicate the external noise characteristics that exactly match the internal noise case. Note that when $\sigma_c < -0.8$ the quality is higher in the focus regime, i.e., opposite to that observed for internal noise NIO. Note also that the quality factor $Q^{99\%}$ is calculated as the peak frequency divided by the difference of the two frequencies at which the spectrum achieves 99% of its maximum power.

for the case $\sigma_c = +0.99$).

With internal noise we found that the quality at a point in the stable node regime ($Q^{99\%}_{node} \approx 2.7$ when $\Lambda \approx 0.053$) is larger than in the focus regime ($Q^{99\%}_{focus} \approx 1.6$ when

$\Lambda = 0.13$, see Fig. 2.12(b)). In Fig. 2.19(b) we show the variation of $Q^{99\%}_{node}$ and $Q^{99\%}_{focus}$ as a function of the noise coupling parameter σ_c when σ_r is fixed to the same value as for the internal noise ($\sigma_r = 1/4$). Note that when $\sigma_c < -0.8$, $Q^{99\%}_{focus} > Q^{99\%}_{node}$, indicating that at this level of external noise coupling the quality of stable focus NIO is better than stable node NIO, a case opposite to that observed for the reaction stimulated by internal noise. Hence it is clear that the counter-intuitive properties of stable node NIO as induced by internal noise stem from the special form of the diffusion matrix enforced by the linear-noise approximation. It is also interesting that this implies that the origin of noise plays a significant role in determining the properties of NIO. For example, for both the node and focus points, increasing positive noise coupling is observed to dramatically weaken and ultimately destroy the NIO. For the focus point, when $\sigma_c > -0.1$ the NIO are so weak that the power spectrum peak is all-but destroyed and the $Q^{99\%}$ value is undefined; the NIO disappear altogether for $\sigma_c \geq +0.1$. Interestingly, for the node point, NIO with defined $Q^{99\%}$ exist for noise coupling values as large as $\sigma_c \approx +0.3$.

An example of a stable node giving rise to NIO has recently been reported by Qian who showed that for a particular numerical choice of the Jacobian and diffusion matrices, one can obtain a peak in the power spectrum for a stable node (see Fig. 3 of [122]). This example falls within the general category of stable node NIO stimulated by external noise since the Jacobian and diffusion matrices are not constrained by means of the linear-noise approximation. The work herein goes further by developing a general theory of stable node NIO in the presence of both internal and external noise and studying the quality of these noisy oscillations in realistic biochemical models.

2.7 Reflecting on the Challenges in Understanding SNO

2.7.1 Is Nonlinearity Required?

Given the apparent ubiquity of NIO when their existence criterion is simply a peak in the power spectrum, it is interesting to reassess any fundamental requirements for oscillations in single cells. Novak and Tyson argued that, along with other requirements, sufficient nonlinearity is required for a motif to exhibit sustained deterministic oscillations [19]. An example of high nonlinearity given is multimeric transcriptions factor binding to a gene regulatory sequence. A way of testing the importance of nonlinearities for weak NIO is by examining whether NIO are possible for a circuit made of zeroth or first order elementary reactions. A table of meaningful biochemical reactions which fit this requirement is shown below:

Reaction Class	Reaction
Import/Production	$\emptyset \to X$, $\emptyset \to Y$, $\emptyset \to X + Y$.
Export/Decay	$X \to \emptyset$, $Y \to \emptyset$.
Transcription / Translation	$X \to X + Y$
Isoform switching	$X \rightleftharpoons Y$
Dimer splitting	$X \to Y + Y$

Table 2.2: The full set of all linear reactions which could be deemed plausible as intracellular biochemical processes.

We find that when the reaction kinetics are elementary-type (i.e., mass action rates and not complicated functions of the concentrations), any combination of these reac-

tions produces a Jacobian matrix of the form:

$$\mathbf{J} = \begin{pmatrix} - & + \\ + & - \end{pmatrix}.$$

Therefore with only zeroth or first order elementary reactions there is no possibility for negative feedback in the system. The parameter $q > 0$ for this type of Jacobian, so any NIO would rely on the precise form of the noise matrix. We conducted a short investigation of some realistic networks constructed from the above reactions, namely:

- Circuit 1: All import reactions, all export reactions, translation.
- Circuit 2: All import reactions, all export reactions, isoform switching.
- Circuit 3: All import reactions, all export reactions, dimer splitting.

We found that NIO are impossible for either species in each of the three circuits. This implies that nonlinearity, either in the form of higher than first order reactions (e.g., bimolecular) or non-elementary reaction kinetics (e.g., Hill-type repression) is required for NIO. This nonlinearity might be crucial for its role in introducing negative feedback in the system, a notion which is supported by the fact that in each of the three nonlinear models presented here (Brusselator, Autocatalysis, Trimerization) the Jacobian structure has a negative value for one of the terms $J_{1,2}$ or $J_{2,1}$.

2.7.2 Interpretation of the Presence of Oscillations in Only One of Two Species

In elucidating the phenomenon of stable node NIO we have highlighted an intriguing example of the non-intuitive effects of intrinsic, i.e., irremovable, noise in biochemical

systems. Thinking of oscillations in a traditional way, either within a deterministic limit-cycle framework or a stochastic amplification framework, these stable node NIO raise many obvious questions. For one, we might ask what it means to have an oscillation in only one of the two variables? How can this make sense? One might argue that this cannot be a true oscillation, but then this is at odds with the view that the peak in the power spectrum indicates a quasi-periodic behaviour, with the frequency of the oscillation taken as the peak frequency in the spectrum. Indeed, if the oscillation is not genuine, what is the meaning of the power spectrum peak?

First, we ask whether the stable node oscillation should be considered a true oscillation? The conditions of what constitutes an oscillation has been challenged in recent years by the discovery of NIO in biochemical systems [30]. The deterministic limit cycle description declares oscillations to be perfectly sustained and perfectly periodic, but as soon as stochasticity is involved there is clearly no possibility for the behaviour to be perfectly periodic. This is not true even of high quality stochastic oscillations, such as the coherent resonance effect very close to the Hopf bifurcation, and yet these oscillations have been widely accepted.

In Fig. 2.10 we showed example time series of the variable X_1 in the Brusselator system for two sets of parameters which yield a stable node in the rate equations. As discussed above, these segments of the stochastic process were the best time series chosen from a sample set of 100 such realisations, and we argued that since the oscillation is reasonably clear by eye, and the proportion of variance in the signal which describes the oscillatory motion is consistent with our quality measure, there is justification for arguing the case of a noisy stochastic oscillation even in the stable node regime. However, some other realisations show considerably poorer behaviour in terms of visually clear oscillations on the peak frequency timescale, and so it would seem that a two

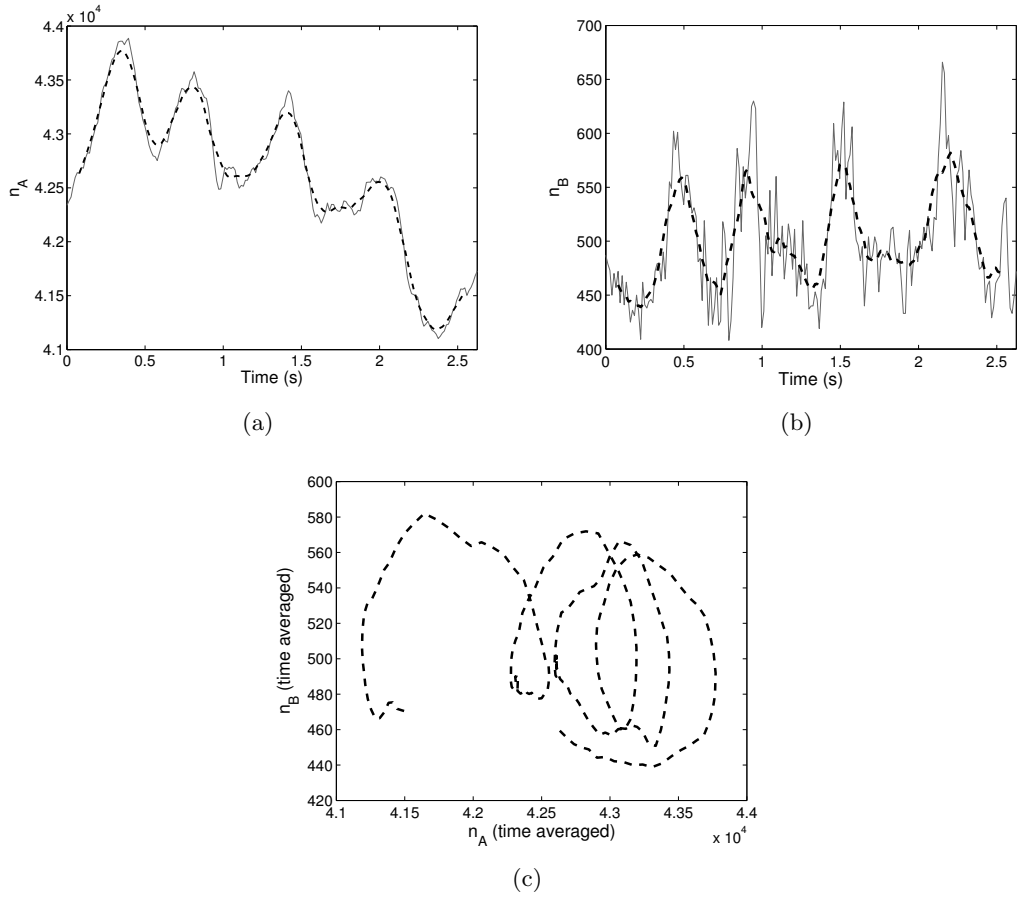


Figure 2.20: Time series and phase portrait of the noise-induced oscillations with the simple autocatalysis model.

species stable node oscillator could not be relied upon as a precise biological timer. Indeed, if the quality of a stochastic oscillations were judged by a visual inspection of time series, our oscillations would likely fail to be considered a 'good' stochastic oscillation. In terms of being a precise biological timer, a two species stable node NIO is not a good candidate. Nevertheless, we observe through the very clear stochastic oscillations of the coupled stable node elements in Fig. 2.16 that the information encoded within the resonant peak of a single, two species stable node oscillator is real and should not be discarded. An alternative way of interpreting the result is that, since the extended coupled system itself possesses only real eigenvalues and does exhibit pronounced and

regular oscillations, one can conclude that for a chemical system with stable node behaviour to behave as a precise biological timer it must involve the interaction of a sufficiently large number of species.

What are we to make of the presence of a resonant peak in one variable and not the other in a two-variable system? Viewed within a deterministic framework this would appear nonsensical, and even in a stochastic framework it is confusing because a one species system cannot exhibit NIO. However, an interesting discovery has helped us to understand this point. We recall that the spectrum is functionally composed of two sub-spectra, one with a peak and the other monotonic decay in frequency. A peak in the total spectrum means that the peaked sub spectrum dominates, and the lack of peak in the total spectrum indicates that the monotonic decay sub spectrum dominates. Using the form of the sub-spectra to infer real physical meaning, we hypothesised that the dominant contribution of the non-peaked sub-spectrum to the fluctuations of the non-oscillatory variable A is representative of low frequency noise that serves to mask any oscillatory behaviour which might be expected due to its interaction with the oscillatory variable B. Indeed, upon examination of the time series of species A in the same time window as the earlier observed oscillatory behaviour of species B in Fig. 2.20(b) (which is a reprint of Fig. 2.10a), species A behaves as in Fig. 2.20(a). The phase portrait of the two variables is shown in Fig. 2.20(c) which, while exhibiting the rotation expected of the stable node oscillation in our cartoon of Fig. 2.2(b), has a noticeable difference. It is evident that the spiralling behaviour in the phase space is akin to an oscillation without a well defined centre of rotation, as the baseline of the variable A wanders slowly. We argue that this low-frequency behaviour dominates any meaningful oscillation in species A in terms of an information carrier to other cellular processes. This is justified since it has previously been argued that removing

high-frequency noise from meaningful signals within cells is far simpler than filtering off low frequency noise [123]. This validates the notion that only one species should be considered oscillatory and justifies the importance that we have placed on identifying the single oscillatory biochemical species across the parameter space for the models in this chapter.

2.8 Discussion and Conclusion

In this chapter, we have shown that NIO can be induced by internal noise in biochemical systems characterized by deterministic stable node steady states. This phenomenon goes beyond the conventional well known case in which NIO are induced by noise for systems with a deterministic focus steady state since stable nodes do not possess an intrinsic frequency which can be stimulated by white noise. Rather the frequency of stable node NIO is determined by the timescales characterizing the non-oscillatory decay of perturbations in stable nodes; in particular for two species systems, the frequency is bounded from above by the geometric mean of the two real eigenvalues of the Jacobian.

These stable node NIO possess properties which are counter-intuitive. For example, for deterministic oscillatory systems and for stable focus NIO close to the Hopf bifurcation oscillations are present in all interacting species, but for stable node NIO we find that this is not generally the case. We note that for internal noise, the linear-noise approximation enforces a complicated dependence of the diffusion matrix on the elements of the Jacobian matrix and on the stoichiometric matrix and that this could be the origin of the counter-intuitive property delineated above. This line of thought is suggested by the fact that if the elements of the diffusion matrix could be freely chosen then the counter-intuitive property can be eliminated by appropriate choices (Section

2.6). This indeed turns out to be the fundamental reason why stable node NIO stimulated by external noise may have different properties than those stimulated by internal noise since the former (unlike the latter) is characterized by a diffusion matrix which can take any values, provided the matrix is positive semi-definite.

It is also the case that one would expect NIO for stable foci to be of better quality than stable node NIO because in the former there exists a clear internal frequency which can be stimulated by white noise while in the latter there is not. However we found that this is not generally the case: there are stable foci regions in parameter space where there are no NIO (region 8 in Figs. 2.4(a), 2.7(a) and 2.12(a)) whereas there are stable node regions where NIO are present (regions 1 and 3 in Figs. 2.4(a), 2.7(a) and 2.12(a)). For the trimerization reaction we found that the NIO quality maximized in the node region and subsequently decreased as one approached the node-focus borderline (Fig. 2.12(b)). Furthermore as we saw for the cascade reaction system of Section 2.5, the quality of stable node NIO increases as the noisy signal makes its way downstream through the network eventually leading to massive NIO, similar to those previously observed for stable foci [30]. In this multi-species cascade model, there are many developments which could be made to the model which are worthy of further investigation. In particular, it would be of interest to study the effect on NIO when there is variability in the cascade modules, and also to look at the effect of other forms of coupling, e.g. bi-directional coupling between stages or coupling stage N to stage 1 and forming a cyclical reaction system.

We have also shown that stable node NIO do occur for chemical systems with mean molecule numbers ranging from the order of one to a few thousand (see Figs. 2.6 and 2.16), i.e., the physiologically relevant range [42]. Furthermore we have identified two chemical systems composed of at most bimolecular reactions and each involving merely

two species, in which noise induces oscillations; this is in contrast to the well known deterministic result that at least three species are needed to obtain limit cycle oscillations from elementary reaction models [124, 24]. It is however the case that oscillations for two species systems are very noisy and their quality improves considerably as the number of interacting species increases.

We conclude from the work in this chapter that noise can induce oscillations over larger regions of parameter space and for simpler chemical systems than currently thought, hence further amplifying the current trend of thought that noise plays a constructive and essential role in cellular regulation [125, 126].

Chapter 3

Effects of Bursty Protein Production on the Oscillatory Properties of Downstream Pathways

3.1 Introduction

In Chapter 2, we discovered that biochemical systems with stable node stability could give rise to NIO. Using internal noise, the NIO were present in one of the two species, and this was identified as the relevant information carrier for further cellular processes. While it was observed that the presence and quality of stable node oscillations could be controlled by selecting specific external noise parameters, the choices of these parameters were left unmotivated from physical principles. In this Chapter we find that there is an important biological phenomenon which, when modelled in terms of an internal

noise framework, allows the tight coupling between the Jacobian and diffusion matrices to be relaxed, resulting in a physically motivated mechanism whereby cells can control the noise dependent NIO.

Given a particular biochemical pathway of interest, noise can be further categorized as that coming from sources external to the pathway and that originating from the individual reactions constituting the pathway. A ubiquitous source of external noise is the mechanism by which molecules are input or injected into a biochemical pathway. The classic model for this is a Poisson process in which a single molecule is injected at random points in time. However, numerous experimental studies over the past decade have shown that such a description is often inaccurate [127, 128, 129, 130, 131].

Injection events have at least two physical interpretations for models of intracellular dynamics; injection can describe protein expression when modelling a biochemical pathway in the cytosol, whereas for pathways in membrane-bound subcellular compartments injection events can describe transport of molecules into the compartment by diffusive or active transport. A number of studies have confirmed that protein expression occurs in sharp and random bursts [127, 128]. The bursts are found to be exponentially distributed and the expression events are temporally uncorrelated. The origin of these bursts can be explained by a simple mechanism. For bacteria and yeast, the lifetime of mRNA is typically short compared to that of proteins. In its short lifetime, each mRNA is translated into a number of protein molecules leading to random uncorrelated bursty events of protein production [129]. Although such protein expression is the best studied example of burstiness in protein production, it is not the only one. It has recently been found that protein translocation to the nucleus in response to an extracellular stimulus in yeast also occurs in sharp bursts; indeed these bursts may be even more influential than those in protein expression since the mean size of the

translocation bursts are estimated to be hundreds of molecules whereas those stemming from protein expressions are of the order of few tens or less [131, 128].

It is interesting to ponder what effects burstiness in protein production has on the steady-state properties and dynamics of the downstream biochemical pathway into which it feeds. Intuitively, a bursty input mechanism introduces a larger degree of noise to the downstream pathway than a non-bursty one. Indeed, this increase in noise has been quantified in very simple scenarios where the downstream pathway involves protein decay via a first-order process; for a bursty production mechanism, it was found that the Fano factor (variance of number fluctuations divided by the mean of molecule numbers) is equal to 1 plus the mean burst size, whereas for a non-bursty mechanism the Fano factor is 1 [131, 121]. It is expected that this noise amplification occurs for all species' concentrations in more complex downstream pathways; from this point of view, bursting appears to be deleterious to the precise orchestration of cellular function. Consequently one might expect the cell to have developed downstream mechanisms to reduce or control such unwanted noise.

In this chapter we challenge this notion by demonstrating the non-intuitive effects of bursty inputs on noise-induced concentration oscillations. We compare the stochastic properties of two identical biochemical pathways, in one of which the protein is produced via a non-bursty input mechanism and in the other via a bursty input mechanism where the number of molecules per burst are distributed according to a general probability distribution. The mean rates of protein production are chosen to be the same in the two pathways and hence, according to the deterministic rate equation formalism, the two systems are characterized by the same steady-state concentrations. However, we show that there exist pronounced differences in the noise-induced oscillations generated by bursty and non-bursty systems, and that the crucial non-dimensional parameter

distinguishing the two is the sum of the mean and the Fano-factor of the burst size probability distribution. By deriving necessary conditions for noise induced oscillations in the two systems we demonstrate a method of classifying simple biochemical circuits by their response to input bursting.

3.2 A General Framework for Assessing the Effects of Bursty Protein Production

In this section we introduce the two-system setup which we will use to study the effects of bursting on the fluctuations in downstream pathways.

Consider a two species system in which both species are injected into a pathway, and subsequently interact via a number R of downstream reactions. The non-bursty version of this system can be schematically represented as:

$$\begin{aligned} &\emptyset \xrightarrow{h_1} X_1, \\ &\emptyset \xrightarrow{h_2} X_2, \\ &s_{1j}X_1 + s_{2j}X_2 \xrightarrow{k_j} r_{1j}X_1 + r_{2j}X_2, \quad j \in [1, R] \end{aligned} \tag{3.1}$$

where X_i denotes species i, s_{ij} and r_{ij} $(i = 1, 2)$ are the integer stoichiometric coefficients and h_j and k_j are the associated rate constants of the j^{th} input and j^{th} processing (downstream) reaction respectively.

A bursty input version of this system can also be envisaged as follows:

$$\begin{aligned}
&\emptyset \xrightarrow{h_1^* q_1(0)} 0X_1, \quad \emptyset \xrightarrow{h_1^* q_1(1)} X_1, \quad \emptyset \xrightarrow{h_1^* q_1(2)} 2X_1, \quad ....., \quad \emptyset \xrightarrow{h_1^* q_1(M_1)} M_1X_1, \\
&\emptyset \xrightarrow{h_2^* q_2(0)} 0X_2, \quad \emptyset \xrightarrow{h_2^* q_2(1)} X_2, \quad \emptyset \xrightarrow{h_2^* q_2(2)} 2X_2, \quad ....., \quad \emptyset \xrightarrow{h_2^* q_2(M_2)} M_2X_2, \\
&s_{1j}X_1 + s_{2j}X_2 \xrightarrow{k_j} r_{1j}X_1 + r_{2j}X_2, \quad j \in [1, R],
\end{aligned} \tag{3.2}$$

where $q_i(m)$ is the probability that the input burst size is m for species X_i and h_i^* is a proportionality constant such that $h_i^* q_i(m)$ is an input rate constant.

For the above two systems chemical master equations take the respective form:

$$\begin{aligned}
\frac{\partial P^s(n_1, n_2, t)}{\partial t} =& \Omega \left[(E_1^{-1} - 1)h_1 + (E_2^{-1} - 1)h_2 \right] P^s(n_1, n_2, t) + \\
& \Omega \sum_{j=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ij}} - 1 \right) \hat{f}_j(n_1, n_2, \Omega) P^s(n_1, n_2, t),
\end{aligned} \tag{3.3}$$

$$\begin{aligned}
\frac{\partial P^b(n_1, n_2, t)}{\partial t} =& \Omega \left[\sum_{m=0}^{M_1} (E_1^{-m} - 1)h_1^* q_1(m) + \sum_{m=0}^{M_2} (E_2^{-m} - 1)h_2^* q_2(m) \right] P^b(n_1, n_2, t) + \\
& \Omega \sum_{j=1}^{R} \left(\prod_{i=1}^{N} E_i^{-S_{ij}} - 1 \right) \hat{f}_j(n_1, n_2, \Omega) P^b(n_1, n_2, t),
\end{aligned} \tag{3.4}$$

where $P^{b/s}(n_1, n_2, t)$ is the probability that there are n_1 molecules of species X_1 and n_2 molecules of species X_2 at time t for the bursty input (b) and non-bursty, i.e., single-molecule input (s) systems, Ω is the volume of the compartment in which the downstream pathway operates, $S_{ij} = r_{ij} - s_{ij}$ are the elements of the stoichiometric matrix, E_i^j is the step operator which when it acts on some function $w(n_i)$ gives $w(n_i+j)$ and $\hat{f}_j$ is the microscopic rate function for the j^{th} processing reaction which is given by [14]:

$$\hat{f}_j(n_1, n_2, \Omega) = k_j \Omega^{-(s_{1j}+s_{2j})} \frac{n_1!}{(n_1 - s_{1j})!} \frac{n_2!}{(n_2 - s_{2j})!}. \tag{3.5}$$

Note that the first term in each of the master equations above describes the input reactions while the second term describes the processing reactions.

As in the previous Chapter we circumvent the problem of the master equations being unsolvable by using the linear noise approximation (LNA), and use the resultant Fokker-Planck equation, with linear drift and diffusion coefficients, to describe the stochastic dynamics. Note that here we prefer the Fokker-Planck representation over the Langevin equation representation simply for clarity of the effect of burstiness on the diffusion matrix. Since we have explained the LNA in previous sections, we now simply state the relevant results of applying the LNA to the master equations (3.3) and (3.4).

Within the LNA, the time evolution of the mean concentrations for the two systems is given by the conventional rate equations:

$$\frac{d\phi_{1,s}}{dt} = h_1(\vec{\phi_s}) + g_1(\vec{\phi_s}), \qquad \frac{d\phi_{2,s}}{dt} = h_2(\vec{\phi_s}) + g_2(\vec{\phi_s}), \tag{3.6}$$

$$\frac{d\phi_{1,b}}{dt} = h_1^*(\vec{\phi_b})\mu_1 + g_1(\vec{\phi_b}), \quad \frac{d\phi_{2,b}}{dt} = h_2^*(\vec{\phi_b})\mu_2 + g_2(\vec{\phi_b}), \tag{3.7}$$

where the first terms describe the input reactions and the second terms describe the processing reactions. The vector $\vec{\phi}_{b/s} = (\phi_{1,b/s}, \phi_{2,b/s})$ is the concentration vector for bursty (b) and non-bursty input (s) systems. The processing rates are given by $g_i(\phi_{1,b/s}, \phi_{2,b/s}) = \sum_{j=1}^{R} S_{ij} k_j \phi_{1,b/s}^{s_{1j}} \phi_{2,b/s}^{s_{2j}}$. The factors μ_1 and μ_2 are the mean input burst size for species X_1 and X_2 respectively, i.e., $\mu_i = \sum_{m=0}^{M_i} m q_i(m)$. Note that the input rates h_i and h_i^* may be constants, e.g., when modelling diffusive transport, or functions of the concentrations, e.g., when modeling repression or activation of gene transcription.

The LNA analysis also shows that the Fokker-Planck equations describing the prob-

ability distribution $\pi^{s/b}(\epsilon_1, \epsilon_2, t)$ of concentration fluctuations about the mean concentration solution of the above rate equations are given by:

$$\frac{\partial \pi^s(\epsilon_1,\epsilon_2,t)}{\partial t} = -\sum_{i,k=1}^{2} J^s_{ik}(\vec{\phi_s})\frac{\partial}{\partial \epsilon_i}\epsilon_k \pi^s(\epsilon_1,\epsilon_2,t) + \frac{\Omega^{-1}}{2}\sum_{i,k=1}^{2}[\delta_{ik}h_i(\vec{\phi_s}) + D^0_{ik}(\vec{\phi_s})]\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_k}\pi^s(\epsilon_1,\epsilon_2,t), \tag{3.8}$$

$$\frac{\partial \pi^b(\epsilon_1,\epsilon_2,t)}{\partial t} = -\sum_{i,k=1}^{2} J^b_{ik}(\vec{\phi_b})\frac{\partial}{\partial \epsilon_i}\epsilon_k \pi^b(\epsilon_1,\epsilon_2,t) + \frac{\Omega^{-1}}{2}\sum_{i,k=1}^{2}[\delta_{ik}h^*_i(\vec{\phi_b})\lambda_i + D^0_{ik}(\vec{\phi_b})]\frac{\partial^2}{\partial\epsilon_i\partial\epsilon_k}\pi^b(\epsilon_1,\epsilon_2,t), \tag{3.9}$$

where ϵ_i is the noise about the mean concentration of species X_i, $\lambda_i = \sum_{m=0}^{M_i} m^2 q_i(m)$ is the second moment of the distribution of bursts, $J^b_{ik}(\vec{\phi_b}) = \partial/\partial\phi^b_k[h^*_i(\vec{\phi_b})\mu_i + g_i(\vec{\phi_b})]$, $J^s_{ik}(\vec{\phi_s}) = \partial/\partial\phi^s_k[h_i(\vec{\phi_s}) + g_i(\vec{\phi_s})]$ are the Jacobian matrices (describing linear stability) of the two rate equations above and $D^0_{ik}(\vec{\phi}_{b/s}) = \sum_{j=1}^{R} S_{ij}S_{kj}k_j\phi^{s_{1j}}_{1,b/s}\phi^{s_{2j}}_{2,b/s}$.

Clearly the bursty and non-bursty systems have generally different mean concentrations and different fluctuations. To more easily and fairly compare the stochastic properties for these two systems we enforce the condition that in each system the mean number of molecules injected per unit time is the same. The mathematical conditions leading to this scenario can easily be deduced from Eqs. (3.6) and (3.7) to be $h_1 = h^*_1\mu_1$, $h_2 = h^*_2\mu_2$. Given these conditions, as well of course as the same initial conditions, both systems have the same mean concentrations at all times. However, while these conditions force the two rate equations to become the same, they do not force an equality of the Fokker-Planck equations Eqs. (3.8) and (3.9), i.e., the two systems have the same mean concentrations but different fluctuations.

An inspection of Eqs. (3.8) and (3.9) shows that when the mean concentrations of

the two systems are the same, the crucial set of non-dimensional parameters determining the differences in the fluctuations between the bursty and non-bursty input systems are:

$$\eta_i = \frac{h_i^* \lambda_i}{h_i} = \frac{\lambda_i}{\mu_i} = \frac{\sigma_i^2}{\mu_i} + \mu_i, \quad i = 1, 2 \tag{3.10}$$

where σ_i^2 is the variance of the probability distribution of the bursts in species X_i. When $\eta_i = 1$ then the differences between the Fokker-Planck equations for the bursty and non-bursty input systems vanish. As expected, this occurs in the limit that the variance approaches zero and the mean burst size is one. As discussed in Section 3.1, experiments show that the mean burst size is larger than one and hence we shall exclusively consider $\eta_i > 1$. The implication of equation (3.10) is that the larger is $\eta_1 - 1$, the larger are the expected differences in the fluctuation properties of the downstream pathways in the two systems. For example for a Poissonian distribution of bursts, it is found $\eta_i - 1 = \mu_i$ whereas for a geometric distribution of bursts we have $\eta_i - 1 = 2\mu_i$ and hence we expect the burstiness-induced effects to be more prominent for systems with the latter burst input.

We finish this section by noting that we now have a convenient analytical setup with which to understand the effects of burstiness on the fluctuations of a downstream pathway. Since the bursty and non-bursty input systems have the same vector of mean concentrations and the same Jacobian, in what follows we shall denote these as $\vec{\phi}$ and $\mathbf{J}$ respectively, for both systems. In the next sections we use the Fokker-Planck equations to understand how the fluctuations from bursty input change the downstream pathway's ability to generate oscillations.

3.2.1 Necessary Conditions for Bursty Input Alteration of the Oscillatory Properties of the Downstream Pathway

In the setup described in the previous section, the bursty and non-bursty input systems are indistinguishable from a rate equation perspective and hence it follows that the deterministic dynamics of the two systems, including their ability to generate deterministic oscillations (limit cycles) are one and the same. However it is well appreciated that noise can induce oscillations in systems whose rate equations predict none. Given that the noise in the bursty and non-bursty input systems is different, it is plausible that the noise-induced oscillations displayed by both systems can also be different. In what follows we use the Fokker-Planck equations of the last section to probe this question.

We consider a general two variable Fokker-Planck equation with linear drift and diffusion coefficients of the form:

$$\frac{\partial \pi(\epsilon_1,\epsilon_2,t)}{\partial t} = -\sum_{i,k=1}^{2} J_{ik}(\vec{\phi})\frac{\partial}{\partial \epsilon_i}\epsilon_k \pi(\epsilon_1,\epsilon_2,t) + \frac{\Omega^{-1}}{2}\sum_{i,k=1}^{2} D_{ik}(\vec{\phi})\frac{\partial^2}{\partial \epsilon_i \partial \epsilon_k}\pi(\epsilon_1,\epsilon_2,t) \quad (3.11)$$

In the previous chapter we showed that the information contained in equation 3.11 is sufficient to derive an equation for the power spectrum of the fluctuations in the number of molecules of species X_i ($\Omega\epsilon_i$), and that this is found to be:

$$S_i(\omega) = \frac{\Omega}{\pi}\frac{\alpha_i(\mathbf{J},\mathbf{D}) + \beta_i(\mathbf{D})\,\omega^2}{p(\mathbf{J}) + q(\mathbf{J})\,\omega^2 + \omega^4}, \quad (3.12)$$

where $\alpha_1 = J_{12}^2 D_{22} - 2D_{12}J_{12}J_{22} + D_{11}J_{22}^2$, α_2 is the same as α_1 but with 1 and 2 interchanged, $\beta_i = D_{ii}$. We emphasised in Chapter 2 that the terms p and q depend only on the eigenvalues λ_1, λ_2. Alternatively we can write $p = [Det(\mathbf{J})]^2$ and $q = [Tr(\mathbf{J})]^2 - 2Det(\mathbf{J})$ where Tr and Det refer to the matrix trace and determinant.

By comparing Eqs. (3.8) and (3.9) with the general form equation (3.11), we can deduce that the power spectrum of the fluctuations in the non-bursty and non-bursty input systems are given by equation (3.12) with $D_{ik} = \delta_{ik}h_i + D^0_{ik}$ and $D_{ik} = \delta_{ik}h_i\eta_i + D^0_{ik}$ respectively. These two spectra we shall refer to as $S^s_i(\omega)$ and $S^b_i(\omega)$ respectively.

By differentiating $S^s_i(\omega)$ and $S^b_i(\omega)$ with respect to ω, we find the sufficient conditions for the power spectra to have a maximum, i.e., for the two systems to exhibit noise-induced oscillations. If $q < 0$, it can be shown that both $S^s_i(\omega)$ and $S^b_i(\omega)$ display a peak in their power spectrum; hence in this case burstiness does not lead to any qualitative change in the oscillatory properties of the downstream pathway. However for $q > 0$ the situation is more interesting. The positive q condition describes downstream pathways which, when parameterized, are far from a Hopf bifurcation [132]; the steady state is described by a node (which satisfies $Tr[\mathbf{J}]^2 > 4Det(\mathbf{J})$) or by a focus close to the node-focus borderline in phase space (which satisfies $2Det(\mathbf{J}) < Tr[\mathbf{J}]^2 \leq 4Det(\mathbf{J})$). In this case (i.e., $q > 0$) the conditions for noise-induced oscillations in the concentration of species X_1 in the non-bursty and bursty input systems are different and given by

$$\theta_1 > \theta^s_1 = \frac{J^2_{12}}{h_1 + D^0_{11}}\left(h_2 + D^0_{22} - \frac{2D^0_{12}J_{22}}{J_{12}}\right), \tag{3.13}$$

$$\theta_1 > \theta^b_1 = \frac{J^2_{12}}{h_1\eta_1 + D^0_{11}}\left(h_2\eta_2 + D^0_{22} - \frac{2D^0_{12}J_{22}}{J_{12}}\right), \tag{3.14}$$

respectively. The parameter θ_1 is a function of the Jacobian elements only and is given by:

$$\theta_1 = \left(\frac{[Det(\mathbf{J})]^2}{((Tr[\mathbf{J}])^2 - 2Det(\mathbf{J}))} - J^2_{22}\right). \tag{3.15}$$

The conditions for noise-induced oscillations in species X_2 are as above but with 1 and 2 interchanged. Note that although not explicitly shown, the elements of the **D** and **J**

matrices in Eqs. (3.14)-(3.15) are functions of the mean concentration vector $\vec{\phi}$.

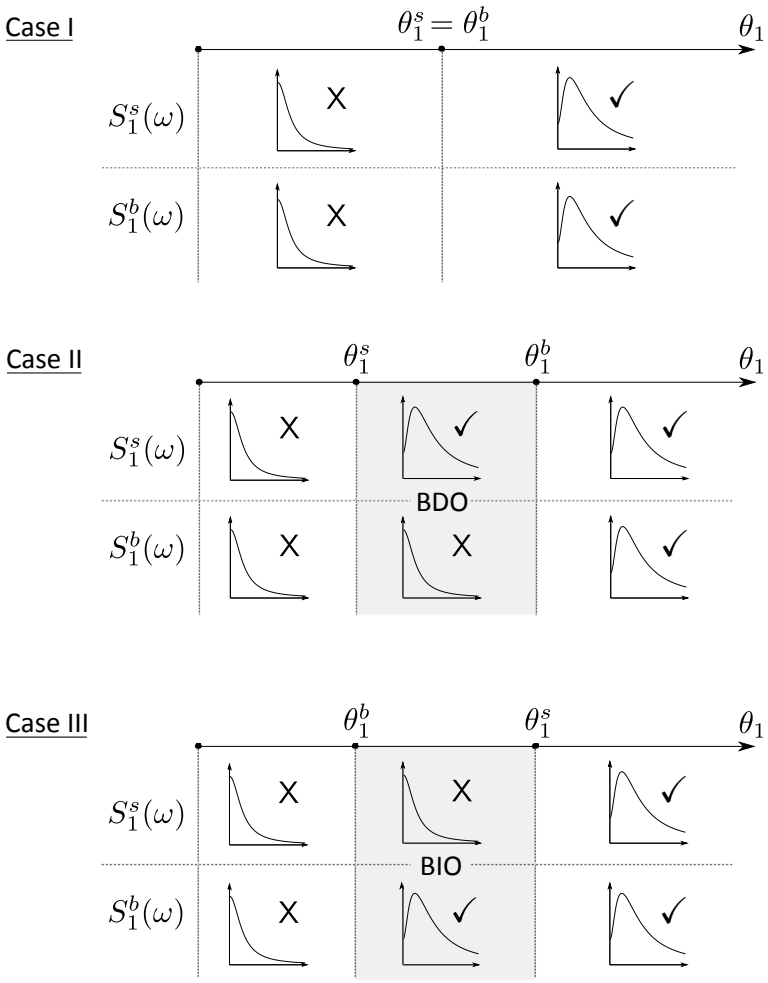


Figure 3.1: The impact of burstiness in the input reaction on the oscillatory properties of species X_1 in a two species downstream pathway. The figure illustrates the three possibilities: Case I in which both bursty and non-bursty input systems have qualitatively similar oscillatory properties, i.e., both $S_1^b(\omega)$ and $S_1^s(\omega)$ have a peak or not; Case II in which there is the possibility that the non-bursty input system displays noise-induced oscillations while the bursty input system does not (peak in $S_1^s(\omega)$ only); Case III in which there is the possibility that the bursty input system displays noise-induced oscillations while the non-bursty input system does not (peak in $S_1^b(\omega)$ only). The parameters θ_1^b, θ_1^s and θ_1 are defined in the main text by Eqs. (3.14)–(3.15) respectively. Burstiness in the input reactions has no effect in Case I, is deleterious in Case II (BDO – bursting destroys oscillations) and promotes noise-induced oscillations in Case III (BIO – bursting induces oscillations).

Hence for $q > 0$ we can identify three distinct cases: (i) $\theta_1^b = \theta_1^s$, (ii) $\theta_1^b > \theta_1^s$ and (iii) $\theta_1^b < \theta_1^s$. These are illustrated in Fig. 3.1. For case (i) either both systems display

no oscillations or they both show noise-induced oscillations. For case (ii), there is the possibility of a special regime ($\theta_1^s < \theta < \theta_1^b$) in which the non-bursty input system displays noise-induced oscillations but the bursty input system does not. For case (iii), there is the possibility of a special regime ($\theta_1^b < \theta < \theta_1^s$) in which the non-bursty input system displays no oscillations but the bursty input system exhibits noise-induced oscillations. Hence burstiness has no effect in case (i), may cause destruction of noise-induced oscillations in case (ii) and may promote noise-induced oscillations in case (iii).

Note that $\theta_1^b > \theta_1^s$ is only a necessary condition for the destruction of noise-induced oscillations by burstiness in the input reactions; sufficient conditions ensue when we further have ($\theta_1^s < \theta_1 < \theta_1^b$) which may not be always possible to satisfy. Similarly $\theta_1^b < \theta_1^s$ should be construed as a necessary condition for the creation of noise-induced oscillations by burstiness in the input reactions.

By inspection of Eqs. (3.14)-(3.13), we can make further specific statements regarding the importance of burstiness in the input reactions to the oscillatory dynamics of the two species pathway:

- If the species X_2 does not activate or inhibit X_1, i.e., $J_{12} = 0$, then $\theta_1^b = \theta_1^s = 0$ and hence burstiness in the inputs of X_1 or X_2 do not cause a qualitative change in the oscillatory dynamics of X_1. Thus it is clear that bursting on its own is insufficient to affect oscillatory dynamics, rather an interplay of bursting with a downstream pathway featuring the interaction of two or more species is required.

- For all other (i.e., $J_{12} \neq 0$) pathways, an increase in the input burstiness of species X_2 (for example by increasing the variance of the burst fluctuations at constant burst size mean) always increases the term $\theta_1^b - \theta_1^s$. Thus, since it is possible to

induce the condition $\theta_1^b > \theta_1^s$ but not the condition $\theta_1^s > \theta_1^b$, bursting in species X_2 may destroy noise-induced oscillations in species X_1 but can never promote noise-induced oscillations in species X_1.

- For pathways that obey the condition $D_{22}^0 - 2D_{12}^0 J_{22} J_{12}^{-1} > 0$, an increase in the input burstiness of species X_1 decreases the term $\theta_1^b - \theta_1^s$. Thus, since it is possible to induce the condition $\theta_1^s > \theta_1^b$, bursting in species X_1 for these pathways may promote noise-induced oscillations in species X_1. An exemplary class of such pathways are those in which the reactions are stoichiometrically uncoupled ($D_{12}^0 = 0$) but kinetically coupled ($J_{12} \neq 0$), i.e., in each reaction there is only a net change in the number of molecules of one species yet the kinetics of the two species are coupled [17].

In the next section we investigate the effects of input bursting in exemplary biochemical circuits, in particular verifying our theoretical prediction that burstiness in the input reactions can both promote and destroy noise-induced oscillations far from the Hopf bifurcation. We conclude here by highlighting a simple four point recipe which can be used to calculate the necessary conditions derived in this section for any two species biochemical circuit:

Four step recipe for calculating the necessary conditions for burstiness-induced effects in a two-species system

Step 1. By comparison of the particular system under study with the general form of the non-bursty input system described by scheme (3.1), one deduces the stoichiometric coefficients s_{ij} and r_{ij} and constructs the elements of the stoichiometric matrix $S_{ij} = r_{ij} - s_{ij}$ of the downstream pathway where i varies between 1 and 2 and j varies

between 1 and the total number R of downstream reactions.

Step 2. Write down the rate equations $d\phi_1/dt = h_1(\vec{\phi}) + g_1(\vec{\phi}), d\phi_2/dt = h_2(\vec{\phi}) + g_2(\vec{\phi})$ where $g_i(\vec{\phi}) = \sum_{j=1}^{R} S_{ij} k_j \phi_1^{s_{1j}} \phi_2^{s_{2j}}$. Solve these equations with the time derivative set to zero to obtain the steady-state concentrations $\vec{\phi} = (\phi_1, \phi_2)$.

Step 3. Calculate the elements of the Jacobian matrix $J_{ik}(\vec{\phi}) = \partial/\partial\phi_k[h_i(\vec{\phi}) + g_i(\vec{\phi})]$. Calculate the elements of the diffusion matrix of the downstream pathway $D^0_{ik}(\vec{\phi}) = \sum_{j=1}^{R} S_{ij} S_{kj} k_j \phi_1^{s_{1j}} \phi_2^{s_{2j}}$. Calculate the elements of the diffusion matrices of the downstream path for bursty and non-bursty input: $D^b_{ik}(\vec{\phi}) = \delta_{ik} h_i(\vec{\phi}) \eta_i + D^0_{ik}(\vec{\phi})$ and $D^s_{ik}(\vec{\phi}) = \delta_{ik} h_i(\vec{\phi}) + D^0_{ik}(\vec{\phi})$.

Step 4. Calculate $\theta_1^b - \theta_1^s$ using their definitions in Eqs. (3.14)–(3.13) and from the sign of this quantity identify which of the three cases illustrated in Fig. 3.1 the system under study falls in.

3.3 Two Species Biochemical Systems with Burst-Induced Oscillation Effects

3.3.1 Modified Brusselator

Here we consider a modified form of the Brusselator [133], which was introduced by Tyson and Kauffman in an early attempt to model dynamics within the process of

mitosis. The non-bursty reaction scheme for this model is:

$$\emptyset \xrightarrow{h_1} X_1, 2X_2 + X_1 \xrightarrow{k_1} 3X_2, X_1 \xrightarrow{k_2} X_2, X_2 \xrightarrow{k_3} \emptyset. \tag{3.16}$$

As previously explained, the bursty input version of this scheme having the same mean concentrations as the non-bursty version involve replacing the input reaction $\emptyset \xrightarrow{h_1} X_1$ by the set of reactions $\emptyset \xrightarrow{h_1 q_1(0)/\mu_1} 0X_1$, $\emptyset \xrightarrow{h_1 q_1(1)/\mu_1} X_1$, $\emptyset \xrightarrow{h_1 q_1(2)/\mu_1} 2X_1$, ..., $\emptyset \xrightarrow{h_1 q_1(M_1)/\mu_1} M_1 X_1$, where M_1 is some positive integer representing the maximum input burst size, $q_1(m)$ is the probability of an input burst of size m in species X_1 and $\mu_1 = \sum_m m q_1(m)$ is the mean burst size.

The quantities $\theta_i^b - \theta_i^s$ (for $i = 1, 2$) which determine the necessary conditions for promotion or destruction of noise-induced oscillations by burstiness can be computed by following the four step recipe outlined above. The steps of this recipe as applied to the modified Brusselator are as follows:

Step 1. Comparison of the modified Brusselator reaction scheme (3.16) with that in (3.1) shows that the stoichiometric coefficients are: $s_{11} = 1$, $s_{21} = 2$, $r_{11} = 0$, $r_{21} = 3$ for the first downstream reaction $2X_2 + X_1 \to 3X_2$; $s_{12} = 1$, $s_{22} = 0$, $r_{12} = 0$, $r_{22} = 1$ for the second downstream reaction $X_1 \to X_2$ and $s_{13} = 0$, $s_{23} = 1$, $r_{13} = 0$, $r_{23} = 0$ for the third downstream reaction $X_2 \to \emptyset$. Hence the stoichiometric matrix of the downstream pathway reads:

$$\mathbf{S} = \begin{pmatrix} -1 & -1 & 0 \\ 1 & 1 & -1 \end{pmatrix}. \tag{3.17}$$

Step 2. Next one uses the stoichiometric information of *Step 1* to write the functions $g_1 = -k_1\phi_1\phi_2^2 - k_2\phi_1$ and $g_2 = k_1\phi_1\phi_2^2 + k_2\phi_1 - k_3\phi_2$ and hence the rate equations are $d_t\phi_1 = h_1 + g_1$ and $d_t\phi_2 = g_2$ which have a steady-state solution $\phi_2 = h_1/k_3$ and $\phi_1 = k_3^2 h_1/(k_1 h_1^2 + k_2 k_3^2)$.

Step 3. Using the functions g_1 and g_2 obtained in *Step 2*, the steady-state concentration solutions also obtained in *Step 2* and the stoichiometric information obtained in *Step 1*, we can now calculate the three relevant matrices:

$$\mathbf{J} = k_3 \begin{pmatrix} -(\Lambda_1 + \Lambda_2) & -\frac{2\Lambda_1}{\Lambda_1+\Lambda_2} \\ \Lambda_1 + \Lambda_2 & \frac{\Lambda_1-\Lambda_2}{\Lambda_1+\Lambda_2} \end{pmatrix}, \quad \mathbf{D^b} = h_1 \begin{pmatrix} 1+\eta_1 & -1 \\ -1 & 2 \end{pmatrix}, \quad \mathbf{D^s} = h_1 \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}, \tag{3.18}$$

where $\Lambda_1 = h_1^2 k_1/k_3^3$ and $\Lambda_2 = k_2/k_3$ are non-dimensional parameters of the system.

Step 4. Using the three matrices calculated in the previous step, and the definitions in Eqs. (3.14)–(3.13) we can finally calculate the two quantities relevant to deduce the necessary conditions:

$$\theta_1^b - \theta_1^s = -\frac{2k_3^2\Lambda_1(\eta_1 - 1)}{(\Lambda_1 + \Lambda_2)(1+\eta_1)}, \tag{3.19}$$

$$\theta_2^b - \theta_2^s = \frac{1}{2}k_3^2(\Lambda_1 + \Lambda_2)^2(\eta_1 - 1), \tag{3.20}$$

Thus we have $\theta_1^b < \theta_1^s$ and $\theta_2^b > \theta_2^s$ for $\eta_1 > 1$, i.e., for all possible distributions of the burst size with mean burst size greater than 1. These are Case III and Case II in Fig. 3.1 respectively, implying necessary conditions for bursting in the input to promote noise-induced oscillations in species X_1 and for it to destroy noise-induced oscillations

in species X_2.

Numerical ID	Classification	
	NIO	Stability of Steady State
0	-	Unstable
1	✓	Node
2	✓	Focus
3	X	Node
4	X	Focus

Table 3.1: Existence of noise-induced oscillations (NIO) and linear stability classifications of the steady-state as used in Figure 3.2.

We investigated these predicted phenomena in further detail as follows. We chose the burst size distribution such that it was geometric with a mean burst size $\mu_1 = 12$ (and hence $\eta_1 = 25$; see equation (3.10) and the discussion which follows it) and varied Λ_1 and Λ_2 over the range 10^{-3} to 10^3. The geometric distribution is the discrete analog of the exponential distribution which has been measured in experiments [128] and has also been predicted from theory [134]. For each parameter set we deduced the nature of the stable steady-state from linear stability analysis (focus, i.e., $Tr[\mathbf{J}] < 0$, $Det[\mathbf{J}] > 0$ and $Tr[\mathbf{J}]^2 < 4Det(\mathbf{J})$ or node, i.e., $Tr[\mathbf{J}] < 0$, $Det[\mathbf{J}] > 0$ and $Tr[\mathbf{J}]^2 > 4Det(\mathbf{J})$ [132]) and also checked if there is a peak at some non-zero frequency in the LNA power spectrum as given by equation (3.12) (which implies noise-induced oscillations). The results for both species X_1 and X_2 are shown in Fig. 3.2 where the various regions have been labeled according to the numeric classification shown in Table 3.1. The red regions in Figs. 3.2 (a) and (b) denote the regions in parameter space where there are noise-induced oscillations in species X_1 for non-bursty and bursty input systems respectively. Similarly the blue regions in Figs. 3.2 (c) and (d) denote the regions in parameter space where there are noise-induced oscillations in species X_2 for non-bursty and bursty input systems respectively. Notice that in accordance with the predictions

based on the necessary conditions discussed in the previous paragraph, we find that the burstiness in the input reaction promotes noise-induced oscillations in X_1 (increased area of red region in Fig. 3.2 (b) compared to Fig. 3.2(a)) and destroys noise-induced oscillations in X_2 (decreased area of blue region in Fig. 3.2 (d) compared to Fig. 3.2 (c)). One also notices that the changes mainly occur in regions of parameter space characterized by a node and not by a focus (dotted region), which is consistent with the earlier prediction that burstiness has an important effect in systems far from the Hopf bifurcation.

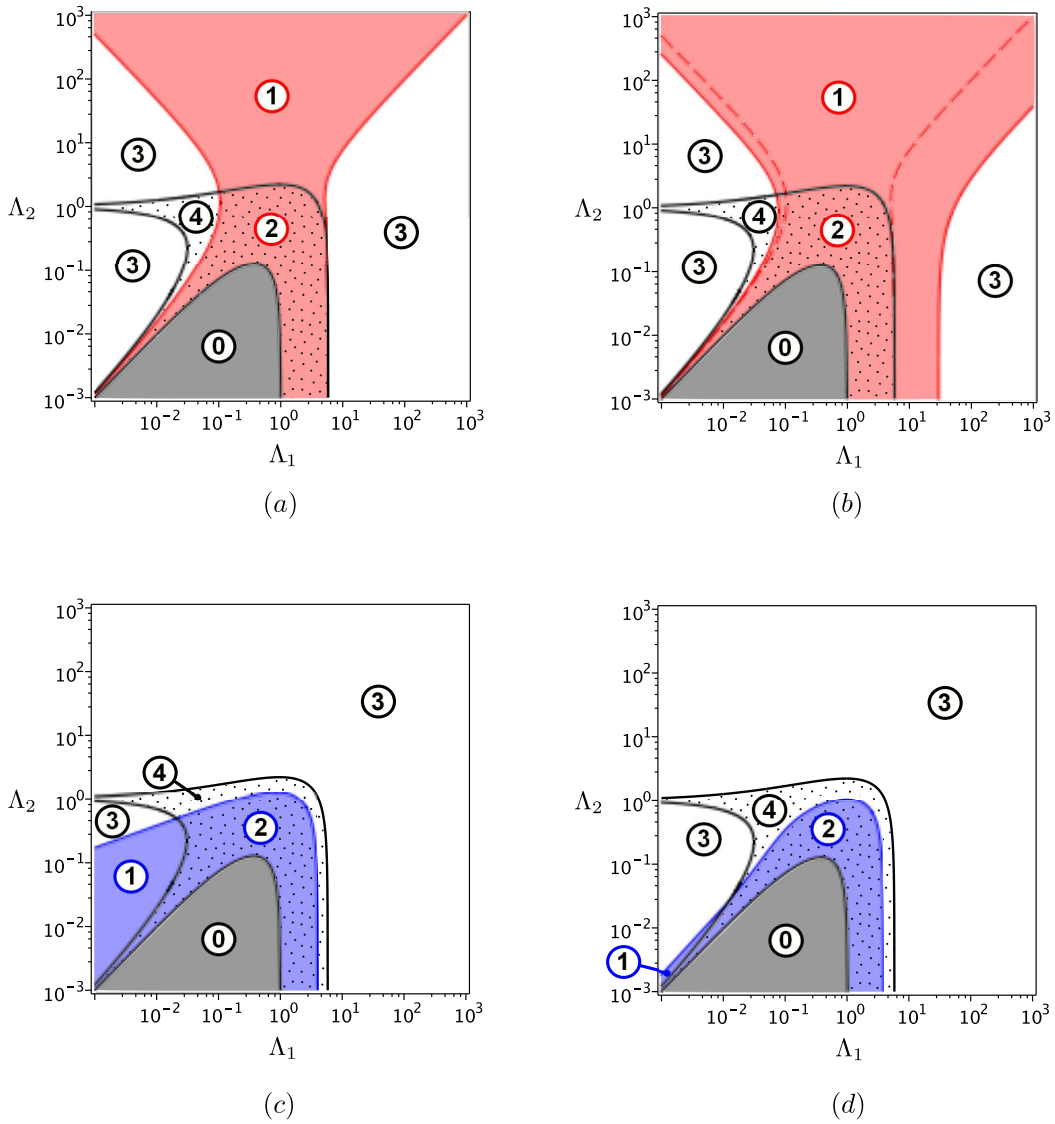


Figure 3.2: The impact of burstiness in the input reaction on the existence of noise-induced oscillations for the modified Brusselator model. The regions in Λ_1-Λ_2 space are classified according to the definitions in Table 3.1. Solid black lines bound the regions of different linear stability: the dotted region corresponds to the stable focus regime; white regions correspond to the stable node regimes and the grey region is where the fixed point is unstable (including the limit cycle regime). The red regions in (a) and (b) show the parameter space region where there are noise-induced oscillations in the concentration of species X_1 for the non-bursty input and bursty-input versions of the modified Brusselator respectively. The blue regions in (c) and (d) imply the same but for species X_2. The burst input distribution is geometric with mean burst size $\mu_1 = 12$. A comparison of (a) and (b) shows that burstiness in the input reaction promotes noise-induced oscillations in X_1 while a comparison of (c) and (d) shows that it destroys them in X_2. Note that the regions where most of these effects occur are not dotted, indicating that they are stable node steady-states.

Power spectra of the Modified Brusselator

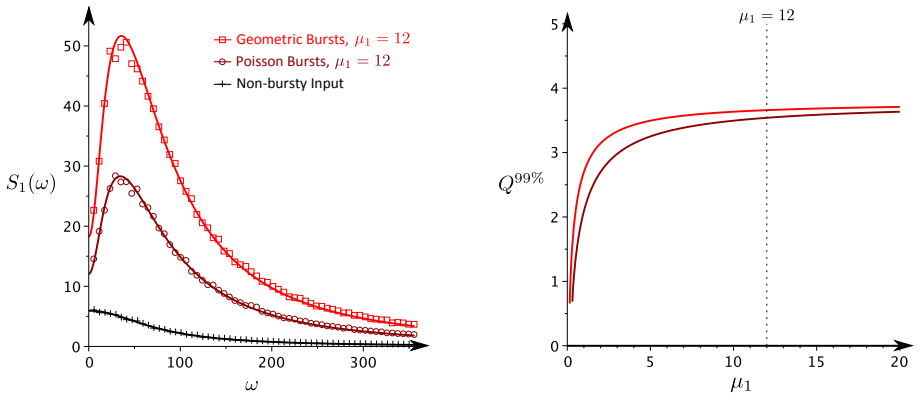
In Fig. 3.3 (a) and (c) we show the power spectra calculated from the LNA and from stochastic simulations for two points in Λ_1–Λ_2 space for which the LNA analysis of Fig. 3.2 predicted that burstiness in the input reaction should promote and destroy noise-induced oscillations, respectively.

For the simulations, SBML reaction models were created describing the non-bursty input and bursty input versions of the modified Brusselator (equation (3.16)). The upper burst size for Poisson or Geometric burst distributions is unbounded, meaning that to be exact an infinite number of input reactions is required in the simulation. For the chosen distributions (with mean burst size equal to 12 molecules) we truncated this to a maximum input burst size of 160 molecules. A simple python script was used to generate such a large reaction scheme. After parameterizing the models with the values given in the legend of Fig. 3.3, the models were simulated using the exact stochastic simulation implementation in the freely available software iNA (intrinsic noise analyser) [112].

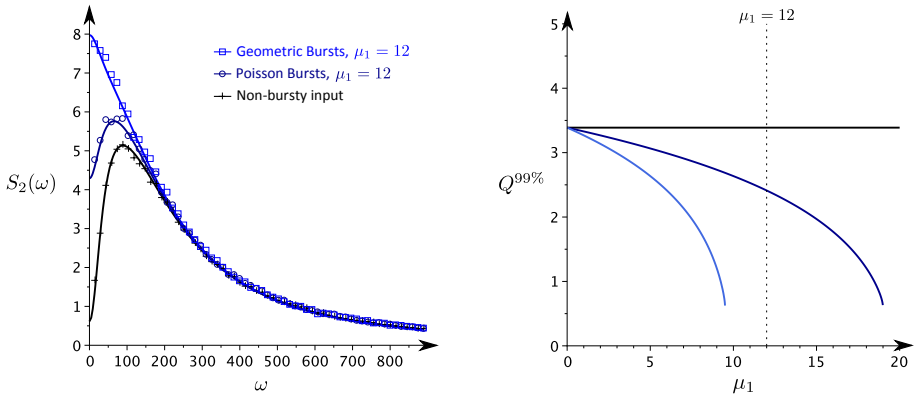
For a single realization r of the stochastic simulation algorithm, the number of molecules $n_i^r(t)$ of species i over some time interval T was regularly sampled at L discrete points separated by Δt, such that $T = (L-1)\Delta t$. The choices of sampling parameters for spectra in Fig. 3.3a and Fig. 3.3c were $\Delta t = 0.0018$s, $L = 600$ and $\Delta t \approx 7 \times 10^{-4}$s, $L = 600$ respectively. Since the variance of the periodogram estimate is known to be high, the final numerical power spectral density estimates plotted in Fig. 3.3 were obtained by averaging over 2000 realizations. See Appendix A.5 for further details of the periodogram method.

The simulations confirm the predicted phenomena by showing that the spectra of

X_1 and X_2 exhibit the appearance and disappearance of a peak at a non-zero frequency respectively, when bursting in the input reaction is turned on. It is also shown that the phenomena are more pronounced for geometric burst size distributions rather than for Poisson ones of the same mean burst size; this is since given the same mean, the width of the former distribution is larger than that of the latter. In Fig. 3.3 (b) and (d) we show the quality factor of the noise-induced oscillations as a function of the mean burst size μ_1. The quality factor is defined as $Q^{99\%} = \hat{\omega}/\Delta\omega^{99\%}$, where $\hat{\omega}$ is the frequency at which maximum power is obtained and $\Delta\omega^{99\%}$ is the difference of the frequencies at which the power takes 99 % of its maximum value; this measure was introduced in the previous Chapter and shown to be highly reflective of the rhythmicity visible in a time series of noise-induced oscillations. For these far-from Hopf oscillations the maximum possible value of $Q^{99\%}$ is ≈ 5 [135]. Of particular interest is the saturation observed in Fig. 3.3 (b) which implies that there are limits to how much burstiness in the input reaction can improve the quality of noise-induced oscillations. This also means that in the limit of large burst sizes the quality of noise-induced oscillations is independent of the precise type of the burst size distribution.



(a) $\Lambda_1 = 5.5$, $\Lambda_2 = 1.5$ (b)



(c) $\Lambda_1 = 5 \times 10^{-3}$, $\Lambda_2 = 10^{-2}$ (d)

Figure 3.3: Verification of LNA predictions by stochastic simulations. In (a) and (c) we plot the power spectra of the concentration fluctuations in X_1 and X_2 respectively using the LNA (solid lines) and also from stochastic simulations (data points). Note that in (a) burstiness in the input reaction promotes noise-induced oscillations (induces a peak in the spectrum of X_1) and in (c) it destroys them (removes the peak in the spectrum of X_2 for non-bursty input). In (b) and (d) we plot the quality of the noise-induced oscillations whose spectra are shown in (a) and (c) respectively. See text for definition of the quality factor $Q^{99\%}$. Note that the quality of the noise-induced oscillations can only be improved by burstiness in the input reaction to a certain extent (saturation of $Q^{99\%}$ with mean burst size μ_1). The constants are (a) $h_1 = 3 \times 10^{-5}\,\mathrm{M\,s^{-1}}$, $k_1 = 2.0625 \times 10^{13}\,\mathrm{M^{-2}\,s^{-1}}$, $k_2 = 22.5\,\mathrm{s^{-1}}$ and $k_3 = 15\,\mathrm{s^{-1}}$; (b) $h_1 = 3 \times 10^{-4}\,\mathrm{M\,s^{-1}}$, $k_1 = 2.34375 \times 10^{13}\,\mathrm{M^{-2}\,s^{-1}}$, $k_2 = 7.5\,\mathrm{s^{-1}}$ and $k_3 = 750\,\mathrm{s^{-1}}$. The compartment volume in each case is $\Omega = 3 \times 10^{-15}\,\mathrm{l}$, which gives mean molecule numbers of (a) $\langle n_{X_1} \rangle \approx 516$ molecules and (b) $\langle n_{X_2} \rangle \approx 723$ molecules. Note that the units for concentration, time and frequency ω are Molar (M), second (s) and radians per second ($\mathrm{rad\,s^{-1}}$) respectively.

3.3.2 Other Simple Circuits, Including Gene Regulation

Next we report the results of a detailed investigation of the effect of burstiness on the oscillatory properties of 8 biochemical pathways. The reaction schemes for the latter including their rate equations and the non-dimensional parameters characterizing the steady-state behaviour are shown in Tables 3.2 and 3.3. Note that for the gene circuits (Tab. 3.3) we have set some of the rate constants to 1; the model's behaviour can then be described by at most three non-dimensional parameters which considerably simplifies our analysis.

Next we used the four point calculational recipe to obtain the quantity $\theta_1^b - \theta_1^s$ for each of these 8 pathways. The sign of this quantity determines which of the three cases shown in Fig. 3.1 each pathway falls into and hence constitutes necessary conditions for promotion and destruction of noise-induced oscillations in species X_1 by bursting. The expressions for $\theta_1^b - \theta_1^s$ are shown in the second column of Table 3.4. It is simple to determine the sign of this quantity since all constants a_1 to a_8 are positive, as are Λ_1, Λ_2 and Λ_3 and $\eta_1, \eta_2 > 1$ (mean burst size is greater than 1). If the sign can take negative values then Case III is possible; if the sign can take positive values then Case II is possible. Which case can be displayed by each pathway is shown in columns 3 and 5 of Table 3.4. Notice that 6 out of 8 pathways can display Case III behaviour, i.e., bursting may induce oscillations; 6 out of 8 pathways can display Case II behaviour, i.e., bursting may destroy oscillations; 4 out of 8 pathways can display both Case II and Case III behaviour, i.e., bursting may promote or destroy oscillations.

As we have shown the simple necessary conditions are very easily determined in practice, and give a quick indication of whether input burstiness could cause a qualitative change in oscillatory behaviour in a system. However these conditions are not

sufficient by themselves to prove that the systems do actually display the burstiness-induced destruction or promotion of noise-induced oscillations. As previously explained and shown in Fig. 3.1, one needs to further determine if θ_1 falls in the correct range of values. This is considerably more involved analytically and hence we determine it numerically by an extensive parameter scan.

The parameter scan algorithm involved the following steps. We randomly picked 10^5 sets of non-dimensional parameters (the Λ_i's in Tables 3.2 and 3.3 are uniformly distributed in log-space over the range $[10^{-3}, 10^3]$ and the burtsiness parameters η_i are uniformly distributed integers in the range $[1, 25]$) for which the system has a steady-state. For each parameter set we calculated the quantities q, θ_1, θ_1^b and θ_1^s. If $q < 0$ then for this parameter set both bursty and non-bursty input systems display noise-induced oscillations. If $q > 0$, then θ_1, θ_1^b and θ_1^s are used to obtain which case and which particular region of the case shown in Fig. 3.1 describes the system's behaviour for the chosen parameter set. These classifications are recorded for each parameter set.

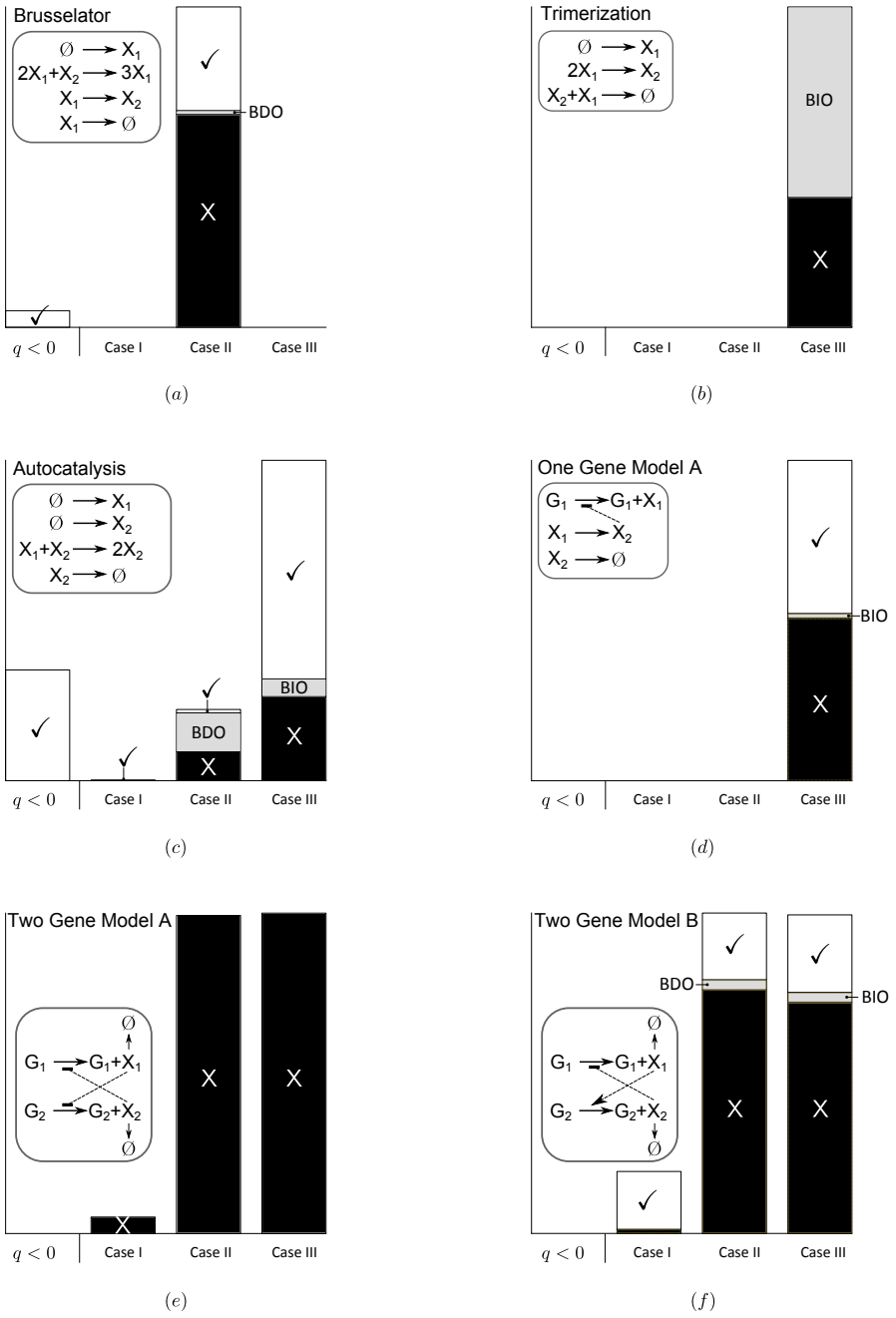


Figure 3.4: Numerical investigation into the effect of bursting in the input reactions on six of the eight biochemical models shown in Tables 3.2 and 3.3. See main text for details of the numerical algorithm used. $q < 0$ refers to the cases close to the Hopf bifurcation where both bursty and non-bursty input systems show noise-induced oscillations. Cases I, II and III are for $q > 0$ (far from Hopf bifurcation) described in Fig. 3.1. Ticks/crosses indicate that noise-induced oscillations are/are not observed for both bursty and non-bursty-input systems. BDO and BIO refer to the cases where burstiness destroys or promotes noise-induced oscillations, i.e., the behaviour of the two systems is different. The heights of the bars for each behaviour are directly proportional to the fraction of the $100,000$ parameter sets which exhibit that behaviour.

Information regarding whether the sufficient conditions were found to be satisfied

or not is reported in columns 4 and 6 of Table 3.4. A more detailed classification is shown in bar chart form in Fig. 3.4 for six of the eight pathways in Tables 3.2 and 3.3. Note that the two remaining pathways (One Gene Model B and Two Gene Model C) are similar in behaviour to Two Gene Model A and hence not shown in the latter figure. At least one of Case II or Case III behaviour was possible for each model. The necessary conditions for bursts destroying or promoting noise-induced oscillations were also sufficient, with three exceptions: One Gene Model B, Two Gene Model A and Two Gene Model C. Interestingly, these three exceptions are unique among the models in that they are the only ones which cannot exhibit noise-induced oscillations for any parameter choices for both bursty and non-bursty systems.

We noticed that the effect of burstiness on each system was strongly linked to two main features: (a) whether burstiness is possible in one or both species; and (b) the pathway's feedback motif, as described by the signs of the off-diagonal elements of the Jacobian matrix (column 3 in Tab. 3.4). The three exceptional models (One Gene Model B, Two Gene Model A and Two Gene Model C) which never exhibited noise-induced oscillations, and for which the necessary conditions for bursts destroying or promoting noise-induced oscillations did not also translate to sufficient conditions, all feature either mutual promotion or mutual inhibition between the two species.

Models with negative feedback, whereby one species promotes the other and that species inhibits the first (indicated by different signs on the off-diagonals of $\mathbf{J}$) were sensitive to burstiness destroying or promoting noise-induced oscillations. When burstiness is possible in both species (Autocatalysis and Two Gene Model B), necessary conditions for both BIO and BDO can be satisfied and BIO and BDO were indeed observed. Therefore, our results suggest that the combination of a negative feedback motif and burstiness in both species allows a wide range of bursting-induced oscillatory behaviour.

When burstiness is only possible in one of the species, the matching of necessary and sufficient conditions is again observed, but here the asymmetry of the Jacobian is important; when J_{12} is positive (Brusselator) the necessary conditions indicate that burstiness tends to destroy noise-induced oscillations in X_1, but when J_{12} is negative (Trimerization and One Gene Model A) the necessary conditions indicate that noise tends to promote noise-induced oscillations in X_1.

Here we have explored in detail the effect of burstiness on species X_1, and it has become clear that asymmetry in the system can be important. For two of the models with similar but asymmetric sign structures, i.e., the Trimerization and Brusselator models, we show in Fig. 3.5 the effect of input burstiness on species X_2. Our theory above suggested that input burstiness in the one species (e.g., X_1) could only serve to destroy noise induced oscillations in the other (e.g., X_2). Indeed, destruction of oscillations were found in each case (Fig. 3.5) in our parameter scan, albeit for a small region of parameter space for the Brusselator. Considering the effect of input burstiness on the biochemical species in these two models, we therefore find that generally the burstiness serves to destroy oscillations in these species, with the exception of X_1 in the Trimerization model for which burstiness induces oscillations for some parameters, and that this can be predicted due to the sign structure of the Jacobian.

Although the regions of parameter space for which bursts promote or destroy noise-induced oscillations is quite small in some models, we note that this region can be considerably enlarged by choosing a smaller range for the burstiness parameters (e.g. if η_1 and η_2 are fixed to 25 and 2 respectively rather than the range $[1, 25]$ used in our parameter scan). The fact that a large proportion of the considered pathways display burstiness alteration of the noisy oscillatory dynamics suggests that such phenomena may be common in many biochemical systems.

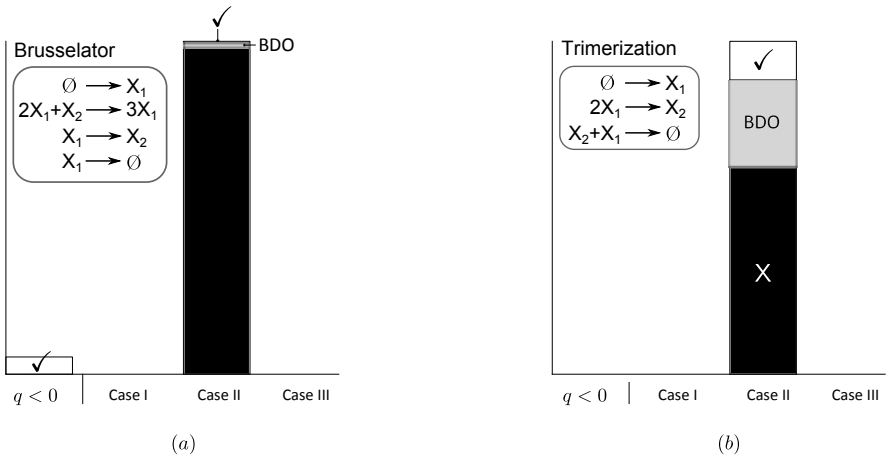


Figure 3.5: Numerical investigation into the effect of bursting in the input reactions on species X_2 in the Brusselator and Trimerization models shown in Tables 3.2 and 3.3. See main text for details of the numerical algorithm used. $q < 0$ refers to the cases close to the Hopf bifurcation where both bursty and non-bursty input systems show noise-induced oscillations. Cases I, II and III are for $q > 0$ (far from Hopf bifurcation) described in Fig. 3.1. Ticks/crosses indicate that noise-induced oscillations are/are not observed for both bursty and non-bursty-input systems. BDO refers to the case where burstiness destroys noise-induced oscillations, i.e., the behaviour of the two systems is different. The heights of the bars for each behaviour are directly proportional to the fraction of the $100,000$ parameter sets which exhibit that behaviour.

3.4 Discussion and Conclusion

In this Chapter we have shown using the LNA that burstiness in the input reactions can have a considerable impact on the oscillatory properties of the downstream pathway. In particular we showed that for two identical pathways, one with bursty and one with non-bursty input, the two pathways may differ in their ability to produce noise-induced oscillations. We derived necessary conditions for the burstiness to promote oscillations and for the burstiness to destroy oscillations and confirmed the existence of these phenomena using stochastic simulations. Our work is the first to investigate the effect of burstiness on the noisy oscillatory dynamics of biochemical pathways; previous work focused on deriving expressions for the steady-state distributions (or the moments) of protein concentrations in the presence of bursting [129, 136, 137, 138].

We note that our analysis is based on the LNA which is a good approximation when describing pathways involving small levels of noise, i.e., pathways characterized by a sufficiently large number of molecules. This is not always the case since a number of species inside cells occur in small molecule numbers [41]. Our theory can be extended to account for such cases by considering higher-order terms than the LNA in the system size expansion of the master equation [70, 139, 140]. We showed in the previous Chapter that in non-bursty systems if the LNA predicts a peak in the power spectrum of fluctuations for systems far from the Hopf bifurcation then the spectrum calculated from stochastic simulations shows a peak even if the molecule numbers are very small; the quality of the oscillations may, however, be lower than that predicted by the LNA. Hence we expect the consideration of terms of higher order than the LNA to have little or no effect on the necessary conditions derived in this chapter since these are specifically for the existence or non-existence of a peak in the power spectrum.

Our study also hints at a novel control mechanism for single cell rhythms. The degree of burstiness in protein production can be controlled through the translation rate or the mRNA degradation rate [131] and the latter can in turn be controlled by small RNAs [141]. Hence it is plausible that the quality and strength of single cell rhythms can be effectively manipulated by controlling the expression levels of small RNAs.

	Reaction Model	Rate Equations	Dimensionless Parameters
Brusselator	$\emptyset \underset{k_3}{\overset{h_1}{\rightleftharpoons}} X_1$ $2X_1 + X_2 \xrightarrow{k_1} 3X_1$ $X_1 \xrightarrow{k_2} X_2$	$\frac{d\phi_1}{dt} = h_1 + k_1\,\phi_1^2\phi_2 - (k_2 + k_3)\,\phi_1,$ $\frac{d\phi_2}{dt} = -k_1\,\phi_1^2\phi_2 + k_2\,\phi_1.$	$\Lambda_1 = h_1^2\,k_1/k_3^3$ $\Lambda_2 = k_2/k_3$
Trimerization	$\emptyset \xrightarrow{h_1} X_1$ $2X_1 \xrightarrow{k_1} X_2$ $X_1 + X_2 \xrightarrow{k_2} \emptyset$	$\frac{d\phi_1}{dt} = h_1 - 2k_1\,\phi_1^2 - k_2\,\phi_1\phi_2,$ $\frac{d\phi_2}{dt} = k_1\,\phi_1^2 - k_2\,\phi_1\phi_2.$	$\Lambda_1 = k_2/k_1$
Autocatalysis	$\emptyset \xrightarrow{h_1} X_1$ $\emptyset \underset{k_2}{\overset{h_2}{\rightleftharpoons}} X_2$ $X_1 + X_2 \xrightarrow{k_1} 2X_2$	$\frac{d\phi_1}{dt} = h_1 - k_1\,\phi_1\phi_2,$ $\frac{d\phi_2}{dt} = h_2 + k_1\,\phi_1\phi_2 - k_2\,\phi_2.$	$\Lambda_1 = h_2/h_1$ $\Lambda_2 = h_1\,k_1/k_2^2$

Table 3.2: Details of three biochemical systems studied in chapter 2 with the notation developed in this chapter for investigating the effects of input burstiness. The symbol ϕ_i denotes the concentration of species X_i.

	Reaction Model	Rate Equations	Dimensionless Parameters
One Gene Model A	$G_1 \xrightarrow{h_1(\phi_2)} G_1 + X_1$ $X_1 \xrightarrow{k_1} X_2$ $X_2 \xrightarrow{k_2} \emptyset$	$\frac{d\phi_1}{dt} = k_0 \left(\frac{1}{1+\phi_2}\right) - k_1\,\phi_1,$ $\frac{d\phi_2}{dt} = k_1\,\phi_1 - k_2\,\phi_2.$	$\Lambda_1 = k_1/k_0$ $\Lambda_2 = k_2/k_0$
One Gene Model B	As above	$\frac{d\phi_1}{dt} = k_0 \left(\frac{\phi_2}{1+\phi_2}\right) - k_1\,\phi_1,$ $\frac{d\phi_2}{dt} = k_1\,\phi_1 - k_2\,\phi_2.$	As above
Two Gene Model A	$G_1 \xrightarrow{h_1(\phi_2)} G_1 + X_1$ $G_2 \xrightarrow{h_2(\phi_1)} G_2 + X_2$ $X_1 \xrightarrow{k_1} \emptyset$ $X_2 \xrightarrow{k_2} \emptyset$	$\frac{d\phi_1}{dt} = k_0 \left(\frac{1}{1+\phi_2}\right) - k_1\,\phi_1,$ $\frac{d\phi_2}{dt} = k'_0 \left(\frac{1}{1+\phi_1}\right) - k_2\,\phi_2.$	$\Lambda_1 = k'_0/k_0$ $\Lambda_2 = k_1/k_0$ $\Lambda_3 = k_2/k_0$
Two Gene Model B	As above	$\frac{d\phi_1}{dt} = k_0 \left(\frac{1}{1+\phi_2}\right) - k_1\,\phi_1,$ $\frac{d\phi_2}{dt} = k'_0 \left(\frac{\phi_1}{1+\phi_1}\right) - k_2\,\phi_2.$	As above
Two Gene Model C	As above	$\frac{d\phi_1}{dt} = k_0 \left(\frac{\phi_2}{1+\phi_2}\right) - k_1\,\phi_1,$ $\frac{d\phi_2}{dt} = k'_0 \left(\frac{\phi_1}{1+\phi_1}\right) - k_2\,\phi_2.$	As above

Table 3.3: Details of the five gene regulation systems studied for effects of input bursting. The symbol ϕ_i denotes the concentration of species X_i.

			Bursts Induce Oscillations (BIO)		Bursts Destroy Oscillations (BDO)	
Model	$\theta_1^b - \theta_1^s$	Feedback Motif (**J**)	Necessary conditions (Case III)	BIO Observed?	Necessary conditions (Case II)	BDO Observed?
Brusselator	$a_1(\eta_1 - 1)$	$\begin{pmatrix} & + \\ - & \end{pmatrix}$	✗	N/A	✓	✓
Trimerization	$-a_2(\eta_1 - 1)$	$\begin{pmatrix} & - \\ + & \end{pmatrix}$	✓	✓	✗	N/A
One Gene Model A	$-a_3(\eta_1 - 1)$	$\begin{pmatrix} & - \\ + & \end{pmatrix}$	✓	✓	✗	N/A
One Gene Model B	$a_4(\eta_1 - 1)$	$\begin{pmatrix} & + \\ + & \end{pmatrix}$	✗	N/A	✓	✗
Autocatalysis	$a_5[(\Lambda_1+1)+\Lambda_1\eta_2 - (2\Lambda_1+1)\eta_1]$	$\begin{pmatrix} & - \\ + & \end{pmatrix}$	✓	✓	✓	✓
Two Gene Model A	$a_6(\eta_2 - \eta_1)$	$\begin{pmatrix} & - \\ - & \end{pmatrix}$	✓	✗	✓	✗
Two Gene Model B	$a_7(\eta_2 - \eta_1)$	$\begin{pmatrix} & - \\ + & \end{pmatrix}$	✓	✓	✓	✓
Two Gene Model C	$a_8(\eta_2 - \eta_1)$	$\begin{pmatrix} & + \\ + & \end{pmatrix}$	✓	✗	✓	✗

Table 3.4: The effect of bursting in the input reactions on the eight chemical systems shown in Table I. The sign of $\theta_1^b - \theta_1^s$ determines whether the model exhibits Case III or Case II behaviour (see Fig. 3.1), i.e., the necessary conditions for burstiness to promote or destroy noise-induced oscillations respectively (fourth and sixth columns of the table). Determining whether input bursting can in fact promote or destroy oscillations requires the full condition that θ_1 is large enough; the columns 'BIO Observed' and 'BDO Observed' are the results of our numerical investigation to this end (see text for details). The upper four circuits feature input burstiness in only one species, the lower four feature input burstiness in both species.

Chapter 4

Intercellular Coupling and Spontaneous Spatiotemporal Patterns

NB: The work in Sections 4.1 to 4.5 of this chapter describes joint research by David Toner (DT) and Benedicte Wenden (BW), published in [142]. Experimental work (plant growth, luminescence imaging) was carried out by BW. Data analysis was carried out by DT, who also led on the data visualization. Some figures were produced by DT, some were produced by BW and some were created collaboratively between DT and BW. Unless otherwise stated in each figure legend, figures in this chapter were produced by DT. Section 4.6 contains unpublished work by DT.

4.1 Introduction

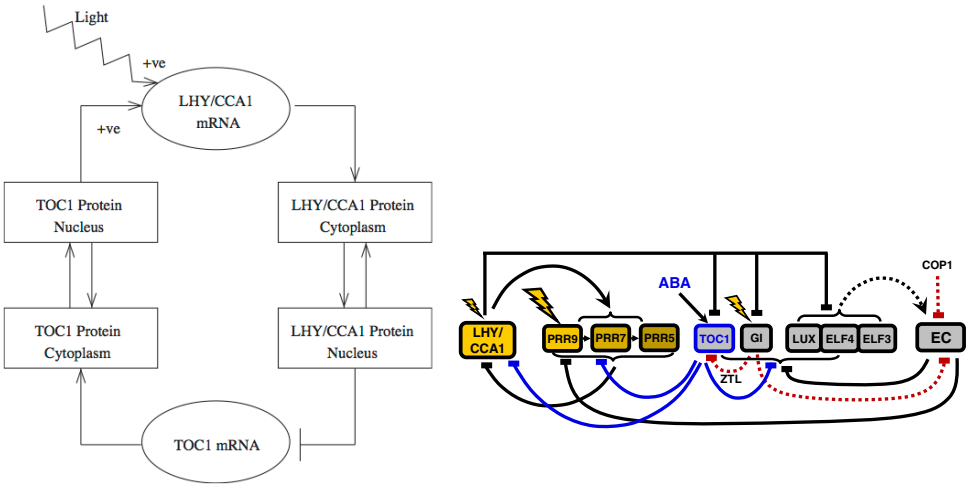
Having investigated intriguing new modes of stochastic rhythmicity in simple models of biochemical oscillators, we now shift our focus to the roles of intercellular coupling in a particular biological oscillator system, namely circadian rhythms in the model plant *Arabidopsis thaliana.* The circadian clock is known to play a fundamental role in plant physiology, in that it regulates an extremely wide range of biological processes including germination, rhythmic leaf movement, photosynthesis, flowering time, hormone responses, Ca^{2+} concentrations, and stomatal opening [143, 144]. Given the importance of the circadian clock for essential plant functions, there is now considerable research into fully understanding the clock's mechanisms. An area which remains particularly poorly understood is the nature of rhythms in multicellular plant organisms, in terms of the heterogeneity of clocks amongst cells and tissues and in terms of the presence and nature of intercellular communication which could drive coherence of the whole organism rhythm.

Circadian rhythms help organisms adapt to their environment, and are defined by three central criteria [145]. Firstly, circadian rhythms are *autonomous* rhythms of approximately 24-hour periodicity, i.e., observed rhythmicity is not simply an effect of daily environmental cycles such as cycles in light or temperature. This is tested by observing whether approximately 24-hour rhythms are observed when an organism is transferred to 'free-running' (constant) conditions (principally in an environment without changes in light and temperature)[145]. The second requirement is that circadian rhythms can be entrained to new environmental stimuli, such as the ability (albeit imperfect) of human body clocks to entrain to new time zones after long distance travel. The third criterion is that circadian rhythms can compensate for environmental tem-

perature changes (such as seasonal changes) to an extent that the period of rhythms is largely unchanged across a wide range of temperatures. Since temperature is well known to be a fundamental factor in determining the rate of biochemical reactions generally, the ability of circadian clocks to perform this extraordinary temperature compensation (an area of active current research [146, 147]) is an indicator of the complexity of the oscillator mechanisms.

Prompted by advances in research into the molecular mechanisms underpinning circadian rhythmicity in other organisms such as *Drosophila melanogaster* and mouse [148], in the past twelve years researchers have attempted to understand the molecular mechanisms of the plant circadian clock in the model plant organism *Arabidopsis thaliana*. Although recent research has shown that circadian rhythms can arise from alternative mechanisms [88], the understanding of plant circadian clocks has been principally within a framework of intracellular transcriptional and translation feedback loops (TTFLs) amongst circadian clock genes. An essential feature of such a model structure is at least one negative feedback loop, by which proteins indirectly repress their own gene expression in a daily cycle. Early work into plant circadian rhythms identified a central role for three genes; TIMING OF CAB EXPRESSION 1 (TOC1), which is expressed in the evening, and two partially redundant transcription factors CIRCADIAN CLOCK ASSOCIATED 1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY), which are expressed in the morning [148]. CCA1 and LHY were both found to directly bind to the *TOC1* promoter and inhibit its expression, whereas TOC1 protein was found to promote the expression of *CCA1* and *LHY* [148]. A first plant circadian mathematical model was built on these discoveries, featuring a single negative feedback loop between the TOC1 gene and a combined pseudo-gene referred to as CCA1/LHY (representing the partially redundant nature of these transcription factors, Fig. 4.1(a))

[51].



(a) (b)

Figure 4.1: (a) The first model of the single transcriptional-translational feedback loop at the heart of the Arabidopsis clock [51]. (b) An overview of the latest multi-loop model of interlocked transcriptional-translational feedback loops, which in fact features more than 30 dynamic chemical species [149]. Lightning-bolt symbols indicate multiple light inputs to the clock.

The single loop model provided a useful starting point for understanding the circadian clock; in the years since a combination of experimental and mathematical research have been used to develop the current three-loop model (Fig. 4.1(b)). Whenever the state-of-the-art mathematical model has been found incapable of reproducing experimentally observed phenotypes this has prompted incremental changes to the model. One such example was the early discovery that the *cca1;lhy* double loss-of-function mutant was found to exhibit damped oscillations in constant light conditions [150], which clearly could not be accounted for by the single loop model with CCA1/LHY as an essential rhythmic component. This prompted the addition of an extra feedback loop to capture this residual rhythmicity, producing a model with two interlocked feedback loops, robust to the loss-of-function mutation. As well as the requirements of being robust in the sense that the clock can sustain rhythms when components are removed,

focus is also given to ensuring that the models can reproduce precise timing of the clock in terms of peak expression of its components under different photoperiods (e.g., 8 hours of light and 16 hours of darkness (short days) or 16 hours of light and 8 hours of darkness (long days)) [150]. The current model (Fig. 4.1(b)) is highly complex, with more than 30 interacting chemical species and multiple light inputs. It is important for us to note that while these biochemical models have been constructed and parameterized with whole seedling data, the intracellular nature of the models imply an ability to quantitatively describe the gene-regulatory mechanisms inside each individual cell in the plant. Furthermore, while the current models of Arabidopsis are complex, and can describe a wide range of experimental behaviour, little consideration has been given to the added complexity of interactions between noisy single cell circadian oscillators. The work in this chapter aims to contribute to current understanding of the multi-cellular nature of the plant clock.

Experimental work over the last decade or so has provided greater insight to how tissue-level rhythms originate from singe cell oscillators. In mammals, in the suprachiasmatic nucleus (SCN) tight coupling by synaptic transmission among SCN neurons is now known to be crucial to sustain rhythmicity [100, 101]. In contrast, circadian research in plants has revealed that plant cells appeared to be coupled more weakly, if at all, though many of these studies have used whole plant assays to reach this conclusion. Circadian time series data in higher plants are now commonly obtained by means of a luminescent marker [145]. In Arabidopsis, the reporter gene *Luc* can be placed under the control of a particular circadian gene's promotor; *Luc* encodes firefly luciferase which quickly catalyzes the oxidation of beetle luciferin (provided in the media) [151, 145], producing light and thereby reporting on the promoter activity. In tobacco, researchers were able to develop a strain which expressed both firefly lu-

ciferase, under the promoter of the circadian gene light-harvesting complex (LHCB), and the photoprotein aequorin [152], which can be used as a reporter of cytosolic free calcium concentrations [153]. By altering incubation conditions, either LHCB activity or $[Ca^{2+}]$ concentrations could be measured over several days. LHCB gene expression was found to spontaneously desynchronise from cytosolic free calcium [153] and it was hypothesised that the cytosolic free calcium may be expressed in specialised cells, and therefore the desynchronisation could mean weak coupling among different cell types. A similar argument for the presence of weakly coupled, heterogeneous, tissue-specific circadian oscillators in Arabidopsis thaliana was made from the observed difference in free-running period of the expression of circadian genes PHYB and CAB in one study [154], and CAB and CHS expression in a separate study [155]. These results derived from whole-plant assays and were interpreted as showing desynchronisation among the different cell types that express the distinct markers. In experimental work which examined Arabidopsis rhythms with spatial resolution, it was shown that different areas of the same plant can be experimentally desynchronised, using light-dark treatments to different locations of the same plant. After transfer to constant light, these regions showed little relative phase change within time series of up to five days [95] and hence no evidence of coupling among cells expressing the same LUC marker. In contrast, more recent spatial-resolution investigations with longer time series have suggested that there is weak coupling among Arabidopsis cells [156].

Due to the continued dominance of modelling Arabidopsis rhythms from a whole plant perspective and the limited research into spatial heterogeneities of Arabidopsis circadian gene expression, our objective was to determine the extent to which oscillations are coherent across Arabidopsis leaves in both natural and constant light conditions, with a view to a better understanding of the nature of communication between

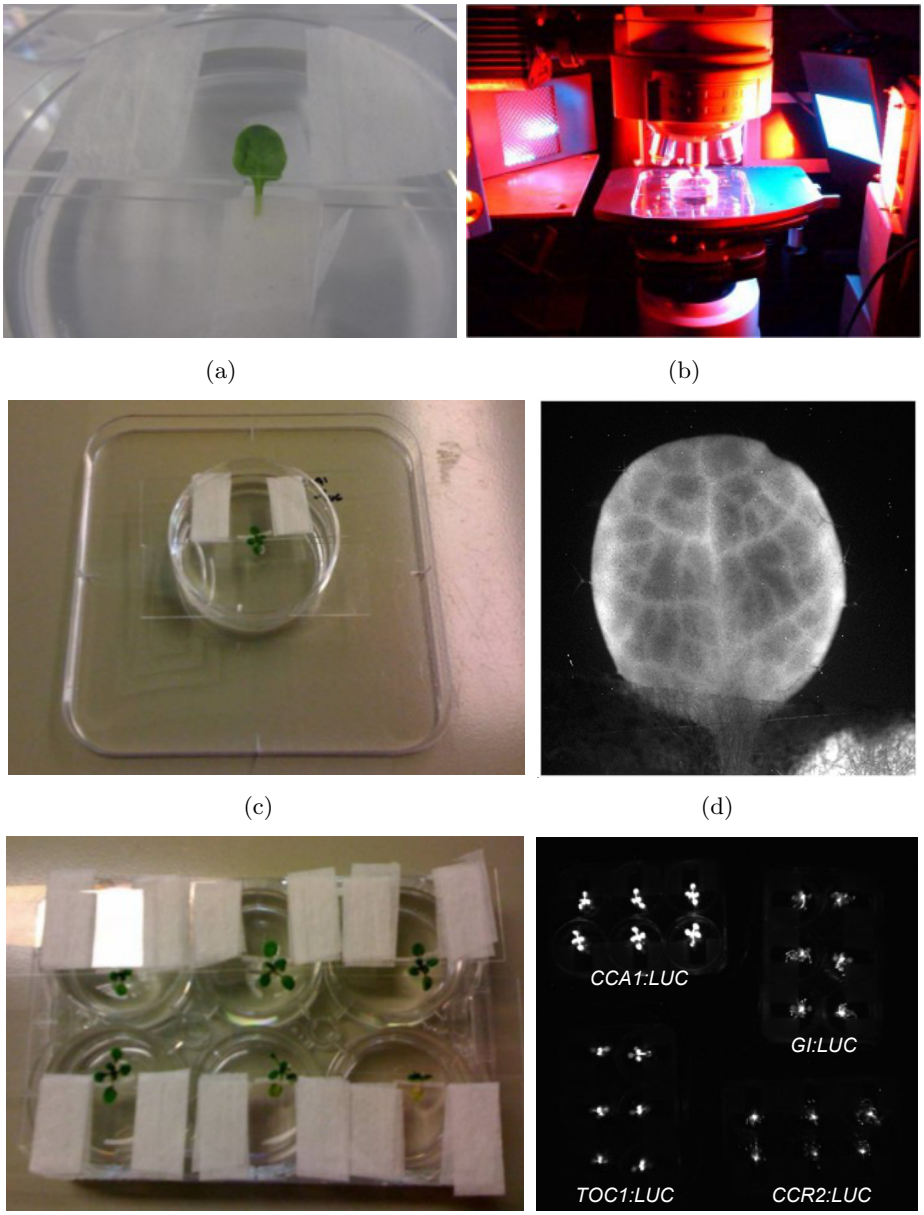
cells. By transferring leaves between different light conditions (such as between constant light conditions and entraining light-dark (LD) cycles) we aimed to capture the relative importance of any intercellular coupling versus coupling between cells and the entraining light signals. We also sought to understand whether any heterogeneities in circadian rhythms across leaves were exhibited with a consistent spatial pattern, and for this a high throughput method was required to yield a sufficient sample of leaves. Our group therefore developed imaging and analysis methods to investigate circadian coupling in short time series of LUC reporter gene expression in plant leaves, using a high throughput imaging cabinet setup (see Appendix A.7), published in [142]. Sections 4.1 to 4.5 of this Chapter give the results of this work. Furthermore, to more precisely investigate local spatial details in the timing in circadian expression, a lower throughput, high-resolution microscope method was used; the results from analysing these data are as yet unpublished and are presented in Section 4.6.

4.2 Setup and Analysis

To monitor luminescence rhythms in plants over several days, a protocol was designed to image healthy leaves of intact transgenic seedlings for up to 6 days (Fig. 4.2). Furthermore, imaging of detached leaves similarly to [156] allowed us to test expression rhythms in older, isolated tissue (Fig. 4.2(a)). Details of the experimental methods regarding plant growth conditions and imaging set-up, as established by Wenden and Millar, are included in Appendix A.7.

Luminescence time series were obtained from 24 12-day-old seedlings or 8 21-day-old detached leaves in a single field of view in the imaging cabinet with tissue-level resolution (image pixel size $230\mu m$, Figs. 4.2(e) and 4.2(f)) or from a single leaf with

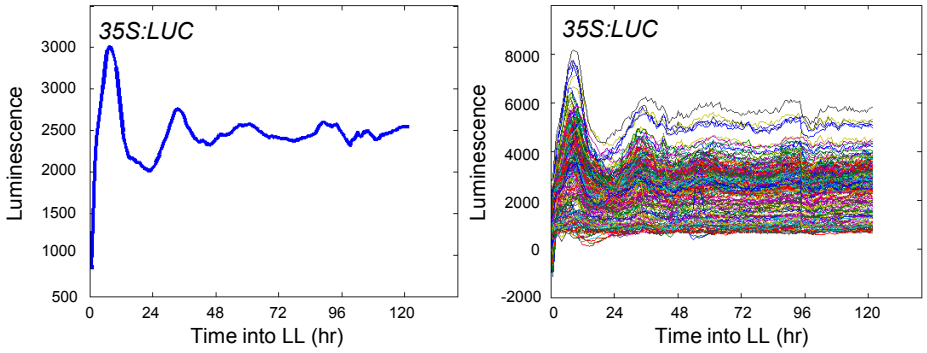
cellular resolution in the microscope (pixel size $5\mu m$, Figs. 4.2(c) and 4.2(d)). In the first part of this chapter we focus on the tissue-level data.



(a) (b) (c) (d) (e) (f)

Figure 4.2: Setup for imaging luciferase in intact and detached leaves over several days. (a) setup for one detached leaf; (b) red and blue light-emitting diode (LED) system for intact plant imaging under the microscope; (c) setup for one intact plant; (d) luminescence output from a leaf from a *CCA1:LUC* intact plant; (e) six-well plate setup for intact plants for open field imaging in a cabinet; (f) luminescence output for four six-well plates as imaged in a cabinet. [Figs. produced by BW and published in [142].]

Preliminary time series for the control marker 35S:LUC confirmed that the signal from plants in imaging cabinets was strong enough to analyse not only the leaf average but also single pixels (Fig. 4.3).



(a) (b)

Figure 4.3: Preliminary results for 35S:LUC (c) and luminescence (d) for all pixels, detrended for a 35S:LUC leaf. Plant was grown in LL for 12 d and imaged under LL conditions. [Figs. produced by BW and published in [142]]

4.2.1 Data Processing

A Matlab (Mathworks) script was developed to extract LUCIFERASE (LUC) luminescence data for each leaf at the pixel level and for each picture frame. For each leaf, luminescence time series were then available for a square array of pixels, containing the leaf and surrounding area. An example is shown in Fig. 4.4(a) and 4.4(b).

The goal was to locate the circadian peaks to define each pixels time-dependent phase, $\theta_n(t)$. For each pixel, n, we use the definition of time-varying phase as a piecewise linear function between successive peak times T_k and T_{k+1}, as used in [156]; i.e.,

$$\theta_n(t) = 2\pi \frac{t - T_n^k}{T_n^{k+1} - T_n^k}, \quad t \in \left[T_n^k, T_n^{k+1}\right). \tag{4.1}$$

Upon visual inspection of the data for all leaves, some commonly found potential issues were quickly identified: i) Extreme luminescence levels in the first few time

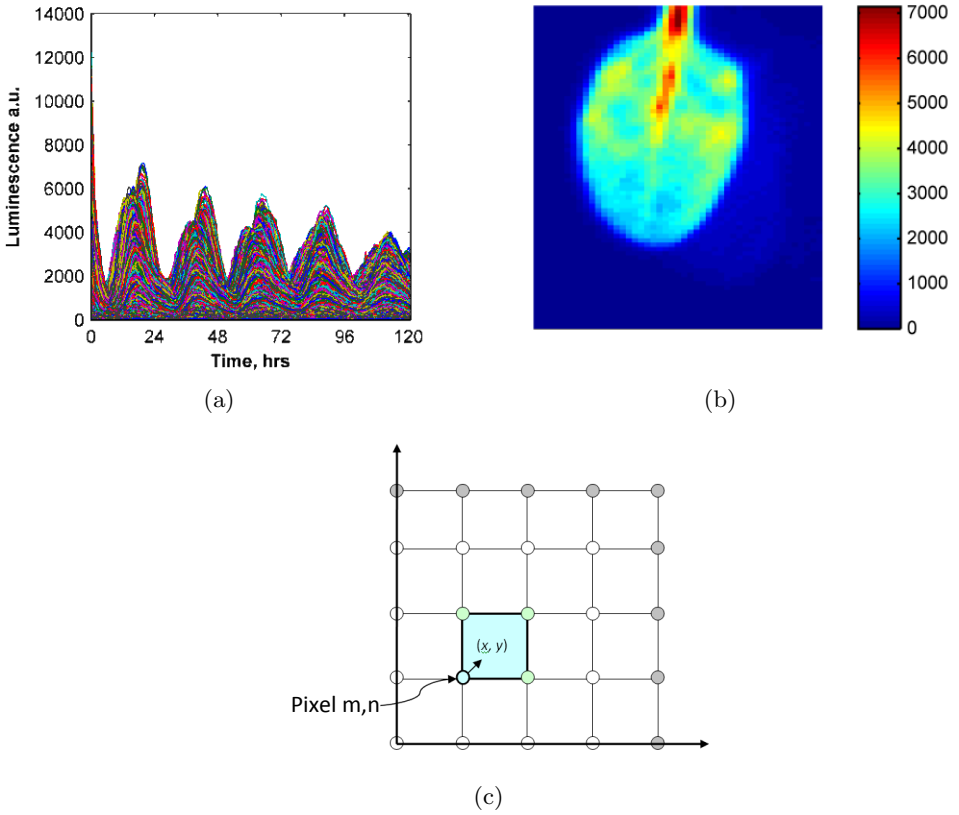
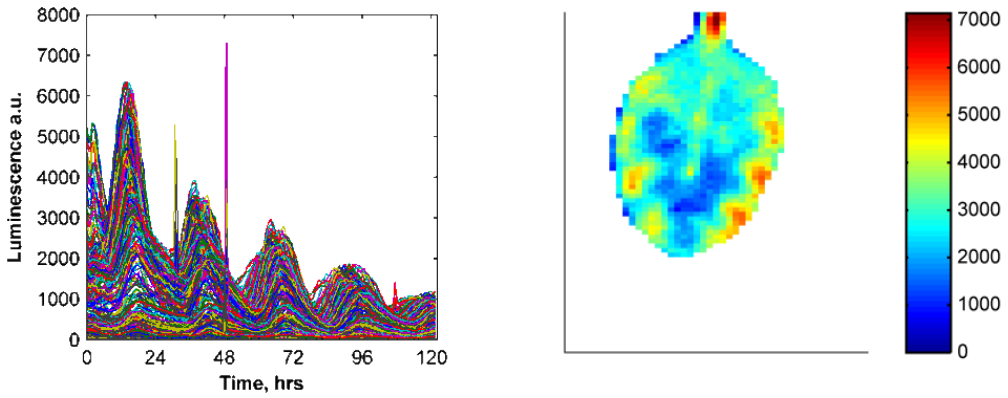


Figure 4.4: (a) All pixels luminescence time series for a single leaf. (b) Luminescence across the square array at ZT20. (c) Assignment of colored image squares from pixel data values using MATLAB's 'surf' function with flat shading.

points (such as in Fig. 4.4(a)); ii) very low luminescence values at longer times; and iii) luminescence spikes, from solar cosmic rays (Fig. 4.5(a)).



(a) (b)

Figure 4.5: (a) Example of solar cosmic rays (e.g., at ZT48, which were identified and removed). (b) Retained pixels in leaf (as in Fig. 4.4(a) and 4.4(b)) following removal of peripheral pixels from the array.

To address issues (i) and (ii), an appropriate data window for each leaf was identified and then globally (i.e., across all pixels) data were omitted at the time points outside this window. Because our determination of phase rests on determining peak times, the start and end points of the data window were typically chosen around a trough time (e.g., a start time of 6 h was chosen for the data in Fig. 4.4(a)). For each time series that contained solar cosmic spikes, the luminescence reading at the identified spike time, $L(t_{sc})$, was replaced by the mean luminescence value from the preceding and subsequent times; i.e., $L(t_{sc}) = \frac{L(t_{sc}-\Delta t)+L(t_{sc}+\Delta t)}{2}$.

Data in peripheral areas were excluded from the square array by specifying a minimum required quality for the data in each pixel. Firstly, pixels around the periphery of the square array, and clearly outwith the leaf area, were omitted according to a sub-threshold mean (over time) luminescence value. Second, because our method of determining phase requires peak finding of oscillatory time series, a more useful mea-

sure of the quality of a pixels data was the amplitude of oscillations around the data's baseline trend. For each pixel, the detrending algorithm for circadian data developed with the mFourfit toolkit [157] was used to subtract the baseline trend from the data, and then the pixel's overall signal size was calculated as the sum of squares of these baseline-detrended data. Further peripheral pixels were then omitted according to a subthreshold signal-size value. Finally, the array of retained pixels was checked for individual or small groups of unconnected pixels, such as those from an overlapping secondary plant leaf, and these were simply removed manually as required. For the leaf shown in Fig. 4.4(b), the resultant retained pixels after the above steps are shown in Fig. 4.5(b).

To determine the phase, we identified peaks in the baseline-detrended time series data. Because the data were noisy, a digital filter was applied to the data before identifying peaks, to remove unwanted high-frequency noise components from the signal. The data were filtered with MATLABs filtfilt routine, using a third-order Butterworth digital filter. The filtfilt routine filters the data twice, in the forward and then the reverse time direction, and in this way avoids any distortion of the signal in time. Eliminating such phase distortion was important for our purposes.

Having selected a third-order Butterworth filter, an additional parameter that was required was the cutoff frequency, f_{co}, of the filter. High f_{co} filters remove less noise than low-f_{co} filters, so giving a tighter fit to the data. From our observations of data fits, we found that f_{co} values between 0.1 and 0.2, representing between 10% and 20% of the Nyquist frequency, generally yielded good fits to the original data. To select the appropriate f_{co} value within this range, we observed that peaks of large amplitude and a sharply peaked profile were better fitted using a tight fit ($f_{co} = 20\%$) whereas smaller-amplitude, broader peaks were better approximated using a looser fit ($f_{co} =$

10%). Criteria were then set for using one of three filters: F1 (loose fit, $f_{co} = 10\%$); F2 (medium fit, $f_{co} = 15\%$); F3 (tight fit, $f_{co} = 20\%$). The full logic of the fitting process used is outlined below, and an example of the final outcome of this procedure is given in Fig. 4.6(a), for a single pixel within the detached leaf presented in Fig. 4.4(a).

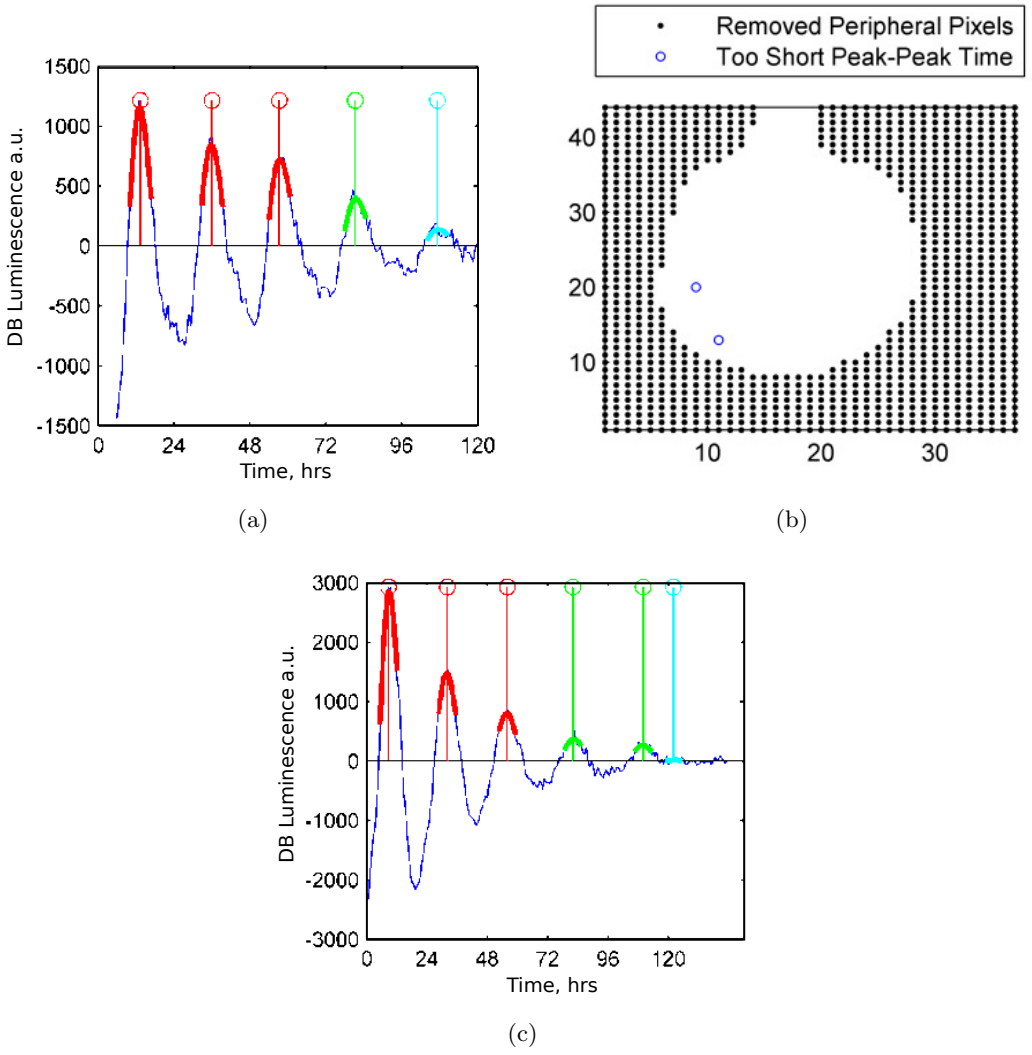


Figure 4.6: (a) Locally filtered data and corresponding peak location using three selected digital filters: blue, luminescence data (baseline-detrended); red, data filtered with filter F3 (20% f_{co}); green, data filtered with filter F2 (15% f_{co}); cyan, data filtered with filter F1 (10% f_{co}). (b) 'Pixel problem map', here showing the location of pixels that contain one or more peak-to-peak times $< 15h$. (c) Peak locations for example 'problem' pixel. Here, the sixth peak is clearly erroneous and was subsequently removed. Leaf: Detached, LL conditions. (DB denotes baseline-detrended data).

Method I: Used for all constant light (LL-LL) data.

1. Baseline-detrended data were filtered using filter F1.

2. Peaks in these filtered data were located using MATLABs findpeaks.m function, with minimum peak height threshold set to the mean value of the baseline-detrended data (approximately zero).

3. Two characteristics of the data around each peak were then used to alter the used fit and therefore refine the peak location:

   - Peak signal size factor (SSF), which we defined as the sum of absolute data values within a 24-h window around the peak.

   - Peak sharpness factor (SHF), which we defined as follows. Data values in a window extending 5 points on either side of the peak were normalised so that the peak height was set to 1. The SSH was then defined as the minimum of $[(1 - h_1), (1 - h_{end})]$, where h_1 is the normalised height of the first point in this 11-point window, and h_{end} is the normalised height of the last point in this window. An SSH value close to 1 indicates a sharp peak, and a value close to 0 indicates a broad peak.

The refinement process then proceeded as follows:

1. If the oscillation was of very large amplitude (SSF$> 10^4$) and sufficiently sharp (SHF > 0.25), filter F3 was applied to the data in this 24-h window and peaks in the window were located. If only one peak was found within the window, the F3 fit was retained and its peak time used. If more than one peak was found using this tight fit, the F3 fit was discarded and this process was repeated using

filter F2. Again, if more than one peak was found by using this medium fit, the original F1 fit was selected.

2. If the oscillation was of medium amplitude ($10^4 > \text{SSF} > 4 \times 10^3$) and sufficiently sharp (SHF$>$ 0.25), filter F2 was applied to the data in this 24-h window and peaks in the window were located. If only one peak was found within the window, the F2 fit was retained and its peak time used. If more than one peak was found using this medium fit, the F2 fit was discarded and the original F1 fit was selected.

3. If the oscillation was either small in amplitude ($\text{SSF} < 4 \times 10^3$) or broad (SHF$<$ 0.25), the original F1 fit was selected.

Method II: Used for data from leaves grown in LL and imaged in lightdark (LD).

1. Baseline-detrended data were filtered using the tight-fit filter F3, because wave profiles in the LD conditions are typically very sharp.

2. Peaks in these filtered data were located using MATLABs findpeaks.m function, with minimum peak height threshold set to the mean value of the baseline-detrended data (approximately zero).

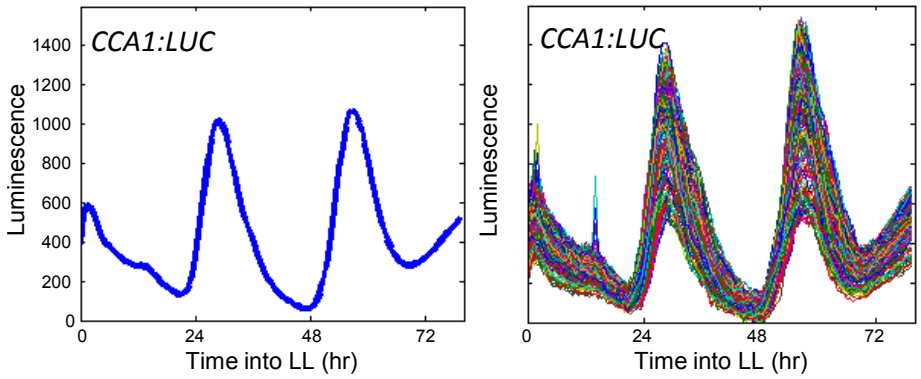
For leaves grown in LD and imaged under LL, one of method I or method II was used, dependent on the very early-time data features. In some leaves there was evidence of a very early- time peak, i.e., a synchronised peak during the LD conditions. In these leaves we used method II, as the tight-fit filter was required to locate this early peak accurately. In the other LD-LL leaves, whose fitted peaks were all in the LL regime, method I was used.

Having performed the fitting and peak finding above, the peak- to-peak times for each pixel were checked against reasonable (15 to 40h) circadian periods. Pixels with

peak-to-peak times out of this range were highlighted for inspection (Fig. 4.6(b)). Visual inspection then allowed, for example, manual removal of spurious peaks (Fig. 4.6(c)).

4.3 Loss of Spatial Synchronisation in Leaves without Wntrainment (LD → LL Experiments)

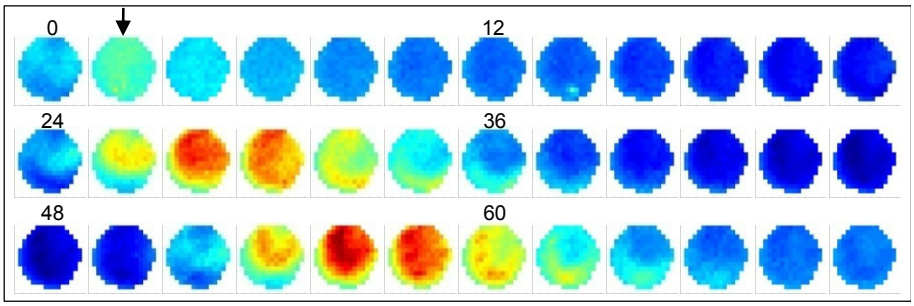
Most circadian studies on seedlings have averaged the luminescence signal across the whole seedling or leaf, in plants grown under LD cycles and then transferred to constant light (LL) for imaging. However, the pixel-level analysis here showed that such averaging masks significant heterogeneity (Fig. 4.7). Similar heterogeneity was observed for multiple circadian clock markers, including GI:LUC (Fig.A.3) and CCR2:LUC (Fig. A.4) (figures in Appendix A.10).



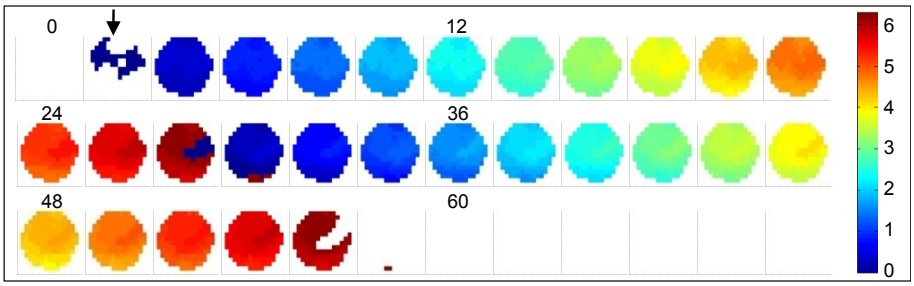
(a) (b)

Figure 4.7: Expression of CCA1:LUC rhythms in an intact leaf entrained under light-dark cycles and imaged under constant light. Plants were entrained under LD 12:12 cycles for 12 d and then transferred and imaged under LL conditions. (a) Average luminescence, detrended, for a CCA1:LUC leaf (Tab. A.2, plant 23); (b) detrended luminescence for all pixels of the leaf in a. Time is in hours; ZT0 corresponds to transfer to LL. [Figs. produced by DT and BW and published in [142].]

To assess the spatial organisation of circadian rhythms across the leaf, rhythmic



(a)



(b)

Figure 4.8: Spatiotemporal analysis of CCA1:LUC rhythms in an intact leaf entrained under lightdark cycles and imaged under constant light. (a) montage showing the spatial pattern of the luminescence in Fig. 4.7(b) (interval between images = 2 h); (b) data in (a) represented as a montage of the circadian phase (in radians) at each pixel. Arrow indicates the petiole. Time is in hours; ZT0 corresponds to transfer to LL. [Figs. produced by DT and BW and published in [142].]

CCA1:LUC luminescence data (Figs. 4.7(b) and 4.8(a)) were processed to identify the circadian phase at the single pixel level. The calculated phases were used to generate image sequences, termed phase maps (Fig. 4.8(b)). The circadian period of this leaf increased in LL conditions (first peak at ZT2, second at ZT28, and third at ZT56, where time is measured as Zeitgeber time (ZT) in hours since the last dark - light transition). In a synchronised leaf all leaf areas are expected to be at the same phase (and hence shown in the same colour) but here (Fig. 4.8(b)), as early as ZT18 there is clear variation in the colours in the map, indicating a range of different phases.

A spatiotemporal pattern was clear at ZT48 (Fig. 4.9). Along the midline, the leaf base phase led the leaf tip by $> 1.5h$. This loss of spatial synchronisation was also observed with GI:LUC and CCR2:LUC markers (Figs. A.3 and A.4). The CCA1:LUC

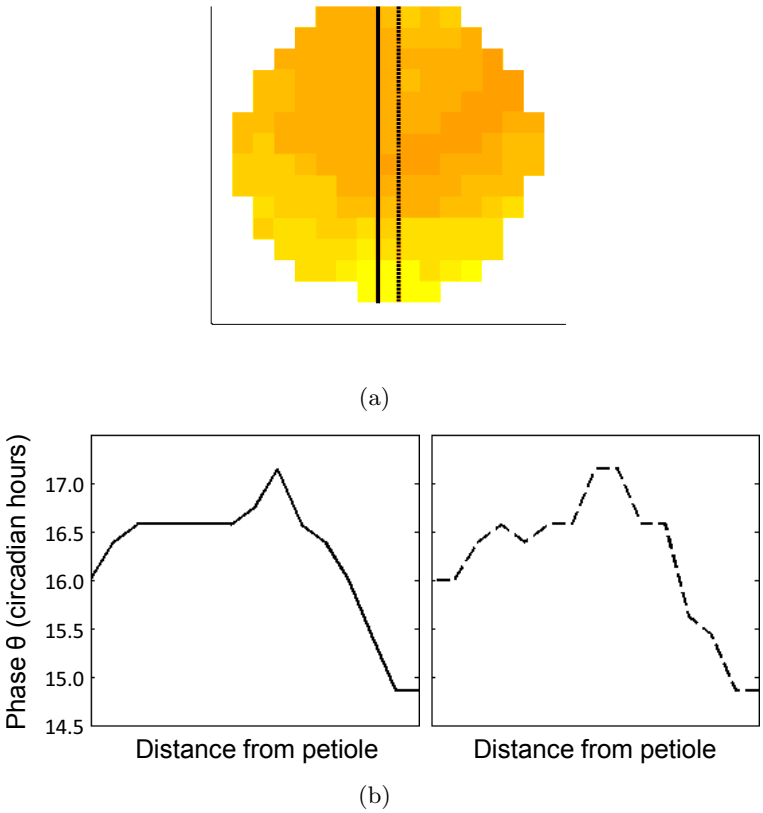


Figure 4.9: Phase at ZT48 along two central lines. The phase values are at the location of the pixel values, as shown by the overlaid lines. No coloured square is created for the uppermost data points, at the leaf petiole. Time is in hours; ZT0 corresponds to transfer to LL.

marker was used for all further analysis, because its high luminescence signals allowed the most precise analysis. Tabs. A.1 and A.2 presents an index of all plants analysed.

We analysed these patterns quantitatively by describing the phase 'coherence', R [158, 159], among the leaf oscillators at each time point t_k:

$$R(t_k) = \frac{1}{N(t_k)} \left| \sum_{n \in Q(t_k)} e^{i\theta_n(t_k)} \right|, \forall t_k \, s.t. N \geq 0.99 N_T. \tag{4.2}$$

where N is the sample size of pixels and θ_n are the phases of the individual oscillators within the population. The phase data are defined only between the first and the last peak time for each pixel, and because these times are different for different pixels within

the array, the size, N, of the set of active pixels, Q, is time dependent. To eliminate the poorly defined R values in early and late times (when only few pixels are active), we defined R only for time points where $N \geq 99\%$ (or 95% for a minority of leaves) of the total number of leaf pixels, N_T.

A pictorial representation of the phase coherence is provided in Fig. 4.10; R, takes values near 0 when phase values of the oscillators are widely dispersed on the unit circle and near 1 when the phases are tightly clustered or coherent. The equation for phase coherence, which comes from the definition of the complex order parameter in Kuramoto's original model for synchronisation of globally coupled phase oscillators:

$$R(t_k)e^{i\phi(t_k)} = \frac{1}{N(t_k)} \sum_{n \in Q(t_k)} e^{i\theta_n(t_k)}, \tag{4.3}$$

describes the same statistical quantity as the mean resultant length, R, of a population of circular data [160]; R is a measure of the dispersion of the data and $R = 1 - CV$, where CV is the circular variance.

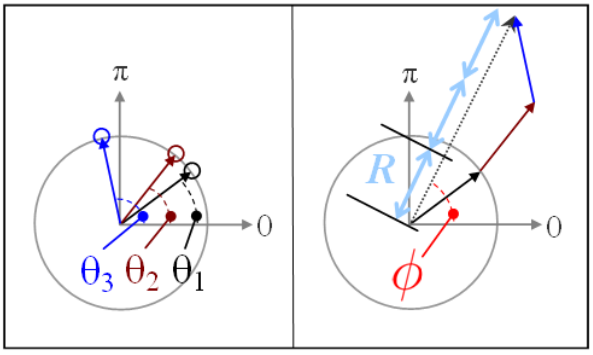


Figure 4.10: Calculation of phase coherence, R, of the phase vectors projected on the unit circle. (Left) At a particular time point, three pixels have phase angles θ_1, θ_2, and θ_3. Addition of the vectors and division by n = 3 gives the mean resultant length, R. In this example R is close to 1, showing that the phases of the three pixels are rather tightly clustered. (Right) indicates the mean phase at this time point.

As expected, R values for both intact and detached leaves of LD-grown plants

started close to 1. However, a loss of synchronisation (in terms of the phase coherence) could be quantitatively observed as R values decreased at various rates after transfer into constant conditions, with similar behaviour in intact plants and detached leaves (Figs. 4.11, 4.12 and 4.13 and A.8(b)). The emerging patterns were similar among some of the leaves. In the examples shown (Figs. 4.12 and 4.13), circadian peaks started first at the leaf margins (from ZT 46-48) and then spread toward the centre of the blade. A similar pattern was observed in the luminescence signal by microscopy (i.e., at the high-resolution level), which we investigate in more detail in Section 4.6. The presence of spatial patterns strongly suggest that the circadian system is heterogeneous among CCA1:LUC-expressing cells and that any intercellular coupling was thus insufficient to prevent the cellular heterogeneity from desynchronising, to some extent, the circadian rhythms among cells, within a few days in constant light.

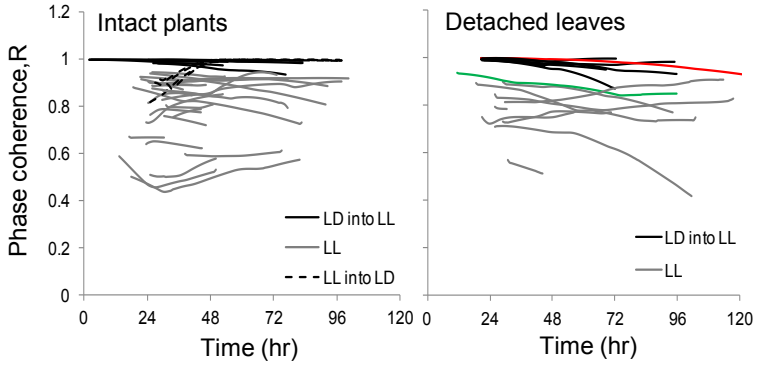


Figure 4.11: R values for intact plants and detached leaves. Red and green lines represent plant 3 (LD - LL conditions) and plant 16 (LL conditions), respectively (Tab. A.1). [Figs. produced by DT and BW and published in [142].]

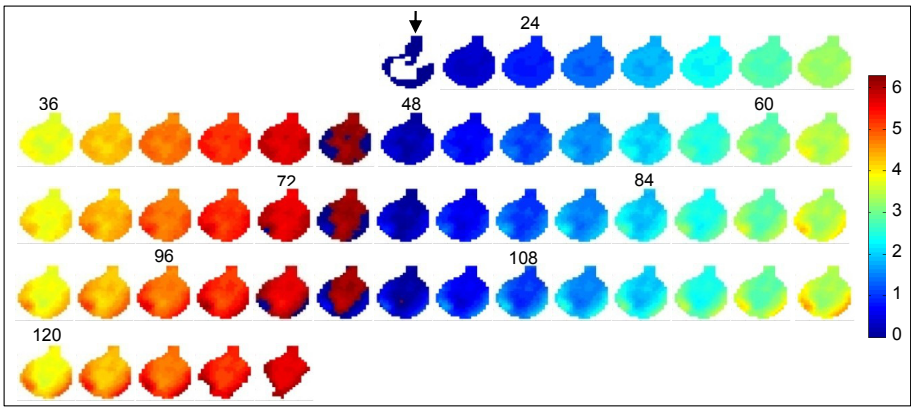


Figure 4.12: Phase montage for a CCA1:LUC detached leaf, grown in LD for 21 d and imaged in LL (Tab. A.1, plant 3); one cycle is 28 h. Interval between two images = 2 h. Time is in hours; ZT0 corresponds to the start of imaging. Arrow indicates the position of the petiole. [Figs. produced by DT and BW and published in [142].]

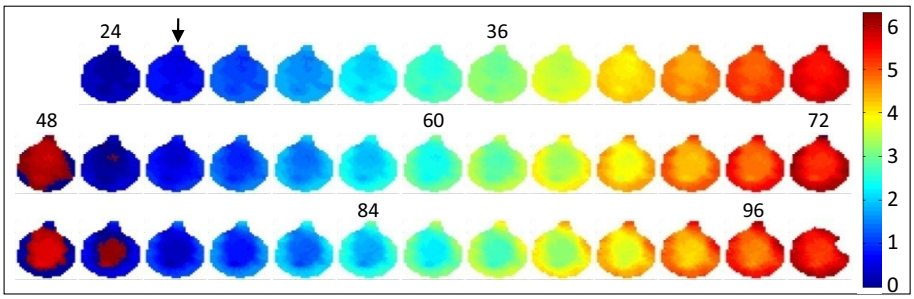


Figure 4.13: Phase montage for a synchronised detached leaf (Tab. A.1, plant 6). The plant was entrained under LD 12:12 cycles for 21 d and then transferred and imaged under LL conditions. ZT0 corresponds to transfer to LL. ZT0 corresponds to the start of imaging. Interval between two images = 2 h. Arrow indicates the position of the petiole. [Figs. produced by DT and BW and published in [142].]

4.4 Spatial Patterns of Circadian Rhythmicity Vary Among Leaves (LL→ LL Experiments)

To investigate the variety of possible spatial patterns of circadian rhythmicity which could be expressed in Arabidopsis leaves, CCA1:LUC expression was imaged, and R values calculated, in leaves of non-synchronised plants that were both grown and imaged in LL. As expected, these leaves were less synchronised and had a wide range of R values, which were < 1 but not < 0.4 (Fig. 4.11), whereas R would be 0 for uniformly

randomised phases. Two LD-grown leaves reached a similar level of asynchrony (R value) to that of LL-grown leaves within 4 days (Fig. 4.11); the data suggest other LD-grown leaves could do so within a few more days. Phase maps showed a range of phases (colours) in LL-grown leaves but their spatial distribution was not random (Figs. 4.14, 4.15 and A.8(c)).

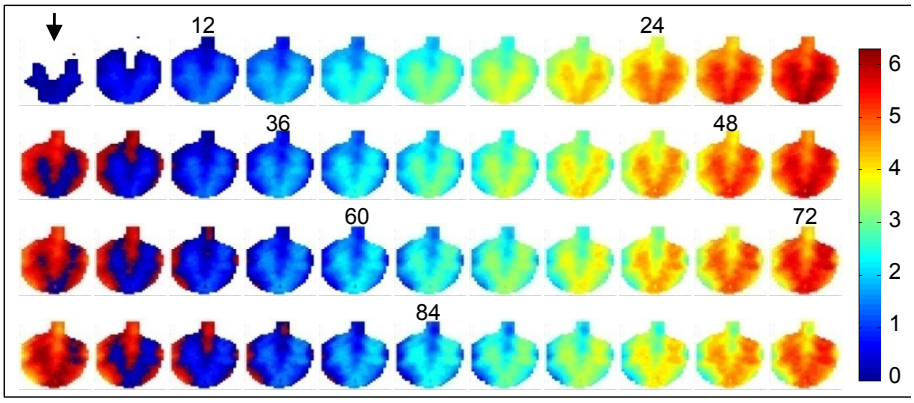
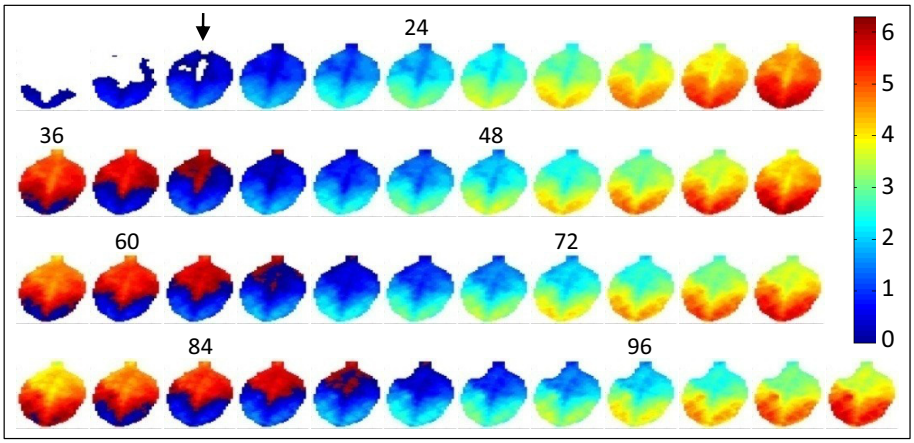
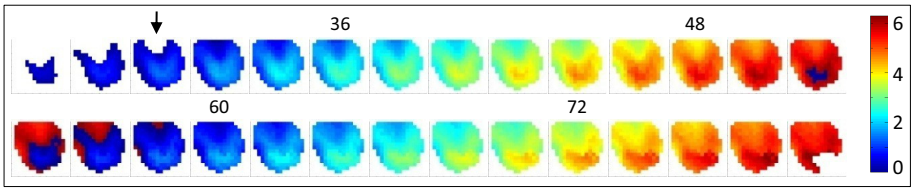


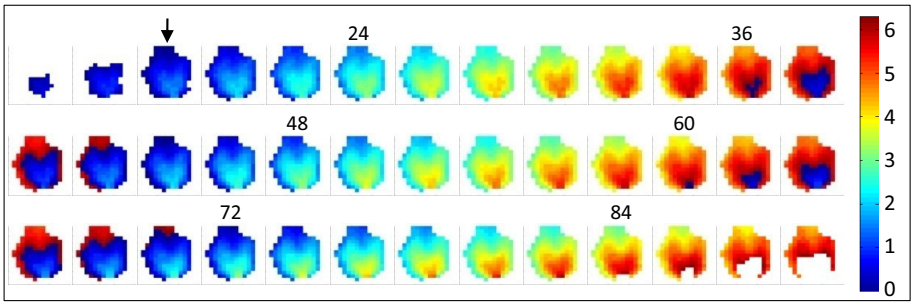
Figure 4.14: Phase montage for a CCA1:LUC detached leaf, grown in LL for 21 d and imaged in LL (Tab. A.1, plant 16); one cycle is 24 h. Interval between two images = 2 h. Time is in hours; ZT0 corresponds to the start of imaging. Arrow indicates the position of the petiole. [Figs. produced by DT and BW and published in [142].]



(a)



(b)



(c)

Figure 4.15: Other examples of phase montages for CCA1:LUC detached and intact leaves. (a) Phase montage for a nonsynchronised detached leaf (Tab. A.1, plant 20). The plant was grown in LL for 21 d and then transferred and imaged under LL conditions. ZT0 corresponds to the start of imaging. (b and c) Phase montages for two nonsynchronised intact leaves (Tab. A.2, plants 56 and 58, respectively). Plants were grown for 12 d and then transferred and imaged under LL conditions. ZT0 corresponds to the start of imaging. Interval between two images = 2 h. Arrow indicates the position of the petiole. [Figs. produced by DT and BW and published in [142].]

4.4.1 Characteristic Length Scale of Phase Coupling

The characteristic length scale at which the intercellular coupling in LL-grown plants is as effective as environmental entrainment was determined by comparing the local degree of phase coherence, i.e., in small clusters of pixels, for leaves in non-entraining (LL) conditions with the global, i.e., whole leaf, degree of phase coherence for leaves in entraining (LD) conditions. Reference typical R values for leaves in entraining conditions were calculated, both for detached leaves and for leaves from intact plants. For detached leaves, R-value data within the time window $0 - 36h$ from plants grown in LD and transferred to LL were ensemble and time averaged (mean), yielding a value of $R = 0.991$ (3 significant figures, Fig. 4.16(b)). For leaves from intact plants two such reference values were obtained: value I, calculated in the same way as above, i.e., time and ensemble-averaged R values over $0 - 36$ h and across LD-LL leaves, resulting in the same value, $R = 0.991$ (3.s.f.); and value II, calculated as the time and ensemble average over $60 - 76$ h and across LL-LD leaves, yielding the slightly higher value of $R = 0.998$ (3.s.f., Fig. 4.17(b)).

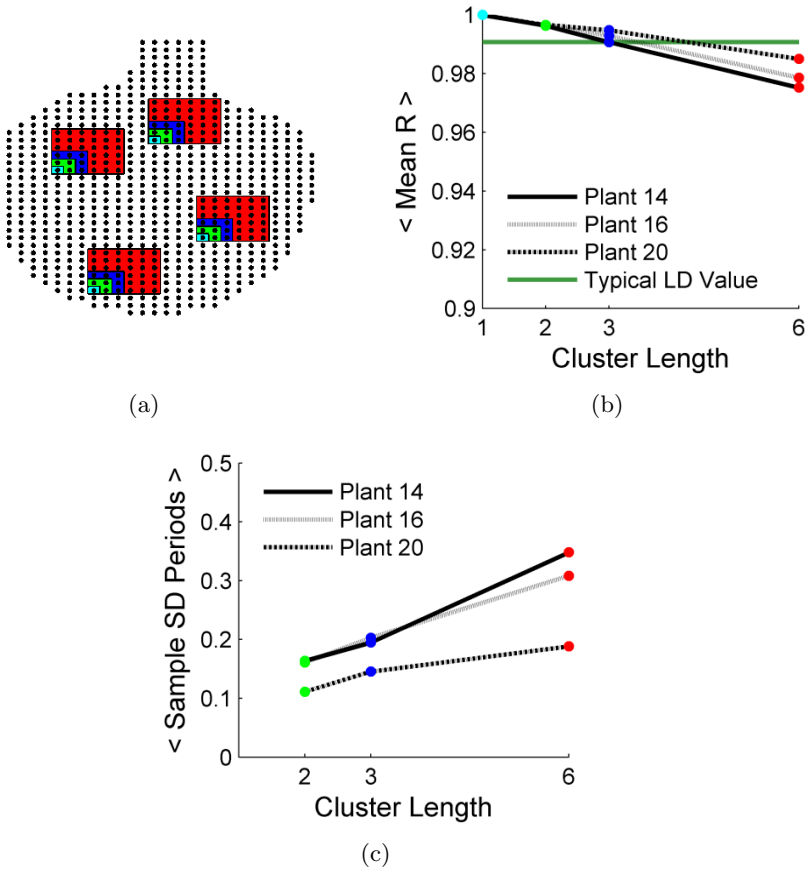


Figure 4.16: Effective length scale of intercellular coupling for phase and period. (a) Schematic of analysis of LL-grown leaf (detached leaf 16). Pixels (black dots) are overlaid with two or four representative pixel patches. The pixels in these patches were then grouped in nonoverlapping clusters of size 1×1 (i.e., single pixel, cyan), 2×2 (green), 3×3 (blue), and 6×6 (red); for clarity only one cluster is illustrated in each patch. (b) Phase coherence, R, as a function of cluster size for three LL detached leaves (plants 14, 16, and 20). Mean R denotes the time average and <> the ensemble average (over clusters). Coloured dots correspond to the size of cluster groups as in (a). Reference R values typical of LD-grown leaves are shown (horizontal lines). (c) Ensemble (across clusters) mean of sample standard deviation in period (mean peak-to-peak time) for clusters of varying size, using the same leaves and pixel patches as in (a,b).

Six leaves from LL-grown plants were chosen for the coupling length-scale study: plant indexes 4, 16, and 20 (detached leaves) and 37, 46, and 58 (intact plants). For the detached leaves, pixels within four randomly selected 6×6 patches were used to represent each leaf (Fig. 4.16(a)). Two patches were used for the smaller leaves of

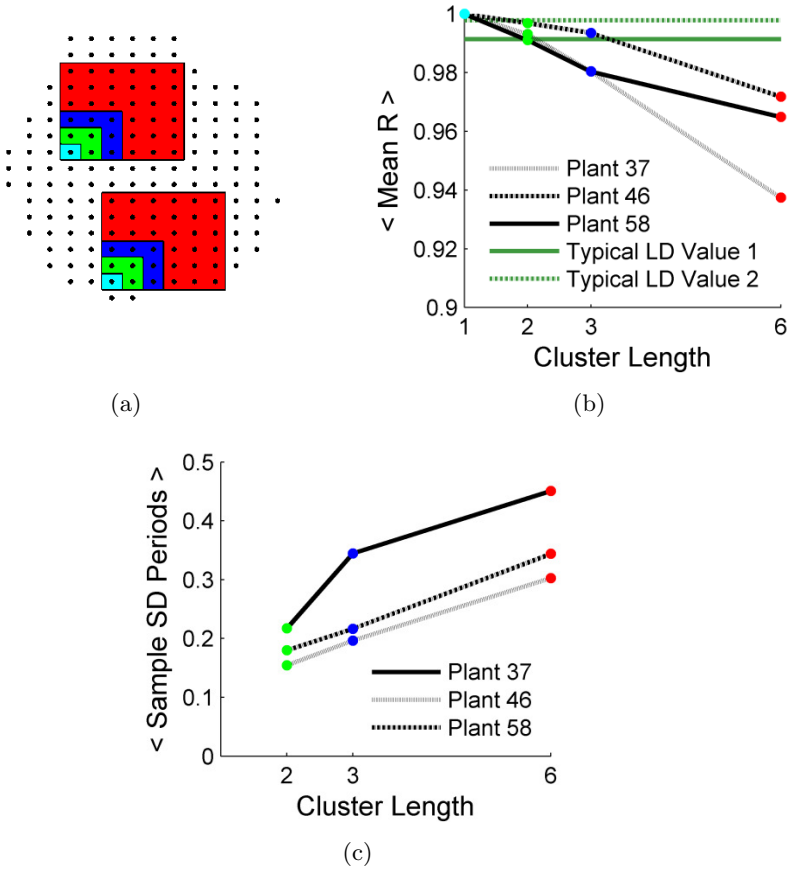


Figure 4.17: Effective length scale of intercellular coupling for phase and period. (a) Schematic of analysis of LL-grown leaf (intact plant 58). Pixels (black dots) are overlaid with two or four representative pixel patches. The pixels in these patches were then grouped in nonoverlapping clusters of size 1×1 (i.e., single pixel, cyan), 2×2 (green), 3×3 (blue), and 6×6 (red); for clarity only one cluster is illustrated in each patch. (b) Phase coherence, R, as a function of cluster size for three leaves from LL intact plants (plants 37, 46, and 58). Mean R denotes the time average and $<>$ the ensemble average (over clusters). Coloured dots correspond to the size of cluster groups as in (a). Reference R values typical of LD-grown leaves are shown (horizontal lines). (c) Ensemble (across clusters) mean of sample SD in period (mean peak-to-peak time) for clusters of varying size, using the same leaves and pixel patches as in (a,b).

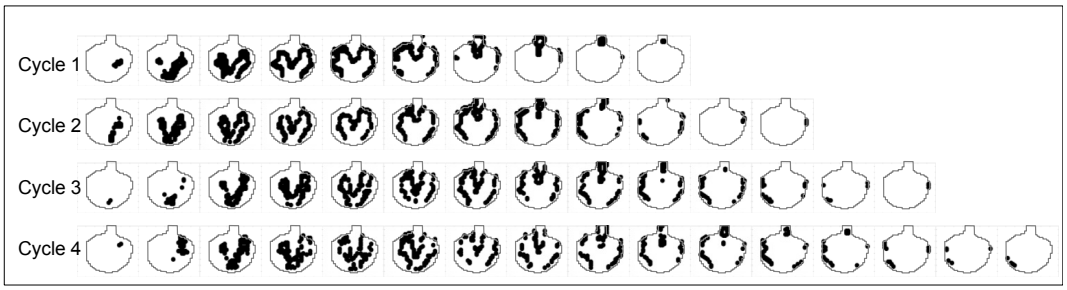
intact plants (Fig. 4.17(a)). All of the pixels in these patches were grouped in square clusters of varying size, 1×1, 2×2, 3×3, and 6×6, and for each choice of cluster size, the time and ensemble (across clusters) mean R value was calculated (Figs. 4.17(b) and 4.16(b)). For detached leaves, the cluster length at which intercellular coupling provides as much phase coherence as environmental entrainment is $3 - 4$ pixels, as judged by

comparison with the reference R value for detached leaves in LD (Fig. 4.16(b)). For intact leaves, the two LD reference values intersect the local coherence plots at slightly different lengths (Fig. 4.17(b)), but the upper bound of 4 pixels again seems reasonable. We conclude that in both detached leaves and intact plants the length scale at which coupling is as effective as entrainment is not >1 mm.

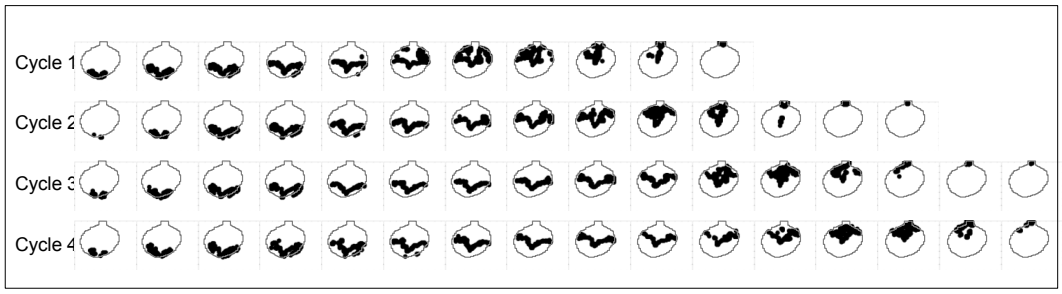
Using the same leaves and pixel patches as in the investigation of the characteristic phase length scale, local similarity of periods was also observed. Here, the sample SD of periods (mean peak-to-peak times) within each cluster of a particular size was calculated and then these values were ensemble averaged (i.e., across clusters), yielding a single value for the typical variation in period at a particular length scale (Figs. 4.17(c) and 4.16(c)). The SD of period increased with cluster size, for all of the leaves. Although there is clear variation in periods among spatially separate regions in the leaf (Fig. 4.19), this result indicates that periods are more spatially similar with decreasing length scale (Figs. 4.17(c) and 4.16(c)).

4.4.2 Peak Firing Maps

To better observe the shape of the traveling waves of CCA1: LUC in LL-grown plants, image sequences were constructed to map only pixels at the peak of rhythm in two independent, detached leaves (Figs. 4.18(a) and 4.18(b)). The leaves had different patterns: (i) Early peaks started in the middle of the leaf blade and spread toward the edges (Fig. 4.18(a)) and (ii) early peaks started at the tip of the leaf and spread toward the petiole (Fig. 4.18(b)). The propagation speeds of the first wave from tip to petiole were 1.3 mm/h (plant 16, Fig. 4.18(a)) and 1.8 mm/h (plant 20, Fig. 4.18(b)). These spatial patterns of peak firing were sustained over four cycles, indicating a degree of stability over time.



(a)



(b)

Figure 4.18: Peak firing patterns for two independent CCA1:LUC detached leaves (Tab. A.1, (a) plant 16 and (b) plant 20) grown and imaged in LL. Black dots represent the leaf areas peaking at the time of the picture. Interval between two images = 40 min. [Figs. produced by DT and BW and published in [142].]

4.4.3 Mean Period Patterns

The firing patterns in Fig. 4.18 could be linked to maps of the mean circadian period. For leaves grown and imaged in constant light conditions, mean period times were calculated per pixel as the sample mean of peak-to-peak times. This method was attempted for those leaves that provided good enough quality of data to allow phase and R value calculations. The mean periods and their sample SDs for the detached leaf featured in Figs. 4.4(a), 4.4(b) and 4.5(b) (plant 20) are shown in Figs. 4.19(g) and 4.19(h). For those leaves with period estimates from only two cycles, the range in the periods was calculated as a simplistic way of measuring data spread in place of SD. Note that SDs in the periods do not account for error in the process of locating data peaks.

The mean period was shown to be lower in the areas where the peaks started (period $\approx$ 22.5 h) and higher in the regions that the wave reached last (period $\approx$ 24.5 h) (Figs. 4.18(a), 4.18(b) and Fig. 4.19). The calculated periods are also locally correlated (Figs. 4.17(c) and 4.16(c)). To investigate the range of spatial patterns among plants, heterogeneity in the mean period was compared among 34 leaves (10 detached and 24 intact) grown and imaged in LL. The period patterns of 14 leaves fell into three groups (Tabs. A.1 and A.2): i) Higher period at the leaf margins (on one side or all around the leaf), sometimes with a slight vascular pattern of increased period: 2 intact leaves, 6 detached leaves (including plant 16; Figs. 4.14 and 4.19(e), 4.19(f). ii) Higher period in the central region of the leaf: 4 intact leaves (including plant 56; Fig. 4.19(a), Fig. 4.19(b)). iii) Higher period toward the petiole: 1 intact leaf (plant 58; Figs. 4.19(c) and 4.19(d)) and 1 detached leaf (plant 20; 4.19(g) and 4.19(h)). Of the remaining leaves, 10 had only one circadian cycle with good quality data over the majority of the leaf and 10 had highly variable periods in some or all of the leaf, so the mean periods gave no useful conclusions. Although phase differences within a leaf could be large (up to the uncharacteristically high value of 17 h in plant 12), leaves never had spatially randomised phases ($R \approx 0$), consistent with a role for intercellular coupling in limiting the extent of asynchrony. The range of R values and spatial patterns of rhythmicity also suggest a dynamic system, rather than a static, spatial pattern of circadian properties among the cells of the leaf.

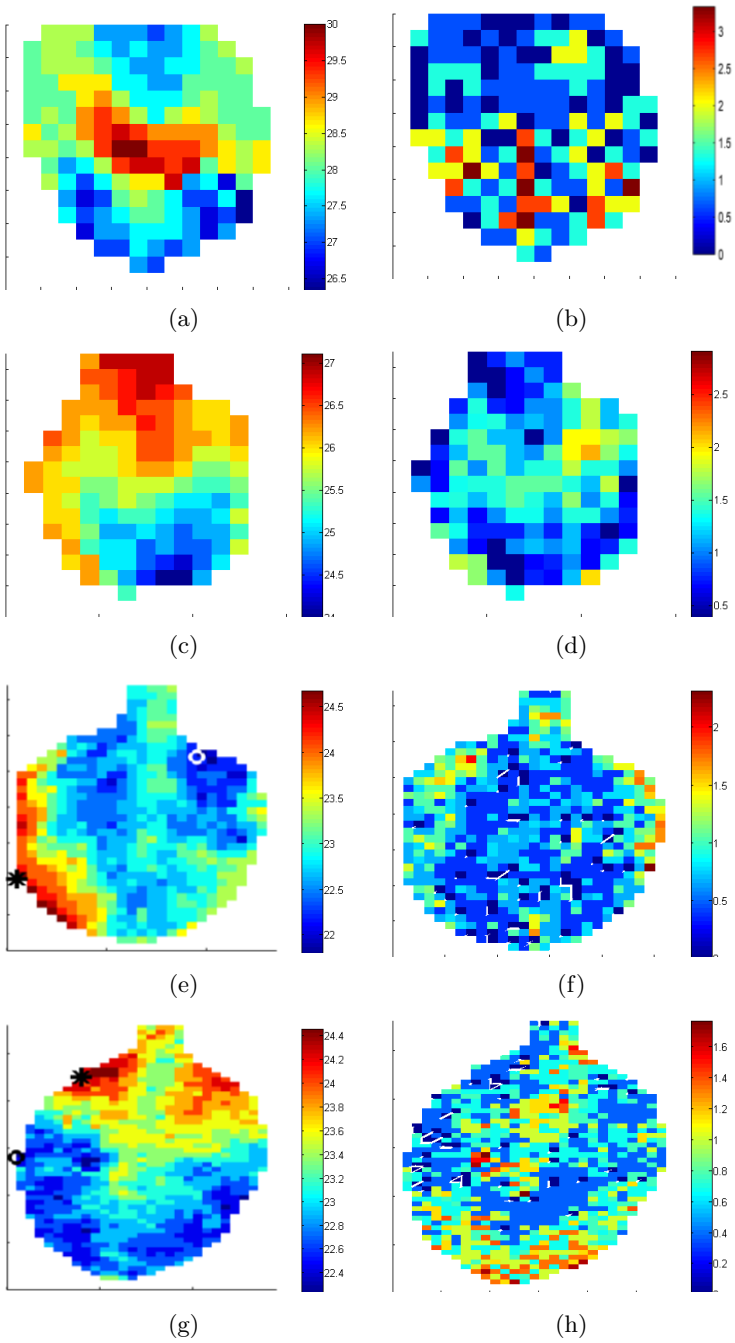
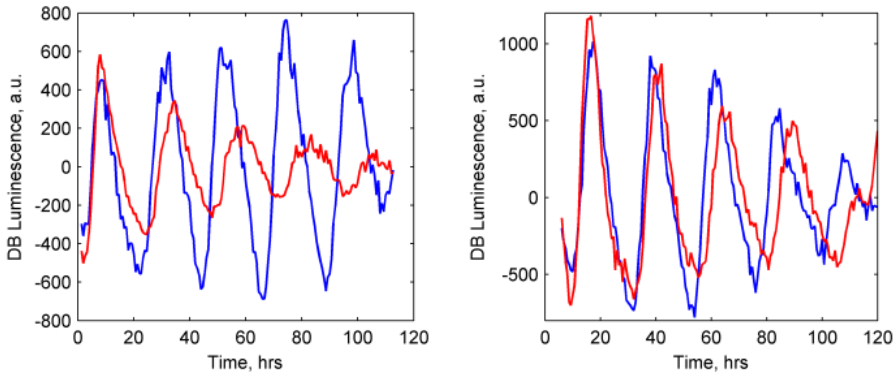


Figure 4.19: Mean period and SD for intact and detached leaves grown and imaged in LL. (a and b) Mean period (a) and range of values (b) for an intact leaf [Tab. A.2, plant 56 (number of periods, $n = 2$)]. (c and d) Mean period (c) and SD (d) for an intact leaf [Tab. A.2, plant 58 ($n = 3$)]. (e, f, g, and h) Mean period (e and g) and SD (f and h) for detached leaves [Tab. A.1, plants 16 ($n = 3$) and 20 ($n = 3$), respectively]. Note that only the first three periods were used as these were common to all pixels in the array.

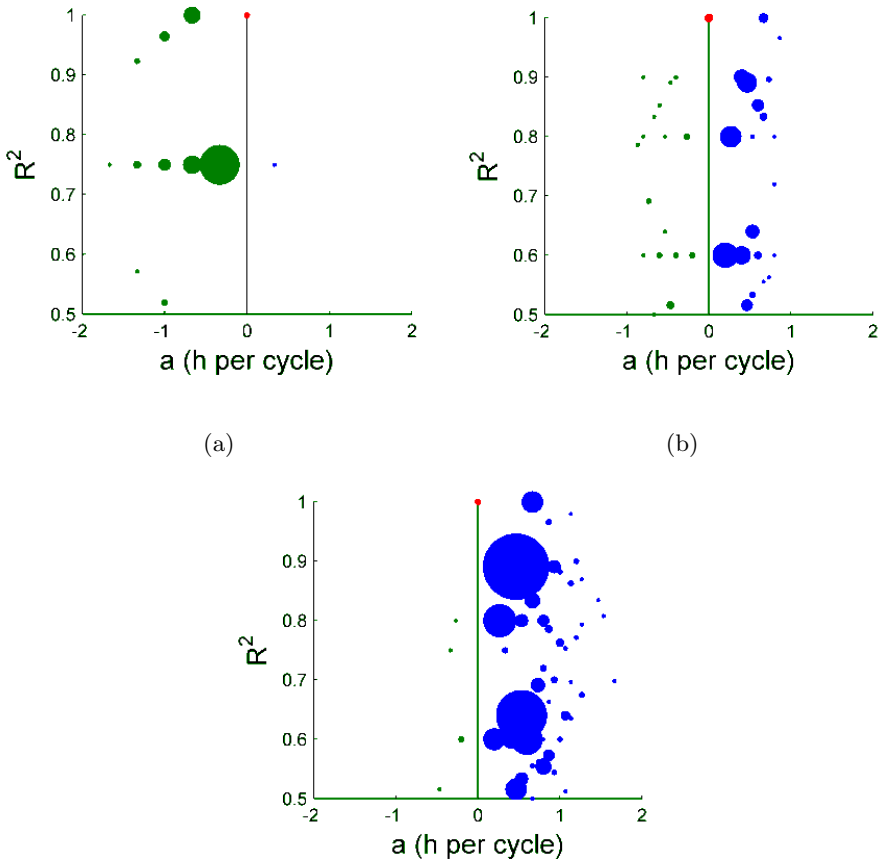


(a) (b)

Figure 4.20: (a) Longest and shortest period pixels for a detached leaf (Tab. A.1, plant 16). Min period pixel $(x, y) = 19, 27$ denoted by white o, left, and blue line, right. Max period pixel $(x, y) = 1, 10$, denoted by *, left, and red line, right. (b) Longest and shortest period pixels for a detached leaf (Tab. A.1, plant 20). Min period pixel $(x, y) = 1, 28$ denoted by o, left, and blue line, right. Max period pixel $(x, y) = 8, 45$, denoted by *, left, and red line, right.

4.4.4 Period Trends

Temporal trends in the period (peak-to- peak) times were investigated for the same six LL-grown leaves as in the investigations of characteristic length scale and local similarity of period. Pixels that gave a period profile reasonably consistent with a simple linear regression (i.e., $R^2 > 0.5$, Fig. 4.21) exhibited period increases or decreases or, in a small minority of cases (3.7%), no period change. Of the subset of pixels with $R^2 > 0.5$ in all six leaves, we tested the mean absolute change in the period length, which was found to be $|0.65|$ h per cycle. Across the six leaves, no pixel changed period by an average of >2.67 h per cycle, although this was an uncharacteristically high value; in four of the six leaves, the maximum period change per cycle was not >1.67 h per cycle on average. These slowly rising or falling periods would progressively alter the spatial phase patterns over longer timescales than our sampling times. Examples of the time series which exhibit period lengthening and period shortening for the leaves in 4.21 are shown in Figs. A.5, A.6 and A.7 in Appendix A.10.



(a)

(b)

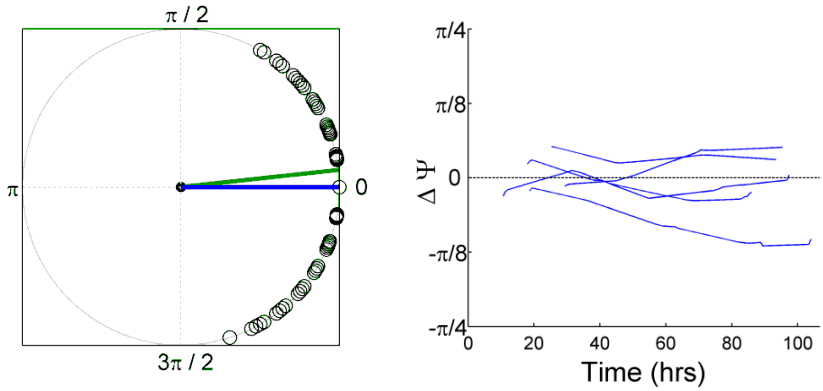
(c)

Figure 4.21: Regression slope (estimated linear rate of change of period), a, against the regression R^2 coefficient for cycle lengths in three leaves. Size of the plot marker is proportional to the number of pixels with that pair of a, R^2 values. For ease of visualisation, negative period trends (period shortening) are marked green, and positive period trends (period lengthening) are marked blue. No period trend (constant period) is marked red.

4.4.5 Effect of Amplitude

To consider the effects of amplitude on the R values, the leaf mean phase, ϕ, calculated in the R-value equation (4.3) was compared with the phase of the single time series obtained by averaging luminescence across the whole leaf (Fig. 4.22). The lat-

ter measure incorporates amplitude effects, because pixels with greater signal strength could dominate the final result. Among our six representative LL-grown plants, examples were found where the angle $\Delta\Psi$ between these two measurements was positive and in other cases $\Delta\Psi$ was negative, with different behaviour for different leaves (Fig. 4.22(b)). Thus, neglecting amplitude effects did not introduce a consistent bias in the description of phase coherence. The whole leaf phase, p, was calculated by averaging the preprocessed luminescence data across the leaf, detrending and then filtering (with filter F3), finding peaks and calculating the phase. The angle $\Delta\Psi$ was determined by simply subtracting the mean phase calculated in the R equation, ϕ, from p; i.e., $\Delta\Psi = p - \phi$. The MATLAB function 'unwrap' was used to locate and correct jumps > 2 in $\Delta\Psi$, which occur when the two angles are describing different circadian cycles.

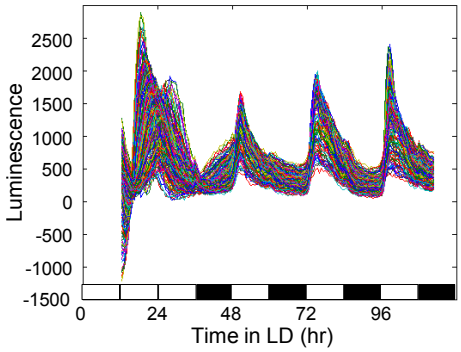


(a) (b)

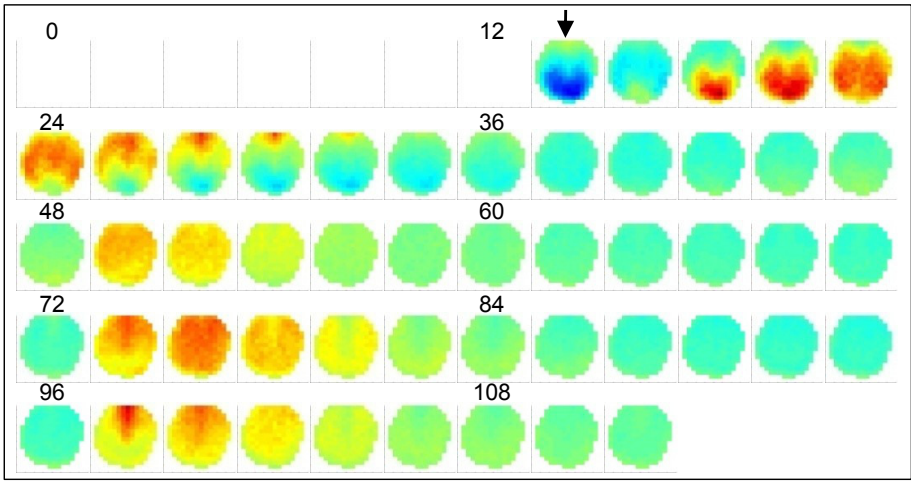
Figure 4.22: Effects of amplitude on the phase measure. (a) Plot of the mean phase for plant 16 at $t = 54.67$ h, as measured by (i) the mean (blue line) of all pixel phases (black circles), as used in R-value calculations, and (ii) the phase of the (detrended) mean luminescence (green line), which is affected by the rhythmic amplitude of each pixel. (b) Plot of the difference in angle, $\Delta\Psi$, between i and ii over time for six representative leaves (14, 16, 20, 37, 46, and 58). Including the amplitude of each pixels rhythm (in measure ii) has no consistent effect across leaves.

4.5 Balance of Internal and External Coupling (LL → LD Experiments)

To test how far these spatiotemporal patterns might affect circadian rhythms in nature, 23 intact, LL-grown plants were transferred to LD and imaged in cabinets. Twenty-two of 23 leaves yielded a good quality CCA1:LUC luminescence signal. The 4 fully analysed leaves lost essentially all phase heterogeneity within three cycles (ZT72), as they synchronised with the LD cycle (Figs. 4.11, 4.23, 4.24, and A.8(d)). The remaining 18 leaves exhibited complex multi-modal expression patterns during synchronisation, due to the variable starting phase and the acute light induction of the CCA1: LUC reporter, which hampered phase analysis. However, 16 showed near-complete synchronisation within four LD cycles, as judged by visually synchronous final peaks in the time series (Fig. 4.24). Thus, the spatiotemporal circadian patterns of leaves in constant light are rapidly erased by entrainment to the external light/dark cycle.



(a)



(b)

Figure 4.23: Resynchronisation of a nonentrained leaf. Plant 31 (Tab. A.2) was grown under LL conditions for 13 d and transferred into LL and then LD for imaging. (a) Luminescence for all pixels, detrended for the CCA1:LUC leaf. (b) Montage of the detrended luminescence, interval between two images = 2 h. Time is in hours; ZT0 corresponds to start of imaging. Arrow indicates the position of the petiole. [Figs. produced by DT and BW and published in [142].]

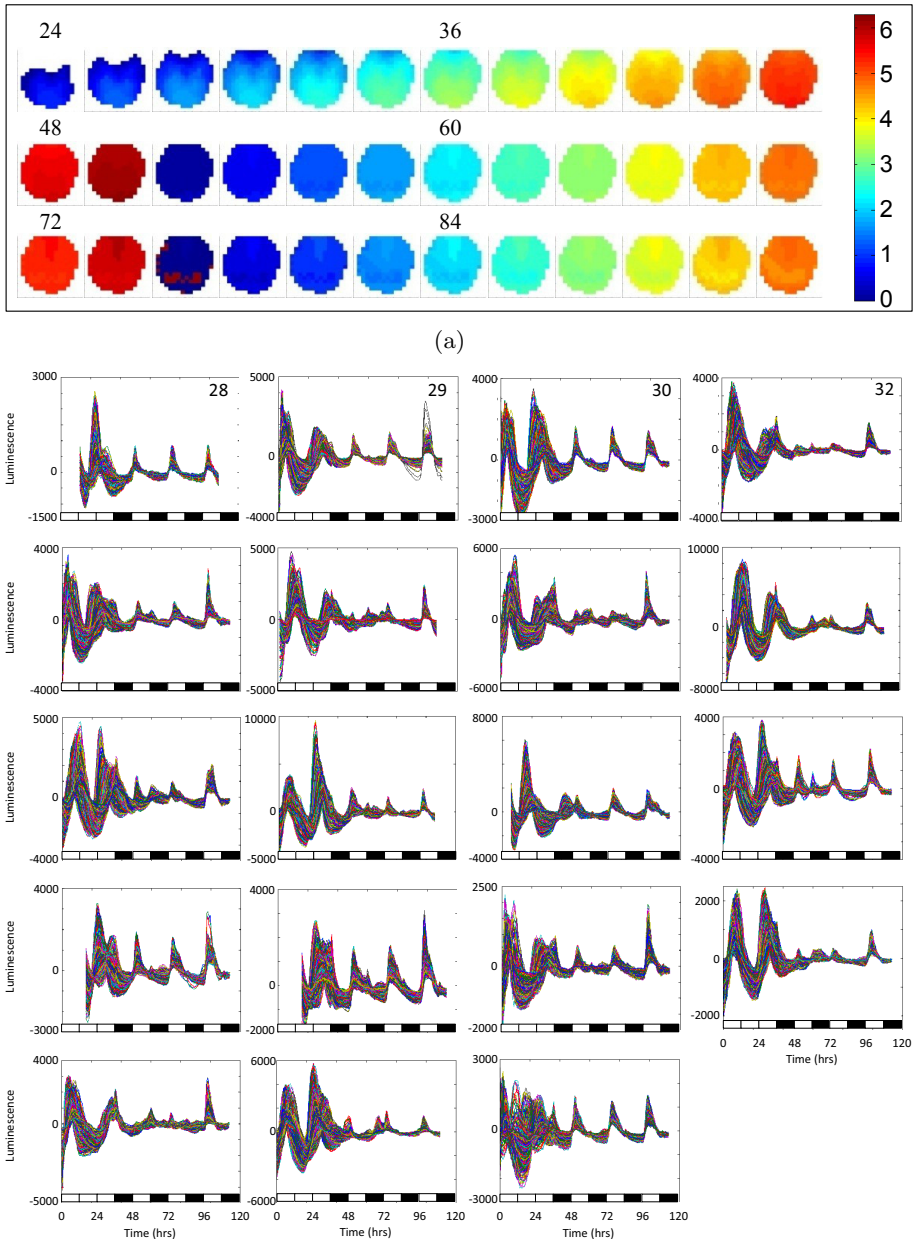


Figure 4.24: Phase montage and luminescence for CCA1:LUC leaves grown under constant light conditions and imaged in lightdark cycles. (a) Phase montage for plant 31 (Tab. A.2). Interval between two images = 2 h. (b) Luminescence for all pixels, detrended, for intact leaves. Plants 28, 29, 30 and 32 (Tab. A.2) are indicated; other plants are not included in Tab. A.2. All plants were grown under LL conditions for 13 d and imaged under 24 h of LL conditions and 96 h of LD cycles. Time in hours, ZT0 corresponds to transfer to imaging. [Figs. produced by DT and BW and published in [142].]

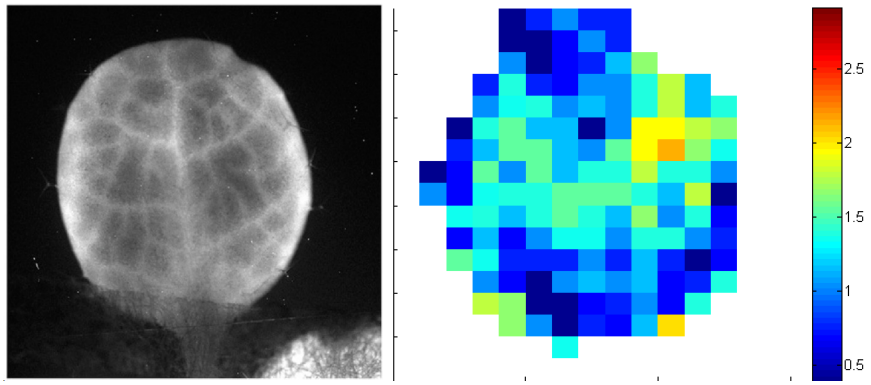
4.6 Local Variations from Macroscopic Phase Waves

By examining circadian luminescence at different areas in leaves we have investigated Arabidopsis circadian rhythms at a resolution not typically considered. However, each of the individual 'oscillators' in the leaf arrays, as described by a single pixel of the 'cabinet images', comprises at least tens, if not hundreds, of cells. In our consideration of coupling length scales thus far, the leaf has been viewed primarily as a continuum of similar cells, from which an emergent macroscopic (i.e., whole leaf level) phase wave is observed. How this wave emerges from interactions between single cells is less clear from the analysis. There are a number of candidates for intercellular coupling, but the connection mechanism for circadian dynamics is still unclear. To extend state-of-the-art Arabidopsis models [87] to meaningful spatial models with cell communication, some knowledge of the biochemical signal will be vital. However, in the absence of knowledge of the communication mechanism, we can still make some advances into early suggestions for the models.

Whatever the biochemical chemical signal involved in communication, a vital starting consideration for any model of circadian phase waves in leaves, which we believe exist through an interaction of cell autonomous rhythm and intercellular coupling, is whether the assumption of a continuum of similar cells is appropriate. Fukuda and others [156] argued the case for two types of cell, main lamina mesophyll cells and cells in the veins, with a phase delay observed in the veins. From this, they developed a coupled phase oscillator model comprised of one layer of mesophyll cells and one layer of a simple vein structure (in the shape of a + sign to describe the main vein and one trans vein). We have observed some cases of a difference in the behaviour between cells within vasculature and lamina regions, in both mean period (suggestion of longer

periods in veins in Fig. 4.19(g)) and phase (suggestions of a delay in the central vein in Fig. 4.18(b)), but clear evidence of a general delay in all vein cells was not found using the resolution provided in the cabinet images. Further, the differences in behaviour between vein and mesophyll cell types that we have observed have only been in the detached leaves - this is likely because the detached leaves are older and larger, therefore the resolution is greater, but alternatively the communication system in each case could be profoundly different.

Fortunately, we also have data from high resolution microscope images, which we have been able to use to probe the question of whether there is indeed a qualitative difference among the rhythmic properties of vein and mesophyll cells.



(a) (b)

Figure 4.25: Comparison of scales in the microscope data (a) and cabinet image data (b). [Fig. (a) produced by BW and published in [142].]

The resolution of the microscope images of intact plants is 1024 $\times$ 1024 pixels - a dramatic advance on the resolution obtained for intact plants in the cabinet images. For images of intact leaves in the cabinet setup, which were no larger than 32 $\times$ 32 pixels, this means that 1024 high resolution pixels are contained within a single low resolution pixel. Any description of the phase of the oscillators at the low resolution is still an average over a large ensemble of cells, and certainly an average over different

cell types.

4.6.1 Investigation of Different Spatial Scales of Dynamic Operation from Microscope Images

We investigated microscope data for a single leaf (Fig. 4.25(a)) at two different spatial scales to try to determine any differences in the phase dynamics among vein and mesophyll cells. This leaf was grown and imaged in constant light conditions. While these data are very rich in spatial resolution, only two full circadian periods were available for this leaf. New microscope data of circadian luminescence in other leaves are currently being collected, and of these some have given clear luminescence signals for $4-5$ days. Therefore, we provide a methodology here which we hope can be pursued and strengthened with the forthcoming longer data.

To make connection with the earlier work, we considered the data of the whole leaf at a 'low resolution' spatial scale, by compressing the full 1024×1024 into 32×32 data. This was achieved by first removing solar cosmics from the 1024×1024 data and then averaging the time series in space, so that pixel (x, y) in the low resolution image comprised the average of the time series $[32(x-1)+1 \ldots 32x, 32(y-1)+1 \ldots 32y]$ from the full high resolution data. We observed that the signals were of very good quality across the leaf for the two periods of data collected and so instead of the involved procedure of detrending and filtering the data around the peaks based on features of the detrended signals, we simply filtered all of the data using the tight fit digital filter (F3) to determine the peaks. Further, we used cubic splines to interpolate the data and achieve greater time resolution for the peak finding; the reason for this step will become apparent.

Having determined the peaks from the time series, we performed the same analysis

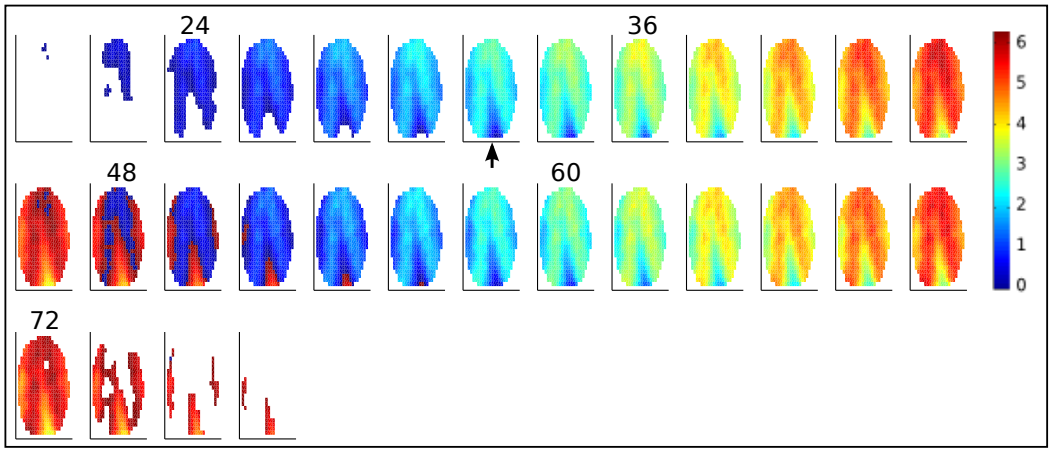


Figure 4.26: Phase montage for the nonsynchronised intact leaf shown in Fig. 4.25(a), at a low resolution level (32×32 pixels). The plant was grown and imaged under LL conditions. ZT0 corresponds to the start of the imaging. Interval between images = 2 h. Arrow indicates the position of the petiole.

as before for determining the spatial phase wave at this spatial resolution, and found a spatial wave propagating from the tip of the leaf towards the petiole and also towards the leaf edges (Fig. 4.26). The phase coherence among the oscillators at this spatial level is quite high ($R \approx 0.9$) and remains constant over the two periods for which data was collected (Fig. 4.27). This is reflected by a relatively stable phase wave in these two periods (Fig. 4.26). Because only two periods are available for this leaf, this behaviour samples from the dynamics in a wider time window which could be quite different.

We then wished to compare the results from data obtained at a higher spatial resolution. For the high resolution comparison, we chose not to use the full available resolution (1024×1024) but rather performed the same spatial averaging process as before but for a target image size of 256×256 pixels. As well as making further analysis more computationally feasible, this level of resolution allowed both clear definition of the vasculature and also a high signal quality in each of the pixels. We used the same simplified time-filtering and peak finding process for the time series in the 256 square array to determine spatial phase wave propagation (Fig. 4.28). Comparison with Fig.

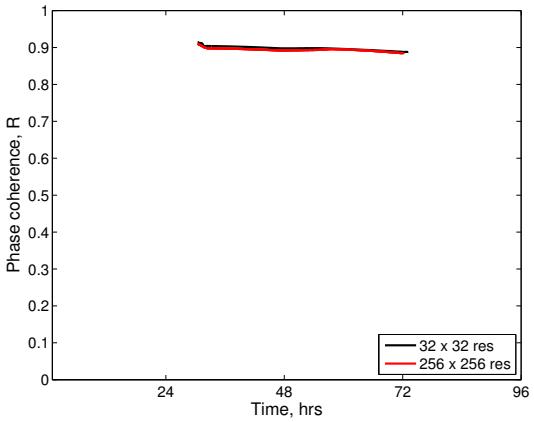


Figure 4.27: Phase coherence (R value) over time for the intact leaf in Fig. 4.25(a), as obtained from the low resolution 32×32 data and the high resolution 256×256 data.

4.26 shows that the same phase wave from central distal point in the leaf to the petiole and leaf edges is observed, but there are some suggestions of local deviations from that phase wave (e.g. at time $t = 28$h). We expected that the R-value for this high-resolution data would be similar to that for the low resolution image, and indeed there is very little difference between the values for low and high resolution data (Fig. 4.27). This validates that the previous analysis of rhythms from cabinet images captured well the phase coherence among all leaf oscillators. However, there are some important subtleties in the phase dynamics which can only be observed at the high resolution level, and these are less obvious from observation of the R values alone.

4.6.2 A General Methodology for Determining High Resolution Conclusions

Because the macroscopic phase wave dominates the behaviour in the leaf, it is at first glance difficult to conclude whether the phase lags in the vasculature, as has been previously proposed [156]. It is precisely the use of two spatial scales which we propose as a methodology to examine the fine details of phase interactions.

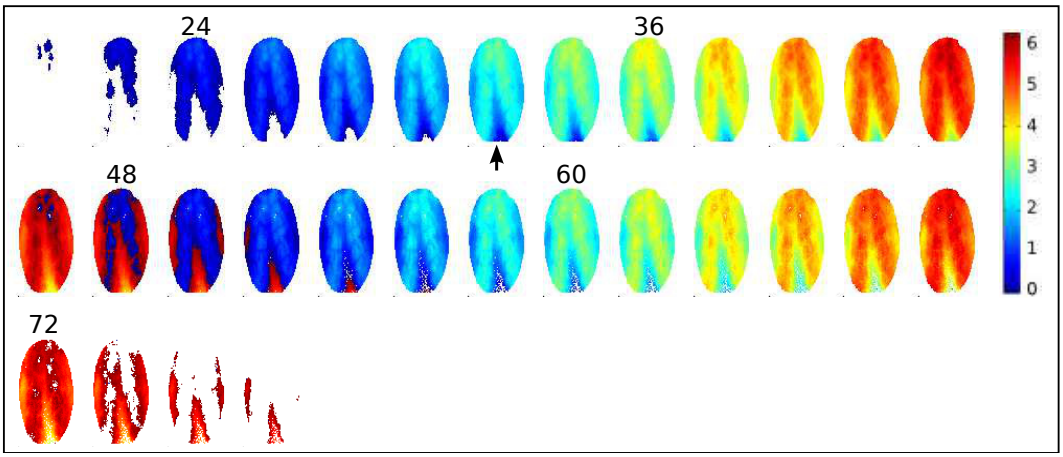


Figure 4.28: Phase montage for the nonsynchronised intact leaf shown in Fig. 4.25(a), at a high resolution level (256 × 256 pixels). The plant was grown and imaged under LL conditions. ZT0 corresponds to the start of the imaging. Interval between images = 2 h. Arrow indicates the position of the petiole.

All of the phase information used throughout our analysis is contained in the timing of peak circadian expression. For the data used here there are three peak times (two periods) available in each pixel. In Fig. 4.29(a) we show the time of expression of the first peak across the leaf from low resolution (32 × 32) data; the equivalent from the high resolution (256 × 256) data is shown in Fig. 4.29(c). To compare the difference in timing between the two resolutions, and therefore focus on the local details of peak timing, we need to subtract the 'low-resolution' peak times from the 'high-resolution' peak times. As a precursor to this step we linearly interpolated in space the 32 × 32 peak-time data (Fig. 4.29(a)) onto a 256 × 256 grid (Fig. 4.29(b)). Fig. 4.29(d) shows the result of subtracting the low resolution data in Fig. 4.29(b) from the high resolution data in Fig. 4.29(c). It appears that there is a spatial pattern in the local details of peak expression (Fig. 4.29(d)) which corresponds rather well with the vasculature (Fig. 4.25(a)). However, we note that this distinction is enabled by the fact that we earlier time-interpolated the time series data of each pixel; the difference in timings between phase leaders (cold colours) and phase laggers (hot colours) in the first peak is only

just detectable at the experimental time resolution available ($\Delta t = 0.67h$). We return to this point below.

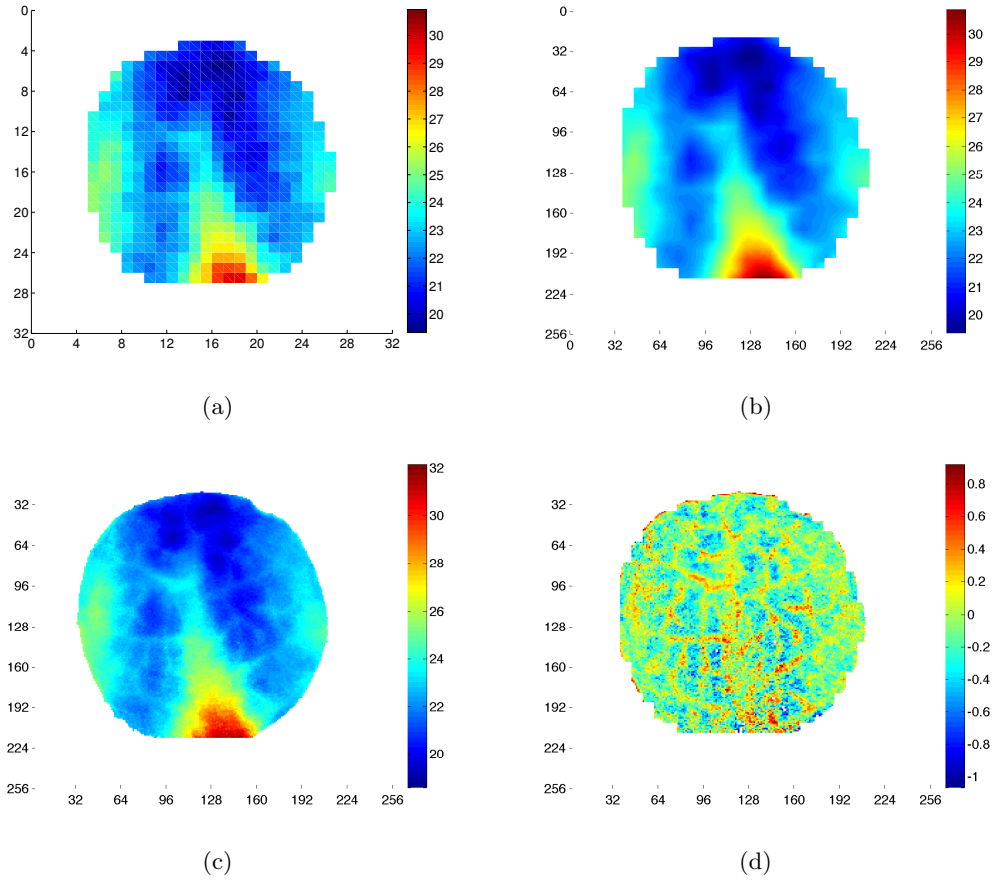


Figure 4.29: Plots of the first peak times across pixels in the (a) 32×32 resolution data; (b) linearly interpolated in space 32×32 resolution data and (c) 256×256 resolution data. Figure (d) shows the difference in timing of the first peak at the high resolution and low resolution levels, i.e. (c) - (b), revealing a spatial pattern. Times on colour bars are shown in hours.

To investigate further the details of peak expression we determined the vasculature at the 256×256 level using two methods based on the level of expression across the leaf. It is clear in Fig. 4.25(a) that the size of the signal is larger in the veins (indeed that is why we can see the vasculature in the microscope image). The level of expression is therefore a natural candidate for detecting edges between vein and mesophyll cells. We used a formal edge detection technique (the 'Canny' method from MATLAB's Image

Processing Toolbox) to find edges based on the expression level (Fig. 4.30(a)), and validated this by grouping cells into two types according to a threshold luminescence level (Fig. 4.30(b)). This process allowed us to test the hypothesis that peak expression in the veins might lag that in the nearby mesophyll cells by selecting three pixels; one within the vein region and two in mesophyll regions on either side of that vein (Fig. 4.30(b)). If there were no difference between mesophyll and vein expression timing we would expect the expression in these cells to follow the global (coarse grained) phase wave expression. Instead, we see in Fig. 4.30(c) that expression in the vein region, especially at the second peak (Fig. 4.30(d)), slightly lags that of the neighbouring mesophyll cells.

Figs. 4.31(a), 4.31(c) and 4.31(e) show the result of the subtracting procedure for all three sets of peak expression times. While leading and lagging cells do seem to be arranged according the vasculature, it is only really valid to draw conclusions about this on the experimental resolution time scale Δt (and not on the finer timescale of the interpolated data). Therefore we highlight in Figs. 4.31(b), 4.31(d) and 4.31(f) those pixels which 'lead the global phase wave' of Fig. 4.26 by $\frac{1}{2}\Delta t$ (coloured red) and those pixels which 'lag the global phase wave' by $\frac{1}{2}\Delta t$ (coloured blue). In these figures we overlay the vein structure and it is thus clear that blue pixels (phase laggers) often correspond to regions within veins and that red regions (phase leaders) often correspond with lamina regions, especially in the second and third peak expression times. The time separation of expression is short, however, and it is certainly not true to say that all cells in the vasculature are expressed after all mesophyll cells. Rather it is the case that the coarse grained phase wave dominates the timing of expression, and that *within* this phase wave there are local differences which seem to correlate well with cell type.

What this means for modelling purposes is that a) mesophyll cells and vein cells

should not be considered to be the same amplitude (since we can see the vasculature, so vein cells have higher expression) and that b) coupling between vein and mesophyll cells may differ from coupling within these cell types, such that the circadian timing of the two cell types differs slightly. Using the same methodology developed here for the high resolution, longer time microscope data now being collected will allow the findings here to be validated. It will be particularly interesting to see whether there are differences in the long term period trends of vein cells and mesophyll cells. We suggest that it would be interesting to also perform the same analysis for a detached leaf at the high resolution level to determine whether there is a qualitative difference in the coupling between vein cells and mesophyll cells when vein transport from the rest of the plant is cut.

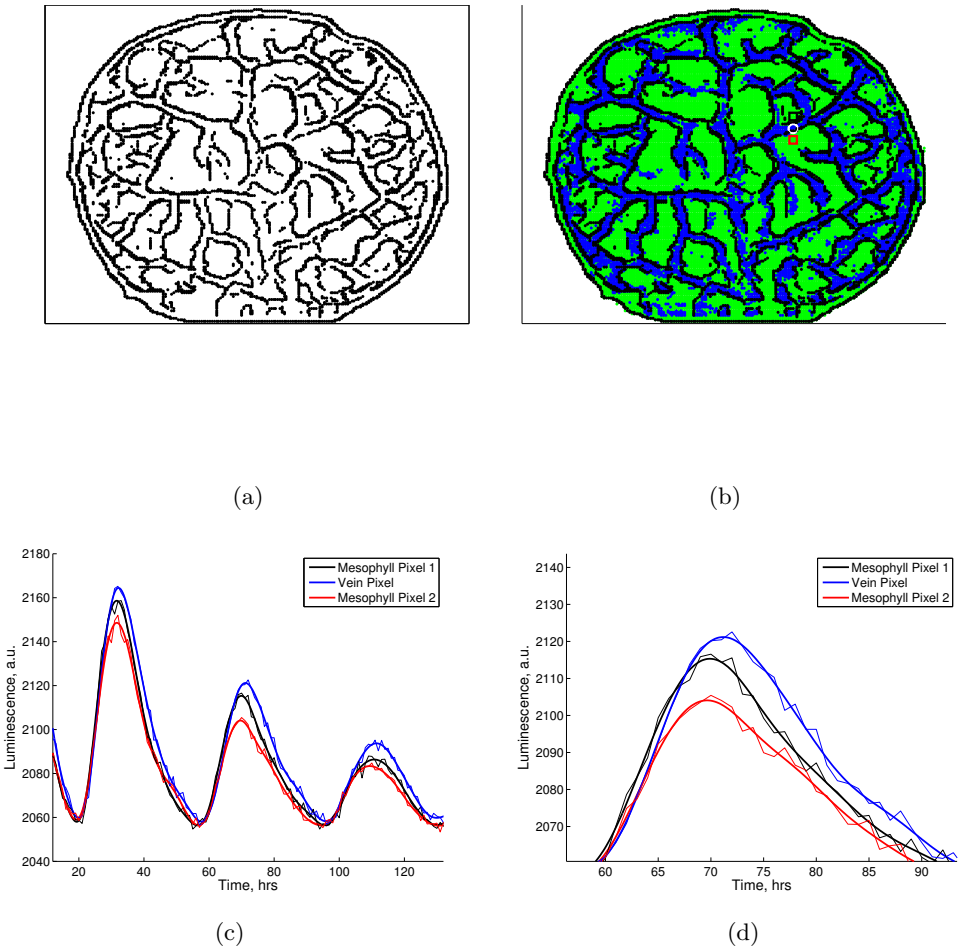


Figure 4.30: Investigating whether expression might lag in vein cells. (a) Leaf veins determined by edge detection using the 'Canny' method of MATLAB's Image Processing Toolbox, (b) Grouped mesophyll (green) and vein (blue) cells, as determined by threshold luminescence levels. Three pixels are highlighted: Mesophyll Pixel 1 (black square); Vein Pixel (white circle); Mesophyll Pixel 2 (red square), (c) Circadian expression of the three pixels in (b), (d) Zoomed view of the second peak showing a slight phase delay in the vein cell, which is spatially located between the two mesophyll cells.

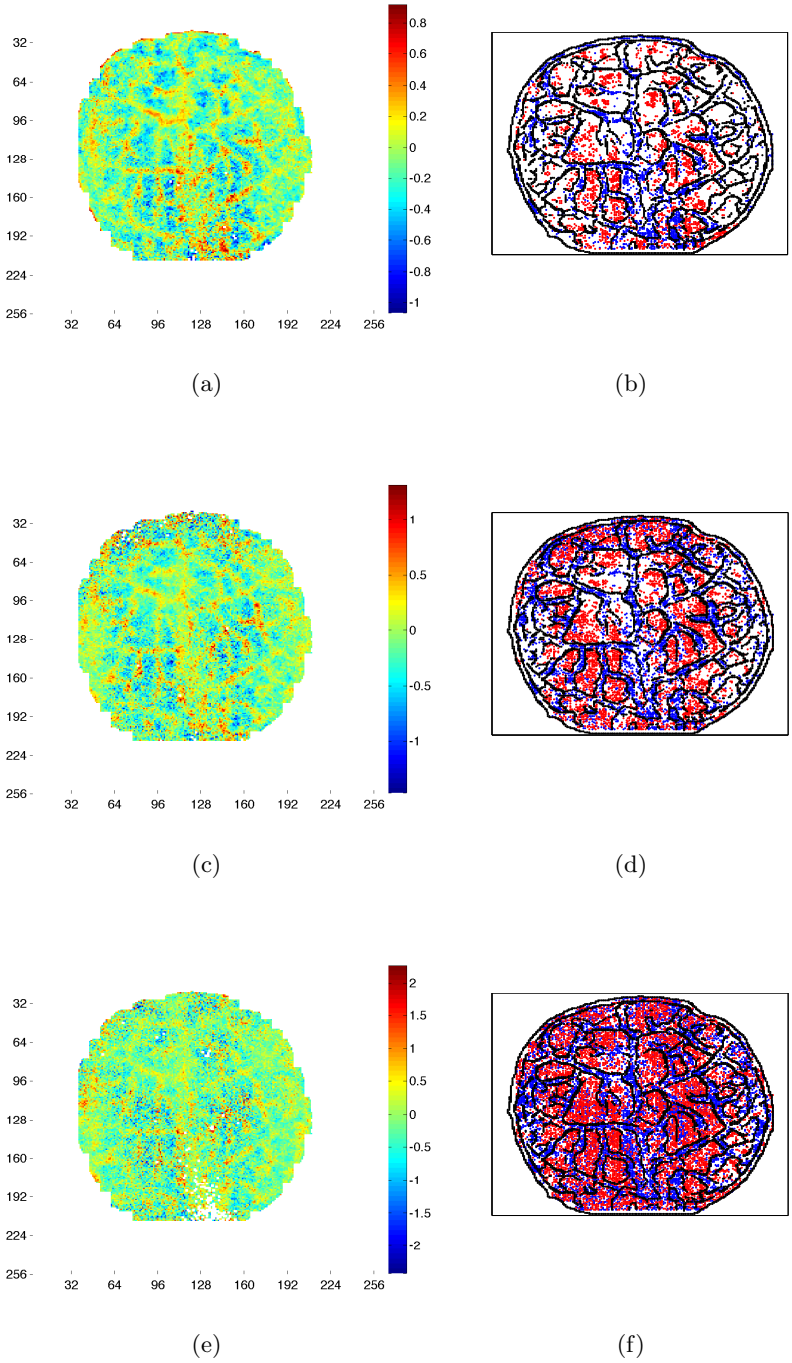


Figure 4.31: Finding cells which lead and lag the global phase wave. (a,b) expression at the first peak; (c,d) expression at the second peak, (e,f) expression at the third peak. (a,c,e) Local details of phase expression determined by subtracting the peak time obtained from the coarse grained (32×32) data from the peak time obtained from the high resolution data (256×256). (b,d,f) Pixels which 'lead' the global phase wave by $\frac{1}{2}\Delta t$ (half an experimental-resolution time point) (red) and pixels which lag the global phase wave by $\frac{1}{2}\Delta t$ (blue), with overlaid vein structure as determined from the 'Canny' edge detection method. Time on colour bars are shown in hours.

4.7 Discussion

Our research group developed a protocol to image LUC reporter expression in young Arabidopsis leaves over several days and analysed spatio-temporal patterns of clock gene expression under different and changing light conditions. It is likely that the data come from multiple different types of cell, with mesophyll cells the most numerous. From imaging the circadian rhythms of this combination of cell types, we observed an increase in phase heterogeneity among the cells just 48 h after transfer from light-dark entraining cycles into constant light conditions. This extends recent work which found desynchronisation between leaf stomatal guard cells and mesophyll cells after 7 days in constant light [161]. In constant light conditions, previous whole plant-level measurements often showed damping in the amplitude, and previous circadian rhythm research has proposed that a part of this damping is from increasing phase heterogeneity among the cells [161, 162, 163]. Population-level damping as a result of single cell desynchronization has also recently been shown from simulations of a stochastic version of the Arabidopsis clock model [164]. To assess the maximum extent of phase heterogeneity in the leaf, we measured leaves which were grown and imaged exclusively in these non-synchronising, constant light conditions. Over the measured time window, we found that the circadian luminescence was expressed in sustained spatio-temporal phase patterns characterized by a range of phase coherence values ($R = 0.4 - 0.95$); this is a larger range than seen in previous measures of phase coherence in SCN slice cultures, which maintained a rather constant level of $R \approx 0.85$ [103]. Intercellular coupling among the plant cells was evidently too weak to maintain synchronisation across the leaf, but was strong enough to avoid phase randomisation and to promote spatiotemporal waves of circadian gene expression. This result is consistent both with the earlier

studies arguing that plant cells were (at least partially) uncoupled [95, 154, 155, 161] and with results that indicated a detectable coupling [156]. Regional phase differences have been documented in the SCN [100, 102] but have rarely been investigated in the leaf, except for the stomatal guard cells [161] and proposals of longer periods in the leaf vasculature [156, 165]. Our analysis of R values, phase patterns, peak firing, and mean period maps revealed substantial variability among the set of leaves tested under constant light conditions, with at least three broad spatiotemporal patterns. The time-varying phase coherence of individual leaves (Fig. 4.11) suggests that the interaction of heterogeneous, stochastic circadian oscillators through intercellular coupling could generate a range of spatiotemporal patterns over time within a single leaf, and this was supported by the presence of different period trends of cells within the same leaf. With this view, the set of leaves and their time-windows measured in this study likely give just a sample of all of the possible leaf phase patterns, such as spiral wave phenomena [156] that we did not observe in any of the leaves.

The mechanisms of clock heterogeneity and intercellular coupling that underlie the observed patterns remain unclear, although our analysis of short time-series data at the microscope resolution level suggests that there may be small differences in the timing of expression of vein and mesophyll cells. We suggest that our methodology developed here, based on defining two functional spatial-scales, could be applied to the leaf data which is currently being collected in the research group to validate this finding. Several signaling components are known to move through the leaf and to alter the circadian clock, including the phytohormones auxin and cytokinins [166, 167]. These species are possible candidates for the intercellular coupling between cellular clocks which could promote wave-like propagation of circadian phase at the rates observed. A combination of experimental uncoupling assays, as demonstrated in the SCN [104],

and models of arrays of circadian oscillators [103] will be essential to further investigate these hypotheses and to test their relevance to any natural condition in the leaf. Such cellular uncoupling could be achieved in two main ways. A simple method would be to physically cut the leaf at certain points to remove direct communication between cells on either side of the incision. Careful experimentation would be needed to ensure that plant tissue is able to survive such intervention. A second, and perhaps more enlightening method, would be to place droplets of signalling inhibitors (e.g., auxin inhibitors) at individual locations on the surface of the leaf. Loss of coherence to some measurable extent among oscillators near the droplet would indicate disruption due to the particular inhibitor and give much needed insight into the method of communication between cells. Unfortunately, first efforts in the group to apply small doses of such chemical inhibitors within droplets of inert lanolin have proved lethal to the leaf's survival, and so more research is required to determine whether a non-lethal but effective dose can be administered. Our results from leaves grown and entrained to natural light-dark cycles, as well as from leaves transferred to these entraining conditions, show that light-dark signals are more important than intracellular coupling signals in synchronising the clocks of leaf cells. In these natural conditions some cell types will have less direct access to light and these may still entrain to chemical proxies for light (such as sugars) that are communicated among cells [168, 169]. It is possible that some plant cells beyond the range of even indirect light signals rely on intercellular circadian coupling, of the type we measure here among leaf cells.

Chapter 5

Summary

Proponents of rate equation models of biochemical oscillations have argued that every model is necessarily an abstraction of reality, and that one should 'avoid complicating from the outset the analysis of systems whose kinetics is already complex' [1]. However the argument of what constitutes a biochemical oscillation has changed by virtue of recognising phenomena such as noise induced oscillations and cellular coupling-sustained oscillations, which then highlights the importance of including these 'details' in the model. Increasingly there is a drive to understand how system-wide oscillatory behaviour arises from dynamic coupling between noisy oscillations in single cells. The work in this thesis contributes to a fuller understanding of both the roles of stochasticity in single cell oscillations and the intriguing complexities observed in non-spatially uniform biochemical oscillations in multi-cellular organisms.

In a stochastic world, oscillations with perfectly regular phase and amplitude are not achievable. While some researchers focused on the ability of biochemical oscillators to be robust to intrinsic noise stemming from the random nature of reactions in cells, others realised that such randomness could be a 'force for good' in terms of its ability to

sustain oscillations which would otherwise die out. In mathematical terms, such noise-induced oscillations have been understood in terms of noise smoothing the transition from (deterministic) limit cycle behaviour to stable steady state behaviour, such that when the noiseless (deterministic) system exhibits a steady state, the noisy (stochastic) system exhibits sustained oscillations. These noise-induced oscillations have commonly been identified by a peak in the fluctuation power spectrum at the dominant frequency, and near the Hopf bifurcation the frequency is closely approximated by the imaginary part of the complex eigenvalues found through deterministic linear stability analysis. A central message in our work is that such a power spectrum peak can exist in a system which does not have complex eigenvalues, i.e. a stable node, a result which seems counterintuitive from a deterministic viewpoint given the lack of an obvious resonant frequency. We have gone on to explore these far-from-Hopf noise-induced oscillations in great depth in this thesis.

In three simple biochemical models, which are abstractions of intracellular autocatalysis and dimerization processes, we used the linear noise approximation to identify the presence and quality of such stable node oscillations induced by internal noise, and verified the fluctuation spectrum peak by means of stochastic simulations. Notably, these models were very simple in their structure, featuring only elementary reactions between two species, and two of the models did not exhibit deterministic limit cycle behaviour for any choice of parameters. To understand the changes in oscillation quality brought about by changes in the kinetic parameters, we developed a novel Q-factor ($Q^{99\%}$) to describe these low-quality stable node oscillations. This technique was validated by using standard LTI filter theory whereby the parameterised biochemical oscillators were treated as a component within a filtering system. This validation method is in fact more than a convenient mathematical technique; by demonstrating

that our stable node oscillators can enhance the quality of other oscillatory signals we also infer their potential importance as constructive modules in a coupled oscillatory biochemical system. Indeed, although the stable node NIO were always of low quality for two species systems, we found that this was not true in a higher dimensional stable node system of coupled Brusselator modules, showing that the presence of complex eigenvalues is not required for visually coherent and sustained oscillations in a noisy biochemical system. Our investigation of these coupled stable node modules was limited to a cascade system of uni-directionally coupled similar modules; further work is required to understand the functionality of stable node modules in heterogeneous, multi-directionally coupled systems.

The simplest stable node NIO in systems with two species prompt interesting debate around what it means to have an oscillation. There is no doubt that there is a clear peak in the power spectrum, but we also showed that, compared to time series data from near-Hopf NIO, stable node oscillations were far less visible by the eyeball norm. By performing running averages of a small number of time series (thereby filtering some of the higher frequency noise) we were able to visualise stable node oscillations at the dominant frequency, and to show that the oscillation in the time series was clearer for higher $Q^{99\%}$ stable node NIO. However, we must be clear that the ability of even the highest quality two-species NIO to provide functionality as a precise biological timer with a well-defined period and phase needs to be questioned. Evidently more advanced methods than mere visual inspection of time series data from simulations of these stable node oscillators will be required to understand the extent to which two species stable node NIO can provide useful, rhythmic biological information to the cell. Having not explored this aspect we highlight this as a central limitation in our research. The technique of recurrence plot analysis from the literature in dynamical systems has

been cited as useful for understanding 'oscillating systems whose oscillations are not easily recognisable' [170] and we believe this offers an exciting prospect for aiding the understanding of our weak stable node oscillations. With this method, recurrences of the dynamical system in regions of phase space are recorded and mapped onto a two-dimensional matrix, providing an information-rich visual representation. Further, formal methods associated with recurrence plots have been developed for quantifying the extent of periodicity, such as the normalised *recurrence period density entropy* (RDPE), which aims to provide a 'straightforward sliding scale' between periodicity ($H = 1$) and aperiodicity ($H = 0$). If our $Q^{99\%}$ measure, derived quickly from the LNA, was found to be directly related to the RDPE measured from simulations then this would be a significant further validation of the work.

In Chapter 3 we investigated the effect of burstiness in the production of proteins on the oscillatory properties of downstream pathways, which is relevant for models of gene expression and of vesicular transport. In essence, this work combined our theory on stable node oscillations with the notion that in a stochastic framework the reaction $\emptyset \to A$ occurring at rate k is an entirely different event from the reaction $\emptyset \to mA$ occurring at rate k/m. We developed a theory for comparing non-bursty input and bursty-input stochastic models and identified a parameter to describe the extent of burstiness to linear noise approximation level. This burstiness parameter was found to act as a control parameter for the presence and quality of the oscillations, and since the burstiness it describes in terms of gene expression can be physically managed by, for example, small RNAs affecting mRNA degradation rates, this is an interesting and important new candidate for the control of oscillations in the cell. In our theory we were able to develop necessary conditions on the model parameters which indicated whether the model was susceptible to burstiness in promoting or destroying a noise-induced

oscillation. Through extensive parameter scanning using the linear noise approximation we tested and confirmed that in general these necessary conditions were powerful in their ability to predict when a biochemical model could indeed exhibit the phenomena of bursting induced oscillations and/or bursting destroyed oscillations. We used this combination of theory and numerical parameter scans to draw out some general rules for the effect of burstiness on noise induced oscillations in two species systems. Firstly, we found that when burstiness was present in the input to both species a wide range of effects on oscillatory properties were possible, highlighting that weak NIO in cellular processes could be strongly influenced by burstiness. Secondly, we found that the sign structure of the Jacobian of the deterministic rate equations was an important feature in determining whether burstiness induced oscillations could be promoted or destroyed. We highlight that in this work on burstiness the same limitations as in our previous work exist, in terms of our emphasis on a peak in the fluctuation power spectrum in identifying a noise induced oscillation and in our lack of an in-depth comparison with advanced time series analysis methods for verifying the quality of rhythmic information. Again, we suggest that this would be an excellent line of work to advance our research.

In Chapter 4 we addressed the open question of the role of intercellular coupling in plants by investigating the expression of the circadian gene CCA1 in Arabidopsis leaves. Given that detailed circadian modelling in Arabidopsis has drawn upon data from whole seedlings, one objective was to understand the extent to which this was a valid approach given the possible heterogeneity in rhythmic expression amongst single cells. Further, our work sought to understand the relative strength of intercellular coupling verses coupling between cells and the entraining signal, by means of observing changes in the spatial expression when leaves were transferred between different entrainment conditions. Further still, our work sought to understand whether intercellular coupling

is arranged in leaves such that circadian rhythms are expressed in a consistent spatial pattern and to answer this question required the analysis of a large data set of leaves. To analyse the luminescent circadian data expressed across leaves, we developed a simple method based on filtering and peak picking to assess the phase of a cluster of single cell oscillators represented by a single pixel in the spatial array. This method allowed a simple quantification of the phase coherence, R, of oscillations across the leaf under different entrainment conditions. We note that analysing short, noisy time series is a notoriously difficult problem and we would suggest that a limitation of our work was in not providing a second method of defining the circadian phase so as to cross-validate our method. Other suitable approaches include finding zero crossing times of the baseline detrended signal, using the analytic signal representation based on Hilbert transforms to define the phase or more advanced methods such as wavelet analysis to denoise the signal. The latter approach might be particularly useful based on the fact that generally both the oscillation amplitudes and baselines decreased over the course of the experiments.

We found that in constant light conditions, sustained but dynamic spatiotemporal patterns were observed across the leaf. The degree of coupling was captured by the phase coherence of oscillators across the leaf in both intact and detached leaves, and this was found to be variable both across leaves and over time within individual leaves. Importantly, coherence levels (R-values) were never less than 0.4, indicating that phase was never uniformly randomised over the leaf. The variation in R-values amongst leaves was reflected by variation in the spatiotemporal patterns of expression, which we studied both by assessing the spread of phase waves across leaves and by calculating mean periods of clusters of cells across the leaf. Through this we were able to show that conclusions in plant chronobiology based on a single endogenous circadian

rhythm across a whole leaf or plant in 'free-running' conditions might be failing to address important heterogeneity amongst cells. From our investigation of entrained leaves which were then transferred to constant light conditions, we concluded from both the spatial patterns which naturally form and the limited decrease in phase coherence over a few days that coupling amongst cells serves to limit the extent of desynchronisation. Conversely, the fast transition to spatial uniformity in the reverse experimental conditions (constant light to entrainment) showed that cellular coupling to the entraining signal dominates intercellular coupling. This point is important in supporting the validity of using data from whole tissues or whole seedlings to infer single cell behaviour in entrained conditions. Specifically, under entrained conditions the relatively weak coupling amongst cells means that rhythms across the whole plant tissue can quickly become extremely coherent; therefore time series data from the whole tissue is indeed highly representative of the rhythmic behaviour of single cells in the ensemble. We have therefore established that the message concerning the validity of using whole tissue data for representing single cell behaviour is strongly dependent on whether the plant is under entraining or free-running conditions. A further avenue of work that we would also recommend is to try to draw together any links between the shape of the waveform from the whole tissue signal and the spatial pattern of expression at a cell-cluster level. Analysis of circadian data from whole tissue or whole seedling data typically places great importance on the precise shape of the expression waveform, and we would suggest that any irregularities in the wave form could be associated with unusual spatial expression patterns caused by abnormal or defective cellular coupling.

Further investigations within the research group into spatial rhythmicity in leaves are being conducted with even higher resolution microscope data, and we provided a starting point for this work by considering the importance of separately considering the

dynamics in vein cells and lamina cells. The amplitude of oscillations in the vasculature is evidently larger than in lamina cells and we provided a framework to assess whether there is also a difference in phase between vein cells and lamina cells by comparing leaf phases at different spatial scales. In the single leaf we investigated, there is some evidence of a short delay of phase in the veins locally, but we were able to conclude that this local variation is within the global phase wave as observed in the lower-resolution leaf data of the main study. Indeed, the delay time between peak expression in lamina and vein cells in this single leaf is so short to be at the limit of what is observable at the time resolution available. We predict that the same analysis performed in different leaves, with sustained oscillations over more periods, will provide a more definite answer to this question. The result will then guide whether a spatial circadian model, which was not developed within the timeframe of our work, should comprise two subsets of dissimilar cells (vein and lamina) or whether a continuum of similar cells would be sufficient.

In this work we have investigated the roles of intracellular noise and intercellular coupling separately but it is likely that stochastic and spatial considerations in combination will be necessary to fully understand the emergence of complex rhythms in multicellular organisms. Indeed, a particularly fascinating avenue of work which we would have liked to have pursued would be to draw together the stable node oscillation aspect of our work with our research of rhythms in a real, coupled multi-cellular system. More specifically, could stable node oscillators really contribute to such complex biological oscillators as the circadian clock? We have explained that the circadian clock is more than just a rhythmic timer because its complexity enables entrainment to dawn and dusk in different photoperiods and also the ability to compensate for temperature changes. Such complexities are certainly beyond our simple biochemical

models exhibiting stable noise NIO - indeed, we have highlighted that even their ability to serve as a precise biological timer is to be challenged. However, it is certainly of great interest to us to investigate the extent to which these stable node oscillations can play a constructive role within a multi-cellular array of heterogeneous oscillators. For example, can an array of mixed strong (near Hopf) and weak (far from Hopf) noise induced oscillators be coupled in a biologically relevant way such that they produce a collective rhythm? Such an exciting question could indeed be approached using our preferred method of the linear noise approximation, as it has recently been shown that this method is highly suitable for assessing the level of synchronisation amongst cells in a stochastic framework [171].

Cellular oscillations such as circadian rhythms are necessarily complex because they need to be both adaptable but also robust to changing external conditions. The emphasis to date in plant circadian rhythms has been to ascribe the complexity to the intracellular working of the clock. We feel that it is important to question whether such flexibility and complexity could not be emergent from a system involving intricate coupling between a mix of different varieties of noisy oscillators. It is clear to us that a combination of further biological and mathematical research will be required to fully understand the relative importance of the intracellular mechanisms in noisy, crowded cellular environments and the extracellular coupling mechanisms in multi-cellular structures in producing life's most fundamental rhythms.

Appendix A

Appendix

A.1 A Simple Example of the Chemical Master Equation Description of Reactions

Consider the univariate system with reactions $\emptyset \rightarrow A, A \rightarrow \emptyset$. The stoichiometric matrix for this system is

$$\mathbf{S} = \begin{pmatrix} 1 & -1 \end{pmatrix}, \tag{A.1}$$

with propensity functions

- $a_1(n_A) = c_1$: The probability, per unit time, that a reaction of type 1 will occur.
- $a_2(n_A) = n_A \times c_2$: The probability, per unit time, that a reaction of type 2 will occur.

In a chemical reaction volume the propensities are better described by the molecule densities than their absolute number [56], so the functions $a(n)$ may be written $\Omega a(\widetilde{\phi})$ by defining the mesoscopic concentration $\widetilde{\phi} \equiv \frac{n}{\Omega}$. Then, the general univariate CME

may be written

$$\frac{dp(n,t)}{dt} = \Omega \sum_{j=1}^{J} \left(\mathbb{E}^{-S_{1,j}} - 1\right) a_j(\widetilde{\phi})\, p(n,t), \tag{A.2}$$

where the step operator $\mathbb{E}^m$ acts on functions of n with the rule $\mathbb{E}^m\left(f(n)\right) = f(n+m)$. Applying A.2 to the current example gives,

$$\frac{dp\left(n_A,t\right)}{dt} = \left(\mathbb{E}^{-1} - 1\right) a_1\left(n_A\right) p(n_A) + \left(\mathbb{E}^{1} - 1\right) a_2\left(n_A\right) p\left(n_A\right), \tag{A.3}$$

the solution of which is straightforward by moment generating functions.

A.2 The System Size Expansion for Multivariate CMEs

The macroscopic rate equation for a multivariate system;

$$\frac{d\vec{\phi}}{dt} = \mathbf{S}\vec{f}(\vec{\phi}), \tag{A.4}$$

can be written in terms of the vector of the molecule numbers:

$$\frac{d\vec{n}^*}{dt} = \Omega \mathbf{S} f(\vec{\phi}), \tag{A.5}$$

where $\vec{n}^*$ is the macroscopic number of molecules.

We recall the equivalent mesoscopic representation; the multivariate Master Equation:

$$\frac{dp(\vec{n},t)}{dt} = \Omega \sum_{j=1}^{J} \left(\prod_{l=1}^{L} \mathbb{E}_l^{-S_{lj}} - 1 \right) a_j\left(\frac{\vec{z}}{\phi}\right) p(\vec{n},t). \tag{A.6}$$

The fundamental Ansatz of Van Kampen's "System Size Expansion", is that we

can change the variable from $\vec{n}$ to the variable $\vec{\xi}$ as follows:

$$\vec{n} = \Omega\vec{\phi}(t) + \Omega^{\frac{1}{2}}\vec{\xi}. \tag{A.7}$$

This says that we *expect* the number of molecules in the system to be approximately equal to the macroscopic value $\Omega\phi(t)$ as obtained from rate equations plus a fluctuation term which scales approximately with the root of the system size. This rule can be rationalised by considering a Poisson process, for which the standard deviation of the fluctuations exactly matches $\sqrt{n}$.

We can rewrite the Ansatz as:

$$\vec{\tilde{\phi}} = \vec{\phi}(t) + \Omega^{-\frac{1}{2}}\vec{\xi}, \tag{A.8}$$

and it is clear that as $\Omega \to \infty$, this Ansatz predicts that $\vec{\tilde{\phi}} = \vec{\phi}$, i.e. in the large-volume limit, the mesoscopic and macroscopic concentrations are equal.

Transformation of the Master Equation

The System Size Expansion converts the mathematically intractable CME in discrete space $\vec{n}$ (representing the number of molecules), to an approximating linear Fokker-Planck equation in continuous space $\vec{\xi}$ (representing the fluctuations around the mean concentration). While $p(\vec{n}, t)$ describes the probability of the system being in state $\vec{n}$ at time t (conditioned on the system being in some state state $\vec{n_0}$ at time t_0), we use $\rho(\vec{\xi}, t)$ to denote the probability that the fluctuation variables are in state $\vec{\xi}$ at time t (conditioned on the system being in some state state $\vec{\xi_0}$ at time t_0).

Following the derivations in [14, 21, 33] the multivariate Master Equation (equation

(A.6)) is transformed as follows:

- ***Left Hand Side***:

$$\frac{dp(\vec{n},t)}{dt} = \frac{\partial\rho(\vec{\xi},t)}{\partial t} + \sum_{j=1}^{J} \frac{\partial\rho(\vec{\xi},t)}{\partial\xi_j}\frac{d\xi_j}{dt}. \tag{A.9}$$

We note that the time derivative of $p(\vec{n},t)$ in the Master Equation is taken with $\vec{n}$ constant, i.e. $\frac{d\vec{n}}{dt} = 0$. Therefore, from equation (A.7) we can write:

$$\frac{d\xi_j}{dt} = -\Omega^{+\frac{1}{2}}\frac{d\phi_j}{dt}, \tag{A.10}$$

and equation (A.9) becomes

$$\frac{dp(\vec{n},t)}{dt} = \frac{\partial\rho(\vec{\xi},t)}{\partial t} - \Omega^{+\frac{1}{2}}\sum_{j=1}^{J} \frac{\partial\rho(\vec{\xi},t)}{\partial\xi_j}\frac{d\phi_j}{dt}. \tag{A.11}$$

- ***Expanding the Step Operator*** $\mathbb{E}$: The multivariate operator $\prod_{l=1}^{L}\mathbb{E}_l^{m_l}$ acts on a function of $\vec{n}$ as follows:

$$\prod_{l=1}^{L}\mathbb{E}_l^{m_l}(f(\vec{n})) = f\left(n_1+m_1, n_2+m_2, \ldots, n_L+m_L\right). \tag{A.12}$$

If f is a smooth function of $\vec{n}$, we can replace this discrete operator by a continuous operator, by means of a Taylor expansion, i.e.

$$\prod_{l=1}^{L}\mathbb{E}_l^{m_l}(f(\vec{n})) = f(\vec{n}) + \sum_{l=1}^{L}\frac{\partial f(\vec{n})}{\partial n_l}m_l + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L}\frac{\partial^2 f(\vec{n})}{\partial n_l\partial n_k}m_l m_k + \ldots \tag{A.13}$$

However, a jump of m may cause a large change in $f(\vec{n})$. Intuitively, if we have a system with the order of millions of molecules of each species, a change of

1 molecule will have little effect on the reaction propenities, but in a system of only tens of molecules this might have a considerable effect. Van Kampen's transformation of variables formalises this point.

In equation (A.12) a jump of m_l in n-space corresponds to a jump of $m_l\Omega^{-\frac{1}{2}}$ in ξ-space. As $\Omega \to \infty$ these jumps in ξ-space become small. Therefore, when Ω is large, and we work with the fluctuation variable ξ, the replacement of the discrete operator with a continuous one can be expected to work well.

We remark here that whether acting upon functions of ξ or $\widetilde{\phi}$, the continuous operator is written:

$$\prod_{l=1}^{L} \mathbb{E}_l^{m_l} = 1 + \sum_{l=1}^{L} \frac{\partial}{\partial \xi_l} m_l \Omega^{-\frac{1}{2}} + \frac{1}{2} \sum_{l=1}^{L} \sum_{k=1}^{L} \frac{\partial^2}{\partial \xi_l \partial \xi_k} m_l m_k \Omega^{-1} + \ldots, \tag{A.14}$$

which can be explained as follows.

For a univariate system, we write the continuous approximation of the step operator for functions of ξ as:

$$\mathbb{E}^m(f(\xi)) = f(\xi + m\Omega^{-\frac{1}{2}}) \tag{A.15}$$

$$= f(\xi) + \frac{df(\xi)}{d\xi} m\Omega^{-\frac{1}{2}} + \frac{1}{2} \frac{d^2 f(\xi)}{d\xi^2} m^2 \Omega^{-1} + \ldots \tag{A.16}$$

In equation (A.6) we need to consider how the step operator acts upon the transition probabilities, which are functions of $\widetilde{\phi}$. Recalling equation (A.8),

$$\mathbb{E}^m(f(\widetilde{\phi})) = f(\widetilde{\phi} + m\Omega^{-1}) \tag{A.17}$$

$$= f(\widetilde{\phi}) + \frac{df(\widetilde{\phi})}{d\widetilde{\phi}} m\Omega^{-1} + \frac{1}{2} \frac{d^2 f(\widetilde{\phi})}{d\widetilde{\phi}^2} m^2 \Omega^{-2} + \ldots \tag{A.18}$$

We can write

$$\frac{df(\widetilde{\phi})}{d\xi} = \frac{df}{d\widetilde{\phi}}\frac{\partial\widetilde{\phi}}{\partial\xi}, \tag{A.19}$$

and from equation (A.8),

$$\frac{\partial\widetilde{\phi}}{\partial\xi} = \Omega^{-\frac{1}{2}}, \tag{A.20}$$

so that

$$\frac{df}{d\widetilde{\phi}} = \Omega^{\frac{1}{2}}\frac{df(\widetilde{\phi})}{d\xi}. \tag{A.21}$$

Therefore equation (A.17) may be written

$$f(\widetilde{\phi} + m\Omega^{-1}) = f(\widetilde{\phi}) + \frac{df(\widetilde{\phi})}{d\xi}m\Omega^{-\frac{1}{2}} + \frac{1}{2}\frac{d^2f(\widetilde{\phi})}{d\xi^2}m^2\Omega^{-1} + \dots \tag{A.22}$$

Comparison of (A.15) and (A.22) shows that the continuous operator has the same form when acting upon functions of ξ of $\widetilde{\phi}$, and it can be written as:

$$\mathbb{E}^m = 1 + \frac{d}{d\xi}m\Omega^{-\frac{1}{2}} + \frac{1}{2}\frac{d^2}{d\xi^2}m^2\Omega^{-1} + \dots \tag{A.23}$$

Finally, returning to the multivariate case, we can write the operator as:

$$\prod_{l=1}^{L}\mathbb{E}_l^{m_l} = 1 + \sum_{l=1}^{L}\frac{\partial}{\partial\xi_l}m_l\Omega^{-\frac{1}{2}} + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L}\frac{\partial^2}{\partial\xi_l\partial\xi_k}m_lm_k\Omega^{-1} + \dots \tag{A.24}$$

- ***Expanding the Propensity Function*** $a_j(\vec{\phi})$:

Finally, approximation of the propensity function $a_j(\vec{\widetilde{\phi}})$ by a Taylor series about

the macroscopic rate equation $f_j(\vec{\phi})$ proceeds as follows [21]:

$$a_j(\vec{\tilde{\phi}}) = a_j(\vec{\phi} + \Omega^{-\frac{1}{2}}\vec{\xi}) \tag{A.25}$$

$$= a_j(\vec{\phi}) + \sum_{l=1}^{L} \frac{\partial a_j(\vec{\phi})}{\partial \phi_l}\Omega^{-\frac{1}{2}}\xi_l + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L} \frac{\partial a_j(\vec{\phi})}{\partial \phi_l \partial \phi_k}\Omega^{-1}\xi_l\xi_k + \dots \tag{A.26}$$

As $\Omega \to \infty$, $a_j(\vec{\phi}) \approx f_j(\vec{\phi})$, so we can write:

$$a_j(\vec{\tilde{\phi}}) = f_j(\vec{\phi}) + \sum_{l=1}^{L} \frac{\partial f_j(\vec{\phi})}{\partial \phi_l}\Omega^{-\frac{1}{2}}\xi_l + \frac{1}{2}\sum_{l=1}^{L}\sum_{k=1}^{L} \frac{\partial f_j(\vec{\phi})}{\partial \phi_l \partial \phi_k}\Omega^{-1}\xi_l\xi_k + \dots \tag{A.27}$$

Performing the above transformations and inserting into equation (A.6) yields the multivariate System Size Expansion.

As with any perturbative approximation scheme the approximation becomes more accurate with the number of terms one considers. In the System Size Expansion, we can collect terms of different orders in Ω, which provide different levels of approximations of the noise. Terms of order $\Omega^{\frac{1}{2}}$ on the left and right hand sides of the Master equation cancel due to the fact that $\vec{\phi}$ satisfies the macroscopic REs (A.4).

- ***To Linear Approximation Order,*** Ω^0:

  The volume independent terms Ω^0 offer the first approximation of the fluctuations, and by considering only these terms we obtain the Linear Noise Approximation. These are often the only terms which are considered to describe the noise, though in recent years there have been notable exceptions which consider higher orders in the expansion [46, 172]. By considering the LNA terms, the intractable Master Equation in discrete space is formulated into a linear Fokker-Planck equation in continuous space. For a multivariate system this linear Fokker-Planck

equation has the form:

$$\frac{\partial \rho(\vec{\xi}, t)}{\partial t} = \sum_{i,j} A_{ij} \frac{\partial (\xi_j \rho)}{\partial \xi_i} + \frac{1}{2} \sum_{i,j} D_{ij} \frac{\partial^2 \rho}{\partial \xi_i \partial \xi_j}, \tag{A.28}$$

where the A_{ij} are the elements of the Jacobian matrix of the macroscopic system of equations, $\mathbf{A}$, and D_{ij} are the elements of the diffusion matrix $\mathbf{D}$.

- ***Next Order Corrections to the Linear Approximation,*** $\Omega^{-\frac{1}{2}}$:

  The procedure of calculating the volume dependent terms $\Omega^{-\frac{1}{2}}$ is considerably more involved than that for the linear noise approximation, although this has been aided by very recent developments of dedicated software to perform these calculations [173]. It has been shown that considering these terms yields, for example, corrections to the mean of ξ in small reaction volumes [174].

A.3 Time Correlation Matrix Derivation

This derivation follows along the lines presented in Gardiner [56].

In a univariate system, the autocorrelation function for species i can be written:

$$R_{\xi_i}(t_1, \tau) = \langle \xi_i(t_1 + \tau)\, \xi_i(t_1) \rangle, \tag{A.29}$$

where <> denotes the expected value. For a stationary process, $R_{\xi_i}(t_1, \tau)$ is independent of t_1 so we can set $t_1 = 0$ without loss of generality:

$$R_{\xi_i}(\tau) = \langle \xi_i(\tau)\, \xi_i(0) \rangle. \tag{A.30}$$

Similarly, for the multivariate system we can write the time-correlation matrix for

the vector of random variables $\vec{\xi}$:

$$\mathbf{R}(\tau) = \left\langle \vec{\xi}(\tau)\vec{\xi}(0)^T \right\rangle, \tag{A.31}$$

where the diagonal element $\left\langle \vec{\xi}(\tau)\vec{\xi}(0)^T \right\rangle_{ii}$ corresponds to the autocorrelation function for species i, i.e. $R_{\xi_i}(\tau)$. The off-diagonals give cross-correlation functions.

Let us denote a particular vector of values that $\vec{\xi}$ can take by $\mathbf{z}$. Then we can write:

$$\begin{aligned}
\left\langle \vec{\xi}(\tau)\vec{\xi}(0)^T \right\rangle &= \int \pi\left([\mathbf{z}_\tau, \tau]\,; [\mathbf{z}_0, t_0]\right) \mathbf{z}_\tau \mathbf{z}_0^T \, d\mathbf{z}_\tau \, d\mathbf{z}_0 \\
&= \int \pi\left(\mathbf{z}_0, t_0\right) \pi\left(\mathbf{z}_\tau, \tau | \, [\mathbf{z}_0, t_0]\right) \mathbf{z}_\tau \mathbf{z}_0^T \, d\mathbf{z}_\tau \, d\mathbf{z}_0 \\
&= \int \pi\left(\mathbf{z}_0, t_0\right) \left\langle \vec{\xi}(\tau) | \, [\mathbf{z}_0, t_0] \right\rangle \mathbf{z}_0^T \, d\mathbf{z}_0.
\end{aligned} \tag{A.32}$$

The first moment dynamics from the LNA are linear, and follow the dynamics from the Jacobian of the macroscopic variables, $\mathbf{A}$, i.e.:

$$\frac{\partial}{\partial \tau} \left\langle \vec{\xi}(\tau) \right\rangle = \mathbf{A} \left\langle \vec{\xi}(\tau) \right\rangle. \tag{A.33}$$

Similarly, we can write:

$$\frac{\partial}{\partial \tau} \left\langle \vec{\xi}(\tau) | \, [\mathbf{z}_0, t_0] \right\rangle = \mathbf{A} \left\langle \vec{\xi}(\tau) | \, [\mathbf{z}_0, t_0] \right\rangle, \tag{A.34}$$

whose solution is:

$$\left\langle \vec{\xi}(\tau) | \, [\mathbf{z}_0, t_0] \right\rangle = e^{-\mathbf{A}\tau} \left\langle \vec{\xi}(0) | \, [\mathbf{z}_0, t_0] \right\rangle. \tag{A.35}$$

This initial condition is trivial; $\left\langle \vec{\xi}(0)|\left[\mathbf{z}_0, t_0\right]\right\rangle = \mathbf{z}_0$, so (A.35) becomes

$$\left\langle \vec{\xi}(\tau)|\left[\mathbf{z}_0, t_0\right]\right\rangle = e^{-\mathbf{A}\tau}\mathbf{z}_0. \tag{A.36}$$

Substitution of (A.36) into (A.32) gives:

$$\left\langle \vec{\xi}(\tau)\vec{\xi}(0)^T\right\rangle = \int \pi\left(\mathbf{z}_0, t_0\right) e^{-\mathbf{A}\tau}\mathbf{z}_0\mathbf{z}_0^T \, d\mathbf{z}_0 \tag{A.37}$$

$$= e^{-\mathbf{A}\tau}\left\langle \vec{\xi}(0)\vec{\xi}(0)^T\right\rangle. \tag{A.38}$$

Therefore, we notice that according to the LNA the evolution of the time correlation functions in the stationary state follow the dynamics of the macroscopic variables, and take as initial conditions the covariance matrix at stationary state, $\left\langle \vec{\xi}(0)\vec{\xi}(0)^T\right\rangle$.

A.4 Proofs Relating to the Connection between Eigenvalues and Quality in the Two-Species Spectrum Equations

A.4.1 The Parameters p and q are Simple Functions of the Eigenvalues of the Jacobian

As discussed in section 1.5.3, the power spectrum matrix equation for number fluctuations resultant from the Linear Noise Approximation is given by

$$S_i(\omega) = \frac{\Omega}{\pi}\left[\left[-\mathbf{J} + \mathbf{I}i\omega\right]^{-1}\mathbf{D}\left[(-\mathbf{J})^T - \mathbf{I}i\omega\right]^{-1}\right]_{ii}, \tag{A.39}$$

where $\mathbf{J}$ and $\mathbf{D}$ are the Jacobian and diffusion matrices shown in Eq. (1.17) and Eq. (1.18).

Writing the matrices fully as

$$\mathbf{J} = \begin{bmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{bmatrix}, \mathbf{D} = \begin{bmatrix} D_{11} & D_c \\ D_c & D_{22} \end{bmatrix}, \tag{A.40}$$

where the form of $\mathbf{D}$ describes its positive semi-definite nature and the subscript c denotes noise coupling, inserting these elements into Eq. (A.39) and writing the equation for $S_1(\omega)$ gives:

$$S_1(\omega) = \frac{\Omega}{\pi} \frac{(D_{11}J_{22}^2 + D_{22}J_{12}^2 - 2D_c J_{12}J_{22}) + D_{11}\omega^2}{(J_{11}^2 J_{22}^2 + J_{12}^2 J_{21}^2 - 2J_{11}J_{22}J_{12}J_{21}) + (J_{11}^2 + J_{22}^2 + 2J_{12}J_{21})\omega^2 + \omega^4} \tag{A.41}$$

The equation for the other variable $S_2(\omega)$ is simply the above with subscripts 1 and 2 reversed. This equation is of the form shown in section 2.2, i.e.,

$$S_i(\omega) = \frac{\Omega}{\pi} \frac{\alpha_i(\mathbf{J}, \mathbf{D}) + \beta_i(\mathbf{D})\,\omega^2}{p(\mathbf{J}) + q(\mathbf{J})\,\omega^2 + \omega^4}, \tag{A.42}$$

where we now see that

$$p = J_{11}^2 J_{22}^2 + J_{12}^2 J_{21}^2 - 2J_{11}J_{22}J_{12}J_{21}, \tag{A.43}$$

$$q = J_{11}^2 + J_{22}^2 + 2J_{12}J_{21}. \tag{A.44}$$

The eigenvalues of **J** are:

$$\lambda_{1,2} = \frac{1}{2}[J_{11} + J_{22} \pm \sqrt{(J_{11} - J_{22})^2 + 4J_{12}J_{21}}]. \tag{A.45}$$

Rearranging Eqs. (A.45) for J_{11} and J_{12} gives

$$J_{11} = \lambda_1 + \lambda_2 - J_{22}, \tag{A.46}$$

$$J_{12} = \frac{1}{J_{21}}(J_{22} - \lambda_1)(J_{22} - \lambda_2). \tag{A.47}$$

Finally, by substituting Eqs. (A.46) into (A.43) it easily follows that

$$p = \lambda_1^2\lambda_2^2, \tag{A.48}$$

$$q = \lambda_1^2 + \lambda_2^2. \tag{A.49}$$

A.4.2 $S_i(\omega)$ Can Have at Most One Peak for $\omega > 0$

To find the peak frequency of $S_i(\omega)$ we simply solve $\frac{d}{d\omega}S_i(\omega)$ for ω. Using the form of the spectrum given in Eq. (A.42), the five solutions of $\frac{d}{d\omega}S_i(\omega)$ are

$$\omega = 0,\ \pm\frac{1}{\beta_i}\sqrt{-\alpha_i\beta_i \pm \beta_i\sqrt{\alpha_i^2 + \beta_i^2 p - \alpha_i\beta_i q}}. \tag{A.50}$$

Of the two positive solutions, the solution $\omega = \frac{1}{\beta_i}\sqrt{-\alpha_i\beta_i - \beta_i\sqrt{\alpha_i^2 + \beta_i^2 p - \alpha_i\beta_i q}}$ is complex valued because $p, \alpha_i, \beta_i \in \mathbb{R}_{>0}$ and $q \in \mathbb{R}$. Hence, there can be at most one (real valued) peak frequency in the range $\omega > 0$.

A.4.3 $Q^{50\%}$ can be written in terms of two non dimensional parameters, $R_{H\alpha\beta}$ andR_ϵ

The traditional quality factor $Q^{50\%} = \frac{\hat{\omega}}{\Delta\omega}$ is intuitive but analytically rather involved. Here we show that it can be written in terms of two non-dimensional parameters: the ratio of eigenvalues $R_\epsilon = \frac{\lambda_1}{\lambda_2}$ and the ratio of maximum heights of the α and β sub-spectra $R_{H\alpha\beta} = \frac{S_i^\alpha(0)}{S_i^\beta(\hat{\omega}_{S^\beta})}$.

Quality of the β sub-spectrum

We begin by analysing the β sub-spectrum to show that in the limit that $\alpha \to 0$, the quality $Q^{50\%}$ is written only as a function of R_ϵ. Starting with the general form of the spectrum equation

$$S_i(\omega) = \frac{\Omega}{\pi} \frac{\alpha_i(\mathbf{J}, \mathbf{D}) + \beta_i(\mathbf{D})\,\omega^2}{p(\mathbf{J}) + q(\mathbf{J})\,\omega^2 + \omega^4}, \tag{A.51}$$

and using the result that $p = \lambda_1^2\lambda_2^2$ and $q = \lambda_1^2 + \lambda_2^2$ we can write the 'β sub-spectrum' as

$$S_i^\beta(\omega) = \frac{\Omega}{\pi} \frac{\beta_i\,\omega^2}{\lambda_1^2\lambda_2^2 + (\lambda_1^2 + \lambda_2^2)\,\omega^2 + \omega^4}. \tag{A.52}$$

For real eigenvalues, the positive real solution of $\frac{d}{d\omega} S_i^\beta(\omega)$ is found easily to be $\hat{\omega}_{S^\beta} = \sqrt{\lambda_1\lambda_2}$. Inserting this value of $\omega = \sqrt{\lambda_1\lambda_2}$ into Eq. (A.52), we find the peak height of this sub spectrum to be

$$S_i^\beta(\hat{\omega}_{S^\beta}) = \frac{\Omega}{\pi} \frac{\beta}{(\lambda_1 + \lambda_2)^2}. \tag{A.53}$$

The bandwidth $\Delta\omega_{S^\beta}$ of the sub-spectrum is found by first solving $S_i^\beta(\omega) = \frac{1}{2} S_i^\beta(\hat{\omega}_{S^\beta})$

to find the two positive frequencies at which the sub spectrum takes its half maximum values, which we denote $\omega_{\frac{1}{2}}$. These frequencies are found to be:

$$\omega_{\frac{1}{2}} = \frac{1}{2}[\sqrt{(\lambda_1+\lambda_2)^2+4\lambda_1\lambda_2} \pm (\lambda_1+\lambda_2)]. \tag{A.54}$$

The bandwidth for this sub spectrum is given by the difference between these half-maximum frequencies, which is simply $\Delta\omega_{S^\beta} = \lambda_1 + \lambda_2$. From this, we find that the $Q^{50\%}$ value in the limit that $\alpha \to 0$ is given by

$$Q^{50\%} = \frac{\hat{\omega}_{S^\beta}}{\Delta\omega_{S^\beta}}, \tag{A.55}$$

$$= \frac{\sqrt{\lambda_1\lambda_2}}{\lambda_1+\lambda_2}. \tag{A.56}$$

For describing stable node NIO, we introduce our first non-dimensional variable (describing the ratio of real valued eigenvalues) via the replacement $\lambda_1 = R_\epsilon\lambda_2$. With this replacement we find that this qualityin the limit that $\alpha \to 0$ can be written only as a function of R_ϵ:

$$Q^{50\%} = \frac{\sqrt{R_\epsilon}}{R_\epsilon+1}. \tag{A.57}$$

Quality of the overall spectrum

Prompted by the usefulness of the first non-dimensional parameter R_ϵ, we write the full spectrum equation as

$$S_i^\beta(\omega) = \frac{\Omega}{\pi}\frac{\alpha_i+\beta_i\,\omega^2}{R_\epsilon^2\lambda_2^4+(R_\epsilon^2+1)\lambda_2^2\,\omega^2+\omega^4}. \tag{A.58}$$

Having now moved away from the limit $\alpha \to 0$, it is natural to introduce some new parameter which describes the relative weighting of the two sub-spectra, which we denote R_0. The candidate $R_0 = \frac{\alpha}{\beta}$ is unsuitable as it is not dimensionless. Progress is made by noting that any linear scaling of frequency has no effect on $Q^{50\%}$, and that we can therefore make the transformation $\omega = w\,\lambda_2$ to find

$$S_i^{\beta}(w) = \frac{\Omega}{\pi}\frac{\beta_i}{\lambda_2^2}\left(\frac{\frac{\alpha_i}{\beta_i\lambda_2^2} + w^2}{R_\epsilon^2 + (R_\epsilon^2+1)\,w^2 + w^4}\right). \tag{A.59}$$

The pre-factor $\frac{\beta_i}{\lambda_2^2}$ now merely serves to scale the above spectrum in power, and this has no effect on $Q^{50\%}$. Therefore it is clear that $Q^{50\%}$ can be described by a combination of R_ϵ and any non-dimensionless parameter of the form

$$R_0 = \frac{\alpha_i}{\beta_i\lambda_2^2}R_1. \tag{A.60}$$

While we could have simply used the non-dimensional parameter $R_0 = \frac{\alpha_i}{\beta_i\lambda_2^2}$ we elected instead to use a non-dimensional parameter with more intuitive meaning, namely the ratio of maximum heights of the two subspectra, i.e. $R_0 = R_{H\alpha\beta} = \frac{S_i^{\alpha}(0)}{S_i^{\beta}(\hat{\omega}_{S^\beta})}$. This is easily found to be

$$R_{H\alpha\beta} = \frac{\alpha_i}{\beta_i\lambda_2^2}\frac{(R_\epsilon+1)^2}{R_\epsilon^2}, \tag{A.61}$$

which is of the form required of our second non-dimensional parameter (Eq. (A.60)). Using the same steps as shown above for finding quality in the limit $\alpha \to 0$, the quality $Q^{50\%}$ of the overall spectrum is then found to be an equation of the form $Q^{50\%} = f(R_{H\alpha\beta}, R_\epsilon)$; this function is shown graphically in Fig. 2.2(a).

A.5 Power Spectrum Estimates from the Stochastic Simulation Algorithm

From a single realization r of the stochastic simulation algorithm, the number of molecules $n^r(t)$ of a particular species over some time interval T was regularly sampled at L discrete points separated by Δt (such that $(L-1)\Delta t = T$).

Using the notation $n_l^r \equiv n^r(l\Delta t), l \in [0, L-1]$, the fluctuation in numbers of molecules about the mean is $\hat{n}_l^r = n_l^r - \frac{1}{L}\sum_{j=0}^{L-1} n_j^r$. The discrete Fourier transform of $\hat{n}^r$, which we denote N^r, was calculated by:

$$N_k^r = \sum_{l=0}^{L-1} \hat{n}_l^r \, e^{-2\pi i l k/L}, \; k \in [0, L-1]. \tag{A.62}$$

From this, the one-sided periodogram estimate, P^r, of the number fluctuation power spectrum from realization r is defined as [175]:

$$P_k^r = \frac{1}{L^2} \begin{cases} |N_k^r|^2 & k = 0, \frac{L}{2} \\ |N_k^r|^2 + |N_{L-k}^r|^2 & k \in [1, \frac{L}{2} - 1]. \end{cases} \tag{A.63}$$

The $\frac{L}{2}+1$ frequencies ω_k over which P^r is defined span the frequency range $[0, \omega_{Nyq}]$ where ω_{Nyq} is the Nyquist rate, $\omega_{Nyq} = \frac{\pi}{\Delta t}$ rad s^{-1}. The frequency interval $\Delta\omega$ between ω_k and ω_{k+1} is $\Delta\omega = \frac{2\pi}{L\Delta t}$ rad s^{-1}.

Since $\hat{n}_l^r$ has zero mean, $P_0^r = 0$. The remaining $\frac{L}{2}$ periodogram values are normalised in Eq. (A.63) such that $\sum_{k=1}^{L/2} P_k^r = \sigma_{\hat{n}^r}^2$ where $\sigma_{\hat{n}^r}^2$ is the sample variance of $\hat{n}^r$, i.e. $\sigma_{\hat{n}^r}^2 = \frac{1}{L}\sum_{l=0}^{L-1}(\hat{n}_l^r)^2$. For comparison with the spectrum from the linear-noise approximation, Eq. (1.23), which is truly a power spectral density, we desire that the integral and not the sum of the numerical spectral density over these $\frac{L}{2}$ frequencies is

equal to $\sigma^2_{\hat{n}^r}$.

The value of P^r at frequency $\omega_k = \frac{2\pi k}{L\Delta t}$ represents some average of the power within the frequency bin $\omega_k \pm \frac{1}{2}\Delta\omega$ [175]. By assuming that the power in a frequency bin $\omega_k \pm \frac{1}{2}\Delta\omega$ is uniformly distributed within that bin, the integral-normalised form of the power spectral density over the frequency range $(\omega_1 - \frac{1}{2}\Delta\omega, \omega_{\frac{L}{2}} + \frac{1}{2}\Delta\omega)$ is given by:

$$S^r(\omega_k \pm \frac{1}{2}\Delta\omega) = \frac{1}{\Delta\omega}P^r_k,\ k \in [1, \frac{L}{2}]. \tag{A.64}$$

For display purposes in Figs. 2.5, 2.8, 2.9 and 2.13, we represent the power spectral density estimate over the frequency bin $\omega_k \pm \frac{1}{2}\Delta\omega$ by a single point of value $S^r(\omega_k \pm \frac{1}{2}\Delta\omega)$ at ω_k. We refer to this as S^r_k.

Finally, since the variance of the estimate S^r at a single frequency k from the periodogram method is known to be high, we obtain the final numerical power spectral density estimate by averaging S^r over R realizations, i.e.,

$$S_k = \frac{1}{R}\sum_{r=1}^{R} S^r_k. \tag{A.65}$$

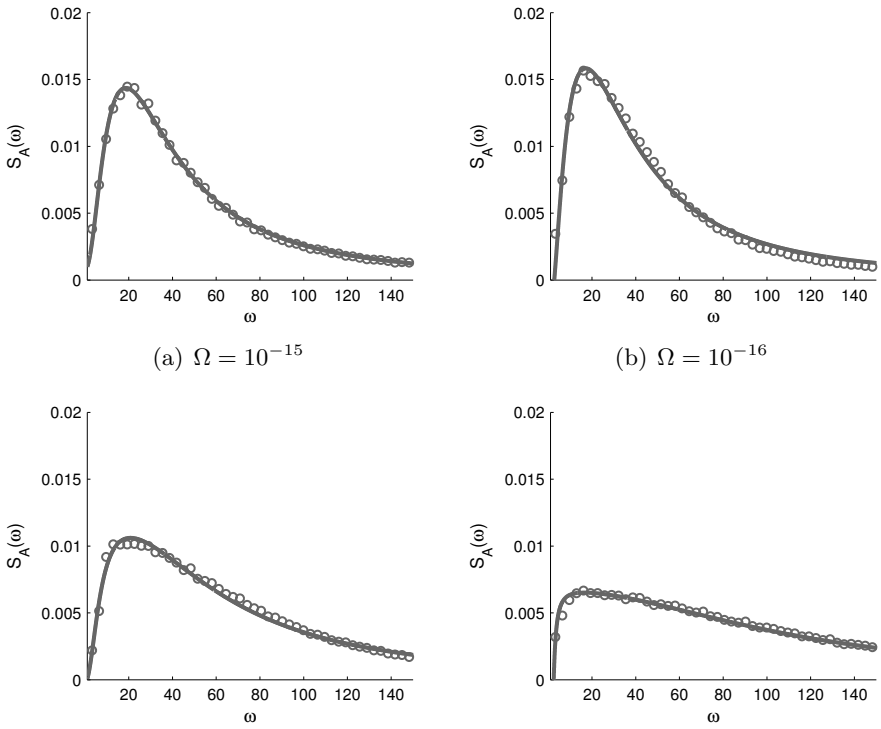
A.6 Suitability of Fitted Functions for Determining Quality of Numerical Fluctuation Power Spectra at Small Volumes

For each of the numerical power spectra S^{SSA} from the stochastic simulation algorithm shown in Fig. 2.6(a), the $Q^{50\%}$ was determined by fitting the numerical power spectra

using a function of the form:

$$S^{fit} = \frac{\alpha + \beta\omega^2}{p + q\omega^2 + \omega^4}.$$

To determine the parameters of S^{fit}, the MATLAB optimization toolbox function fminsearch was used to perform an unconstrained parameter search which minimized the Euclidean norm between the data and the fitted function, i.e., $C = \sqrt{\sum_{k=1}^{L/2} \left(S_k^{fit} - S_k^{SSA}\right)^2}$. The fitted functions are shown in Fig. A.1.



(a) $\Omega = 10^{-15}$ (b) $\Omega = 10^{-16}$

(c) $\Omega = 10^{-17}$ (d) $\Omega = 10^{-18}$

Figure A.1: The close fit between fitted spectrum functions using fminsearch and normalised fluctuation spectra obtained by the SSA for the Brusselator reaction at small volumes.

A.7 Experimental Setup and Luminescence Image Analysis

NB: The following section, describing experimental protocol developed by Wenden and Millar for plant growth and imaging, was written by Benedicte Wenden and is included here for reference.

The 35S:LUC, CCA1:LUC+, GI:LUC+, and CCR2:LUC+ lines in the Wassilewskija accession have been described [176, 157]. Seeds were surface sterilized, sown on Gilroy-agar media (no added sucrose,1.5% (wt/vol) agar), and stratified at 4°C for 96 h in darkness. Seedlings were then grown at 22 C under LL conditions or (12 h/12 h) LD cycles of 75 μmol $m^{-2}s^{-1}$ cool white fluorescent light for 11 days (intact plants) or 20 days (detached leaves) in Sanyo MLR350 environmental test chambers. 11-day-old intact leaves and 20-day-old detached leaves were kept healthy between a slide and a coverslip and kept immobile both horizontally and vertically (Figs. 4.2(c), 4.2(e) and Fig. 4.2(a)).

Seedlings were sprayed with 5 mM luciferin solution in 0.01% Triton-X100 24 h before the imaging setup and provided with a liquid Murashige and Skoog (MS) solution and luciferin during imaging. Luminescence images were captured as described in ref. [157]. For the intact plants, 11-day-old leaves were kept still both horizontally and vertically between a slide and a coverslip lying on two layers of porous fiber tape (Micropore), thus allowing an airflow around the leaves. Horizontal movements were prevented by taping the petiole to the slide with a thin piece of porous fiber tape. The rest of the seedling was supported by a second slide and nutrients and water were provided to the root from a liquid Murashige and Skoog (MS) medium (Figs. 4.2(c) and 4.2(e)). 21-day-old detached leaves were also placed between a slide and a coverslip

lying on two layers of porous fiber tape and provided with water and nutriments through a thin piece of precision wipe tissue in contact with both the petiole and the liquid MS medium (Fig. 4.2(a)). Intact plants were prepared in individual plates (Fig. 4.2(c)) or in six-well plates (Fig. 4.2(e)) and detached leaves were prepared in individual plates (Fig. 4.2(a)). This setup did not allow imaging of older rosette leaves as other leaves from the rosette would move on top of the imaged leaf. As acclimatization to the setup, plants and detached leaves were then left in the Sanyo chambers under their growth conditions for 24 h.

For imaging, seedlings and detached plants were then transferred to the microscope or open-field Percival I-30BLL growth chambers (CLF Plant Climatics) at 22 C under an equal mix of red and blue light-emitting diodes (LEDs). This setup was developed for six-well plates, allowing the imaging of 24 plants in open-field imaging cabinets and single plants for microscope imaging (Fig. 4.2(c) and Figs. 4.2(b) and 4.2(f)). Older detached leaves were also imaged (Fig. 4.2(a)). Luminescence outputs were obtained for both scales: at the tissue level in 24, 12-d-old seedlings in open-field imaging cabinets (Fig. 4.2(f)) and also at the cellular level under the microscope (Fig. 4.2(d)). Luminescence images were captured as described [157, 177], at 40 minute intervals with 15 minute exposure.

A.8 Creating MATLAB 'surf' Images

The spatial images of luminescence, phase and mean periods were all constructed using MATLAB's surf function, using the default flat shading. The operation of this function is described below.

We define **A** as a matrix of values, such as phase. The entries of this matrix are

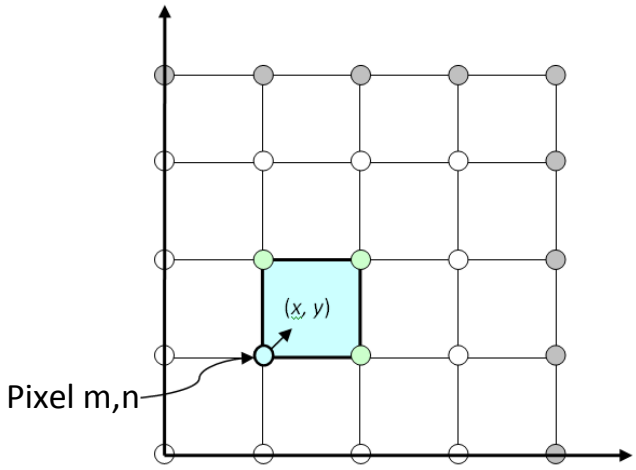


Figure A.2: Assignment of coloured image squares from pixel data values using MATLABs surf function with flat shading.

$\mathbf{A}_{m,n}$ where, as usual, m is the column index and n the row index, and to each entry is assigned a value from the camera pixel m, n. Each MATLAB image created is an array of squares, whose centre point we can index (x, y), with x the horizontal and y the vertical. The surf function with flat shading takes the value of the pixel m, n and maps it to the square x, y, assigning a colour according to the pixel's value as shown in each figure's colour scale.

An additional feature of this function is that for the square x, y to be assigned a colour, it is required that all of the pixels at the other three corners of that square (marked in light green in Fig. A.2) must themselves have associated data values. This requirement does mean that around the edge of the leaf, where this outcome may not be the case, some pixel values are not represented in the coloured image.

A.9 Plant Index Tables

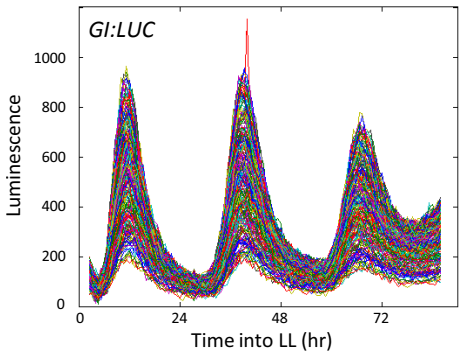
Plant Index	Marker	Detached/ Intact	Growth Conditions	Imaging Conditions	Imaging Date	Plant Number	Period Classification
1	*CCA1:LUC*	Detached	LD	LL	04/10/2010	3	
2	*CCA1:LUC*	Detached	LD	LL	04/10/2010	4	
3	*CCA1:LUC*	Detached	LD	LL	04/10/2010	7	
4	*CCA1:LUC*	Detached	LD	LL	12/10/2010	1	
5	*CCA1:LUC*	Detached	LD	LL	12/10/2010	2	
6	*CCA1:LUC*	Detached	LD	LL	16/10/2010	3	
7	*CCA1:LUC*	Detached	LD	LL	16/10/2010	4	
8	*CCA1:LUC*	Detached	LD	LL	23/11/2010	1	
9	*CCA1:LUC*	Detached	LD	LL	23/11/2010	4	
10	*CCA1:LUC*	Detached	LD	LL	23/11/2010	5	
11	*CCA1:LUC*	Detached	LL	LL	09/11/2010	1	None (c)
12	*CCA1:LUC*	Detached	LL	LL	09/11/2010	2	None (c)
13	*CCA1:LUC*	Detached	LL	LL	16/11/2010	1	None (c)
14	*CCA1:LUC*	Detached	LL	LL	16/11/2010	2	i
15	*CCA1:LUC*	Detached	LL	LL	16/11/2010	3	i
16	*CCA1:LUC*	Detached	LL	LL	16/11/2010	5	i
17	*CCA1:LUC*	Detached	LL	LL	16/11/2010	6	i
18	*CCA1:LUC*	Detached	LL	LL	16/11/2010	7	i
19	*CCA1:LUC*	Detached	LL	LL	16/11/2010	8	i
20	*CCA1:LUC*	Detached	LL	LL	18/08/2010	3	iii

Table A.1: Index table of detached leaves in the study. [Tab. derived from that produced by DT and BW and published in [142]. Leaves are described for growth and imaging conditions and mean period pattern classification: (i) Higher period at the leaf margins (predominantly on one side or all around the leaf), sometimes with a slight vascular pattern of increased period; (ii) higher period in the central region of the leaf; (iii) higher period toward the petiole. None (a): No period classification, leaf data deemed of insufficient quality to produce a reliable R value. None (b): No period classification, good data available over majority of leaf for only one circadian cycle. None (c): No period classification, SD/range values either too large or too spatially correlated with means. Plants shaded correspond to plants examined in detail in this thesis.]

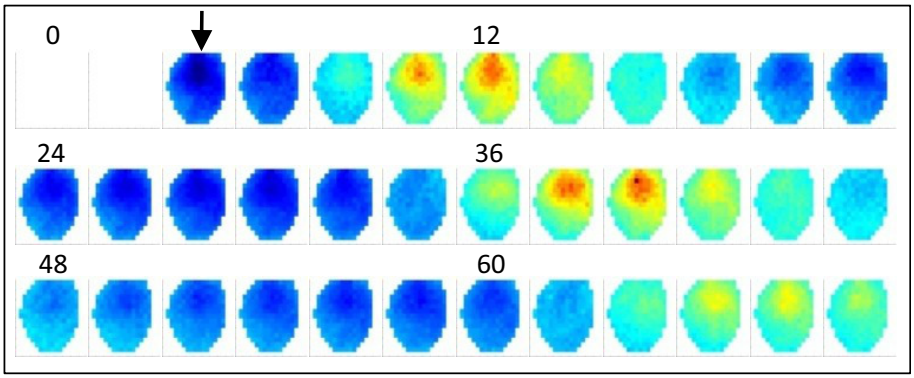
Plant Index	Marker	Detached/ Intact	Growth Conditions	Imaging Conditions	Imaging Date	Plant Number	Period Classification
21	*CCA1:LUC*	Intact	LD	LL	10/09/2010	2	
22	*CCA1:LUC*	Intact	LD	LL	10/09/2010	3	
23	*CCA1:LUC*	Intact	LD	LL	10/09/2010	4	
24	*CCA1:LUC*	Intact	LD	LL	25/02/2010	1	
25	*CCA1:LUC*	Intact	LD	LL	25/02/2010	2	
26	*CCA1:LUC*	Intact	LD	LL	25/02/2010	3	
27	*CCA1:LUC*	Intact	LD	LL	25/02/2010	4	
28	*CCA1:LUC*	Intact	LL	LD	09/06/2010	1	
29	*CCA1:LUC*	Intact	LL	LD	09/06/2010	11	
30	*CCA1:LUC*	Intact	LL	LD	09/06/2010	21	
31	*CCA1:LUC*	Intact	LL	LD	09/06/2010	22	
32	*CCA1:LUC*	Intact	LL	LD	09/06/2010	23	
33	*CCA1:LUC*	Intact	LL	LL	09/06/2010	1	ii
34	*CCA1:LUC*	Intact	LL	LL	09/06/2010	2	None (a)
35	*CCA1:LUC*	Intact	LL	LL	09/06/2010	4	None (c)
36	*CCA1:LUC*	Intact	LL	LL	09/06/2010	5	None (c)
37	*CCA1:LUC*	Intact	LL	LL	09/06/2010	6	ii
38	*CCA1:LUC*	Intact	LL	LL	13/09/2009	1	None (b)
39	*CCA1:LUC*	Intact	LL	LL	13/09/2009	2	None (b)
40	*CCA1:LUC*	Intact	LL	LL	13/09/2009	3	None (b)
41	*CCA1:LUC*	Intact	LL	LL	13/09/2009	4	None (a)
42	*CCA1:LUC*	Intact	LL	LL	13/09/2009	5	None (b)
43	*CCA1:LUC*	Intact	LL	LL	13/09/2009	6	None (b)
44	*CCA1:LUC*	Intact	LL	LL	13/09/2009	7	None (a)
45	*CCA1:LUC*	Intact	LL	LL	14/10/2009	3	i
46	*CCA1:LUC*	Intact	LL	LL	14/10/2009	4	i
47	*CCA1:LUC*	Intact	LL	LL	17/04/2010	1	None (c)
48	*CCA1:LUC*	Intact	LL	LL	17/04/2010	2	None (c)
49	*CCA1:LUC*	Intact	LL	LL	17/04/2010	3	None (c)
50	*CCA1:LUC*	Intact	LL	LL	17/04/2010	4	None (b)
51	*CCA1:LUC*	Intact	LL	LL	17/04/2010	5	None (b)
52	*CCA1:LUC*	Intact	LL	LL	17/04/2010	6	None (b)
53	*CCA1:LUC*	Intact	LL	LL	21/04/2010	1	None (b)
54	*CCA1:LUC*	Intact	LL	LL	21/04/2010	3	None (c)
55	*CCA1:LUC*	Intact	LL	LL	21/04/2010	5	None (c)
56	*CCA1:LUC*	Intact	LL	LL	21/04/2010	6	ii
57	*CCA1:LUC*	Intact	LL	LL	25/02/2010	2	None (b)
58	*CCA1:LUC*	Intact	LL	LL	25/02/2010	4	iii
59	*CCA1:LUC*	Intact	LL	LL	25/02/2010	5	ii
60	*GI:LUC*	Intact	LD	LL			
61	*CCR2:LUC*	Intact	LD	LL			

Table A.2: Index table of intact plants in the study. [Tab. derived from that produced by DT and BW and published in [142]. Leaves are described for growth and imaging conditions and mean period pattern classification: (i) Higher period at the leaf margins (predominantly on one side or all around the leaf), sometimes with a slight vascular pattern of increased period; (ii) higher period in the central region of the leaf; (iii) higher period toward the petiole. None (a): No period classification, leaf data deemed of insufficient quality to produce a reliable R value. None (b): No period classification, good data available over majority of leaf for only one circadian cycle. None (c): No period classification, SD/range values either too large or too spatially correlated with means. Plants shaded correspond to plants examined in detail in this thesis.]

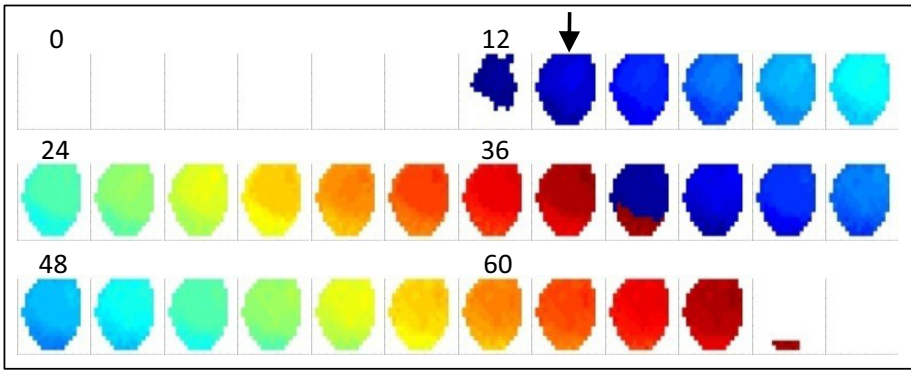
A.10 Additional Luminescence Data



(a)

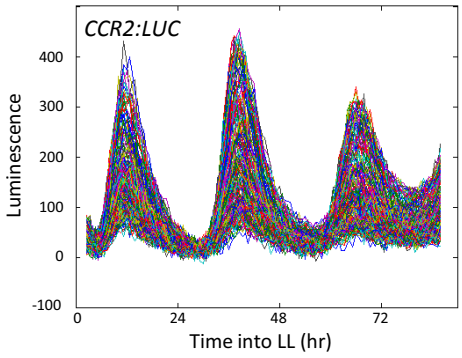


(b)

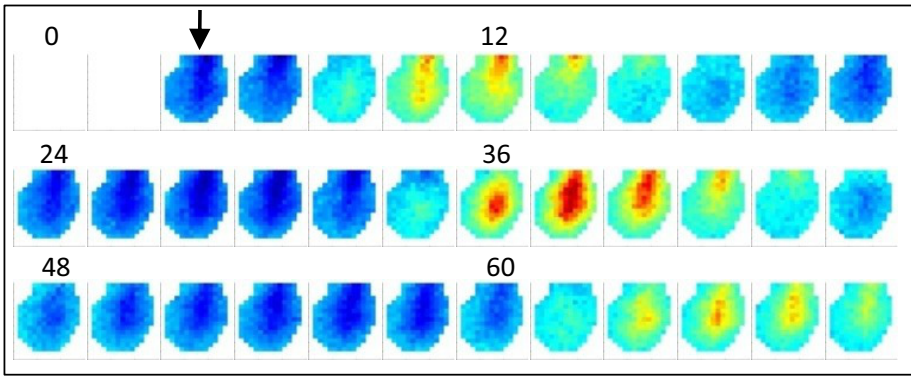


(c)

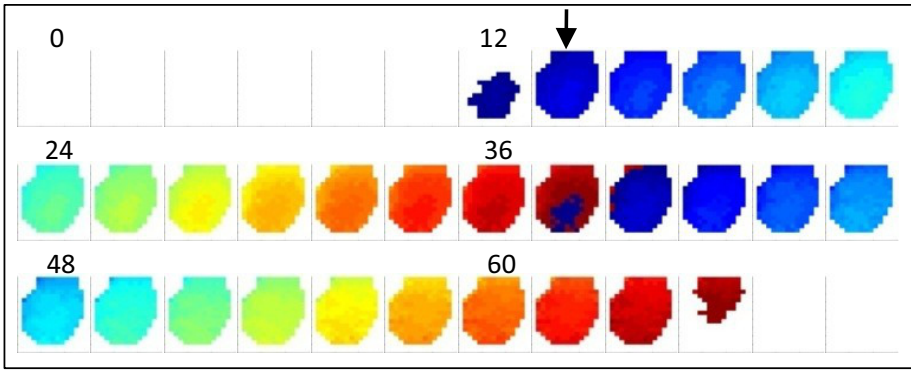
Figure A.3: Spatiotemporal analysis of GI:LUC in an intact leaf (Tab. A.2, plant 60) entrained under lightdark cycles and imaged under constant light. (a) Luminescence for all pixels, detrended; (b) montage for the detrended luminescence; (c) phase montage. The plant was entrained under LD 12:12 cycles for 12 d and then transferred and imaged under LL conditions. Zeitgeber time (ZT)0 corresponds to transfer to LL. Interval between two images = 2 h. Arrow indicates the position of the petiole. [Fig. produced by BW and DT and published in [142]]



(a)

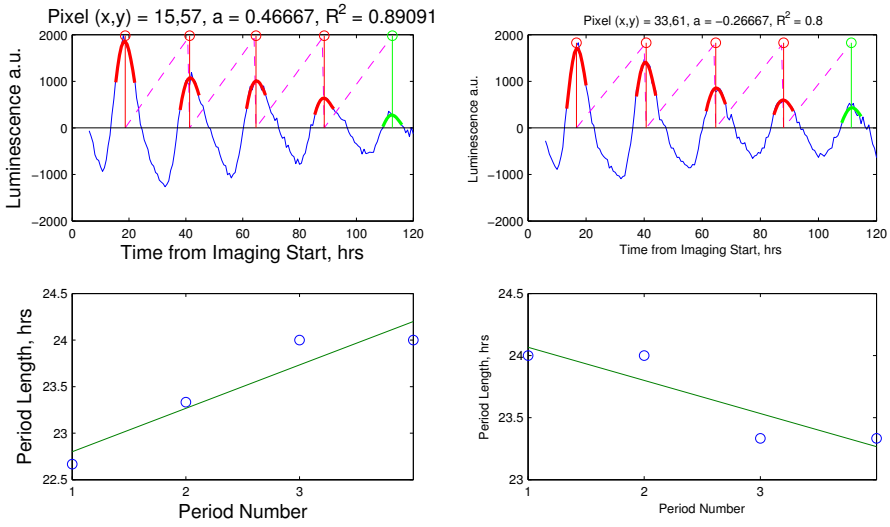


(b)



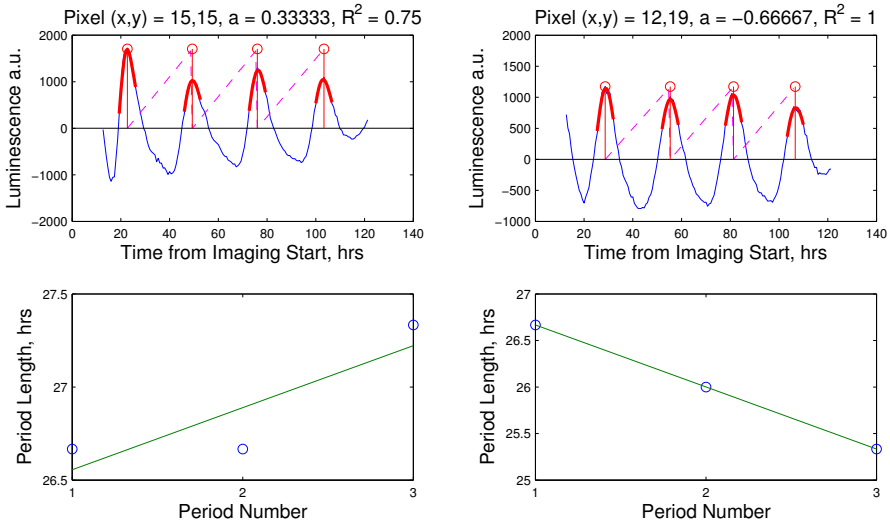
(c)

Figure A.4: Spatiotemporal analysis of CCR2:LUC in an intact leaf (Tab. A.2, plant 61) entrained under lightdark cycles and imaged under constant light. (a) Luminescence for all pixels, detrended; (b) montage for the detrended luminescence; (c) phase montage. The plant was entrained under LD 12:12 cycles for 12 d and then transferred and imaged under LL conditions. Zeitgeber time (ZT)0 corresponds to transfer to LL. Interval between two images = 2 h. Arrow indicates the position of the petiole. [Fig. produced by BW and DT and published in [142]]



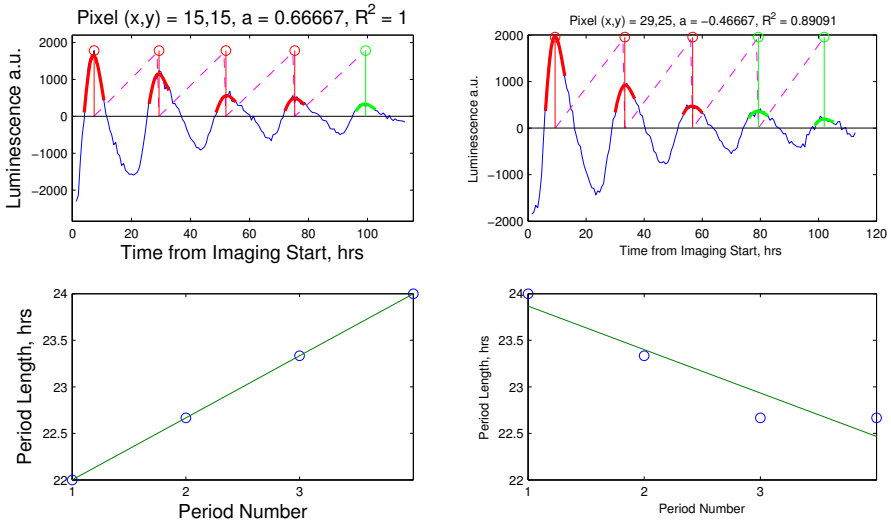
(a) (b)

Figure A.5: Examples of expression data from two pixels in the leaf from plant 20 (Tab. A.1); (a) a pixel which exhibits period lengthening and (b) a pixel which exhibits period shortening.



(a) (b)

Figure A.6: Examples of expression data from two pixels in the leaf from plant 37 (Tab. A.2); (a) a pixel which exhibits period lengthening and (b) a pixel which exhibits period shortening.



(a) (b)

Figure A.7: Examples of expression data from two pixels in the leaf from plant 16 (Tab. A.1); (a) a pixel which exhibits period lengthening and (b) a pixel which exhibits period shortening.

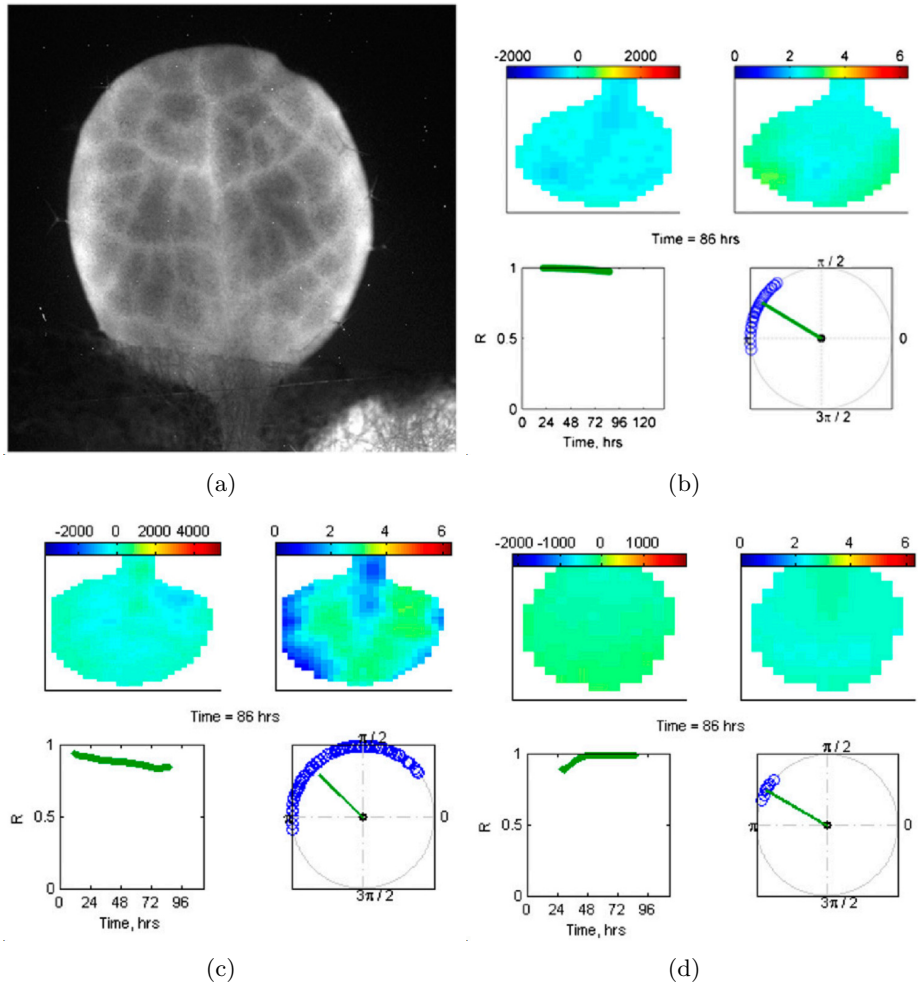


Figure A.8: Snapshots of circadian expression movies, available for reference at [178]. (a) Luminescence output of a leaf from a CCA1:LUC intact plant. The plant was grown for 12 days and then transferred and imaged under LL conditions, (b) Processed output for a detached leaf grown in LD and imaged in LL. (Upper Left) Movie of the luminescence, baseline-detrended. (Upper Right) Movie of the phase of each pixel, as defined by the timing of peak expression. (Lower Right) Movie of all leaf pixels phase values projected onto the unit circle (blue circles). The green vector has length R and mean phase angle ϕ (anticlockwise from the horizontal). (Lower Left) Phase coherence, R, of the leaf oscillators. Interval between two images = 40 min. (c) Processed output for an intact leaf grown in LD and imaged in LL. (d) Processed output for an intact leaf grown in LL and imaged in LD. [Fig. (a) only - produced by BW and published in [142]]

Appendix B

Bibliography

Bibliography

[1] Albert Goldbeter. *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, Cambridge, 1996.

[2] R E Dolmetsch, K Xu, and R S Lewis. Calcium oscillations increase the efficiency and specificity of gene expression. *Nature*, 392(6679):933–936, 1998.

[3] P Richard, B Teusink, M B Hemker, K Van Dam, and H V Westerhoff. Sustained oscillations in free-energy state and hexose phosphates in yeast. *Yeast*, 12(8):731–740, 1996.

[4] J C Dunlap. Molecular Bases for Circadian Clocks. *Cell*, 96:271–290, 1999.

[5] P Richard. The rhythm of yeast. *FEMS Microbiology Reviews*, 27(4):547–557, 2003.

[6] L N M Duysens and J Amesz. Fluorescence Spectrophotometry of Reduced Phosphopyridine Nucleotide in Intact Cells in the Near-Ultraviolet and Visible Region. *Biochimica et Biophysica Acta*, 24:19–26, 1957.

[7] K Pye and B Chance. Sustained sinusoidal oscillations of reduced pyridine nucleotide in a cell-free extract of Saccharomyces Carlsbergensis. *Proc. Nat. Acad.*

Sci. USA, 55(4):888–894, 1966.

[8] B Chance, K Pye, and J Higgins. Waveform generation by enzymatic oscillators. *IEEE Spectrum*, 4(8):79–86, 1967.

[9] E E Sel'kov. Self–Oscillations in Glycolysis. *European Journal of Biochemistry*, 4:79–86, 1968.

[10] A Goldbeter and R Lefever. Dissipative structures for an allosteric model. Application to glycolytic oscillations. *Biophysical Journal*, 12(10):1302–1315, 1972.

[11] A Boiteux, A Goldbetert, and B Hess. Control of oscillating glycolysis of yeast by stochastic, periodic, and steady source of substrate: A model and experimental study. *Proc. Nat. Acad. Sci. USA*, 72(10):3829–3833, 1975.

[12] D A McQuarrie. Stochastic Approach to Chemical Kinetics. *Journal of Applied Probability*, 4(3):413–478, 1967.

[13] L Wilhelmy. The law by which the action of acid on cane sugar occurs. *Annalen Physik Chemie*, 81:413–433, 1850.

[14] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. North Holland, 1992.

[15] C M Guldberg and P Waage. Studies concerning affinity. *CM Forhandlinger: Videnskabs-Selskabet i Christiana*, 35:1864, 1864.

[16] A Goldbeter. Patterns of spatiotemporal organization in an allosteric enzyme model. *Proc. Nat. Acad. Sci. USA*, 70(11):3255–3259, 1973.

[17] J J Higgins. The Theory of Oscillating Reactions. *Ind. Eng. Chem.*, 59:19, 1967.

[18] C P Fall, E S Marland, J M Wagner, and Tyson J J. *Computational Cell Biology*, volume 20 of *Interdisciplinary Applied Mathematics*. Springer, New York, 2002.

[19] B Novák and J J Tyson. Design principles of biochemical oscillators. *Nature Reviews Molecular Cell Biology*, 9(12):981–991, 2008.

[20] Stefan Schuster, Marko Marhl, and Thomas Höfer. Modelling of simple and complex calcium oscillations. From single-cell responses to intercellular signalling. *European Journal of Biochemistry*, 269(5):1333–1355, 2002.

[21] J Elf and M Ehrenberg. Fast evaluation of fluctuations in biochemical networks with the linear noise approximation. *Genome Research*, 13(11):2475–2484, 2003.

[22] S Danø, P G Sorensen, and F Hynne. Sustained oscillations in living cells. *Nature*, 402(6759):320–322, 1999.

[23] B C Goodwin. Oscillatory behavior in enzymatic control processes. *Advances in Enzyme Regulation*, 3:425–437, 1965.

[24] T. Wilhelm and R Heinrich. Smallest chemical reaction system with Hopfbifurcation. *Journal of Mathematical Chemistry*, 17:1–14, 1995.

[25] P Richard, B Teusink, H V Westerhoff, and K van Dam. Around the growth phase transition S. cerevisiae's make-up favours sustained oscillations of intracellular metabolites. *FEBS Letters*, 318(1):80–82, 1993.

[26] M A Aon, S Cortassa, H V Westerhoff, and K Van Dam. Synchrony and mutual stimulation of yeast cells during fast glycolytic oscillations. *Journal of General Microbiology*, 138:2219–2227, 1992.

[27] P Richard, B M Bakker, B Teusink, K Van Dam, and H V Westerhoff. Acetaldehyde mediates the synchronization of sustained glycolytic oscillations in populations of yeast cells. *European Journal of Biochemistry*, 235(1/2):238–241, 1996.

[28] S Danø, M F Madsen, and P G Sorensen. Quantitative characterization of cell synchronization in yeast. *Proc. Nat. Acad. Sci. USA*, 104(31):12732–12736, 2007.

[29] A Gustavsson, D D van Niekerk, C B Adiels, F B du Preez, M Goksör, and J L Snoep. Sustained glycolytic oscillations in individual isolated yeast cells. *FEBS Journal*, 279(16):2837–2847, 2012.

[30] A J McKane, J D Nagy, T J Newman, and M O Stefanini. Amplified biochemical oscillations in cellular systems. *Journal of Statistical Physics*, 128(1/2):165–191, 2007.

[31] A Einstein. On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat. *Annalen der Physik*, pages 549–560, 1905.

[32] M Von Smoluchowski. Zur kinetischen theorie der brownschen molekularbewegung und der suspensionen. *Annalen der physik*, 326(14):756–780, 1906.

[33] M Scott. *Applied Stochastic Processeses in Science and Engineering, http://www.math.uwaterloo.ca/~mscott/Little_Notes.pdf (accessed: 30/05/13)*. 2013.

[34] H A Kramers. Brownian motion in a field of force and the diffusion model. *Physica VII*, 4:284–304, 1940.

[35] M Delbruck. Statistical Fluctuations in Autocatalytic Reactions. *The Journal of Chemical Physics*, 8(1):120–124, 1940.

[36] E Schrodinger. *What is Life?: With Mind and Matter and Autobiographical Sketches*. Cambridge University Press, Cambridge, 1992.

[37] D A McQuarrie. Kinetics of small systems. I. *The Journal of chemical Chysics*, 38(2):433–436, 1963.

[38] K Singer. Application of the Theory of Stochastic Processes to the Study of Irreproducible Chemical Reactions and Nucleation Processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 15(1):92–106, 1953.

[39] A F Bartholomay. Stochastic models for chemical reactions: I. Theory of the unimolecular reaction process. *The Bulletin of Mathematical Biophysics*, 20(3):175–190, 1958.

[40] A F Bartholomay. Enzymatic reaction-rate theory: a stochastic approach. *Annals of the New York Academy of Sciences*, 96(4):897–912, 1962.

[41] Y Ishihama, T Schmidt, J Rappsilber, M Mann, F U Hartl, M J Kerner, and D Frishman. Protein abundance profiling of the Escherichia coli cytosol. *BMC Genomics*, 9:102, 2008.

[42] R Grima and S Schnell. Modelling reaction kinetics inside cells. *Essays in Biochemistry: Systems Biology*, 45:41–56, 2008.

[43] D Gonze, J Halloy, and A Goldbeter. Deterministic versus stochastic models for circadian rhythms. *Journal of Biological Physics*, 28(4):637–653, 2002.

[44] T G Kurtz. The Relationship between Stochastic and Deterministic Models for Chemical Reactions. *The Journal of Chemical Physics*, 57(7):2976, 1972.

[45] D T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22:403–434, 1976.

[46] R Grima. Investigating the robustness of the classical enzyme kinetic equations in small intracellular compartments. *BMC Systems Biology*, 3:101, 2009.

[47] A P Minton. The influence of macromolecular crowding and macromolecular confinement on biochemical reactions in physiological media. *The Journal of Biological Chemistry*, 276(14):10577–10580, 2001.

[48] A P Minton. How can biochemical reactions within cells differ from those in test tubes? *Journal of Cell Science*, 119(14):2863–2869, 2006.

[49] S B Zimmerman and A P Minton. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annual Review of Biophysics and Biomolecular Structure*, 22:27–65, 1993.

[50] A Goldbeter. A model for circadian oscillations in the Drosophila period protein (PER). *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 261(1362):319–324, 1995.

[51] J C W Locke, A J Millar, and M S Turner. Modelling genetic networks with noisy and varied experimental data: the circadian clock in Arabidopsis thaliana. *Journal of Theoretical Biology*, 234(3):383–393, 2005.

[52] J D Challenger, A J McKane, and J Pahle. Multi-compartment linear noise approximation. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(11):P11010, 2012.

[53] M B Priestley. *Spectral Analysis and Time Series (Probability and mathematical statistics)*. Academic Press Limited, London, 1994.

[54] A Papoulis and S Unnikrishna Pillai. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, New York, 2002.

[55] G M Jenkins and D G Watts. *Spectral Analysis and its Applications*. Holden-Day, San Francisco, 1968.

[56] C W Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences*. Springer, 2004.

[57] T S Rao, M B Priestley, O Lessi, and E J Hannan. *Applications of time series analysis in astronomy and meteorology*. Chapman & Hall, 1997.

[58] R Kleeman. Spectral Analysis of Multidimensional Stochastic Geophysical Models with an Application to Decadal ENSO Variability. *Journal of the Atmospheric Sciences*, 68(1):13–25, 2011.

[59] A C Harvey. Spectral Analysis in Economics. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 24(1):1–36, 1975.

[60] A Demir, A Mehrotra, and J Roychowdhury. Phase noise in oscillators: a unifying theory and numerical methods for characterization. *IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications*, 47(5):655–674, 2000.

[61] H H McAdams and A Arkin. Stochastic mechanisms in gene expression. *Proc. Nat. Acad. Sci. USA*, 94(3):814–819, 1997.

[62] D T Gillespie. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.*, 58:35–55, 2007.

[63] I J Laurenzi. An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics. *The Journal of Chemical Physics*, 113(8):3315–3322, 2000.

[64] D T Gillespie. A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and its Applications*, 188(1):404–425, 1992.

[65] D T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Chemical Physics*, 81(25):2340–2361, 1977.

[66] R Grima, D R Schmidt, and T J Newman. Steady-state fluctuations of a genetic feedback loop: an exact solution. *The Journal of Chemical Physics*, 137(3):035104, 2012.

[67] C Gadgil, C H Lee, and H G Othmer. A stochastic analysis of first-order reaction networks. *Bulletin of Mathematical Biology*, 67(5):901–946, 2005.

[68] J P Hespanha. Moment closure for biochemical networks. *Control and Signal Processing, 2008. ISCCSP 2008. 3rd International Symposium on. IEEE*, pages 142–147, 2008.

[69] M Ullah and O Wolkenhauer. Investigating the two-moment characterisation of subcellular biochemical networks. *Journal of Theoretical Biology*, 260, 2009.

[70] R Grima. Noise-induced breakdown of the Michaelis-Menten equation in steady-state conditions. *Physical Review Letters*, 102(21):218103, 2009.

[71] M Scott and B Ingalls. Using the Linear Noise Approximation to Characterize Molecular Noise in Reaction Pathways. *Proceedings of the AIChE Conference on Foundations of Systems Biology in Engineering (FOSBE), Santa Barbara, California.*, 2005.

[72] F Hayot and C Jayaprakash. The linear noise approximation for molecular fluctuations within cells. *Physical Biology*, 1:205–210, 2004.

[73] J M G Vilar, H Y Kueh, N Barkai, and S Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proc. Nat. Acad. Sci. USA*, 99(9):5988–5992, 2002.

[74] D B Forger and C S Peskin. Stochastic simulation of the mammalian circadian clock. *Proc. Nat. Acad. Sci. USA*, 102(2):321–324, 2005.

[75] H Li, Z Hou, and H Xin. Internal noise stochastic resonance for intracellular calcium oscillations in a cell system. *Physical Review E*, 71(6):061916, 2005.

[76] Z Hou and H Xin. Internal noise stochastic resonance in a circadian clock system. *The Journal of Chemical Physics*, 119(22):11508, 2003.

[77] T Dauxois, F Di Patti, D Fanelli, and A J McKane. Enhanced stochastic oscillations in autocatalytic reactions. *Physical Review E*, 79(3):036112.1–7, 2009.

[78] D Gonze, J Halloy, and A Goldbeter. Robustness of circadian rhythms with respect to molecular noise. *Proc. Nat. Acad. Sci. USA*, 99(2):673–678, 2002.

[79] P Gaspard. The correlation time of mesoscopic chemical clocks. *The Journal of Chemical Physics*, 117(19):8905–8916, 2002.

[80] H Gang, T Ditzinger, C Z Ning, and H Haken. Stochastic Resonance without External Periodic Force. *Physical Review Letters*, 71(6):807–810, 1993.

[81] P Jung and P Hanggi. Amplification of Small Signals via Stochasatic Resonance. *Phys. Rev. A*, 44(12):8032, 1991.

[82] S K Han, T G Yim, D E Postnov, and O V Sosnovtseva. Interacting Coherence Resonance Oscillators. *Physical Review Letters*, 83(9):1771–1774, 1999.

[83] A Neiman, P Saparin, and L Stone. Coherence resonance at noisy precursors of bifurcations in nonlinear dynamical systems. *Physical Review E*, 56(1):270–273, 1997.

[84] K L Davis and M R Roussel. Optimal observability of sustained stochastic competitive inhibition oscillations at organellar volumes. *FEBS Journal*, 273(1):84–95, 2006.

[85] R Kuske, L F Gordillo, and P Greenwood. Sustained oscillations via coherence resonance in SIR. *Journal of Theoretical Biology*, 245(3):459–469, 2007.

[86] H Qian and M Qian. Pumped biochemical reactions, nonequilibrium circulation, and stochastic resonance. *Physical Review Letters*, 84(10):2271–2274, 2000.

[87] A Pokhilko, A P Fernández, K D Edwards, M M Southern, K J Halliday, and A J Millar. The clock gene circuit in Arabidopsis includes a repressilator with additional feedback loops. *Molecular Systems Biology*, 8(574), 2012.

[88] J S O'Neill, G van Ooijen, L E Dixon, C Troein, F Corellou, F Bouget, A B Reddy, and A J Millar. Circadian rhythms persist without transcription in a eukaryote. *Nature*, 469(7331):554–558, 2011.

[89] E Nagoshi, C Saini, C Bauer, T Laroche, F Naef, and U Schibler. Circadian Gene Expression in Individual Fibroblasts : Oscillators Pass Time to Daughter Cells. *Cell*, 119:693–705, 2004.

[90] D K Welsh, S H Yoo, A C Liu, J S Takahashi, and S A Kay. Bioluminescence imaging of individual fibroblasts reveals persistent, independently phased circadian rhythms of clock gene expression. *Current Biology*, 14(24):2289–2295, 2004.

[91] A C Liu, D K Welsh, C H Ko, H G Tran, E E Zhang, A A Priest, E D Buhr, O Singer, K Meeker, I M Verma, F J Doyle 3rd, J S Takahashi, and S A Kay. Intercellular coupling confers robustness against mutations in the SCN circadian clock network. *Cell*, 129(3):605–616, 2007.

[92] O E Akman, M L Guerriero, L Loewe, and C Troein. Complementary approaches to understanding the plant circadian clock. *Arxiv preprint arXiv:1002.4661*, 2010.

[93] A J Millar, S R Short, N H Chua, and S A Kay. A novel circadian phenotype based on firefly luciferase expression in transgenic plants. *The Plant Cell*, 4(9):1075–1087, 1992.

[94] J D Plautz, M Kaneko, J C Hall, and S A Kay. Independent Photoreceptive Circadian Clocks Throughout Drosophila. *Science*, 278(5343):1632–1635, 1997.

[95] S C Thain, A Hall, and A J Millar. Functional independence of circadian clocks that regulate plant gene expression. *Current Biology*, 10(16):951–956, 2000.

[96] M W Merrow, N Y Garceau, and J C Dunlap. Dissection of a Circadian Oscillation into Discrete Domains. *Proc. Nat. Acad. Sci. USA*, 94(8):3877–3882, 1997.

[97] P O Westermark, D K Welsh, H Okamura, and H Herzel. Quantification of circadian rhythms in single cells. *PLoS Computational Biology*, 5(11):e1000580, 2009.

[98] T L Leise, C W Wang, P J Gitis, and D K Welsh. Persistent cell-autonomous circadian oscillations in fibroblasts revealed by six-week single-cell imaging of PER2::LUC bioluminescence. *PloS One*, 7(3):e33334, 2012.

[99] S H Strogatz. Exploring complex networks. *Nature*, 410(6825):268–276, 2001.

[100] D K Welsh, J S Takahashi, and Steve A Kay. Suprachiasmatic nucleus: cell autonomy and network properties. *Annual Review of Physiology*, 72:551–577, 2010.

[101] S J Aton and E D Herzog. Come together, right...now: synchronization of rhythms in a mammalian circadian clock. *Neuron*, 48(4):531–534, 2005.

[102] J A Evans, T L Leise, O Castanon-Cervantes, and A J Davidson. Intrinsic regulation of spatiotemporal organization within the suprachiasmatic nucleus. *PloS One*, 6(1):e15869, 2011.

[103] H Fukuda, I Tokuda, S Hashimoto, and N Hayasaka. Quantitative analysis of phase wave of gene expression in the mammalian central circadian clock network. *PloS One*, 6(8):e23568, 2011.

[104] S Yamaguchi, H Isejima, T Matsuo, R Okura, K Yagita, M Kobayashi, and H Okamura. Synchronization of cellular clocks in the suprachiasmatic nucleus. *Science*, 302(5649):1408–1412, 2003.

[105] M Scott, B Ingalls, and M Kaern. Estimations of intrinsic and extrinsic noise in models of nonlinear genetic networks. *Chaos*, 16(2):026107, 2006.

[106] I Prigogine and R Lefever. Symmetry Breaking Instabilities in Dissipative Systems. II. *The Journal of Chemical Physics*, 48(4):1695–1700, 1968.

[107] R Lefever, G Nicolis, and P Borckmans. The Brusselator: it does oscillate all the same. *Journal of the Chemical Society, Faraday Transactions 1*, 84(4):1013–1023, 1988.

[108] J J Tyson. Some further studies of nonlinear oscillations in chemical systems. *The Journal of Chemical Physics*, 58(9):3919–3930, 1973.

[109] J J Tyson. Properties of two-component bimolecular and trimolecular chemical reaction systems. *The Journal of Chemical Physics*, 59(8):4164–4173, 1973.

[110] R T Pack, R B Walker, and B K Kendrick. Three-body collision contributions to recombination and collision-induced dissociation. II. Kinetics. *The Journal of Chemical Physics*, 109(16):6714–6724, 1998.

[111] P Thomas, A V Straube, and R Grima. Communication: limitations of the stochastic quasi-steady-state approximation in open biochemical reaction networks. *The Journal of Chemical Physics*, 135(18):181103, 2011.

[112] P Thomas, A V Straube, and R Grima. The slow-scale linear noise approximation: an accurate, reduced stochastic description of biochemical networks under timescale separation conditions. *BMC Systems Biology*, 6:39, 2012.

[113] G B Cook, P Gray, D G Knapp, and S K Scott. Bimolecular Routes to Cubic Autocatalysis. *The Journal of Chemical Physics*, 93(7):2749–2755, 1989.

[114] N Kellershohn and M Laurent. Prion diseases: dynamics of the infection and properties of the bistable transition. *Biophysical Journal*, 81(5):2517–2529, 2001.

[115] J J Tyson. Modeling the cell division cycle : cdc2 and cyclin interactions. *Proc. Nat. Acad. Sci. USA*, 88:7328–7332, 1991.

[116] C Choudhary and M Mann. Decoding signalling networks by mass spectrometry-based proteomics. *Nature Reviews Molecular Cell Biology*, 11(6):427–439, 2010.

[117] M Akerfelt, R I Morimoto, and L Sistonen. Heat shock factors: integrators of cell stress, development and lifespan. *Nature Reviews Molecular Cell Biology*, 11(8):545–555, 2010.

[118] C Chatfield. *The Analysis of Time Series: An Introduction*. Chapman and Hall/CRC, 2004.

[119] T Shibata. Amplification of noise in a cascade chemical reaction. *Physical Review E*, 69(5):056218, 2004.

[120] R Ramaswamy and I F Sbalzarini. Intrinsic noise alters the frequency spectrum of mesoscopic oscillatory chemical reaction systems. *Scientific Reports*, 1:154, 2011.

[121] V Shahrezaei, J F Ollivier, and P S Swain. Colored extrinsic fluctuations and stochastic gene expression. *Molecular Systems Biology*, 4(1):196, 2008.

[122] H Qian. Nonlinear stochastic dynamics of mesoscopic homogeneous biochemical reaction systems – an analytical theory. *Nonlinearity*, 24(6):R19–R49, 2011.

[123] Michael L Simpson, Chris D Cox, and Gary S Sayler. Frequency domain analysis of noise in autoregulated gene circuits. *Proceedings of the National Academy of Sciences of the United States of America*, 100(8):4551–6, April 2003.

[124] P Hanusse. On the existence of a limit cycle in the evolution of open chemical systems. *Comptes Rendus, Acad. Sci. Paris, (C)*, 1972.

[125] C V Rao, D M Wolf, and A P Arkin. Control, exploitation and tolerance of intracellular noise. *Nature*, 420(6912):231–237, 2002.

[126] A Eldar and M B Elowitz. Functional roles for noise in genetic circuits. *Nature*, 467(7312):167–173, 2010.

[127] E M Ozbudak, M Thattai, I Kurtser, A D Grossman, and A van Oudenaarden. Regulation of noise in the expression of a single gene. *Nature Genetics*, 31(1):69–73, 2002.

[128] L Cai, N Friedman, and X S Xie. Stochastic protein expression in individual cells at the single molecule level. *Nature*, 440(7082):358–362, 2006.

[129] N Friedman, L Cai, and X Xie. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical Review Letters*, 97(16):168302, 2006.

[130] L Cai, C K Dalal, and M B Elowitz. Frequency-modulated nuclear localization bursts coordinate gene regulation. *Nature*, 455(7212):485–490, 2008.

[131] M Thattai and A van Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Nat. Acad. Sci. USA*, 98(15):8614–8619, 2001.

[132] S H Strogatz. *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering*. Westview Press, 2001.

[133] J J Tyson and S Kauffman. Control of mitosis by a continuous biochemical oscillation: synchronization; spatially inhomogeneous oscillations. *Journal of Mathematical Biology*, 1(4):289–310, 1975.

[134] Vahid Shahrezaei and Peter S Swain. Analytical distributions for stochastic gene expression. *Proc. Nat. Acad. Sci. USA*, 2(28):1–9, 2008.

[135] D L K Toner and R Grima. Molecular noise induces concentration oscillations in chemical systems with stable node steady states. *The Journal of Chemical Physics*, 138(5):055101, 2013.

[136] T Jia and R V. Kulkarni. Intrinsic Noise in Stochastic Models of Gene Expression with Molecular Memory and Bursting. *Physical Review Letters*, 106(5):058102, 2011.

[137] J M Pedraza and J Paulsson. Effects of molecular memory and bursting on fluctuations in gene expression. *Science*, 319(5861):339–343, 2008.

[138] R Ramaswamy, N González-Segredo, I F Sbalzarini, and R Grima. Discreteness-induced concentration inversion in mesoscopic chemical systems. *Nature Communications*, 3:779, 2012.

[139] R Grima, P Thomas, and A V Straube. How accurate are the nonlinear chemical Fokker-Planck and chemical Langevin equations? *The Journal of chemical physics*, 135(8):084103, 2011.

[140] M Scott. Non-linear corrections to the time-covariance function derived from a multi-state chemical master equation. *IET Systems Biology*, 6(4):116–124, 2012.

[141] M A Valencia-Sanchez, J Liu, G J Hannon, and R Parker. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development*, 20(5):515–524, 2006.

[142] B Wenden, D L K Toner, S K Hodge, R Grima, and A J Millar. Spontaneous spatiotemporal waves of gene expression from biological clocks in the leaf. *Proc. Nat. Acad. Sci. USA*, 109(17):6757–6762, 2012.

[143] E Yakir, D Hilman, Y Harir, and R M Green. Regulation of output from the plant circadian clock. *The FEBS Journal*, 274(2):335–345, 2007.

[144] H G McWatters and P F Devlin. Timing in plants – a rhythmic arrangement. *FEBS Letters*, 585(10):1474–1484, 2011.

[145] S L Harmer. The circadian system in higher plants. *Annual review of plant biology*, 60:357–377, 2009.

[146] P D Gould, J C W Locke, C Larue, M M Southern, S J Davis, S Hanano, R Moyle, R Milich, J Putterill, A J Millar, and A Hall. The molecular basis of temperature compensation in the Arabidopsis circadian clock. *The Plant Cell*, 18(5):1177–1187, 2006.

[147] Peter D Gould, Nicolas Ugarte, Mirela Domijan, Maria Costa, Julia Foreman, Dana Macgregor, Ken Rose, Jayne Griffiths, Andrew J Millar, Bärbel Finkenstädt, Steven Penfield, David a Rand, Karen J Halliday, and Anthony J W Hall. Network balance via CRY signalling controls the Arabidopsis circadian clock over ambient temperatures. *Molecular Systems Biology*, 9:650, 2013.

[148] D Alabadí, T Oyama, M J Yanovsky, F G Harmon, P Más, and S A Kay. Reciprocal regulation between TOC1 and LHY/CCA1 within the Arabidopsis circadian clock. *Science*, 293(5531):880–883, 2001.

[149] A Pokhilko, P Mas, and A J Millar. Modelling the widespread effects of TOC1 signalling on the plant circadian clock and its outputs. *BMC Systems Biology*, 7:23, 2013.

[150] J C W Locke, M M Southern, L Kozma-Bognár, V Hibberd, P E Brown, M S Turner, and A J Millar. Extension of a genetic network model by iterative experimentation and mathematical analysis. *Molecular Systems Biology*, 1(1), 2005.

[151] A J Millar, S R Short, K Hiratsuka, N H Chua, and S A Kay. Firefly luciferase as a reporter of regulated gene expression in higher plants. *Plant Molecular Biology Reporter*, 10(4):324–337, 1992.

[152] J M Kendall and M N Badminton. Aequorea victoria bioluminescence moves into an exciting new era. *Trends in Biotechnology*, 16(5):216–224, 1998.

[153] J Sai and C H Johnson. Different circadian oscillators control Ca(2+) fluxes and Lhcb gene expression. *Proc. Nat. Acad. Sci. USA*, 96(20):11659–11663, 1999.

[154] A Hall, L Kozma-Bognár, R M Bastow, F Nagy, and A J Millar. Distinct regulation of CAB and PHYB gene expression by similar circadian clocks. *The Plant journal : for cell and molecular biology*, 32(4):529–537, 2002.

[155] S C Thain, G Murtas, J R Lynn, R B Mcgrath, and A J Millar. The Circadian Clock That Controls Gene Expression in Arabidopsis Is Tissue Specific 1. *Plant Physiology*, 130(September):102–110, 2002.

[156] H Fukuda, N Nakamichi, M Hisatsune, H Murase, and T Mizuno. Synchronization of plant circadian oscillators with a phase delay effect of the vein network. *Physical Review Letters*, 99(9):98102, 2007.

[157] K D Edwards, O E Akman, K Knox, P J Lumsden, A W Thomson, P E Brown, A Pokhilko, L Kozma-Bognar, F Nagy, D A Rand, and A J Millar. Quantitative analysis of regulatory flexibility under changing environmental conditions. *Molecular Systems Biology*, 6(424):1–11, 2010.

[158] Y Moreno and A F Pacheco. Synchronization of Kuramoto oscillators in scale-free networks. *Europhysics Letters*, 68(4):603–609, 2004.

[159] J Acebrón, L Bonilla, C Pérez Vicente, F Ritort, and R Spigler. The Kuramoto model: A simple paradigm for synchronization phenomena. *Reviews of Modern Physics*, 77(1):137–185, 2005.

[160] K V Mardia and P E Jupp. *Directional Statistics*. John Wiley & Sons Inc., 2000.

[161] E Yakir, M Hassidim, N Melamed-Book, D Hilman, I Kron, and R M Green. Cell autonomous and cell-type specific circadian rhythms in Arabidopsis. *The Plant Journal*, 68(3):520–531, 2011.

[162] A Hall, L Kozma-Bognár, T Reka, F Nagy, and A J Millar. Conditional Circadian Regulation of PHYTOCHROME A Gene Expression. *Plant Physiology*, 127(December):1808–1818, 2001.

[163] U Rascher, M T Hütt, K Siebke, B Osmond, F Beck, and U Lüttge. Spatiotemporal variation of metabolism in a plant circadian rhythm: the biological clock as an assembly of coupled individual oscillators. *Proc. Nat. Acad. Sci. USA*, 98(20):11801–11805, 2001.

[164] M L Guerriero, A Pokhilko, A P Fernández, K J Halliday, A J Millar, and J Hillston. Stochastic properties of the plant circadian clock. *Journal of the Royal Society Interface*, 9(69):744–756, 2012.

[165] A Para, E M Farré, Takato Imaizumi, J L Pruneda-Paz, F G Harmon, and S A Kay. PRR3 Is a vascular regulator of TOC1 stability in the Arabidopsis circadian clock. *The Plant Cell*, 19(11):3462–3473, 2007.

[166] M F Covington and S L Harmer. The circadian clock regulates auxin signaling and responses in Arabidopsis. *PLoS Biology*, 5(8):e222, 2007.

[167] S Hanano, M A Domagalska, F Nagy, and S J Davis. Multiple phytohormones influence distinct parameters of the plant circadian clock. *Genes to Cells*, 11(12):1381–1392, 2006.

[168] A B James, J A Monreal, G A Nimmo, C L Kelly, P Herzyk, G I Jenkins, and G Nimmo. The circadian clock in Arabidopsis roots is a simplified slave version of the clock in shoots. *Science*, 322(5909):1832–1835, 2008.

[169] F Bischoff, A J Millar, S A Kay, and M Furuya. Phytochrome-induced intercellular signalling activates cab::luciferase gene expression. *The Plant Journal*, 12(4):839–849, 1997.

[170] N Marwan, M C Romano, M Thiel, and J Kurths. Recurrence plots for the analysis of complex systems. *Physics Reports*, 438:237–329, 2007.

[171] Joseph D Challenger and Alan J Mckane. Synchronisation of stochastic oscillators in biochemical systems. *arXiv preprint*, arXiv:1302:1–10, 2013.

[172] Paul C Bressloff. Stochastic neural field theory and the system-size expansion. *SIAM J. Appl. Math.*, 70(5):1488–1521, 2009.

[173] P Thomas, H Matuschek, and R Grima. Intrinsic noise analyzer: a software package for the exploration of stochastic biochemical kinetics using the system size expansion. *PloS One*, 7(6):e38518, 2012.

[174] R Grima. An effective rate equation approach to reaction kinetics in small volumes: theory and application to biochemical reactions in nonequilibrium steady-state conditions. *The Journal of Chemical Physics*, 133(3):035101, 2010.

[175] W H Press, S A Teukolsky, W T Vetterling, and B P Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, 1992.

[176] M R Doyle, S J Davis, R M Bastow, F Nagyk, H G Mcwatters, A J Millar, and R M Amasino. The ELF4 gene controls circadian rhythms and flowering time in Arabidopsis thaliana. *Nature*, 419(6902):74–77, 2002.

[177] K D Edwards, J R Lynn, P Gyula, F Nagy, and A J Millar. Natural allelic variation in the temperature-compensation mechanisms of the Arabidopsis thaliana circadian clock. *Genetics*, 170(1):387–400, 2005.

[178] http://www.pnas.org/content/109/17/6757/suppl/DCSupplemental, accessed: 30/05/13.