Mathematical Aspects of Coagulation and Fragmentation Processes

D. Allen

Department of Mathematics and Statistics

University of Strathclyde, Glasgow

June 4, 2020

Acknowledgements

I would like to thank my supervisors Drs Michael Grinfeld and Paul Mulheran for their support throughout my PhD. Their guidance and support has helped me all through the research and writing of this thesis and for this I am truly grateful.

Additionally, I would like to thank my family: Linda Allen, Charlotte Moseley and Artjoms Zmakins for supporting me emotionally throughout writing and preparing this thesis.

Abstract

In this thesis, we develop a number of approaches to investigate coagulation and fragmentation processes.

We initially use visibility graphs as a tool to analyse the results of kinetic Monte Carlo (kMC) simulations of submonolayer deposition in a one dimensional point island model. We introduce an efficient algorithm for the computation of the visibility graph resulting from a kMC simulation and show that from the properties of the visibility graph one can determine the critical island size, thus demonstrating that the visibility graph approach, which combines island size and spatial distribution data, can provide insights into island nucleation and growth mechanisms.

We then consider the dynamics of point islands during submonolayer deposition, in which the fragmentation of subcritical size islands is allowed. To understand asymptotics of solutions, we use methods of centre manifold theory, and for globalisation, we employ results from the theories of compartmental systems and of asymptotically autonomous dynamical systems. We also compare our results with those obtained by making the quasi-steady state assumption.

Finally, we demonstrate the versatility of the coagulation-fragmentation framework by considering the asymptotics of the average Erdős number. We also compare our results with those obtained by using a Gillespie type algorithm.

Contents

Contents

Contents

Contents

List of Figures

List of Figures

Chapter 1

Introduction

The topic of this thesis is coagulation and fragmentation processes. Coagulation is synonymous with aggregation, coalescence, and agglomeration in Roget's Thesaurus, and each of these words has a variety of definitions. However, when we refer to coagulation in this thesis, we mean the creation of larger clusters (of adatoms or monomers) through collisions of smaller clusters. We call a cluster of size one a monomer. As it is rare that multiple collisions happen simultaneously, we will assume that coagulation occurs only through one binary collision at a time. Figure 1.1 shows an example of a coagulation process between two clusters of sizes 5 and 15 to create a cluster of size 20.
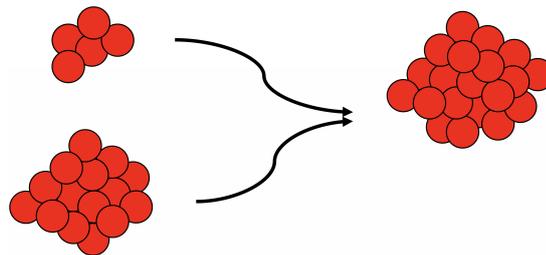
Figure 1.1: A coagulation process for a $5 - mer$ and a $15 - mer$ to create a $20 - mer$.

When referring to the process of fragmentation in this thesis, we mean the splitting of clusters into smaller clusters. Figure 1.2 shows an example of a fragmentation process where an unstable cluster of size 18 fragments into 3 clusters of sizes 3, 6 and 9.
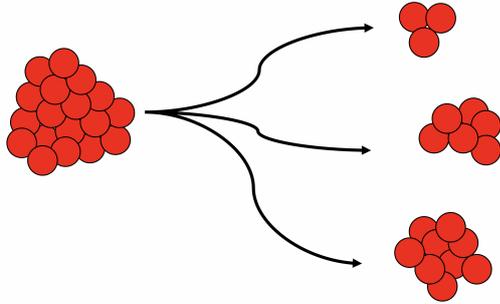
Figure 1.2: A multiple fragmentation process for an $18-mer$ into a $3-mer$, $6-mer$ and $9-mer$.

The processes of coagulation and fragmentation have been observed in a variety of different settings in the natural world. Examples of such settings include:

- **Aerosols**, i.e. suspensions of small solid or liquid particles in air, which have played an important role in the global dynamics of the Earth's atmosphere. The creation of the ozone hole, where particles act as catalysts of ozone depletion, shows the importance of understanding aerosols [8]. Additionally, in large cities where pollution levels are high, understanding the coagulation and fragmentation of soot particles is an intense area of research. Flame aerosol reactors, i.e. gas-gas combustion synthesis, are considered to be a potentially cheap solution to the pollution problem and as a result research has intensified in recent years [80].

- **The formation of stars and galaxies**, in which dust particles coagulate, is a common occurrence in many settings in astrophysics. Coagulation occurs in the formation of protoplanetary disks, which are rotating disks

of dense gas and particles surrounding a recently formed star in the early stages of planet formation [104]. Models for the formation of protoplanetary disks have been considered in [15] which describe the evolution of a system of masses developing by completely inelastic collisions and spontaneous fragmentations.

- **Biological systems** where groups form, such as in fish schooling and in the time-evolution of phytoplankton, i.e. microscopic marine algae. Such systems have been analysed in [75] where school size statistics of fish have been considered in the context of fish schools breaking up and uniting with other schools and in [9] where a model of phytoplankton dynamics is introduced.

- **Chemical systems**, such as in polymers. Polymer chains can break due to high shear, chemical attack or radiation-induced chain scission; an understanding of these processes is of importance industrially [13, 102].

- **Formation of clouds and precipitation**. An understanding of the atmosphere is important as clouds help to regulate the energy that reaches the Earth by scattering solar radiation and absorbing Earth's infrared energy. Additionally, clouds are required for precipitation to occur and to distribute water. Such systems have been discussed in [81, 86] where a model is introduced for the coagulation of cloud droplets.

Coagulation and fragmentation processes have also been used in the development of technologies such as **Molecular Beam Epitaxy** (MBE), a process for growing thin films. Applications of MBE include optical coatings, corrosion protection, fabrication of semiconductor devices and the self-assembly of nanostructures [10, 54, 67].

The early stage of MBE, i.e. atomic surface coverage of less than one monolayer,

where monomers or clusters are deposited onto a surface, diffuse and form large-scale structures, is called **submonolayer deposition** (SD).

1.1   What is Submonolayer Deposition?

We begin by outlining some fundamental concepts in the area of SD.

Although it is possible to view images of processes like MBE using devices such as a scanning tunnelling microscope (an instrument for imaging surfaces at the atomic level), these processes are often slow and the process can be examined at fixed times rather than in real time [93]. Thus, we are interested in mathematical modelling of MBE in order to consider the time-evolution of the process and, ultimately, to observe how the statistics of our simulations compare with the experimental observations and to allow us to observe the process as time evolves. A mathematical model has the potential of reducing the cost of experiments, which can be significant.

To describe SD, we introduce the following concepts. Any cluster of two or more monomers will be known as an **island**. An island is a **stable island** if no monomers or clusters can dissociate from the island. If the number of monomers contained in the smallest stable island is $i + 1$ then the **critical island size** is defined to be i. **Coverage**, θ, is the percentage of the substrate that is occupied by monomers or islands.

Using our concepts, we can now outline a simple description of SD for a lattice-gas model [41]: monomers are deposited randomly with flux α (in units of monomers per site per unit time) onto an initially empty two dimensional substrate, e.g. a lattice of sites L^2. Monomers undergo isotropic diffusion, hopping with Arrhenius rate; $D_r = ve^{-E_d/(k_b T)}$, where

- v is the vibration frequency for hopping,

- $k_b \approx 1.3807 \times 10^{-23}$ joules per Kelvin is Boltzmann's constant,

- T is the substrate temperature,

- E_d is the activation barrier for diffusion,

per direction to adjacent empty sites. Islands nucleate upon the formation of an island of at least $i + 1$ monomers.

Realistic models of SD may include additional processes such as desorption (i.e. evaporation of monomers from the substrate at high temperatures, island dissociation by release of monomers), island mobility, non-constant critical island sizes where i is dependent on temperature (in general i increases as temperature increases), direct impingement (i.e. the dynamics of monomers deposited on top of clusters), and the action of electric fields.

In our work we assume constant temperature, immobile islands, and irreversible aggregation. In addition, in our simulations we assume that monomers hop between the nearest neighbour sites at a constant hopping rate D, and the lattice is equally spaced. Surprisingly, under these assumptions the dynamics of the system, in our simulations are only dependent on:

- $i \in \mathbb{N}$,

- $R := \frac{D}{\alpha}$, where $\alpha, D > 0$,

- $\theta = \alpha t$, where t is time.

In terms of coverage, we can divide the process, in our simulations, into the following four regimes [7, 76, 94]:

1. **Low regime** - the average island size is typically small, and the density of monomers far outweighs the density of islands. The density of monomers in the system increases linearly, the likelihood of nucleation is small.

2. **Intermediate regime** - the density of islands approximately reaches the density of monomers. The density of monomers decreases proportionately to the increase in the island density due to the large number of nucleations of new islands.

3. **Aggregation regime** - the island density increases slowly, and the monomer density decreases rapidly. The monomers are now much more likely to be captured by existing islands than help nucleate new islands.

4. **Coalescence regime** - the islands become large and begin to coalesce with other islands. The joining of islands ultimately causes the creation of a single island that spans the entire lattice. Eventually an additional layer of growth will occur.

1.2   Kinetic Monte Carlo Simulations

Stanisław Ulam is often credited with the invention of Monte Carlo (MC) methods, due to remarks he made in 1946 to von Neumann about the card game solitaire [33]. The story is as follows: Ulam was playing Canfield solitaire. He asked what the chances were that a shuffled deck laid out with 52 cards will come out successfully. First, Ulam tried to solve the problem using combinatorics; this led to little success due to the large number of computations required. As an alternative, Ulam asked if a more sensible approach would be to lay out a hundred games and count the number of successful outcomes. The story describes the essence of MC methods. Broadly speaking, MC methods are algorithms that

perform a large number of repeated random sampling of a process to obtain its statistics.

At this point it is worth mentioning what we mean by "random". The question of whether or not true randomness exists is still very much open [26]. Many people are of the opinion that true randomness cannot exist, not even at the quantum scale [48]. However, we may consider a number to be **statistically random** if the number does not contain any recognisable patterns [50], e.g. at the time of writing the digits of π exhibit statistical randomness. So, for clarity, throughout this thesis when referring to a random number we will be referring to a number generated by a pseudo-random number generator, an algorithm that produces a string of numbers that are not truly random but are statistically random (see e.g. the algorithm in [99]). Pseudo-random numbers are used in computer simulations, in finance, statistics and many other areas of mathematical modelling, and are generally widely accepted.

There are many competing definitions of MC simulations, see e.g. [18, 84]. That being said, generally one identifies random variables that describe a real world system, then runs the resulting model numerous times on a computer. Once the run is completed, the data generated from the simulations can be analysed [84].

In processes that include coagulation, fragmentation and deposition, kinetic Monte Carlo (kMC) simulations are driven by diffusion and deposition steps [20, 52]:

- in the deposition step, clusters are randomly deposited onto an n dimensional, $1 \le n \le 3$ lattice L^n at a deposition rate of $\alpha(t)$ monolayers per unit time.

- In the diffusion step, a cluster is selected at random and moves a random length in a random direction on the lattice.

Chapter 1. Introduction

Additionally, a fundamental concept that has emerged from experimental data is the formation of stable islands above a certain critical size (the critical island size i) [6]. In many kMC simulations the property of a critical island size is added to the model. Depending on the system we are trying to model, we may also include processes such as evaporation, direct impingement and the action of electric fields that may bias diffusion. Finally, we can consider the shape of islands which can be **extended** (dendritic, spherical) or **point** islands:

- In **extended island** models[1], islands grow by capturing monomers that diffuse to their edges.

- In **point island** models, each island occupies only a single site.

kMC simulations have been invaluable in generating numerical data for coagulation, fragmentation and deposition processes, which is expensive to obtain from experiments, as they have provided numerical data for both **island size distributions** (ISDs) and **capture zone distributions** (CZDs):

- the **island size** is the number of monomers at a particular site on the lattice (in the point island model).

- The **capture zone** associated with an island is the lattice region surrounding the island that consists of all points closer to the island than to any other island [71].

Obviously, the ultimate comparison is with experiments. However, although numerical data from kMC simulations of coagulation-fragmentation systems has

---

[1]Note, in models with impingement, islands grow both vertically and laterally so an island of size j will not occupy j sites on the lattice.

been compared to experimental data [20], these comparisons have been conducted by visual inspection and a more rigorous approach such as the two sample Kolmogorov-Smirnov test is needed to quantify how these kMC simulations compare with experimental data. Therefore, it is generally accepted that kMC simulations of coagulation-fragmentation systems are only a reasonable first approximation to the experimental situation [69].

Although it is true that kMC simulations of coagulation-fragmentation systems have provided useful numerical data, it is important to note kMC simulations are restricted by the memory and processors of computers, e.g. the time complexity of our simulations grows exponentially as i increases, therefore, at the time of writing, simulating $i \geq 4$ on a quad core desktop computer is not possible.

1.3   Modelling Methodology

Early attempts to create mathematical models of coagulation and fragmentation began in the early 1960s (see e.g. [94]) with the aim of developing a theory that would explain the ISDs, i.e. the scaling of the monomer and island densities. A theory that describes the ISD accurately is considered an important challenge in the overall objective of developing a general theory of SD.

Although rate equations have been shown to produce the average behaviour of islands, i.e. the island densities, they have failed to produce ISDs accurately as they depend only on time and cluster sizes and neglect the clusters' local environments. The rate equations employ a **mean field** assumption that the surroundings of islands are independent of their shapes and sizes [16]. The failure of rate equations to produce ISDs accurately has prompted the use of alternative frameworks that could potentially predict the experimental results. We will

discuss such frameworks in Section 1.3.2 – 1.3.2.1.

1.3.1 Rate Equations

A rate equation for a particular chemical species X is a differential equation, or an integro-differential equation, which represents the rate of change in the concentration of X to the rate of the reactions it participates in. Using the law of mass kinetics (i.e. the principle that the rate of a reaction is proportional to the concentrations of the reacting substances) on reaction systems produces systems of ordinary differential equations (ODEs) for the various species, see, for example, [36, 57].

As a large proportion of this thesis contributes to rate equations, in this section we will introduce a general framework to describe the time-evolution of cluster sizes [22, 27, 97, 101] via rate equations.

1.3.1.1 Smoluchowski's Coagulation Equations

We begin by considering the pure coagulation reaction originally introduced as a model for colloid formation [90, 91]. Considering the reaction process from Figure 1.1:

$$x - mer + y - mer \overset{a(x,y)}{\to} (x + y) - mer, \qquad x, y \in \mathbb{R}^+,$$
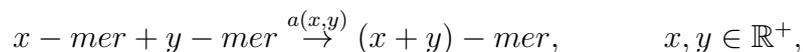
where $a(x, y) = a(y, x)$ is the coagulation coefficient, i.e. the rate of the coagulation reaction among an $x - mer$ and a $y - mer$. When the system only contains clusters of discrete size, the clusters are denoted by $j, k, \ldots$ instead of $x, y, \ldots$. As we will only consider the discrete equations, we adopt this notation throughout this thesis (see e.g. [27, 61, 95, 102] for the continuous equations).

Chapter 1. Introduction

$$j - mer + k - mer \overset{a(j,k)}{\rightsquigarrow} (j+k) - mer, \qquad j,k \geq 1.$$

If we denote by $C_j(t) := C_j$ the concentration of $j - mers$ at time t, and use primes for differentiation with respect to t, the laws of mass kinetics give us the following infinite system of ODEs:

$$C_j' = \frac{1}{2} \sum_{k=1}^{j-1} a(k, j-k) C_k C_{j-k} - C_j \sum_{k=1}^{\infty} a(j,k) C_k, \quad j \geq 1. \qquad (1.1)$$

The first term on the right-hand side of (1.1) represents the creation of clusters of size j by coagulation of clusters of sizes k and $j - k$. The factor of $1/2$ is included to avoid double-counting since the first sum includes every possible way of constructing the clusters of size j twice. The second term represents the depletion of clusters of size j due to their coagulation with other clusters.

If we further assume clusters grow through collisions between monomers and clusters and let $J_0(t)$ represent a source term, i.e. an external supply of monomers into the system [30, 96], then the simplest case where $J_0(t) = \alpha$ leads to models that have been widely studied in the literature and in many real world situations [11, 56].

Under the above assumptions, a realistic system of rate equations where the critical island size $i = 1$ can be obtained, by setting $a(1,1) = 2Da_1$ and, $a(j,1) = Da_j$, $j > 1$ [21, 82]:

$$C_1' = \alpha - 2Da_1 C_1^2 - DC_1 \sum_{k=2}^{\infty} a_k C_k,$$

$$C_j' = DC_1(a_{j-1} C_{j-1} - a_j C_j), \quad j > 1, \qquad (1.2)$$

where a_j represents the capture rate of monomers from clusters of size j, and D is a diffusion rate.

11

Chapter 1. Introduction

We will return to system (1.2) in Section 3.1 when we consider the work of Blackman and Wilding.

The simplest case of system (1.2) occurs when we let all clusters be **point islands** i.e. clusters have no spatial extent. Additionally, assuming all the coagulation constants can be taken to be equal, we can set $Da_j = 1$ i.e. by scaling the time variable. The assumption of point islands and $Da_j = 1$ is a good approximation to Brownian coagulation and linear chain polymerisation [16, 36]. Hence, we have the following infinite system of ODEs:

$$
\begin{aligned}
C_1' &= \alpha - 2C_1^2 - C_1 \sum_{k=2}^{\infty} C_k, \\
C_j' &= C_1 C_{j-1} - C_1 C_j, \quad j > 1.
\end{aligned}
\tag{1.3}
$$

We will return to system (1.3) in Section 3.1.1 when we consider the work of da Costa, van Roessel and Wattis.

1.3.1.2 Fragmentation Equations

We now deal with a fragmentation system. This time, we consider the reaction process from Figure 1.2:

$$
j - mer \overset{b_j}{\to} k_1 - mer + k_2 - mer + \ldots,
$$

where b_j is the rate of fragmentation of $j - mer$.

An assumption that is often made for fragmentation, that we will adopt throughout this thesis, and is valid in many cases (see e.g. [102]), is that fragmentation is binary, i.e. each fragmentation produces only two clusters. Additionally, we assume that one of the clusters is always a monomer [28]:
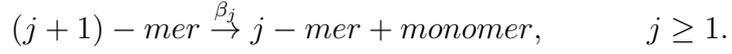
$$(j+1) - mer \overset{\beta_j}{\rightarrow} j - mer + monomer, \qquad j \geq 1.$$

Using the same notation as in (1.1), the laws of mass kinetics give us the following infinite system of ODEs:

$$C_1' = 2\beta_2 C_2 + \sum_{k=3}^{i} \beta_k C_k,$$

$$C_j' = -\beta_j C_j + \beta_{j+1} C_{j+1}, \quad j > 1,$$

where (for the C_j equation) the first term represents the rate at which clusters of size j fragment into clusters of size $j - 1$ and the second term represents the rate at which clusters of size j grow by fragmentation of clusters of size $j + 1$.

1.3.1.3 Fragmentation of Subcritical Islands

There are a number of ways we can model clusters of size $1 < j \leq i$. In Section 3.1.2, we consider the work of Costin, Grinfeld, O'Neill and Park where it is assumed clusters of size $1 < j \leq i$ simply do not arise [31]. Alternatively, the other physically relevant possibility is clusters of size $1 < j \leq i$ fragment (at some rate independent of the cluster size, which is consistent with the point-island assumption). To model the process of clusters fragmenting below a given critical size, we combine our reactions for coagulation and fragmentation, however, only allowing fragmentation of islands below a certain critical size. Under these assumptions we can assume that the following reactions occur:

$$j - mer + monomer \underset{\beta}{\overset{1}{\rightleftarrows}} (j+1) - mer, \qquad 1 \leq j < i,$$

$$j - mer + monomer \overset{1}{\rightarrow} (j+1) - mer, \qquad j \geq i.$$
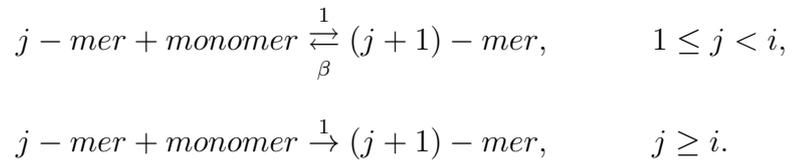
In other words, we allow clusters of size less than $i + 1$ to fragment at a rate

13

$\beta > 0.$

If we set $J_0(t) = \alpha$ to be an external supply of monomers into the system, using the same notation as in (1.1), the laws of mass kinetics give us the following infinite system of ODEs:

$$C_1' = \alpha - 2C_1^2 + 2\beta C_2 - C_1 \sum_{k=2}^{\infty} C_k + \beta \sum_{k=3}^{i} C_k,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j - \beta C_j + \beta C_{j+1}, \quad 1 < j < i,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j - \beta C_j, \quad j = i,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j, \quad j > i.$$

We will discuss fragmentation of subcritical islands in more detail in Sections 3.1, 3.1.2 and Chapter 5.

1.3.1.4 Existence and Uniqueness of Solutions

In this section we will discuss the existence and uniqueness of solutions to discrete coagulation-fragmentation systems. However, in the majority of cases, equivalent rigorous relations between the existence and uniqueness of solutions in both the discrete and continuous equations can be established, see e.g. [28].

If we consider the system of ODEs from (1.1), then it is reasonable to impose the restriction that solutions of (1.1) must have a finite mass, which implies that, for $t \geq 0$, a solution of (1.1) must be an element of the Banach space $X_1 \subset \ell^1$, where

$$X_1 := \left\{ C = (C_j) \in \mathbb{R}^{\mathbb{N}} \text{ such that } \|C\|_1 := \sum_{n=1}^{\infty} n|C_n| < \infty \right\}.$$

For different coagulation-fragmentation equations it may be necessary to consider

Chapter 1. Introduction

other Banach spaces, namely

$$X_a := \left\{ C = (C_j) \in \mathbb{R}^{\mathbb{N}} \text{ such that } \|C\|_a := \sum_{n=1}^{\infty} n^a |C_n| < \infty \right\}, \quad a \geq 0. \quad (1.4)$$

Note, some of these spaces (1.4) have physical interpretations, such as X_0 since the corresponding norm is proportional to the total number of clusters. Due to the physical meaning of coagulation-fragmentation equations, we only consider non-negative solutions:

$$X_{\alpha}^{+} := \left\{ C \in X_\alpha \text{ such that } C_j \geq 0, \text{ for all } j \right\}.$$

When proving **existence** of discrete coagulation-fragmentation equations the most useful approach has been to consider an approximation to the infinite systems of ODEs by making a finite n dimensional truncation, where one can prove that solutions of the truncated system of ODEs approach a function which can subsequently be proven to be a solution of the infinite dimensional system of ODEs (see [12] for a rigorous argument of this process). It is worth noting that the method of a finite n dimensional truncation cannot be used to prove existence in all circumstances, e.g. when the rate coefficients decay rapidly, the method will fail [28].

Alternatively, for both the discrete and continuous coagulation-fragmentation equations, a different approach using fixed-point theorems or operator semigroup theory has been successful in proving existence, see e.g. [1, 14, 61, 89].

As with the case of existence, **uniqueness** of discrete coagulation-fragmentation equations has been proved under certain assumptions about the rate coefficients. The approach assumes the existence of two distinct solutions, say C_1 and C_2, and then proves that some moment of the function $|X| := |C_1 - C_2|$ satisfies a

differential inequality which implies $X = 0$, (see e.g. [28] for further details).

1.3.1.5   Asymptotic Behaviour of Solutions

Armed with existence and uniqueness results we can now consider qualitative results about solutions of coagulation-fragmentation equations. Explicit solutions of coagulation-fragmentation equations are important to understand the behaviour of cluster sizes. However, finding explicit solutions using elementary analytical methods is not always possible.

In cases where the coagulation-fragmentation equations are invariant under certain transformations we can reduce the number of independent variables, allowing useful simplifications to find exact solutions [45]. Such solutions are known as **similarity solutions** (sometimes referred to as scaling solutions or self-similar solutions).

The crucial property of this scaling approach is that when time is large most variables can be combined using a single variable (a "typical size") which grows or shrinks with time [62]. For example, in [37,64,65,97] it is shown that solutions with a variety of different initial conditions will approach the same similarity solution as time tends to infinity. However, similarity solutions have only been found in a limited number of cases. For a discussion of similarity solutions for the continuous equations see [37].

In the case of pure discrete coagulation models (see e.g. (1.1)) for homogeneous kernels i.e. $a_{j,k} = j^\lambda p(j/k)$, where $\lambda \in \mathbb{R}$, a similarity solution is expected to exist in the form

$$C_j = \frac{1}{r(t)^\tau} \phi\left(\frac{j}{r(t)}\right),$$

where $\tau > 0$,

$$r(t) = \frac{\sum_{k=1}^{\infty} k C_k}{\sum_{k=1}^{\infty} C_k}.$$

$r(t)$ represents the average cluster size and ϕ is a scaling of the distance of a cluster of size j to the average $r(t)$, and is invariant over time in the asymptotic limit, see [32,62] for further details. ϕ is often referred to as a scaling function. Note, the average size $r(t)$ is the only size scale, therefore behaviour relative to size cannot show any other size scale.

When considering discrete coagulation-fragmentation models, similarity solutions do not exist in all cases. Indeed, in cases where the growth of the coagulation kernel for large sizes is sufficiently fast, "a cluster of infinite size" will form in finite time, this phenomenon is usually referred to as gelation [28]. Additionally, in systems that contain both coagulation and fragmentation, similarity solutions do not always exist, as we may find equilibrium solutions or solutions that tend to infinity, such as in the constant mass Becker-Döring equations:

$$
\begin{aligned}
C_1' &= R_0(t) - \sum_{k=1}^{\infty} R_j, \\
C_j' &= R_{j-1} - R_j, \quad j > 1,
\end{aligned}
\tag{1.5}
$$

where $R_0(t) = 0$ and $R_j = -C_1 C_j - C_{j+1}$ (see [96] for further details).

However, systems that contain both coagulation and fragmentation can produce similarity solutions in certain circumstances as discussed in [96], where similarity solutions are found for the Becker-Döring equations with a time-dependent input of monomers i.e. $R_0(t) = \alpha t^w$, $w > -1$ in (1.5). The most important case for the work presented in this thesis, is when $w = 0$ for the case of irreversible coagulation, where a similarity solution is found:

$$C_j = C t^{\gamma} \phi(\eta), \quad t \to \infty,$$

where $\phi(0) = 1$, and $\eta = j/t^b$, with $b > 0$.

We will return to this similarity solution in Sections 3.1.2 and 5.5.

Thought this thesis we will use the following definition by Wasow [98, pp. 30-32] on asymptotic power series:

Definition 1.3.1. *[The asymptoticity condition] Let $f(t)$ be defined in a point-set S of the complex $t-$plane having $t = \infty$ as an accumulation point. As $t \to \infty$, the power series $\sum_{r=0}^{m} a_r(t)t^{-r}$ represents $f(t)$ asymptotically in S if*

$$t^m \left[f(t) - \sum_{r=0}^{m} a_r(t)t^{-r} \right] \to 0$$

for all $m \geq 0$.

1.3.2 Alternative Frameworks

As previously discussed the failure of rate equations to produce ISDs accurately has prompted the use of alternative frameworks that could potentially predict the experimental results. What follows is by no means an exhaustive systematic review, but broadly introduces a number of alternative frameworks:

in 1996 Blackman and Mulheran [71] suggested modelling nucleation and coagulation of clusters in various dimensions via the CZD (see Section 1.2). Subsequently, in the same year, for a one dimensional point island model when $i = 1$, Blackman and Mulheran [19] proposed the Blackman-Mulheran theory, a fragmentation process to analyse the **gap size distribution** (GSD), i.e. the one dimension CZD. In the model, an island is the start/end of a gap and nucleation of new islands (during deposition) causes the fragmentation of a gap (see [19] for further details).

Chapter 1. Introduction

In 2007, Pimpinelli and Einstein proposed the Generalised Wigner Surmise (GWS) [78] as an extension of the Blackman-Mulheran theory to accurately describe the CZD in any dimension and for any critical island size. However, the GWS is widely discredited, even by Pimpinelli and Einstein themselves [46], and the extension of the Blackman-Mulheran theory to higher dimensions and critical island sizes remains an open problem.

In 2000, Mulheran and Robbie used the concept of the CZD to make a **joint probability distribution** (JPD) for cluster sizes and capture zone areas to reconstruct the rate equations that govern their evolution (see [73] for further details). This approach has been explored further by Bartelt and Evans [39, 40].

In 2007, Seba introduced a one dimensional model to describe the spacing distribution between parked cars in an infinitely long street to ensure the parking of as many cars as possible. Seba described the spacing distribution approximately by the **distributional fixed point equation** (DFPE) (see [87] for further details). Subsequently, Mulheran *et al.* [72] and Krcelic *et al.* [58] have adapted the DFPE model proposed by Seba for parked cars to the case of nucleation and growth on a one dimensional lattice, by interpreting the distance between neighbouring cars as the gap between any two neighbouring clusters (see [72] for further details).

1.3.2.1 The Visibility Graph

In Sections 1.3.1 – 1.3.2 we discussed previous work on SD models. All previous work has focused on the islands' size statistics (see Section 1.3.1 for details) and the spatial distribution of islands (see Section 1.3.2 for details). However, one would like to combine the information contained in the spatial distribution of islands and in their size statistics. A suitable tool is offered by **Visibility Graphs** (VGs), introduced by Lacasa *et al.* [60], originally to bring the tools of

graph theory to bear on time-series analysis. Subsequently, VGs have been used to analyse exchange rate series [100] and to make solar cycle predictions [103], among other things.

Briefly, in a VG we connect each point P with coordinates (location, size) (the top of our grey bars in Figure 1.3) to all other points that "are visible" from P and analyse the resulting graph.
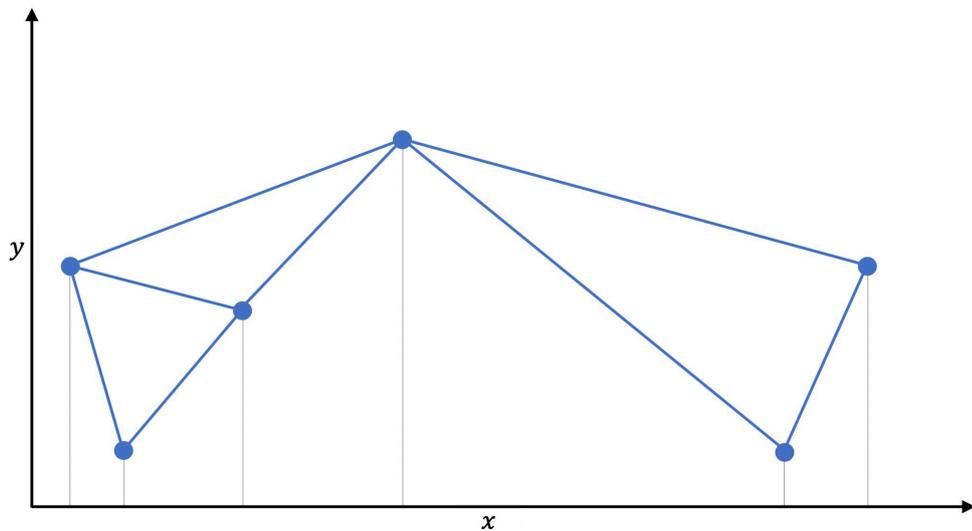
Figure 1.3: An example of a VG where the blue lines and dots represent the edges and vertices in the network respectively.

We will introduce the graph theory required for such a construction in Section 2.2 and will discuss applications of VGs in SD in Chapter 4.

1.4 Overview of Thesis

This thesis consists of seven chapters. Below we give a summary of each of the subsequent chapters.

Chapter 1. Introduction

In Chapter 2, we discuss techniques in graph theory, centre manifold theory and the quasi-steady state assumption that we will use in subsequent chapters, e.g. the adjacency matrix for Chapter 4 and the centre manifold for determining the asymptotic behaviour of solutions for Chapter 5.

In Chapter 3, we discuss several established contributions to this area of SD for future reference.

In Chapter 4, we use visibility graphs as a tool to analyse the results of kinetic Monte Carlo (kMC) simulations of submonolayer deposition in a one dimensional point island model. We introduce an efficient algorithm for the computation of the visibility graph resulting from a kMC simulation and show that from the properties of the visibility graph one can determine the critical island size, thus demonstrating that the visibility graph approach, which implicitly combines size and spatial data, can provide insights into island nucleation and growth processes.

In Chapter 5, we consider the dynamics of point islands during submonolayer deposition, in which the fragmentation of subcritical size islands is allowed. To understand asymptotics of solutions, we use methods of centre manifold theory, and for globalisation, we employ results from the theories of compartmental systems and of asymptotically autonomous dynamical systems. We also compare our results with those obtained by making the quasi-steady state assumption.

In Chapter 6, we consider the asymptotics of the average Erdős number using a rate equations approach. We also compare our results with those obtained by using a Gillespie type algorithm.

In Chapter 7, a summary of the conclusions is presented along with future directions.

Chapter 2

Summary of Methods

2.1 Introduction

In this chapter, we will discuss several methods that have been crucial in our analysis of the VG in relation to SD (see Chapter 4) and our investigation of the ISD for the fragmentation of subcritical islands (see Chapter 5).

In Section 2.2 we begin by introducing some notation and definitions from graph theory. Subsequently, we will discuss methods for analysing graphs which we will use in Chapter 4.

In Section 2.3 we will discuss centre manifold theory, one of the cornerstones of the theory of dynamical systems. An important application of centre manifold theory is the ability to rigorously simplify a dynamical system to reduce the dimension of the system, at the very least, near equilibria. In Section 3.1.1 we will discuss the work of da Costa, van Roessel and Wattis that uses centre manifold theory to determine the asymptotic behaviour of solutions to a system of ODEs, modelling coagulation with Becker-Döring type interactions and a time-independent input

of monomers. In addition, in Chapter 5 we use centre manifold theory to find the asymptotic behaviour of solutions to a system of ODEs modelling point islands during submonolayer deposition, in which the fragmentation of subcritical islands is allowed.

After discussing centre manifold theory, in Section 2.4 we will discuss the quasi-steady-state assumption (QSSA), sometimes referred to as the pseudo-steady-state hypothesis. The basic principle underlying the QSSA reduction heuristics is the elimination of certain variables by setting their rates of change equal to zero and then utilising the resulting algebraic equations. We introduce the QSSA to investigate whether the asymptotics of solutions obtained in Section 5.5 based on the centre manifold analysis of Section 5.4 can be recovered more easily by combining centre manifold reasoning with the QSSA.

2.2   Graph Theory

A **graph** is an ordered pair $G = (V, E)$ used to model pairwise relations between objects where,

- V is the vertex set whose elements are the vertices, or nodes of the graph.

- E is the edge set whose elements are the edges of the graph.

Developments in computing since the 1970s has led to more research in the area of graphs and their applications. Graph theory has been shown to have applications in disciplines including biology, statistical physics, medicine, cognitive science, particle physics, economics, finance and sociology [85]. In this section we introduce the graph theory required to analyse the VG (see Section 1.3.2.1) in relation to SD (see Chapter 4 for the analysis).

Chapter 2.   Summary of Methods

We begin by introducing some notation and definitions [38], as illustrated in Figure 2.1 and the accompanying list:
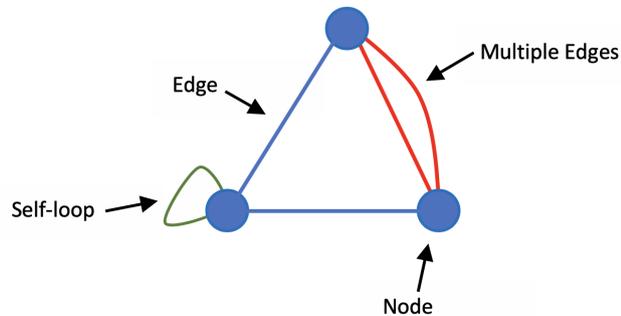
Figure 2.1: An illustration of a graph $G = (V, E)$, where V and E represent the vertices and edges respectively.

1. An **undirected graph** is a graph where all edges are bidirectional.

2. A **simple graph** is an undirected graph, contains no duplicate edges and no self loops.

3. An undirected graph is **connected** if there is a path between every pair of vertices.

4. The **vertex degree** is the number of edges incident to a given vertex.

5. An **unweighted graph** is a graph where no edges have any associated costs or weights.

One of the simplest ways to analyse a graph is to consider the probability distribution of the vertex degree, the **vertex degree distribution**. If we let n be the number of vertices in our graph and $m(k)$ be the number of vertices in our graph having vertex degree k, we can define $q(k) := m(k)/n$ [38].

Chapter 2. Summary of Methods

Clearly the vertex degree distribution only captures a small amount of information from the graph. However, the information contained in the vertex degree distribution still leads to some insight into the underlying system, e.g. in random graphs, where each vertex is connected with probability p, the vertex degree distribution follows a binomial distribution. Comparably, in social media graphs, we see that the vast majority of vertices have a relatively small degree while a small number of vertices have a very large degree so the degree distribution approximately following a power law [85].

Alternatively, we may consider matrix representations of graphs to enable techniques from linear algebra to be applied to graphs. There are many matrix representations of graphs. One of the simplest is the **adjacency matrix** [38]:

Definition 2.2.1. *Let* $G = (V, E)$ *be an unweighted simple graph where* $V = \{v_1, v_2, v_3, \ldots, v_n\}$. *For* $1 \leq i, j \leq n$ *we define*

$$a_{ij} = \begin{cases} 1, & \text{if there exists an edge from vertex } v_j \text{ to vertex } v_i, \\ 0, & \text{otherwise.} \end{cases}$$

The $n \times n$ *matrix* $A = (a_{ij})$ *is the adjacency matrix of* G.

If A is symmetric, then all the eigenvalues are real and there is an orthogonal basis $v_1, \ldots, v_n$ of the space $\mathbb{R}^n$ consisting of eigenvectors of A. Note, the adjacency matrix of an undirected graph is always symmetric.

We can now consider some characteristics of the eigenvalues of the adjacency matrix in relation to VGs. As, by definition, all VGs are undirected connected graphs (see Section 1.3.2.1), all eigenvalues are real and the Perron-Frobenius Theorem implies that the largest eigenvalue λ_{max} of the corresponding adjacency matrix has multiplicity 1. Additionally, λ_{max} can be connected to the average

vertex degree

$$\max\{\bar{d}, \sqrt{d_{\max}}\} \leq \lambda_{\max} \leq d_{\max}$$

where $\bar{d}$ is the average vertex degree and $d_{\max}$ is the maximum vertex degree [66].

The spectral gap, i.e. the gap between the largest and second largest eigenvalue, is an important parameter in many areas of mathematics. We may expect the spectral gap of the adjacency matrix to be related to the connectivity of the graph, i.e. the minimum number of elements (vertices or edges) that need to be removed to separate the remaining vertices into isolated subgraphs. Indeed, fundamental results due to Alon-Milman and Jerrum-Sinclair relate the spectral gap to connectivity properties of graphs (see e.g. [4, 5, 53] for further details).

There are many other ways to characterise a graph; these include criteria based on the spectrum of matrices defined from graphs, communicability and centrality indices [66]. In Chapter 4, we only analyse the vertex degree distribution, the spectrum and spectral gap in the adjacency matrix as these are sufficient to differentiate between VGs corresponding to different critical island sizes i in SD.

2.3 Centre Manifold Theory

In this section, using the machinery in the book by Carr [25], we will introduce centre manifold theory for finite dimensional systems to enable us to analyse the stability of non-hyperbolic fixed points. Further applications of centre manifold theory can be found in [25].

We begin by considering a linear system for $X(t) := X \in \mathbb{R}^N$,

$$X' = AX, \; X(0) = X_0 \in \mathbb{R}^N, \; A \in \mathbb{R}^{N \times N}, \tag{2.1}$$

where, throughout this section, we use primes for the differentiation with respect to time t.

The solution of (2.1) is $X(t) = e^{At}X_0$, where we define

$$e^{At} := \sum_{k=0}^{\infty} = \frac{t^k}{k!}A^k = I + tA + \frac{t^2 A^2}{2!} + \dots .$$

When A has N distinct eigenvalues λ_i and corresponding normalised eigenvectors v_i, the general solution of (2.1) is $X(t) = \sum_{i=1}^{N} e^{\lambda_i t} v_i v_i^T X_0$. Hence, we can decompose the space $\mathbb{R}^N$ into a direct sum of the stable subspace E^s, the unstable subspace E^u and the centre subspace E^s:

$$E^s = Sp\{v_j : Re(\lambda_j) < 0\},$$
$$E^u = Sp\{v_j : Re(\lambda_j) > 0\},$$
$$E^c = Sp\{v_j : Re(\lambda_j) = 0\}.$$

We next consider a general non-linear system of ODEs,

$$X' = f(X), \quad X(0) = X_0 \in \mathbb{R}^N, \ X \in \mathbb{R}^N, \tag{2.2}$$

where we let $\phi(t, X_0)$ be the solution of (2.2) and U be some neighbourhood of a fixed point $\bar{X}$. We first need some definitions.

Definition 2.3.1. *A set $S \subset \mathbb{R}^N$ is a local **invariant manifold** for (2.2) if $\phi(t, X_0) \in S$ for all $|t| < T$, where $T > 0$.*

Definition 2.3.2. *Let U be a topological space. The local **stable manifold** W^s_{loc} and local **unstable manifold** W^u_{loc} of the fixed point $\bar{X}$ for (2.2) are defined as*

follows:

$$W_{loc}^s(\bar{X}) = \{X \in U : \phi(t, X) \to \bar{X} \text{ as } t \to \infty, \phi(t, X) \in U \text{ for all } t \geq 0\},$$

$$W_{loc}^u(\bar{X}) = \{X \in U : \phi(t, X) \to \bar{X} \text{ as } t \to -\infty, \phi(t, X) \in U \text{ for all } t \leq 0\}.$$

$$(2.3)$$

Definition 2.3.3. *A fixed point is said to be **hyperbolic** if its linearisation has no eigenvalues with zero real parts.*

The Stable Manifold Theorem connects the local stable and unstable manifolds of a hyperbolic rest point $\bar{X}$ of the non-linear system (2.2) with stable and unstable subspaces of the origin of the corresponding linearised system:

Theorem 2.3.4. *(Stable Manifold Theorem) Let (2.2) have a hyperbolic fixed point $\bar{X}$. Then W_{loc}^s and W_{loc}^u exist and are of the same dimension as E^s and E^u of the linearised equations and are tangential to the linearised manifolds of $\bar{X}$.*

What Theorem 2.3.4 says is that, in a neighbourhood U of a hyperbolic fixed point $\bar{X}$ of (2.2), the non-linear system (2.2) is topologically equivalent to its linearisation. Hence, for a hyperbolic fixed point, if we understand the linear system then we know the stability properties of the non-linear system close to equilibrium.

To be able to consider stability properties of non-hyperbolic fixed points, and to be able to construct, a centre manifold, we consider system (2.2) in a special form. Suppose (2.2) has a fixed point at $\bar{X} = 0$ and that its variables X can be decomposed into two parts, $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$, where $n + m = N$. Hence, (2.2) has the form

$$x' = Ax + f(x, y), \ \ y' = By + g(x, y), \ \ (x, y) \in \mathbb{R}^n \times \mathbb{R}^m, \qquad (2.4)$$

where $A \in \mathbb{R}^{n \times n}$ and $B \in \mathbb{R}^{m \times m}$, all the eigenvalues of A have zero real parts while all the eigenvalues of B have negative real parts. The functions f and g are sufficiently smooth, $f(0,0) = 0, g(0,0) = 0, Jf(0,0)$ and $Jg(0,0) = 0$. (Thoughout this section, Jf is the Jacobian matrix of f).

If $f = g = 0$ then (2.4) has two local invariant manifolds, namely $x = 0$ and $y = 0$. The local invariant manifold $x = 0$ is a local stable manifold of (2.4), since if we restrict initial data to $x = 0$, all solutions tend to zero. The invariant manifold $y = 0$ is called a centre manifold of (2.4). In general

Definition 2.3.5. *A local **centre manifold** is a local invariant manifold $y = h(x)$, such that $h(0) = 0$ and h is tangent to E^c at the origin i.e. $Jh(0) = 0$*

We now give the results of centre manifold theory to enable us to analyse stability of non-hyperbolic fixed points, these results are proved in [25]:

Theorem 2.3.6. *Let $f(\cdot)$ and $g(\cdot)$ in (2.4) be C^r i.e. r times continuously differentiable in all its variables, then there exists a local **centre manifold** $y = h(x)$, $||x|| < \delta$, where $h \in C^{r-1}$ and the flow on the centre manifold is given by*

$$x' = Ax + f(x, h(x)), \quad x \in \mathbb{R}^n. \tag{2.5}$$

Theorem 2.3.7. *If the fixed point $\bar{x} = 0$ of (2.5) is:*

- *stable/asymptotically stable/unstable, then the fixed point $(\bar{x}, \bar{y}) = 0$ of (2.4) is stable/asymptotically stable/unstable.*

- *stable. Let (x, y) be a solution of (2.4) with $(x(0), y(0))$ sufficiently small. Then there exists a solution of (2.5), $u(t) := u$, such that as $t \to \infty$,*

$$x = u + O(e^{-\gamma t}),$$
$$y = h(u) + O(e^{-\gamma t}),$$

29

where $\gamma > 0$ is a constant.

To use Theorems 2.3.6 – 2.3.7, we need to have enough information about the centre manifold $y = h(x)$ in order to determine the local dynamics of (2.5). If we substitute $y = h(x)$ into the second equation in (2.4) we obtain

$$Jh(x)[Ax + f(x, h(x))] - Bh(x) - g(x, h(x)) = 0. \tag{2.6}$$

Equation (2.6) together with Definition 2.3.5 is the system that must be solved to obtain the centre manifold; in general this cannot be done. However, Theorem 2.3.8 shows the centre manifold can be approximated to any degree of accuracy:

Theorem 2.3.8. *Let u be a C^1 mapping of a neighbourhood of the origin from $\mathbb{R}^n$ to $\mathbb{R}^m$ such that $u(0) = 0$ and $Ju(0) = 0$. Suppose as $||x|| \to 0$, $Ju(x)[Ax + f(x, u(x))] - Bu(x) - g(x, u(x)) = O(||x||^q)$, where $q > 1$, then as $x \to 0$, $||u(x) - h(x)|| = O(||x||^q)$.*

Example 2.3.9. *Consider the system*

$$\begin{aligned} x' &= xy + ax^3 + bxy^2, \\ y' &= -y + cx^2 + dx^2y. \end{aligned} \tag{2.7}$$

The linearised problem has eigenvalues 0 and -1 hence (2.7) has a non-hyperbolic fixed point $\bar{z} := (\bar{x}, \bar{y}) = 0$.

By Theorem 2.3.6, we know the system has a one dimensional centre manifold $y = h(x)$. From definition 2.3.5 of the centre manifold we know h must pass through the origin and be tangent to E^c at the origin; in one dimension that condition is equivalent to $h(0) = h'(0) = 0$.

If $u(x) = O(x^2)$ then $0 = u(x) - cx^2 + O(x^4)$. Hence, if $u(x) = O(x^2)$ then

$u(x) = cx^2 + O(x^4)$. *So, by Theorem 2.3.8 $h(x) = cx^2 + O(x^4)$, and by Theorem 2.3.7, the equation which determines the stability of the fixed point $\bar{z} = 0$ is the scalar equation*

$$u' = uh(u) + au^3 + buh^2(u) = (a + c)u^3 + O(u^5).$$

Hence, the fixed point $\bar{z} = 0$ is:

- *asymptotically stable if $a + c < 0$*

- *unstable if $a + c > 0$.*

We could continue this process to determine the properties when $a + c = 0$ by finding a better approximation to the centre manifold.

2.4   Quasi-Steady State Assumption

In this section, we introduce the QSSA for finite dimensional systems of ODEs. In Section 5.6 we compare the results obtained by making a QSSA with those obtained by a much more laborious, but rigorous, centre manifold technique.

In reaction processes it is often observed, that at some "relevant" time the rate of change of certain reactants is "negligible" compared to the rate of change of other reactants. In such situations it makes sense to set the rate of change of slowly varying reactants to zero, i.e. to make the assumption that they are in a "quasi-steady state" [83]. The QSSA is a largely heuristic technique for reducing the dimension of ODEs that govern the system; the QSSA is often used in the engineering community (see [44, 77, 88]). The best studied example to understand the validity of the QSSA is the famous Michaelis-Menten model for

enzyme kinetics analysed by Segel and Slemrod [88].

In general, conditions that determine the errors associated with making a QSSA, and how to identify a small parameter associated with making a QSSA do not exist. Furthermore, even for the Michaelis-Menten model the validity of the QSSA is still under investigation. However, it is generally accepted that the validity of the QSSA relies on the existence of some "slow and fast variables" in the model [44].

To introduce the QSSA formally we begin by considering a system of ODEs for $x(t) := x \in \mathbb{R}_+^n$,

$$x' = g(x, p), \ p \in \mathbb{R}_+^m, \quad x(0) := x_0, \tag{2.8}$$

where $g : \mathbb{R}_+^{n+m} \to \mathbb{R}^n$ and we use primes for the differentiation with respect to time t.

In order to define the QSSA, we assume there is a distinction between the "fast variables" $x^f \in \mathbb{R}_+^{n-k}$ and the "slow variables" $x^s \in \mathbb{R}_+^k$, where $k < n$.

The variables are defined as linear combinations of the original variables x:

$$\begin{bmatrix} x^s \\ x^f \end{bmatrix} = Tx,$$

where $T \in \mathbb{R}^n \times \mathbb{R}^n$.

The introduction of the fast and slow variables means system (2.8) can be expressed as

$$x'^s = g^s(x^s, x^f, p), \quad x^s(0) := x_0^s,$$
$$x'^f = g^f(x^s, x^f, p), \quad x^f(0) := x_0^f.$$

Chapter 2.   Summary of Methods

The application of the QSSA is based on the hypothesis that the fast variables instantaneously adapt to changes in the slow variables, i.e. $x'^f = 0$. Mathematically the QSSA is valid if after a transition the systems dynamics are restricted to a manifold of lower dimension, (see [44] for further details):

$$x'^s = f^s(x^s, x^f, p), \quad x^s(0) := x_0^s,$$
$$x^f = \Psi(x^s, p).$$

A key step in the above process is explicitly solving $0 = g^f(x^s, x^f, p)$ using a finite number of operations of addition, subtraction, multiplication, division, and radicals, it is worth noting that there are many systems for which an explicit reduction does not exist [77].

Prior to the 1960s there did not appear to be a way of formulating the underlying assumptions in the QSSA mathematically, or even an a priori reason for the QSSA. However, after many ad hoc arguments, justifications for the QSSA have begun to be formulated, such as for a singular perturbation [43] and by using a Computational Singular Perturbation [44].

Below we give an example of the QSSA, when we know the fast variable(s):

Example 2.4.1. *We assume that the following reactions occur:*

$$A \xrightarrow{k_1} I \xrightarrow{k_2} B,$$
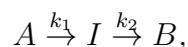
where $k_2 \gg k_1$ and I is the fast variable.

Chapter 2. Summary of Methods

The laws of mass kinetics give us the following finite system of ODEs:

$$A' = -k_1 A,$$
$$I' = k_1 A - k_2 I, \tag{2.9}$$
$$B' = k_2 I.$$

where $A(0) = A_0$, $I(0) = 0$ and $B(0) = 0$.

We can easily explicitly solve system (2.9):

$$A = A_0 e^{-k_1 t},$$
$$I = A_0 \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t}),$$
$$B = A_0 [1 - e^{-k_1 t} - \frac{k_1}{k_2 - k_1} (e^{-k_1 t} - e^{-k_2 t})].$$

Alternatively, as I is a fast variable, we can make a QSSA:

$$A_1' = -k_1 A_1,$$
$$0 = k_1 A_1 - k_2 I_1, \tag{2.10}$$
$$B_1' = k_2 I_1.$$

where $A_1(0) = A_0$ and $B_1(0) = 0$.

Solving system (2.10)

$$A_1 = A_0 e^{-k_1 t},$$
$$I_1 = \frac{k_1}{k_2} A_0 e^{-k_1 t},$$
$$B_1 = A_0 - A_0 e^{-k_1 t}$$

We see, as $k_2 \gg k_1$, the QSSA (2.10) gives a good approximation to (2.9), after

Chapter 2. Summary of Methods

an initial transition.

Chapter 3

Previous Work

Analysis of the Island Size Distribution via Rate Equations

In this section we will discuss several papers that have played a key role in the motivation and investigation of the ISD and have led to the development of ideas presented in this thesis.

In Section 3.1 we will discuss the work of Blackman and Wilding [21] where a successful scaling analysis of rate equations that enables the asymptotics of clusters is proposed.

In Sections 3.1.1 and 3.1.2 we will discuss the work of da Costa, van Roessel and Wattis [30] and Costin, Grinfeld, O'Neill and Park [31], where the case of a constant capture rate is considered. In contrast to the work of Blackman and Wilding the authors were able to establish the asymptotic behaviour of clusters in a mathematically rigorous manner but only in the limited case when $Da_j = 1$ in (3.1).

3.1   The Work of Blackman and Wilding

In the case of pure coagulation, if we assume clusters grow by capturing a single monomer, then (as discussed in Section 1.3.1.1) a realistic system of rate equations with critical island size $i = 1$, constant deposition rate α and diffusion rate D is given by:

$$C_1' = \alpha - 2Da_1C_1^2 - DC_1\sum_{k=2}^{\infty} a_kC_k,$$
$$C_j' = DC_1(a_{j-1}C_{j-1} - a_jC_j), \quad j > 1. \tag{3.1}$$

Blackman and Wilding [21] made the assumption that the capture rates a_j are dependent on j, where a_j has a power law dependence of the form

$$a_j = j^p, \quad p < 1. \tag{3.2}$$

Blackman and Wilding [21] proposed that the asymptotic growth of monomers was governed by a power law:

$$C_1 \sim t^{-w}, \quad t \to \infty, \tag{3.3}$$

where the exponent w is to be determined and we use $\sim$ to mean satisfies the Definition 1.3.1.

Additionally, assuming a unique similarity solution (see Section 1.3.1.5) exists in the form

$$C_j = \frac{1}{j^\tau}\phi\left(\frac{j}{t^z}\right), \tag{3.4}$$

it was shown in [21] that when $0 \leq p < 1/2$,

$$z = 2(3 - 2p)^{-1},$$
$$\tau = \frac{1}{2} + p,$$
$$w = (3 - 2p)^{-1}.$$

Hence, as we will consider the case of a constant capture rate in Section 3.1.1, it is worth taking note of the result for a constant capture rate, i.e. when $p = 0$:

Proposition 3.1.1. *The asymptotics of C_1 and C_j from (3.1) when $a_j = 1$ and assuming a unique similarity solution exists in the form (3.4) are governed by*

$$C_1 \sim t^{-\frac{1}{3}}, \quad t \to \infty,$$
$$C_j = \frac{1}{j^{\frac{1}{2}}} \phi\left(\frac{j}{t^{\frac{2}{3}}}\right).$$

In the same paper Blackman and Wilding [21] considered the effects of fragmentation of subcritical islands (see Section 1.3.1.3), where clusters below a given critical size $i + 1$ are allowed to dissociate.

Assuming the same scaling analysis as in the pure coagulation case namely (3.2)-(3.4) Blackman and Wilding proposed, that:

$$z = (i + 1)[(i + 2) - (i + 1)p]^{-1},$$
$$\tau = \frac{i}{i + 1} + p,$$
$$w = [(i + 2) - (i + 1)p]^{-1}.$$

Hence, as we will consider fragmentation of subcritical islands with a constant capture rate in Sections 3.1.2 and Chapter 5, it is worth taking note of the result

for a constant capture rate, i.e. when $p = 0$:

Proposition 3.1.2. *The asymptotics of C_1 and C_j in the case of fragmentation of subcritical islands when $a_j = 1$ and assuming a unique similarity solution exists in the form (3.4) are governed by:*

$$C_1 \sim t^{-\frac{1}{i+2}}, \quad t \to \infty,$$

$$C_j = \frac{1}{j^{\frac{i}{i+1}}} \phi \left(\frac{j}{t^{\frac{i+1}{i+2}}} \right).$$

It is worth noting that subsequently to the work by Blackman and Wilding [21], Bartelt and Evans [16] have shown rate equations will fail to reproduce the ISD from kMC simulations if the wrong capture rates are chosen.

3.1.1 The Work of da Costa, van Roessel and Wattis

The simplest case of system (3.1) occurs when we let all clusters be point islands and assume all coagulation constants can be taken to be equal (see Section 1.3.1.1 for further details). Hence, setting $Da_j = 1$ for all j in (3.1):

$$C_1' = \alpha - 2C_1^2 - C_1 \sum_{k=2}^{\infty} C_k,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j, \quad j > 1. \tag{3.5}$$

In [30] da Costa, van Roessel and Wattis began by introducing a new variable, the zeroth-moment, i.e. the total number of clusters,

$$T = \sum_{k=1}^{\infty} C_k,$$

where it is assumed $T(0) < \infty$.

Chapter 3.   Previous Work

In terms of T, equations (3.5) can be expressed as

$$T' = \alpha - C_1 T,$$
$$C_1' = \alpha - C_1^2 - C_1 T, \qquad (3.6)$$
$$C_j' = C_1 C_{j-1} - C_1 C_j, \quad j \geq 2.$$

The advantage of system (3.6) over system (3.5) is the first two equations in (3.6) only involve the variables T and C_1 so they can be decoupled from the rest.

In [30], Poincaré compactification and centre manifold methods were then used to prove

Theorem 3.1.1. *The asymptotics of T and C_1 from (3.6) are given by*

$$T \sim (3\alpha^2 t)^{\frac{1}{3}}, \quad t \to \infty,$$
$$C_1 \sim \left(\frac{\alpha}{3t}\right)^{\frac{1}{3}}, \quad t \to \infty.$$

Finally, da Costa, van Roessel and Wattis [30] used a scaling analysis on the final equation from (3.6) (by a change of variable: t to

$$\tau = \int_0^r C_1(s)\,ds$$

and letting $C_j(t) := \tilde{c}_j(\tau))$ to linearise the final equation from (3.6)

$$\tilde{c}_j{}' = \tilde{c}_{j-1} - \tilde{c}_j, \quad j > 1. \qquad (3.7)$$

Hence, da Costa, van Roessel and Wattis showed, using variation of parameters, that the linear system (3.7) can be solved to give

Chapter 3. Previous Work

Theorem 3.1.2. *The asymptotics of C_j from (3.6), when $t \to \infty$, are given by*

$$C_j \sim \left(\frac{\alpha}{3t} \right)^{\frac{1}{3}}, \quad j \geq 1.$$

To formulate the similarity solution da Costa, van Roessel and Wattis first computed the asymptotics of the average cluster size $\langle j \rangle$ using the information in Theorem 3.1.2:

$$\langle j \rangle = \frac{\sum_{j=1}^{\infty} jC_j}{\sum_{j=1}^{\infty} C_j} \sim \left(\frac{\alpha}{3} \right)^{\frac{1}{3}} t^{\frac{2}{3}}, \quad t \to \infty.$$

Next, they defined the function ϕ by

$$\phi(\eta) = \begin{cases} (1-\eta)^{-\frac{1}{2}}, & \eta < 1, \\ \\ 0, & \text{otherwise.} \end{cases}$$

Finally, they defined the similarity variable η by

$$\eta = \frac{2}{3} \frac{j}{\langle j \rangle}.$$

Then they had that the solutions of (3.5) converge to a (discontinuous) similarity profile:

Theorem 3.1.3.

$$C_j = \langle j \rangle^{-\frac{1}{2}} \phi(\eta), \quad t \to \infty.$$

Note, the rigorous results presented by da Costa, van Roessel and Wattis e.g. Theorem 3.1.2 are in agreement with Proposition 3.1.1 of [21].

3.1.2 The Work of Costin, Grinfeld, O'Neill and Park

As discussed in Section 1.3.1.3, there are a number of ways we can model clusters of size $1 < j \leq i$. In [31], Costin, Grinfeld, O'Neill and Park assumed clusters of size $1 < j \leq i$ do not arise. Under such an assumption we can assume the following reactions occur:
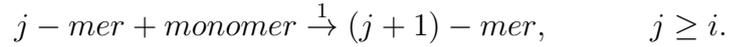
$$j - mer + monomer \xrightarrow{1} (j+1) - mer, \qquad j \geq i.$$

If we set $J_0(t) = \alpha$ to be an external supply of monomers into the system, using the same notation as in (1.1), the laws of mass kinetics give us the following infinite system of ODEs:

$$
\begin{aligned}
C_1' &= \alpha - (i+1)C_1^{i+1} - C_1 \sum_{k=i+1}^{\infty} C_k, \\
C_{i+1}' &= C_1^{i+1} - C_1 C_{i+1}, \\
C_j' &= C_1 C_{j-1} - C_1 C_j, \quad j > i+1.
\end{aligned}
\tag{3.8}
$$

Similarly to the work of da Costa, van Roessel and Wattis [30] (see Section 3.1.1), Costin, Grinfeld, O'Neill and Park employed a compactness argument to rigorously show (by introducing a new variable, $X := \sum_{k=i+1}^{\infty} C_k$) that the analysis of the asymtotics of solutions can be reduced to the study of a two dimensional system of ODEs:

$$
\begin{aligned}
C_1' &= \alpha - (i+1)C_1^{i+1} - C_1 X, \\
X' &= C_1^{i+1}.
\end{aligned}
\tag{3.9}
$$

It was shown in [31] that

Chapter 3. Previous Work

Theorem 3.1.4. *The asymptotics of C_1 and X from (3.9) are given by*

$$C_1 \sim \left(\frac{\alpha}{(i+2)t} \right)^{\frac{1}{i+2}}, \quad t \to \infty,$$

$$X \sim [(i+2)\alpha^{i+1}t]^{\frac{1}{i+2}}, \quad t \to \infty.$$

Finally, using the same scaling analysis used by da Costa, van Roessel and Wattis [30] (see Section 3.1.1) on the final equation from (3.8), Costin, Grinfeld, O'Neill and Park showed:

Theorem 3.1.5. *The asymptotics of C_j from (3.8), when $t \to \infty$, are given by*

$$C_j \sim \left(\frac{\alpha}{(i+2)t} \right)^{\frac{i}{i+2}}, \quad j > i.$$

To formulate the similarity solution Costin, Grinfeld, O'Neill and Park first computed the asymptotics of the average cluster size $\langle j \rangle$ using the information in Theorem 3.1.5:

$$\langle j \rangle = \frac{\sum_{j=1}^{\infty} jC_j}{\sum_{j=1}^{\infty} C_j} \sim \left(\frac{\alpha}{(i+2)t} \right)^{\frac{i+1}{i+2}}, \quad t \to \infty.$$

Next, they defined the function ϕ by

$$\phi(\eta) = \begin{cases} (1-\eta)^{-\frac{i}{i+1}}, & \eta < 1, \\ \\ 0, & \text{otherwise.} \end{cases}$$

Finally, they defined the similarity variable η by

$$\eta = \frac{(i+1)}{i+2} \frac{j}{\langle j \rangle}.$$

Then they had that the solutions of (3.5) converge to a (discontinuous) similarity profile:

Theorem 3.1.6.

$$C_j = \langle j \rangle^{-\frac{1}{2}} \phi(\eta), \quad t \to \infty.$$

Note, the rigorous results presented by Costin, Grinfeld, O'Neill and Park e.g. Theorem 3.1.5 are in agreement with Proposition 3.1.2 of [21].

As discussed in Section 1.3.1.3, when considering the fragmentation of subcritical islands, the other physically relevant possibility is clusters of size $1 < j \leq i$ fragment (at a rate independent of the cluster size), we will investigate this possibility in Section 5.

Chapter 4

Characterising Submonolayer Deposition via Visibility Graphs

This section is an expanded version of our publication, characterising submonolayer deposition via visibility graphs [2].

4.1   Introduction

In Section 1.3 we discussed previous models of SD and noted that all previous work has focused on the island size statistics (see Section 3 and Chapter 5 for details) and the spatial distribution of islands (see Section 1.3.2 for details). However, as noted in Section 1.3.2.1, one would like to combine the information contained in the spatial distribution of islands and in their size statistics. VGs are a suitable tool for this.

Below, we will be considering point islands, i.e. islands whose extent and internal structure have been neglected. We consider a one dimensional model. Both of

these choices have been made for simplicity as the goal of this chapter is a "proof of concept" to demonstrate the ability of VGs to extract mechanism information from kMC simulations. That said, point islands are often used in SD models as they approximate SD accurately when the islands are "well separated" [72]. In Section 4.4 and Chapter 7 we will discuss generalisations of our method to extended islands and to higher dimensional settings.

Thus, we consider the situation where monomers are randomly deposited onto an initially empty one dimensional lattice L at a deposition rate of α monolayers per unit time (t). The monomers diffuse at a rate D and islands nucleate when $i+1$ monomers coincide at a lattice site. We assume no monomers can evaporate from the lattice and direct impingement does not occur so the coverage θ can be defined as $\theta = 100\alpha t\%$. θ is chosen large enough for us to be in the aggregation regime (where scale-invariance is found), i.e. where the monomers are much more likely to be incorporated into existing islands than nucleate into new islands. The appropriate value of θ where the aggregation regime starts is dependent on i and the ratio $R = D/\alpha$.

4.2   The Visibility Graph Algorithm

First, we would like to describe an efficient algorithm for the computation of a VG (thoughtout this section Claim 4.2.1. will act as our definition of a VG):

Claim 4.2.1. *Given n points in the plane, $S = \{P_1, P_2, \ldots, P_n\}$, where $P_j = (x_j, y_j)$, $j \in \{1, \ldots, n\}$. Let P_a, $P_b \in S$ (assuming without loss of generality that $x_a < x_b$); then P_a and P_b are visible from each other if all points $P_c \in S$ such that $x_a < x_c < x_b$, satisfy*

$$y_c < y_b + \frac{y_b - y_a}{x_b - x_a}(x_c - x_b).$$

Proof. The equation of the line connecting the points P_a and P_b is

$$y - y_a = \frac{y_b - y_a}{x_b - x_a}(x - x_a),$$

or equivalently

$$y - y_b = \frac{y_b - y_a}{x_b - x_a}(x - x_b).$$

Hence, P_b will be visible from P_a if all points $P_c \in S$ satisfy the following condition

$$y_c < y_a + \frac{y_b - y_a}{x_b - x_a}(x_c - x_a),$$

or equivalently

$$y_c < y_b + \frac{y_b - y_a}{x_b - x_a}(x_c - x_b).$$

$\square$

To construct the VG we need to consider all two-point subsets of S (as we must ask if each node is connected to all nodes to the right), which gives us an algorithm with time complexity of

$$C = \sum_{k=1}^{n} k(n - k) = \frac{1}{6}(n - 1)n(n + 1) = \frac{1}{6}n^3 + \mathcal{O}(n^2), \quad n = |S|.$$

As our kMC simulations produce up to 10^5 nucleated sites per kMC simulation, this algorithm is impractical as one VG takes nearly two hours to produce on a quad core desktop PC. Hence, we aim to find an algorithm that is faster than the naïve one. We collect the results needed for the construction of such an algorithm in the following claims. Throughout, we let $P_a, P_b, P_c \in S$ be such that $x_a < x_b < x_c$.

Claim 4.2.2. *Let $A = (a_{jk})$ where $j, k \leq n$ be the adjacency matrix of the VG. Then $a_{jj} = 0$, $a_{j,j+1} = 1$ when $j < n$, $a_{jk} = a_{kj}$.*

Chapter 4. Characterising Submonolayer Deposition via Visibility Graphs

Proof. Our graph is a simple graph, i.e. it is an undirected graph with neither multiple edges nor loops, and each vertex is connected to its nearest neighbours. See Figure 4.3. □

Claim 4.2.3. *Let P_a and P_b be connected and $y_a < y_b$. Then all points P_c such that $x_c > x_b$ and $y_c < y_b$ are not visible from P_a.*

Proof. See Figure 4.1. □
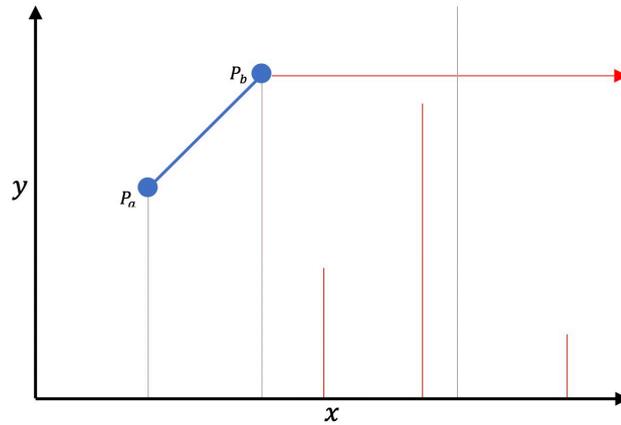
Figure 4.1: An illustration of Claim 4.2.3.

Claim 4.2.4. *Let P_a be connected to P_b and P_b be connected to P_c. Then the slopes of the line segments connecting P_a to P_b and P_b to P_c are given by*

$$m_1 = \frac{y_b - y_a}{x_b - x_a} \text{ and } m_2 = \frac{y_c - y_b}{x_c - x_b}, \text{ respectively.}$$

Thus

1. *if $m_2 > m_1$, P_c is visible from P_a,*

2. *if $m_2 \leq m_1$, P_c is not visible from P_a.*

48

Proof. If the slope of the line connecting P_b and P_c is larger than the gradient of the line connecting P_a and P_b, then P_b will be visible from P_a.See Figure 4.2.   □
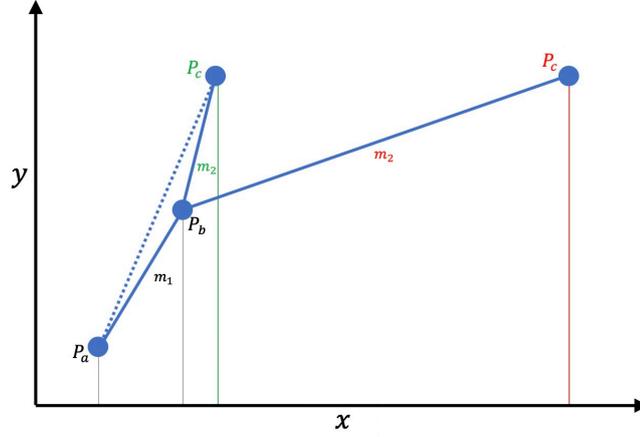
Figure 4.2: An illustration of Claim 4.2.4.

We define the exchange matrix as

$$J_2 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Claim 4.2.5. *Let P_a be connected to P_b (where no point occurs between P_a and P_b) and define $\tilde{P}_c := (x_c, 0)$. Then P_c is visible from P_a if and only if*

$$t_2 = \frac{\boldsymbol{\gamma_1} \cdot \boldsymbol{\gamma_2}}{\boldsymbol{\gamma_3} \cdot \boldsymbol{\gamma_1}}, \quad \in [0, 1],$$

where $\boldsymbol{\gamma_1} = P_b - P_a$, $\boldsymbol{\gamma_2} = [J_2(P_c - P_a)] \odot (1, -1)$, and $\boldsymbol{\gamma_3} = J_2(P_c - \tilde{P}_c)$.See Figure 4.3.

Proof. In parametric form, the ray and the line segment become:

$$X_1(t) = P_a + (P_b - P_a)t_1, \quad t_1 \in [0, \infty),$$
$$X_2(t) = P_c + (\tilde{P}_c - P_c)t_2, \quad t_2 \in [0, 1].$$

49

We consider when the line segment and the ray cross, equating $X_1 = X_2$ and considering the (x, y) component separately:

$$x_a + (x_b - x_a)t_1 = x_c,$$

$$y_a + (y_b - y_a)t_1 = y_c - y_c t_2.$$

Hence, the line and the ray cross if, by solving this system for (t_1, t_2) we obtain $t_1 > 0$ and $t_2 \in [0, 1]$, which is true by the assumption that $x_b > x_a$ and $x_c > x_a$, and

$$t_2 = \frac{(y_c - y_a)(x_b - x_a) - (y_b - y_a)(x_c - x_a)}{y_c(x_b - x_a)} = \frac{\boldsymbol{\gamma_1 \cdot \gamma_2}}{\boldsymbol{\gamma_3 \cdot \gamma_1}} \in [0, 1].$$

$\square$
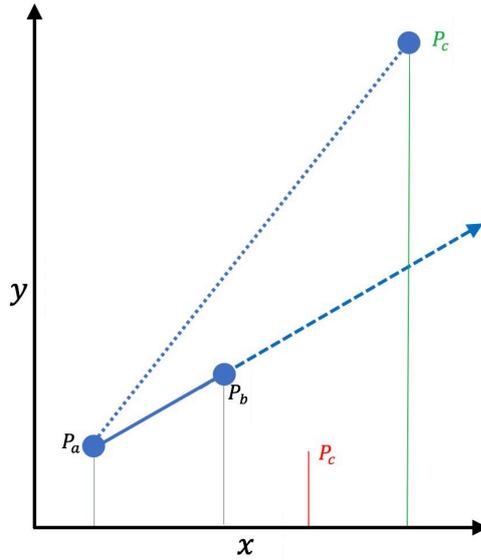
Figure 4.3: An illustration of Claim 4.2.5.

4.2.1 The New Visibility Graph Algorithm

We can now use Claims 4.2.2 − 4.2.5 to construct the adjacency matrix of the VG. We consider an arbitrary point $P_j \in S$ and the vector of elements to the

right of the main diagonal in the j-th row of the adjacency matrix of the VG $[a_{j,j+1},\, a_{j,j+2}, \ldots,\, a_{j,n}]$. Letting $2 \leq k \leq n - j$, we have:

- $a_{j,j+1} = 1$, by Claim 4.2.2.

- If $a_{j,j+k-1} = 1$ and $y_{j+i} < y_{j+k-1}$ where $k \leq i \leq n - j$, for all k, then $a_{j,j+i} = 0$ by Claim 4.2.3.

- If $a_{j,j+k-1} = 1$ then $a_{j,j+k} = 1$ if $m_2 > m_1$ and $a_{j,j+k} = 0$ otherwise by Claim 4.2.4.

- If $a_{j,j+k-1} = 0$, then $a_{j,j+k} = 1$ if $t_2 \in [0, 1]$ and $a_{j,j+k} = 0$ otherwise by Claim 4.2.5.

We continue this process for all $P_j \in S$ and then use the final property from Claim 4.2.2 to complete our adjacency matrix.

Our new algorithm is around 15 times faster than the original (when the number of nodes in our system is approximately 50000) and hence, typically reduces computation time for the construction of one VG from nearly two hours to eight minutes on a quad core desktop PC.

4.3 Characterising the Visibility Graph

Thus, we start with a kMC simulation of SD in one space dimension. Once the kMC simulation is complete, we mark the location and the size ("height") of each nucleated island and construct the resulting VG.

Our kMC simulations were performed on lattices with $L = 10^6$ sites, $R = 10^6$ up to coverage of $\theta = 200\%$ for different critical island sizes i. (For $i = 0$ we set the

spontaneous nucleation probability, i.e. the chance a monomer becomes fixed to the lattice, to $p = 10^{-6}$). We choose these conditions to guarantee that we are in the aggregation regime, where variations in L, R, p and θ do not effect the island size and gap size distributions [72], and throughout the remainder of this chapter we refer to these conditions as our 'standard conditions'.

There are many ways to characterise a graph; these include criteria based on vertex degree, spectrum of the adjacency and other matrices defined from the graph, communicability and centrality indices [66]. Below, we only analyse the vertex degree distribution, the spectrum and the spectral gap in the adjacency matrix as these are sufficient to differentiate between VGs corresponding to different critical island sizes i.

4.3.1   Vertex Degree Distribution

We begin our characterisation of the VG by considering the vertex degree distribution. Let n be the number of vertices in our VG and $m(k)$ be the number of vertices in our VG with k connectivity. For simplicity, we define $q(k) := m(k)/n$. The vertex degree distributions of 10 VGs generated from kMC simulations (under our standard conditions) are shown in Figure 4.4.
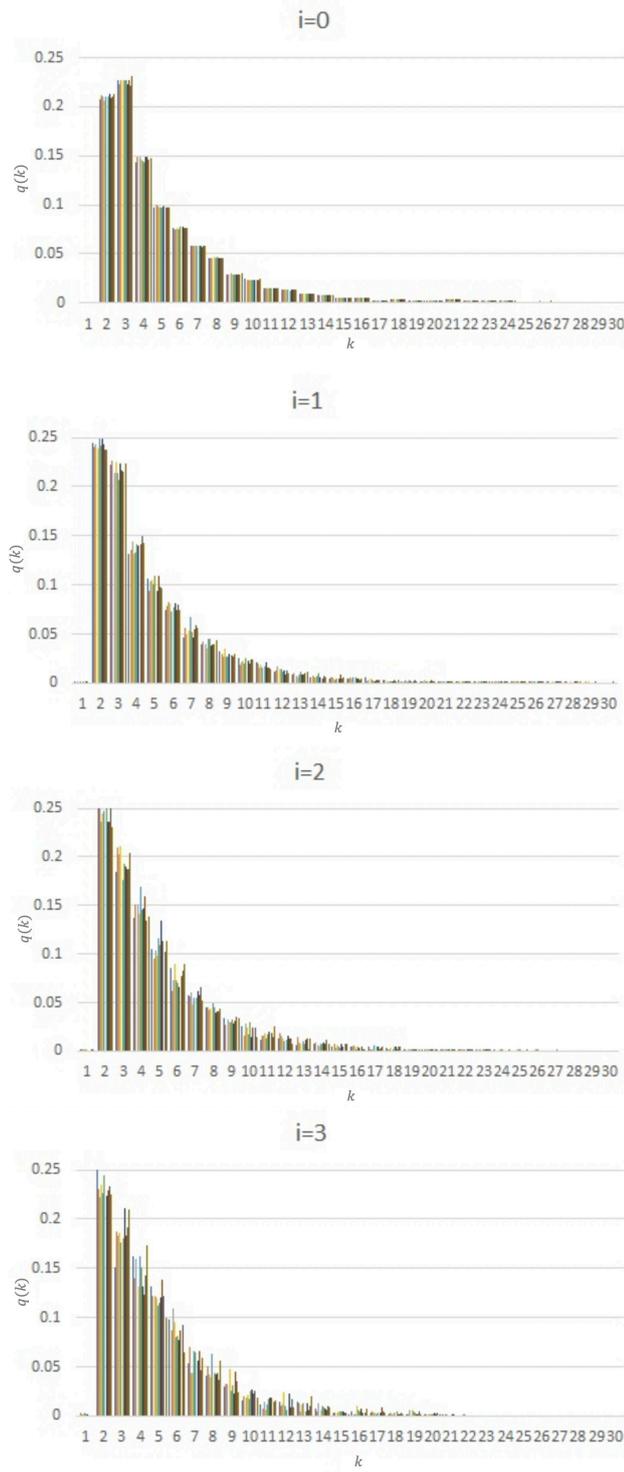
Figure 4.4: The vertex degree distributions of 10 VGs generated from kMC simulations on lattices with $L = 10^6$ sites, with $R = 10^6$ and up to coverage of $\theta = 200\%$ for $i = 0, 1, 2$ and 3 and in the $i = 0$ case we let $p = 10^{-6}$, where each colour corresponds to a different simulations.

The histograms in Figure 4.4 show consistent behaviour for each i. Note, the increase in noise as i increases is due to the number of vertices n decreasing as i increases. Given that we have consistent behaviour for each i, we consider the above process again, this time averaging the vertex degree distributions produced from 50 kMC simulations.
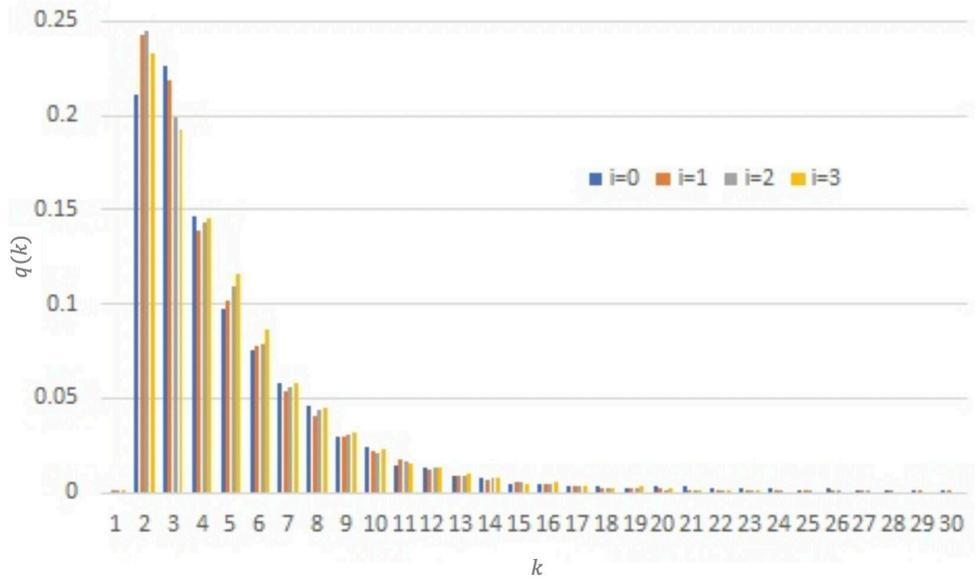
Figure 4.5: The vertex degree distributions of VGs generated from kMC simulations on lattices with $L = 10^6$ sites, with $R = 10^6$ and up to coverage of $\theta = 200\%$ when $i = 0, 1, 2$ and 3 and in the $i = 0$ case we let $p = 10^{-6}$, averaging results over 50 runs.

From Figure 4.5, we see that graphs corresponding to different values of i differ in the statistics of vertices having degree k, particularly for $3 \leq k \leq 8$. To investigate this finding further, we consider this specific region, as shown in Figure 4.6. To emphasise the differences, we connect the points with straight lines.
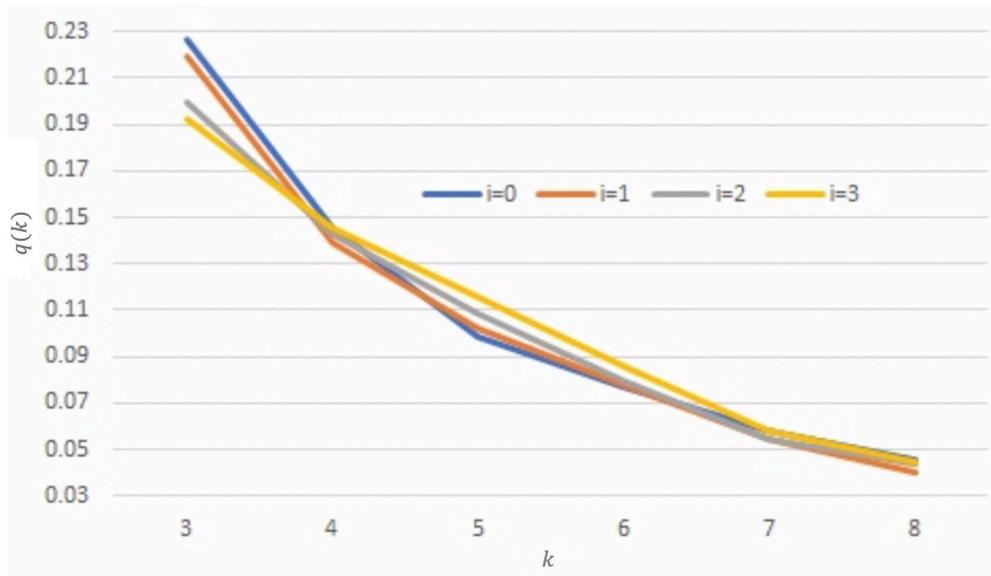
Figure 4.6: The vertex degree distributions of VGs generated from kMC simulations on lattices with $L = 10^6$ sites, with $R = 10^6$ and up to coverage of $\theta = 200\%$ when $i = 0, 1, 2$ and 3 and in the $i = 0$ case we let $p = 10^{-6}$, averaging results over 50 runs, for $3 \leq k \leq 8$.

As expected, for every i considered, the degree distributions are monotonically decreasing, however, there are noticeable differences, particularly for $q(3)$ for different i. Changes in R (when $R = 10^7$), L (when $L = 10^7$) and θ (when $\theta = 100\%$) have a negligible effect on the degree distributions, see Figure 4.7; this is consistent with the work on gap size, island size and spatial distributions [72].
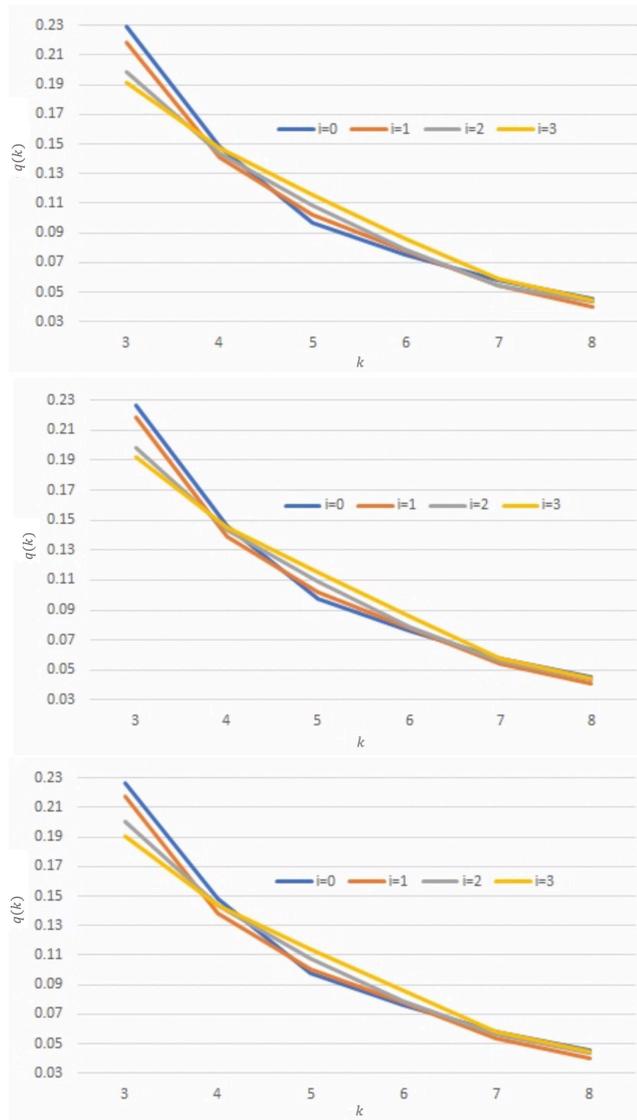
Figure 4.7: The vertex degree distributions of VGs generated from kMC simulations under our standard conditions where: $L = 10^7$ sites, $R = 10^7$, and up to coverage of $\theta = 100\%$ (from top to bottom respectively), when $i = 0, 1, 2$ and 3, averaging results over 50 runs, for $3 \leq k \leq 8$.

To test the $q(3)$ as predictor of i, we generate a VG from a kMC simulation where $1 \leq i \leq 3$ is chosen randomly and then compute $q(3)$ to predict i. Our kMC simulations were performed on lattices with $L = 10^6, 10^7$ or 10^8 sites, $R = 10^6, 10^7, 10^8$ or 10^9, $\theta = 100\%$ or 200% (these conditions guarantee that we are

in the aggregation regime; we will refer to them as "operational conditions") and i was randomly chosen between 0 and 3, averaging results over 50 runs. We performed this process 100 times. We found that i was correctly predicted in 92% of cases. In addition, in all cases the predicted i was within 1 of the true value of i. See Figure 4.8 for the maximum and minimum values achieved for each i.
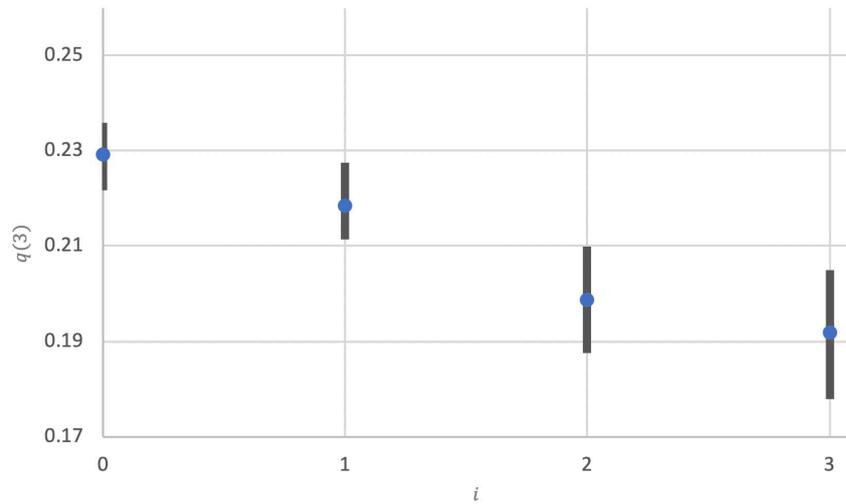
Figure 4.8: The maximum and minimum values achieved generated from kMC simulations under our standard conditions.

4.3.2 Spectrum of the Adjacency Matrix

As discussed in Section 2.2, as the vertex degree distribution only captures a small amount of information from the graph, next we consider the adjacency matrix of the VG. We consider the first five eigenvalues of the adjacency matrix of our VGs generated from kMC simulations under our standard conditions. As with the vertex degree distribution, we find consistent behaviour for each i. As our

Chapter 4. Characterising Submonolayer Deposition via Visibility Graphs

network is connected, the largest eigenvalue, $\lambda_{\max}$, has multiplicity 1 and

$$\max\{\bar{d}, \sqrt{d_{\max}}\} \leq \lambda_{\max} \leq d_{\max}$$

where $\bar{d}$ is the average degree and $d_{\max}$ is the maximum degree [66]. We average the eigenvalues over 50 runs, as shown in Figure 4.9.
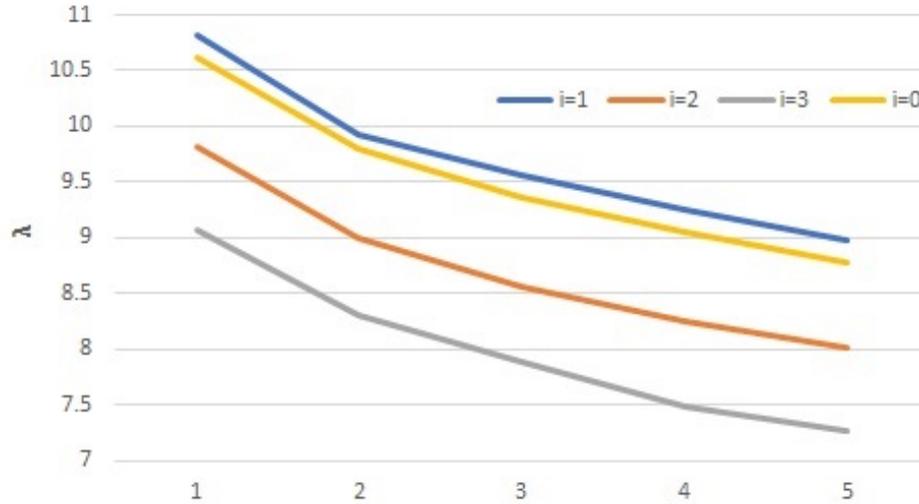
Figure 4.9: The eigenvalues of VGs generated from kMC simulations on lattices with $L = 10^6$ sites, with $R = 10^6$ and up to coverage of $\theta = 200\%$ when $i = 0, 1, 2$ and 3 and in the $i = 0$ case we let $p = 10^{-6}$, averaging results over 50 runs for the adjacency matrix.

As the $i = 0$ case is practically indistinguishable from the $i = 1$ case, to separate these two we consider the gap between the largest eigenvalue and the second largest eigenvalue of the adjacency matrix, i.e. the spectral gap, which has been shown to be related to the connectivity of the graph [66]; these results are shown in Figure 4.10.
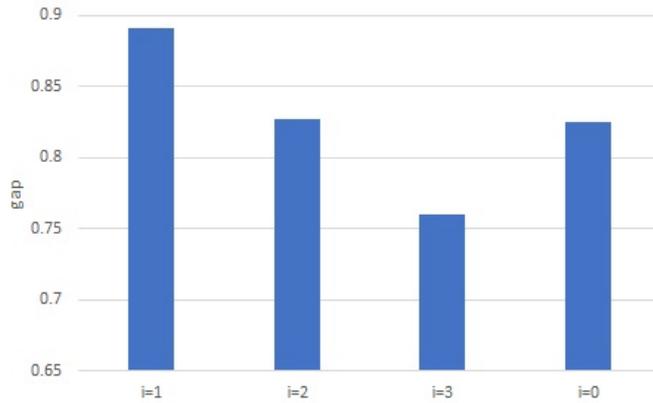
Figure 4.10: The gap between the first and second eigenvalue of the adjacency matrix from VGs generated from kMC simulations when $i = 0, 1, 2$ and 3, $R = 16 \times 10^6$, $L = 10^6$, $\theta = 200\%$ and in the $i = 0$ we let $p = 10^{-6}$, averaging results over 50 runs.

Once again, we find that changes in R, L and θ have a negligible effect on the eigenvalues and the gaps between the eigenvalues.

As in the case of using $q(3)$ to distinguish between nucleation mechanisms, we generate a VG from a kMC simulation under our operational conditions averaging results over 50 runs. In order to identify the value of i from the adjacency matrix, we use the largest eigenvalue to separate the case of $i = 2, 3$ from the rest, and then the spectral gap to differentiate between $i = 0$ and $i = 1$. We performed this process 50 times. We found that i was correctly predicted in all cases. Note, in contrast to Figure 4.6, the excellent separation of the $i = 2$ and $i = 3$ case.

4.4   Conclusions

We have shown that the analysis of some of the properties of the VG generated from a kMC simulation allow us to determine the underlying nucleation mechanism. Both the degree distribution ($q(3)$) and the spectrum of the adjacency

matrix reliably allow us to identify the value of i used in the kMC simulation. We have also created an efficient algorithm for processing the kMC position/size data. Therefore, we have created an effective characterisation process that can be applied to experimental data for SD in one dimension, such as island nucleation and growth on a stepped substrate [79].

The VG method has the potential to deal with more complicated mechanisms include evaporation, mobile islands, the action of electric fields and any level of coverage within the scaling regime as discussed above. The generalisation of our work to extended islands is also straightforward as we can create the vectors P, used in the construction of VG, by using the position of the centre of mass of an island and its mass as coordinates. We leave these versions of SD to future work.

It is true that at this stage there is no *a priori* reason why information about the critical island size i should be contained in $q(3)$ or in the spectrum of the adjacency matrix, as demonstrated here. For that reason, the VG framework used here falls in the domain of "equation-free" approaches (for a general philosophy of this see [55]), as do the applications in complex (in particular, biological and financial) systems of topological data analysis [24] and Minkowski functionals [17]. Such an exploratory stage is necessary to verify, as we do here, that the tool is up to the task.

An important question is how to extend this methodology to two and three space dimensions. In [59] a method is proposed to extend one dimensional VGs to higher dimensions which enables the construction of VGs of large-scale spatially-extended surfaces. The method uses one dimensional VGs along different straight lines in the multidimensional lattice to construct a single VG (only dependent on the number of lines one considers).

Of course, other ways of correctly identifying i from data, such as from the

scaled distribution of island sizes, already exist, and it is not clear whether a VG offers any immediate advantages in terms of robustness against noise or clarity of interpretation when the growth rules evolve over time. Nevertheless, we have successfully demonstrated that the VG approach usefully combines spatial and size data in a physically meaningful way, relating SD to network theory, thereby opening up new approaches to understanding more complex SD processes and their classification.

Chapter 5

Point Island Dynamics Under Fixed Rate Deposition

This section is an expanded version of our publication, point island dynamics under fixed rate deposition [3].

5.1 Introduction

In Section 3, we described how ODE models of SD have been the subject of rigorous mathematical analysis with the aim of determining the ISD. As mentioned in Section 3.1.1, studies of this type of rate equations have been initiated by da Costa, van Roessel and Wattis; see also [29] and Section 3.1.2, all of which are relevant to the present work.

As in Section 3.1.2, here we assume that there exists a critical island size i such that islands of size $j \geq i + 1$ are immobile and can only grow by attachment of a single monomer.

As discussed in Section 1.3.1.3, there are a number of possibilities of how to model clusters of size $1 < j \leq i$. The possibility considered in Section 3.1.2, is that clusters of size $1 < j \leq i$ simply do not exist. There is one other physically relevant possibility, i.e. that clusters of every size $1 < j \leq i$ are allowed to fragment (at some rate independent of the cluster size, which is consistent with the point-island assumption). This possibility has been considered formally, see Section 3.1 and [70]. In this chapter we consider this mechanism using centre manifold techniques and globalising the results.

In Section 3.1.1 and 3.1.2, it was possible by a change of variables to decouple the infinite system of ODEs in a way that reduced its analysis to an analysis of a two dimensional system. In our case, the reduction is to $i + 1$ equations, and the remarkable property of these equations is that the complexity of the calculations is independent of i. Additionally, we show that making the QSSA results in the same leading term behaviour as the centre manifold computation and emphasise the differences between the two approaches.

5.2   Governing Equations

As in Section 1.3.1.3, we consider a system containing clusters of any number $j \geq 1$ or monomers, where subcritcal islands are allowed to fragment. We make the additional assumption that the fragmentation rate, $\beta > 0$, is independent of the cluster size. Hence, we assume that the following reactions occur

$$ j - mer + monomer \underset{\beta}{\overset{1}{\rightleftarrows}} (j+1) - mer, \qquad 1 \leq j < i, $$

$$ j - mer + monomer \overset{1}{\rightarrow} (j+1) - mer, \qquad j \geq i. $$
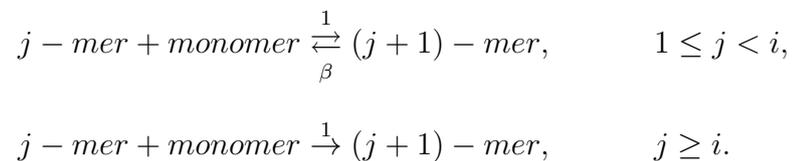
If we set $\hat{\alpha}$ to be the deposition rate, denote by $C_j(t) := C_j$ the concentration of

$j - mers$ at time t and use primes for differentiation with respect to t, the laws of mass kinetics give us the following infinite system of ODEs:

$$C_1' = \hat{\alpha} - 2C_1^2 + 2\beta C_2 - C_1 \sum_{k=2}^{\infty} C_k + \beta \sum_{k=3}^{i} C_k,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j - \beta C_j + \beta C_{j+1}, \quad 1 < j < i,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j - \beta C_j, \quad j = i,$$

$$C_j' = C_1 C_{j-1} - C_1 C_j, \quad j > i.$$

(5.1)

It makes sense to scale the variables and the deposition rate to remove β from the equations. Thus scaling $t \to T := \beta t$, retaining primes for differentiation with respect to the new time scale, setting $C_j(t) = \beta c_j(T)$ and $\alpha = \hat{\alpha}/\beta^2$, we obtain the system

$$c_1' = \alpha - 2c_1^2 + 2c_2 - c_1 \sum_{k=2}^{\infty} c_k + \sum_{k=3}^{i} c_k,$$

$$c_j' = c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i,$$

$$c_j' = c_1 c_{j-1} - c_1 c_j - c_j, \quad j = i,$$

$$c_j' = c_1 c_{j-1} - c_1 c_j, \quad j > i.$$

(5.2)

5.3 Globalisation

In this section we consider the global dynamics of c_j, $1 \le j \le i$, satisfying (5.2) and of $v = \alpha - c_1 \sum_{k=2}^{\infty} c_k$, and establish that all solutions of these equations with non-negative initial data approach the origin. This will show that the flow on the centre manifold, as given by Theorem 5.4.1, describes the asymptotics of every non-negative solution of this system of equations.

For that purpose, it is more convenient to rewrite equations (5.2) formally as

follows:

$$c_1' = \alpha - 2c_1^2 + 2c_2 - c_1 \sum_{k=2}^{i} c_k + \sum_{k=3}^{i} c_k - c_1 y,$$

$$c_j' = c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \qquad 1 < j < i, \tag{5.3}$$

$$c_i' = c_1 c_{i-1} - c_1 c_i - c_i,$$

$$y' = c_1 c_i,$$

where we have put $y(t) = \sum_{k=i+1}^{\infty} c_k$.

First of all, we have

Theorem 5.3.1. *If $\sum_{k=1}^{\infty} c_k(0) < \infty$, a solution of (5.2) for $j \geq 1$ is also a solution of (5.3).*

Proof. The argument of the proof is similar to that of [30, Theorem 2.1]. We indicate the main steps.

Let $(c_j)_{j=1}^{\infty}$ be a solution of (5.2). To show that this is also a solution of (5.3) we must prove that $\sum_{k=i+1}^{\infty} c_k$ converges to y for all T. We change time from T to

$$\rho = \int_0^T c_1(s)\, ds.$$

This change of variable (also used in [30, Theorem 2.1]) makes the c_j equations of (5.2) linear in c_j for $j > i$. Keeping primes for differentiation with respect to the new time variable ρ and letting $c_j(T) =: \tilde{c}_j(\rho)$, $y(T) =: \tilde{y}(\rho)$, these equations become

$$\tilde{c}_j{}' = \tilde{c}_{j-1} - \tilde{c}_j, \quad j > i, \text{ and } \tilde{y}' = \tilde{c}_i. \tag{5.4}$$

This system of ODEs for $\tilde{c}_j$, $j > i$, can now be solved in terms of $\tilde{c}_i$ recursively

by variation of parameters starting at $j = i + 1$, to give

$$\tilde{c}_j(\rho) = e^{-\rho} \sum_{k=i+1}^{j} \frac{\rho^{j-k}}{(j-k)!} \tilde{c}_k(0) + \frac{1}{(j-(i+1))!} \int_0^\rho \tilde{c}_i(\rho - s) s^{j-(i+1)} e^{-s} ds. \quad (5.5)$$

Introducing the generating function

$$F(\rho, z) := \sum_{n=i+1}^{\infty} \tilde{c}_n(\rho) z^n,$$

we can use (5.5) to rewrite F as $F(\rho, z) := G(\rho, z) + H(\rho, z)$, where

$$G(\rho, z) = e^{-\rho} \sum_{n=i+1}^{\infty} \sum_{k=i+1}^{n} \frac{\rho^{n-k} z^n}{(n-k)!} \tilde{c}_k(0),$$

and

$$H(\rho, z) = \sum_{n=i+1}^{\infty} \frac{z^n}{(n-(i+1))!} \int_0^\rho \tilde{c}_i(\rho - s) s^{n-(i+1)} e^{-s} ds.$$

We now consider these two expressions separately. For G we obtain

$$G(\rho, z) = e^{-\rho} \sum_{n=i+1}^{\infty} \sum_{k=0}^{n-(i+1)} \frac{\rho^{n-k-(i+1)} z^n}{(n-k-(i+1))!} \tilde{c}_{k+i+1}(0)$$

$$= e^{-\rho} \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} \frac{\rho^m z^{m+k+i+1}}{m!} \tilde{c}_{k+i+1}(0)$$

$$= e^{-\rho} \sum_{k=0}^{\infty} \left(\sum_{m=0}^{\infty} \frac{(\rho z)^m}{m!} \right) z^{k+2} \tilde{c}_{k+i+1}(0)$$

$$= e^{-\rho(1-z)} \sum_{k=i+1}^{\infty} z^k \tilde{c}_k(0).$$

Since $\sum_{k=1}^{\infty} c_k(0) < \infty$ by assumption, the above series converges when $|z| \le 1$,

and we have

$$G(\rho, z) = e^{-\rho(1-z)}(F(0, z) - \tilde{c}_i(0)z) \quad \text{for} \quad |z| \le 1.$$

For H, by interchanging the order of summation and integration, we have that

$$
\begin{aligned}
H(\rho, z) &= z^{i+1} \int_0^\rho \tilde{c}_i(\rho - s) \bigg(\sum_{n=i+1}^\infty \frac{(sz)^{n-(i+1)}}{(n - (i+1)!} \bigg) e^{-s} ds \\
&= z^{i+1} \int_0^\rho \tilde{c}_i(\rho - s) e^{sz} e^{-s} ds \\
&= z^i \int_0^\rho \tilde{c}_i(s) e^{-(\rho-s)(1-z)} ds.
\end{aligned}
$$

The expression for F at $z = 1$ now becomes

$$F(\rho, 1) = F(0, 1) - \tilde{c}_i(0) + \int_0^\rho \tilde{c}_i(s) ds. \tag{5.6}$$

Hence, by differentiating with respect to ρ, we see that $F(\rho, 1)$ given by (5.6) satisfies the same differential equation as $\tilde{y}$ in (5.4) which proves that $F(\rho, 1) = \tilde{y}$. Thus, in the T variables $\sum_{k=i+1}^\infty c_k$ converges to y. $\square$

As a result of Theorem 5.3.1, we can use finite dimensional techniques to discuss the dynamics of $c_j(T)$, $1 \le j \le i$.

We begin our analysis of long-time dynamics of (5.3) by considering the system without outflows through higher clusters, i.e. we let $y \equiv 0$ in system (5.3).

$$c_1' = \alpha - 2c_1^2 + 2c_2 - c_1 \sum_{k=2}^{i} c_k + \sum_{k=3}^{i} c_k,$$

$$c_j' = c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i,$$

$$c_i' = c_1 c_{i-1} - c_1 c_i - c_i,$$

(5.7)

which can be expressed as the following system of reactions:

$$monomer + i - mer \rightarrow \emptyset \rightarrow monomer,$$

$$j - mer + monomer \rightleftarrows (j+1) - mer, \qquad j < i.$$
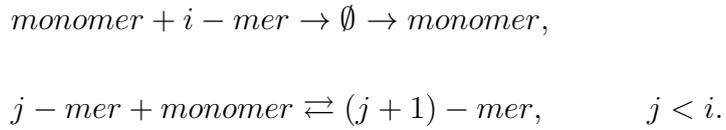
We begin by defining a compartmental system in the sense of Jacquez and Simon [51]. Let the flows into the compartment from outside the system, or inflows, be represented by $I_i \geq 0$; the outflows to the environment and therefore out of the system by g_{0i}; the transfers from compartment i to compartment j by g_{ji} and the transfers from j to i by g_{ij}. All the g_{hk}'s are ≥ 0. If the general equations for such a system can be obtained by writing the instantaneous mass balance equation:

$$q_i' = \sum_{k \neq j} -g_{ji} + g_{ij} + I_i - g_{0i},$$

(5.8)

then the system is a compartmental system.

Let us show that the system (5.7) is a compartmental system. Let I_j represent the inflows from outside the system into c_j, F_{0j} represent the outflow from c_j to outside of the system, F_{kj} represent the transfer from c_j to c_k and F_{jk} represent the transfer from c_k to c_j, when $1 \leq j, k \leq i$. Below is an illustration for visualisation

purposes:
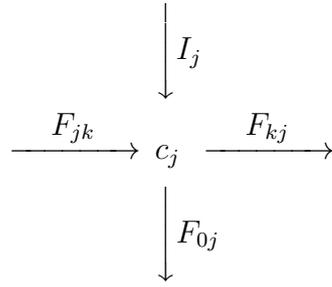
Let $I_1 = \alpha$ and let $I_j = 0$ for all $2 \leq j \leq i$. Now put

$$F_{j1} = c_1 c_{j-1}, \quad j = 2, \ldots, i;$$

$$F_{12} = 2c_2 \text{ and } F_{1j} = c_j \quad j = 3, \ldots, i.$$

For $k = j - 1$, $2 \leq k \leq i - 1$ put $F_{kj} = c_j$, $F_{jk} = c_1 c_k$ and for $k = j + 1$, $2 \leq j \leq i - 1$, put $F_{kj} = c_j$. Finally, let $F_{0k} = 0$ if $k \neq 1, i$ and $F_{0i} = F_{01} = c_1 c_i$, the only outflows from the system. In other words,

$$I_j = \begin{bmatrix} \alpha & 0 & \ldots & 0 \end{bmatrix},$$

$$F_{jk} = \begin{bmatrix} 0 & 2c_2 & c_3 & c_4 & c_5 & \ldots & c_i \\ c_1^2 & 0 & c_3 & & & & \\ c_1 c_2 & c_1 c_3 & 0 & c_4 & & & \\ c_1 c_3 & & c_1 c_4 & 0 & c_5 & & \\ \vdots & & & \ddots & \ddots & \ddots & \\ c_1 c_{i-1} & & & & c_1 c_i & 0 & c_i \end{bmatrix},$$

$$F_{0j} = \begin{bmatrix} c_1 c_i & 0 & \ldots & 0 \end{bmatrix}.$$

Then clearly for each $j = 1, \ldots i$ we can write

$$c'_j = \sum_{k \neq j}^{i} -F_{kj} + F_{jk} + I_j - F_{0j}, \qquad (5.9)$$

where all the F's and I's are positive, which shows that (5.7) is a compartmental system in the sense of [51].

Also note that
$$\frac{\partial F_{jk}}{\partial c_m} \geq 0 \text{ for all } 1 \leq j, k, m \leq i, \quad j \neq k. \qquad (5.10)$$

Hence, we can use the Theorem of Maeda, Kodama and Ohta [63], (see also part (i) of Theorem 9 of [51]):

Theorem 5.3.2 ([63]). *Given a compartmental system (5.9) with time-independent inputs I_j that satisfies the monotonicity condition (5.10), every non-negative solution of the system is bounded if and only if the system has a positive rest point.*

We will make use of the following lemma:

Lemma 5.3.3. *The system* (5.7), *with non-negative initial data, admits a positive unique rest point*

$$(c_1, c_2, \ldots, c_i) = (\alpha^{\frac{1}{i+1}}, \alpha^{\frac{2}{i+1}}, \ldots, \alpha^{\frac{i}{i+1}}).$$

Proof. Letting, $c'_j = 0$ when $1 \leq j \leq i$ in the system (5.7):

$$
\begin{aligned}
0 &= \alpha - 2c_1^2 + 2c_2 - c_1 \sum_{k=2}^{i} c_k + \sum_{k=3}^{i} c_k, \\
0 &= c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i, \\
0 &= c_1 c_{j-1} - c_j, \quad j = i.
\end{aligned}
\qquad (5.11)
$$

Expressing the equations for c_j, $1 < j \leq i$, from (5.11) in the following form

$$c_j = c_1 c_{j-1}, \quad j = i,$$

$$c_j = c_1 c_{j-1} - c_1 c_j + c_{j+1}, \quad 1 < j < i,$$

and substituting the c_i equation into the equation for c_{i-1}

$$c_{i-1} = c_1 c_{i-2}.$$

Continuing this recurrence

$$c_i = c_1 c_{i-1},$$

$$c_j = c_1 c_{j-1}, \quad 1 < j < i. \tag{5.12}$$

Hence, from (5.12)

$$c_j = c_1^j, \quad 1 < j \leq i. \tag{5.13}$$

Substituting the expressions for c_j from (5.13) into the first equation from (5.11) and considering only non-negative initial data

$$0 = \alpha - 2c_1^2 + 2c_1^2 - c_1 \sum_{k=2}^{i} c_1^k + \sum_{k=3}^{i} c_1^k,$$

$$0 = \alpha - \sum_{k=3}^{i+1} c_1^k + \sum_{k=3}^{i} c_1^k, \tag{5.14}$$

$$c_1 = \alpha^{\frac{1}{i+1}}.$$

Hence, combining the results from (5.13)-(5.14) we find, with non-negative initial data, the only rest point is

$$(c_1, c_2, \ldots, c_i) = (\alpha^{\frac{1}{i+1}}, \alpha^{\frac{2}{i+1}}, \ldots, \alpha^{\frac{i}{i+1}}), \tag{5.15}$$

which is positive as $\alpha > 0$. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Since Lemma 5.3.3 shows (5.7) admits a unique positive equilibrium we conclude using Theorem 5.3.2 that all non-negative solutions of (5.7) are bounded.

Now we consider the first i equations of the system (5.3). Since the system (5.3) preserves non-negativity and y is a positive function, by comparison with solutions of (5.7) it follows that the $(c_1, \ldots, c_i)$ components of non-negative solutions of (5.3) are bounded for any positive initial condition.

In this section we will make use of the Theorem of Thieme [92, Theorem 4.2] on asymptotically autonomous dynamical systems.

Let ϕ be an asymptotically autonomous continuous semiflow on the metric space X and Θ its continuous limit-semiflow and $e \in X$ such that $\Theta(t, e) = e$ for all $t \geq 0$.

Theorem 5.3.4. *Let the equilibria of Θ be an isolated compact Θ-invariant sub-sets of X and the point (s, x), $s \geq t_0, x \in X$, have a pre-compact ϕ-orbit. Then the following alternative holds:*

- *$\phi(t, s, x) \to e$, $t \to \infty$ for some Θ-equilibrium e.*

- *The $\omega - \phi$-limit set of (s, x) contains finitely many Θ-equilibria which are chained to each other in a cyclic way.*

Now consider the dynamics of the last component of (5.3), y. As it is monotone-increasing it can either converge to some limit $l < \infty$ or it can go to infinity.

Let us show that the first possibility cannot occur. For, if it did, we could the Theorem 5.3.4 combined with the fact that all non-negative solutions of (5.3)

are bounded and the uniqueness of the positive equilibrium, to conclude that the ω-limit set of every orbit of (5.3) would be the same as that of the system

$$
\begin{aligned}
c_1' &= \alpha - 2c_1^2 + 2c_2 - c_1 \sum_{k=2}^{i} c_k + \sum_{k=3}^{i} c_k - c_1 l, \\
c_j' &= c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i, \\
c_i' &= c_1 c_{i-1} - c_1 c_i - c_i.
\end{aligned}
\tag{5.16}
$$

But if $y \to l$ as $T \to \infty$, we must have that either $c_1 \to 0$ or $c_i \to 0$. If we suppose, for example, that $c_i \to 0$, we see from the c_i' equation of (5.16) that either c_1 or c_{i-1} must go to zero. Continuing in this way, we see that all c_j must go to zero as $T \to \infty$, but the origin is not a rest point of the first i equations of (5.16). Hence, we conclude that $y \to \infty$.

Furthermore, since the positive orthant of $\mathbb{R}^{i+1}$ is invariant under the flow of (5.3), this must mean that $c_1 \to 0$ as $T \to \infty$.

Now, from the equations for $c_i, c_{i-1}, \ldots, c_2$ it follows consecutively that for all $2 \leq k \leq i$, $c_k \to 0$ as $T \to \infty$, again using the same result of Thieme [92] for asymptotically autonomous dynamical systems. Applying these results to the c_1 equation in (5.3) we finally conclude that $c_1 y \to \alpha$ as $T \to \infty$. If we now set

$$
v := \alpha - c_1 \sum_{k=2}^{\infty} c_k,
\tag{5.17}
$$

this is equivalent to saying that $v \to 0$ as $T \to \infty$.

We collect these results as a theorem:

Theorem 5.3.5. *As $T \to \infty$, for all non-negative initial data, $c_k \to 0$, $1 \leq k \leq i$, and $v \to 0$.*

To understand the dynamics of c_j as $T \to \infty$ for all $j \geq 1$, we first use centre manifold techniques to understand the rate of approach of c_j to zero, $1 \leq i$ as $T \to \infty$.

5.4 Centre Manifold Analysis

We next compactify the ODEs from (5.2) by setting $z = \sum_{k=2}^{\infty} c_k$ and $v = \alpha - c_1 z$. Hence, the equation for c_1 becomes

$$c_1' = v - 2c_1^2 + 2c_2 + \sum_{k=3}^{i} c_k.$$

Differentiating v with respect to T we find that,

$$v' = -c_1 z' - c_1' z =$$

$$= -c_1 \sum_{k=2}^{\infty} c_k' - c_1' z =$$

$$= -c_1^3 + c_1 c_2 - \alpha z + 2z c_1^2 - 2z c_2 + c_1 z^2 - z \sum_{k=3}^{i} c_k =$$

$$= -\frac{1}{c_1} \left[c_1^4 - c_1^2 c_2 + \alpha^2 - \alpha c_1 z - \alpha^2 + 2\alpha c_1 z - c_1^2 z^2 - 2\alpha c_1^2 + 2\alpha c_1^2 \right.$$

$$\left. - 2c_1^3 z + 2\alpha c_2 - 2\alpha c_2 + 2c_1 c_2 z + \alpha \sum_{k=3}^{i} c_k - \alpha \sum_{k=3}^{i} c_k + c_1 z \sum_{k=3}^{i} c_k \right]$$

$$= -\frac{1}{c_1} \left[c_1^4 - c_1^2 c_2 + \alpha v - v^2 - 2\alpha c_1^2 + \right.$$

$$\left. + 2c_1^2 v + 2\alpha c_2 - 2c_2 v + \alpha \sum_{k=3}^{i} c_k - v \sum_{k=3}^{i} c_k \right].$$

Hence, we can now consider the following system:

$$
\begin{aligned}
c_1' &= v - 2c_1^2 + 2c_2 + \sum_{k=3}^{i} c_k, \\
c_j' &= c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i, \\
c_j' &= c_1 c_{j-1} - c_1 c_j - c_j, \quad j = i, \\
v' &= -\frac{1}{c_1}\Bigg[c_1^4 - c_1^2 c_2 + \alpha v - v^2 - 2\alpha c_1^2 \\
&\quad + 2c_1^2 v + 2\alpha c_2 - 2c_2 v + \alpha \sum_{k=3}^{i} c_k - v \sum_{k=3}^{i} c_k \Bigg].
\end{aligned}
\tag{5.18}
$$

We now change time from T to τ:

$$
\tau = \int_0^T \frac{1}{c_1(s)} ds.
\tag{5.19}
$$

This change of variable (also used in [30, p. 377] and [29, (3.3)]) is needed to desingularise the v equation when $c_1 = 0$. Note that by the result of Theorem 5.3.5, $\tau \to \infty$ as $T \to \infty$.

Letting dots represent the differentiation with respect to τ and defining $c_j(\tau) :=$

c_j, (5.18) becomes:

$$\dot{c}_1 = c_1\left(v - 2c_1^2 + 2c_2 + \sum_{k=3}^{i} c_k\right),$$

$$\dot{c}_j = c_1(c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}), \quad 1 < j < i,$$

$$\dot{c}_j = c_1(c_1 c_{j-1} - c_1 c_j - c_j), \quad j = i,$$

$$\dot{v} = -\left[c_1^4 - c_1^2 c_2 + \alpha v - v^2 - 2\alpha c_1^2\right.$$

$$\left. + 2c_1^2 v + 2\alpha c_2 - 2c_2 v + \alpha \sum_{k=3}^{i} c_k - v \sum_{k=3}^{i} c_k\right].$$

(5.20)

Note that $0 \in \mathbb{R}^{i+1}$ is a rest point of the system of equations (5.20), the rest point $(0,\ldots,0,\alpha)$ is generated by compactification and is an artefact of the change of variables. The object of interest is to establish stability properties of the rest point 0 and the way in which it is approached.

At this stage it is useful to make another change of variables. We set

$$w = v + 2c_2 + \sum_{k=3}^{i} c_k.$$

(5.21)

Hence in the $(c_1, c_2, \ldots, c_i, w)$ variables (5.20) becomes:

$$\dot{c}_1 = c_1(w - 2c_1^2),$$

$$\dot{c}_j = c_1(c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}), \quad 1 < j < i,$$

$$\dot{c}_j = c_1(c_1 c_{j-1} - c_1 c_j - c_j), \quad j = i,$$

(5.22)

and

$$\dot{w} = \dot{v} + 2\dot{c}_2 + \sum_{k=3}^{i} \dot{c}_k. \tag{5.23}$$

Substituting the equations from (5.22) into (5.23), we have

$$\dot{w} = \dot{v} + 2c_1(c_1^2 - c_1c_2 - c_2 + c_3) + (c_1^2c_2 - c_1c_3 - c_1^2c_i)$$
$$= \dot{v} + 2c_1^3 - 2c_1c_2 - c_1^2c_i - c_1^2c_2 + c_1c_3.$$

Using the equation of $\dot{v}$ from (5.20), this becomes

$$\dot{w} = -c_1^4 - \alpha v + v^2 + 2\alpha c_1^2 - 2c_1^2 v - 2\alpha c_2 + 2c_2 v - \alpha \sum_{k=3}^{i} c_k$$

$$+ v \sum_{k=3}^{i} c_k + 2c_1^3 - 2c_1c_2 - c_1^2c_i + c_1c_3.$$

Solving for v from (5.21), we obtain

$$\dot{w} = -2c_2 w - w \sum_{k=3}^{i} c_k + 2c_1^2 \sum_{k=3}^{i} c_k - \alpha w + w^2 - 2c_1^2 w + c_1c_3$$

$$- c_1^2c_i - 2c_1c_2 + 2c_1^3 - c_1^4 + 4c_1^2c_2 + 2\alpha c_1^2.$$

Thus, using the above computation for $\dot{w}$, we have the equations

$$\dot{c}_1 = c_1(w - 2c_1^2),$$

$$\dot{c}_j = c_1(c_1c_{j-1} - c_1c_j - c_j + c_{j+1}), \quad 1 < j < i, \tag{5.24}$$

$$\dot{c}_j = c_1(c_1c_{j-1} - c_1c_j - c_j), \quad j = i$$

and

$$\dot{w} = -\, 2c_2 w - w \sum_{k=3}^{i} c_k + 2c_1^2 \sum_{k=3}^{i} c_k - \alpha w + w^2 - 2c_1^2 w + c_1 c_3$$
$$-\, c_1^2 c_i - 2c_1 c_2 + 2c_1^3 - c_1^4 + 4c_1^2 c_2 + 2\alpha c_1^2. \tag{5.25}$$

Now we appeal to centre manifold theory [25] (see Section 2.3). In the language of that theory, for the equations (5.24), (5.25), the variables c_j, $1 \le j \le i$ are "centre" variables while w is a "stable" variable. Therefore, according to centre manifold theory, in a neighbourhood of the origin in $\mathbb{R}^{i+1}$, equations (5.24), (5.25) admit an i dimensional centre manifold, $w = h(c_1, c_2, \ldots, c_i)$. Furthermore, from Theorem 5.3.5 it follows that the centre manifold attracts all solutions in a neighbourhood of the origin in $\mathbb{R}^{i+1}$.

On this centre manifold, the flow is given by

$$\begin{aligned}
\dot{c}_1 &= c_1(h(c_1, c_2, \ldots, c_i) - 2c_1^2), \\
\dot{c}_j &= c_1(c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}), \quad 1 < j < i, \\
\dot{c}_i &= c_1(c_1 c_{i-1} - c_1 c_i - c_i).
\end{aligned} \tag{5.26}$$

Remarkably, we can reparameterise time by going back to the T variable to obtain on the centre manifold $w = h(c_1, c_2, \ldots, c_i)$ the equations

$$\begin{aligned}
c_1' &= h(c_1, c_2, \ldots, c_i) - 2c_1^2, \\
c_j' &= c_1 c_{j-1} - c_1 c_j - c_j + c_{j+1}, \quad 1 < j < i, \\
c_i' &= c_1 c_{i-1} - c_1 c_i - c_i.
\end{aligned} \tag{5.27}$$

Since according to centre manifold theory the asymptotic expansion of $h(c_1, c_2, \ldots, c_i)$ contains only quadratic terms and above in c_j, $j \ge 1$, the $i \times i$ Jacobian matrix

Chapter 5. Point Island Dynamics Under Fixed Rate Deposition

$J(0)$ of equations (5.27) around the origin in $\mathbb{R}^{i+1}$ has the following structure:

$$J(0) = \left[\begin{array}{c|c} 0 & 0 \\ \hline 0 & A \end{array}\right],$$

with the first row being made of zeros and the $(i-1)\times(i-1)$ bi-diagonal matrix A having -1 on the main diagonal and 1 in the $(j, j+1)$ positions, $2 \le j \le i-1$. It is easily seen that all eigenvalues of A are negative. Such structure of the Jacobian matrix means that, for the equations of the flow on the centre manifold $w = h(c_1, c_2, \ldots, c_i)$, c_j, $2 \le j \le i$ are "stable" variables and c_1 is a "centre" variable, so that inside the i dimensional centre manifold there is another, one dimensional centre manifold parameterised by c_1, i.e a curve with components $c_j = g_j(c_1)$, $1 < j \le i$. We will write $g_w(c_1) = h(c_1, g_2(c_1), \ldots, g_i(c_1))$. Furthermore, we also know by centre manifold theory that as $c_1 \to 0$,

$$g_j(c_1) \sim \sum_{k=2}^{\infty} \gamma_{j,k} c_1^k, \tag{5.28}$$

We also have

$$g_w(c_1) \sim \sum_{k=2}^{\infty} \gamma_{w,k} c_1^k. \tag{5.29}$$

Hence (see [25]) the flow on the one dimensional centre manifold is given by

$$c_1' = g_w(c_1) - 2c_1^2, \tag{5.30}$$

and as the rest point at the origin of the one dimensional ODE (5.30) is asymptotically stable by Theorem 5.3.5, the one dimensional centre manifold $(g_2(c_1), \ldots, g_i(c_1))$ attracts nearby solutions, so all solutions approach the origin along this curve (apart possibly from sets of zero $i + 1$ dimensional Lebesgue measure).

We have

Theorem 5.4.1. c_1 *asymptotically satisfies the differential equation*

$$c_1' = \frac{1}{\alpha}(-c_1^{i+3} + c_1^{i+4} - c_1^{2i+3}) + O\left(\frac{c_1^{2i+4}}{\alpha^2}\right), \quad t \to \infty.$$

Proof. As we are interested in the asymptotics of (5.28)-(5.29) we can differentiate term by term. Hence, differentiating (5.28)-(5.29) and from the definition of c_1' from (5.27)

$$
\begin{aligned}
w' = g_w'(c_1) &\sim c_1' \sum_{k=2}^{\infty} k\gamma_{w,k} c_1^{k-1} \\
&= \sum_{k=2}^{\infty} k\gamma_{w,k} c_1^{k-1}(w - 2c_1^2) \\
&= \sum_{k=2}^{\infty} k\gamma_{w,k} c_1^{k-1}\left(\sum_{k=2}^{\infty} \gamma_{w,k} c_1^k - 2c_1^2\right).
\end{aligned}
\tag{5.31}
$$

$$
\begin{aligned}
c_j' = g_j'(c_1) &\sim c_1' \sum_{k=2}^{\infty} k\gamma_{j,k} c_1^{k-1} \\
&= \sum_{k=2}^{\infty} k\gamma_{j,k} c_1^{k-1}(w - 2c_1^2) \\
&= \sum_{k=2}^{\infty} k\gamma_{j,k} c_1^{k-1}\left(\sum_{k=2}^{\infty} \gamma_{w,k} c_1^k - 2c_1^2\right),
\end{aligned}
\tag{5.32}
$$

when $1 < j \leq i$.

Alternately we can find w' from the equation (5.25) by changing time from τ to

T from (5.19) then

$$w' = g'_w(c_1) \sim \frac{1}{c_1}\Bigg[g_w(c_1)(-2g_2(c_1) - \sum_{k=3}^{i} g_k(c_1) -$$
$$- \alpha + g_w(c_1) - 2c_1^2) +$$
$$+ 2c_1^2 \sum_{k=3}^{i} g_k(c_1) + c_1 g_3(c_1) - c_1^2 g_i(c_1) -$$
$$- 2c_1 g_2(c_1) + 2c_1^3 - c_1^4 + 4c_1^2 g_2(c_1) + 2\alpha c_1^2 \Bigg].$$

We have

$$g'_w(c_1) \sim \frac{1}{c_1}\Bigg[\sum_{k=2}^{\infty} \gamma_{w,k} c_1^k \Big(-2g_2(c_1) - \sum_{k=3}^{i} g_k(c_1) -$$
$$- \alpha + \sum_{k=2}^{\infty} \gamma_{w,k} g_k(c_1) - 2c_1^2 \Big)$$
$$+ 2c_1^2 \sum_{k=3}^{i} g_k(c_1) + c_1 g_3(c_1) - c_1^2 g_i(c_1) -$$
$$- 2c_1 g_2(c_1) + 2c_1^3 - c_1^4 + 4c_1^2 g_2(c_1) + 2\alpha c_1^2 \Bigg]$$

and then using (5.28), we obtain

$$g'_w(c_1) \sim \frac{1}{c_1}\Bigg(\sum_{k=2}^{\infty}\bigg[\gamma_{w,k} c_1^k \Big(\sum_{m=2}^{\infty}(-2\gamma_{2,m} - \sum_{n=3}^{i}\gamma_{n,m} + \gamma_{w,m})c_1^m - 2c_1^2 - \alpha \Big)$$
$$+ (\gamma_{3,k} - 2\gamma_{2,k})c_1^{k+1} + (4\gamma_{2,k} + 2\sum_{n=3}^{i}\gamma_{n,k} - \gamma_{i,k})c_1^{k+2} \bigg] \qquad (5.33)$$
$$+ 2\alpha c_1^2 - c_1^4 + 2c_1^3 \Bigg).$$

In addition, we can find c'_j, $1 < j \leq i$ by substituting equations (5.28) into

equations (5.27):

$$
g_j'(c_1) \sim -c_1 \sum_{k=2}^{\infty} \gamma_{j,k} c_1^k - \sum_{k=2}^{\infty} \gamma_{j,k} c_1^k +
\begin{cases}
c_1^2 + \sum_{k=2}^{\infty} \gamma_{3,k} c_1^k, & j = 2, \\[2ex]
c_1 \sum_{k=2}^{\infty} \gamma_{j-1,k} c_1^k + \\
\quad + \sum_{k=2}^{\infty} \gamma_{j+1,k} c_1^k, & 2 < j < i, \\[2ex]
c_1 \sum_{k=2}^{\infty} \gamma_{j-1,k} c_1^k, & j = i.
\end{cases}
\tag{5.34}
$$

We are now in a position to be able to find the coefficients.

From (5.33) with (5.31) we obtain

$$
\begin{aligned}
0 \equiv \sum_{k=2}^{\infty} \Bigg[& \gamma_{w,k} c_1^k \Bigg(\sum_{m=2}^{\infty} \big(-2\gamma_{2,m} - \sum_{n=3}^{i} \gamma_{n,m} + \\
& + \gamma_{w,m}(1-k) \big) c_1^m + 2c_1^2(k-1) - \alpha \Bigg) \\
& + (\gamma_{3,k} - 2\gamma_{2,k}) c_1^{k+1} + \Big(4\gamma_{2,k} + 2\sum_{n=3}^{i} \gamma_{n,k} - \gamma_{i,k} \Big) c_1^{k+2} \Bigg] \\
& + 2\alpha c_1^2 - c_1^4 + 2c_1^3.
\end{aligned}
\tag{5.35}
$$

In addition, from (5.34) and (5.32) there are three cases:

Case 1: $j = 2$

$$
\begin{aligned}
0 \equiv c_1^3 + \sum_{k=2}^{\infty} \Bigg(& \gamma_{2,k}(2k-1) c_1^{k+2} + \\
& + (\gamma_{3,k} - \gamma_{2,k}) c_1^{k+1} - k\gamma_{2,k} c_1^k \sum_{n=2}^{\infty} \gamma_{w,n} c_1^n \Bigg).
\end{aligned}
\tag{5.36}
$$

Case 2: $2 < j < i$

$$
\begin{aligned}
0 \equiv \sum_{k=2}^{\infty} \Bigg(& (\gamma_{u-1,k} + \gamma_{u,k}(2k-1))c_1^{k+2} \\
& + (\gamma_{u+1,k} - \gamma_{u,k})c_1^{k+1} - k\gamma_{u,k}c_1^{k} \sum_{n=2}^{\infty} \gamma_{w,n}c_1^{n} \Bigg).
\end{aligned}
\tag{5.37}
$$

Case 3: $j = i$

$$
\begin{aligned}
0 \equiv \sum_{k=2}^{\infty} \Bigg(& (\gamma_{i-1,k} + \gamma_{i,k}(2k-1))c_1^{k+2} - \\
& - \gamma_{i,k}c_1^{k+1} - k\gamma_{i,k}c_1^{k} \sum_{n=2}^{\infty} \gamma_{w,n}c_1^{n} \Bigg).
\end{aligned}
\tag{5.38}
$$

Hence, we can now calculate $\gamma_{j,k}$ and $\gamma_{w,k}$, from (5.35)-(5.38) as expressed in the following propositions.

Proposition 5.4.1.

$$
\begin{aligned}
& \gamma_{j,2} = 0, \quad 2 < j \leq i, \\
& \gamma_{2,2} = 1, \gamma_{w,2} = 2, \gamma_{w,3} = 0.
\end{aligned}
\tag{5.39}
$$

Proof. From (5.38)

$$
-\gamma_{i,2}c_1^3 + O(c_1^4) \equiv 0 \implies \gamma_{i,2} = 0.
\tag{5.40}
$$

From (5.37) when $2 < j < i$.

$$
(\gamma_{j+1,2} - \gamma_{j,2})c_1^3 + O(c_1^4) \equiv 0 \implies \gamma_{j,2} = 0, \text{ from (5.40)}.
\tag{5.41}
$$

(5.36) leads to

$$(1 + \gamma_{3,2} - \gamma_{2,2})c_1^3 + O(c_1^4) \equiv 0 \implies \gamma_{2,2} = 1, \text{ from (5.41)}. \tag{5.42}$$

From (5.35) we obtain

$$(-\alpha\gamma_{w,2} + 2\alpha)c_1^2 + O(c_1^3) \equiv 0 \implies \gamma_{w,2} = 2,$$

$$O(c_1^2) + (-\alpha\gamma_{w,3} + \gamma_{3,2} - 2\gamma_{2,2} + 2)c_1^3 + O(c_1^4) \equiv 0$$

$$\implies \gamma_{w,3} = 0, \text{ from (5.40)-(5.42)}.$$

$\square$

Proposition 5.4.2. *For $1 < j \leq i$ we have the following recurrences:*

$$\gamma_{i,k} = \gamma_{i-1,k-1} + (2k-1)\gamma_{i,k-1} - \sum_{n=2}^{k} n\gamma_{i,n-1}\gamma_{w,k-n+1}, \quad k > 1.$$

$$\gamma_{j,k} = \gamma_{j-1,k-1} + (2k-1)\gamma_{j,k-1} + \gamma_{j+1,k} - \sum_{n=2}^{k} n\gamma_{j,n-1}\gamma_{w,k-n+1}, \quad k > 1.$$

$$\gamma_{2,k} = (2k-1)\gamma_{2,k-1} + \gamma_{3,k} - \sum_{n=2}^{k} n\gamma_{2,n-1}\gamma_{w,k-n+1}, \quad k > 1.$$

$$\gamma_{w,3} = \frac{1 - (\gamma_{3,2} - 2\gamma_{2,2})}{\alpha}.$$

$$\gamma_{w,4} = \frac{1}{\alpha}\left[\gamma_{3,3} - 2\gamma_{2,3} + 4\gamma_{2,3} + 2\sum_{n=3}^{i}\gamma_{n,2} - \gamma_{i,2} + 2\gamma_{w,2}+\right.$$

$$\left. + \sum_{m=2}^{3}\gamma_{w,m-1}\left(-2\gamma_{2,4-m} - \sum_{n=3}^{i}\gamma_{n,4-m} + \gamma_{w,4-m}\right) - 1\right].$$

$$\gamma_{w,k} = \frac{1}{\alpha}\left[\gamma_{3,k-1} - 2\gamma_{2,k-1} + 4\gamma_{2,k-2} + 2\sum_{n=3}^{i}\gamma_{n,k-2} - \gamma_{i,k-2} + 2\gamma_{w,k-2}+\right.$$

$$\left. + \sum_{m=2}^{k-1}\gamma_{w,m-1}\left(-2\gamma_{2,k-m} - \sum_{n=3}^{i}\gamma_{n,k-m} + \gamma_{w,k-m}\right)\right], \quad k > 4.$$

Hence, we have that:

Proposition 5.4.3.

$$c_j = g_j(c_1) = c_1^j - c_1^{i+1} + c_1^{i+j} + O\left(\frac{c_1^{i+j+2}}{\alpha}\right), \quad 1 < j < i,$$

$$c_j = g_j(c_1) = c_1^j - c_1^{j+1} + c_1^{2j} - c_1^{2j+1} + O\left(\frac{c_1^{2j+2}}{\alpha}\right), \quad j = i, \tag{5.43}$$

$$w = g_w(c_1) = 2c_1^2 - \frac{1}{\alpha}c_1^{i+3} + \frac{1}{\alpha}c_1^{i+4} - \frac{1}{\alpha}c_1^{2i+3} + O\left(\frac{c_1^{2i+4}}{\alpha^2}\right).$$

Thus, from Proposition 5.4.3 and (5.30) we have that

$$c_1' = \frac{1}{\alpha}(-c_1^{i+3} + c_1^{i+4} - c_1^{2i+3}) + O\left(\frac{c_1^{2i+4}}{\alpha^2}\right).$$

$\square$

Note that beyond terms of $O(c_1^{i+j})$ the interplay among $g_j(c_1)$, $1 < j \leq i$, and $g_w(c_1)$ becomes complex and that the later coefficients of these functions depend on α. Computations using the MAPLE code in the Appendix indicate that the radius of convergence of the expansions is 0 for all $\alpha > 0$.

5.5   Asymptotics of Solutions

Armed with Theorem 5.4.1 which holds for any non-negative solution of (5.2) by the globalisation results of Section 5.3, we can discuss asymptotics of solutions of (5.1) using the methods of [29, 30], which were also used in [31]. As proofs are similar to those used in the above papers, we indicate only the main ideas. Further terms in the expansions in this section can be computed using the machinery of [29], we illustrate this for the computation of c_1; in all other cases we only

determine the leading terms, denoting higher order terms by "h.o.t". Going back
to our original variables $C_j(t)$ to exhibit the complicated dependence of the results
on β, from Theorem 5.4.1 we have the following statement:

Theorem 5.5.1. *The asymptotics of C_1 are given by*

$$C_1 \sim \left(\frac{\hat{\alpha}\beta^{i-1}}{(i+2)t} \right)^{\frac{1}{i+2}} + \frac{1}{i+1} \left(\frac{\hat{\alpha}\beta^{\frac{i-5}{2}}}{(i+2)t} \right)^{\frac{2}{i+2}} + h.o.t.. \tag{5.44}$$

Proof. Following the method of [29] we begin by considering the approximation
to the centre manifold $g_w(c_1)$, (5.43), given by taking the first two non-zero terms.

$$g_w(c_1) \sim 2c_1^2 - \frac{1}{\alpha}c_1^{i+3} + O(c_1^{i+4}).$$

The flow governing the asymptotics of $c_1(t)$ from (5.30) is

$$c_1' = -\frac{1}{\alpha}c_1^{i+3} + O(c_1^{i+4}),$$

which, as the equation is separable, leads us to

$$\frac{d}{dT} \left(\frac{\alpha}{(i+2)c_1^{i+2}} \right) = 1 + O(c_1^2). \tag{5.45}$$

Using that $c_1 \to 0$ as $T \to \infty$, this implies that $1 + O(c_1^2) \to 1$ as $T \to \infty$
.Hence, for all $\epsilon > 0$, there exists a $\zeta > T_0$ such that, for all $T > \zeta$, the following
inequalities hold

$$1 - \epsilon \le \frac{d}{dT} \left(\frac{\alpha}{(i+2)c_1^{i+2}} \right) \le 1 + \epsilon. \tag{5.46}$$

Integrating (5.46) between ζ and T

$$(1-\epsilon)(T-\zeta) \le \frac{\alpha}{(i+2)c_1^{i+2}} - \frac{\alpha}{(i+2)c_{1\zeta}^{i+2}} \le (1+\epsilon)(T-\zeta), \tag{5.47}$$

where $c_{1\zeta} = c_1(\zeta)$. Dividing (5.47) by T and taking $\liminf_{T\to\infty}$ and $\limsup_{T\to\infty}$

leads us to

$$1 - \epsilon \leq \liminf_{T \to \infty} \frac{\alpha}{(i+2)T{c_1}^{i+2}} \leq \limsup_{T \to \infty} \frac{\alpha}{(i+2)T{c_1}^{i+2}} \leq 1 + \epsilon,$$

which, as ϵ is arbitrary implies that $\liminf_{T \to \infty} \frac{i+2}{\alpha} T {c_1}^{i+2} = 1$, and thus

$$c_1^{i+2}(T) = \frac{\alpha}{i+2} \frac{1}{T}(1 + o(1)). \tag{5.48}$$

We now consider a better approximation to $g_w(c_1)$ by taking the first three non-zero terms

$$g_w(c_1) = 2c_1^2 - \frac{1}{\alpha}c_1^{i+3} + \frac{1}{\alpha}c_1^{i+4} + O(c_1^{2i+3}).$$

The flow governing the asymptotics of c_1 is now given by

$$c_1' = -\frac{1}{\alpha}c_1^{i+3} + \frac{1}{\alpha}c_1^{i+4} + O(c_1^{2i+3}). \tag{5.49}$$

Writing this differential equation in a separable form

$$\frac{\alpha c_1'}{-c_1^{i+3} + c_1^{i+4}} = 1 + O(c_1^{i+2}).$$

Using partial fractions, we have

$$\frac{\alpha}{-s^{i+3} + s^{i+4}} = -\frac{\alpha}{s^{i+3}} - \frac{\alpha}{s^{i+2}} - \alpha \left[\sum_{k=1}^{i+1} \frac{1}{s^k} - \frac{1}{1-s} \right].$$

For convenience we set

$$\psi_i(s) := \int -\alpha \left[\sum_{k=1}^{i+1} \frac{1}{s^k} - \frac{1}{1-s} \right] ds.$$

Hence, (5.49) can be written as

$$\frac{d}{dT}\left(\frac{\alpha}{(i+2)c_1{}^{i+2}} + \frac{\alpha}{(i+1)c_1{}^{i+1}} + \psi_i(c_1)\right) = 1 + c_1{}^{i+2}O(1). \tag{5.50}$$

We now aim to estimate $1 + c_1{}^{i+2}O(1)$. We use the fact that $c_1 \to 0$ as $T \to \infty$, and the information from (5.48). There exist constants $K^* \geq K_*$ such that $1 + c_1{}^{i+3}O(1)$ can be bounded by

$$1 + K_* c_1{}^{i+2} \leq 1 + c_1{}^{i+2}O(1) \leq 1 + K^* c_1{}^{i+2}.$$

Hence,

$$1 + K_* c_1{}^{i+2} \leq \frac{d}{dT}\left(\frac{\alpha}{(i+2)c_1{}^{i+2}} + \right. $$
$$\left. + \frac{\alpha}{(i+1)c_1{}^{i+1}} + \psi_i(c_1)\right) \leq 1 + K^* c_1{}^{i+2}, \quad \forall T > \zeta. \tag{5.51}$$

Integrating (5.51) between ζ and $T > \zeta$, we obtain

$$T - \zeta + K_* \int_\zeta^T c_1{}^{i+2}(s)ds \leq$$
$$\leq \frac{\alpha}{(i+2)c_1{}^{i+2}} + \frac{\alpha}{(i+1)c_1{}^{i+1}} + \psi_i(c_1) -$$
$$\left(\frac{\alpha}{(i+2)c_{1\zeta}{}^{i+2}} + \frac{\alpha}{(i+1)c_{1\zeta}{}^{i+1}} + \psi_i(c_{1\zeta})\right) \tag{5.52}$$

and

$$T - \zeta + K^* \int_\zeta^T c_1{}^{i+2}(s)ds \geq$$
$$\geq \frac{\alpha}{(i+2)c_1{}^{i+2}} + \frac{\alpha}{(i+1)c_1{}^{i+1}} + \psi_i(c_1) -$$
$$\left(\frac{\alpha}{(i+2)c_{1\zeta}{}^{i+2}} + \frac{\alpha}{(i+1)c_{1\zeta}{}^{i+1}} + \psi_i(c_{1\zeta})\right). \tag{5.53}$$

We now use (5.48) to estimate the integral of c_1^{i+2}: let $\epsilon > 0$ be a fixed arbitrary number and redefine ζ such that $\frac{(i+2)}{\alpha}Tc_1^{i+2} \in [1 - \epsilon, 1 + \epsilon]$, for all $T > \zeta$. Thus

$$\frac{\alpha}{i+2}(1 - \epsilon)(\log T - \log \zeta) \leq \int_\zeta^T c_1^{i+2}(s)ds \leq \frac{\alpha}{i+2}(1 + \epsilon)(\log T - \log \zeta).$$

Noting that $\lim_{T\to\infty} \frac{\psi_i(c_1)}{T} = 0$ we can now estimate (5.52)-(5.53) for large values of T. Dividing (5.52) by T and taking infinite as $\liminf$ as $T \to \infty$, we have

$$
\begin{aligned}
&\liminf_{T\to\infty} \left(\frac{\alpha}{(i+2)Tc_1^{i+2}} + \frac{\alpha}{(i+1)Tc_1^{i+1}} \right) \\
&\geq \liminf_{T\to\infty} \frac{1}{T}\left(T - \zeta + \frac{K_*\alpha}{i+2}(1 - \epsilon)(\log T - \log \zeta) - \psi_i(c_1) \right) \\
&= 1 + \lim_{T\to\infty}\left(\frac{K_*\alpha}{i+2}(1 - \epsilon)\frac{\log T}{T} - \frac{1}{T}\left(\zeta + \frac{K_*\alpha}{i+2}(1 - \epsilon)\log \zeta \right) - \frac{\psi_i(c_1)}{T} \right) \\
&= 1.
\end{aligned}
\tag{5.54}
$$

Similarly, dividing (5.53) by T and taking taking $\limsup$ as $T \to \infty$, gives us

$$
\begin{aligned}
&\limsup_{T\to\infty} \left(\frac{\alpha}{(i+2)Tc_1^{i+2}} + \frac{\alpha}{(i+1)Tc_1^{i+1}} \right) \\
&\leq 1 + \lim_{T\to\infty}\left(\frac{K^*\alpha}{i+2}(1 + \epsilon)\frac{\log T}{T} - \frac{1}{T}\left(T + \frac{K^*\alpha}{i+2}(1 + \epsilon)\log \zeta \right) - \frac{\psi_i(c_1)}{T} \right) \\
&= 1.
\end{aligned}
$$

$$\tag{5.55}$$

Thus, (5.54)-(5.55) imply there exists a function F_i such that $F_i(T) = O(1)$ for large T, and

$$\left(\frac{\alpha}{(i+2)Tc_1^{i+2}} + \frac{\alpha}{(i+1)Tc_1^{i+1}} \right)\frac{1}{T} \sim 1 + \frac{\log T}{T}F_i(T). \tag{5.56}$$

Chapter 5. Point Island Dynamics Under Fixed Rate Deposition

Setting, $B := \frac{1}{i+2}$ and $A := \frac{1}{i+1}$, we have

$$\alpha B + \alpha A c_1 \sim T c_1{}^{i+2}\left(1 + \frac{\log \tau}{\tau} F_i(\tau)\right). \tag{5.57}$$

Since,

$$\left(1 + \frac{\log T}{T} F_i(T)\right)^{-1} = 1 - \frac{\log T}{T} F_i(T) + O\left(\left(\frac{\log T}{T}\right)^2\right).$$

We can express (5.57) as

$$c_1{}^{\frac{1}{B}} = \left(\frac{\alpha B}{T}\right)\left(1 + \frac{A}{B} c_1\right)\left(1 - \frac{\log T}{T} F_i(T) + h.o.t\right). \tag{5.58}$$

From (5.48) we have that

$$c_1 \sim \left(\frac{\alpha B}{T}\right)^{B},$$

hence, (5.58) becomes

$$c_1{}^{\frac{1}{B}} = \left(\frac{\alpha B}{T}\right)\left(1 + \frac{A}{B}\left(\frac{\alpha B}{T}\right)^{B}\right)\left(1 - \frac{\log T}{T} F_i(T) + h.o.t\right). \tag{5.59}$$

Using the binomial expansion we can express (5.59) as

$$c_1 \sim \left(\frac{\alpha B}{T}\right)^{B}\left(1 + A\left(\frac{\alpha B}{T}\right)^{B} - B\frac{\log T}{T} F_n(T) + h.o.t\right). \tag{5.60}$$

Hence, from the definition of A and B, (5.60) can be expressed as

$$c_1 \sim \left(\frac{\alpha}{(i+2)T}\right)^{\frac{1}{i+2}} + \frac{1}{i+1}\left(\frac{\alpha}{(i+2)T}\right)^{\frac{2}{i+2}} + h.o.t. \tag{5.61}$$

Going back to our time variable t we obtain the required result. □

Note that if we set $\beta = 1$ in the equation above, from Theorem 5.5.1, we obtain the same result as in [31], to leading order. Already at the level of $C_1(t)$ one sees that the influence of the fragmentation rate β is not intuitive. Once we know the

asymptotics of $C_1(t)$ from Theorem 5.5.1, the asymptotics of $C_j(t)$ when $1 \leq j \leq i$ follow from Proposition 5.43.

Lemma 5.5.2. *The asymptotics of C_j when $1 \leq j \leq i$ are given by*

$$C_j \sim \left(\frac{\hat{\alpha} \beta^{\frac{ij-3j+2}{j}}}{(i+2)t} \right)^{\frac{j}{i+2}} + h.o.t, \quad 1 < j \leq i.$$

Hence, we are now in a position to express the asymptotics of $C_j(t)$ when $j > i$ by solving linear nonhomogeneous ODEs using the same change of variable as in the proof of Theorem 5.3.1.

Lemma 5.5.3. *The asymptotics of C_j when $j > i$ are given by*

$$C_j \sim \left(\frac{\hat{\alpha} \beta^{\frac{i^2-3i+2}{i}}}{(i+2)t} \right)^{\frac{i}{i+2}} + h.o.t, \quad j > i.$$

From this information we have the equivalent of [30, Theorem 5.1] and, to which it is more directly comparable, [31, Theorem 6], concerning similarity solutions of (5.1). These references should be consulted for the required computations. To formulate the theorem, we first compute the asymptotics of the average cluster size j using the information in Theorem 5.5.1 and Lemmas 5.5.2–5.5.3:

$$\langle j \rangle = \frac{\sum_{j=1}^{\infty} j C_j}{\sum_{j=1}^{\infty} C_j} \sim \left(\frac{\hat{\alpha} \beta^{i-1}}{i+2} \right)^{\frac{1}{i+2}} t^{\frac{i+1}{i+2}} + h.o.t.$$

Next, we define the function ϕ by

$$\phi(\eta) = \begin{cases} (1-\eta)^{-\frac{i}{i+1}}, & \eta < 1, \\\\ 0, & \text{otherwise.} \end{cases}$$

Finally, we define the similarity variable η by

$$\eta = \frac{(i+1)\beta^{-\frac{i+1}{i+2}}}{i+2}\frac{j}{\langle j \rangle}.$$

Then we have that the solutions of (5.1) converge to a (discontinuous) similarity profile:

Theorem 5.5.4.

$$C_j = \langle j \rangle^{-\frac{1}{i+1}}\phi\big(\eta\big), \quad t \to \infty.$$

The profile obtained in this theorem can be further analysed by the methods of [31, Section 6].

5.6 Quasi-Steady State Assumption

In this section we will to investigate whether the asymptotics of solutions obtained in Section 5.5, based on the centre manifold analysis of Section 5.4, can be recovered more easily by combining centre manifold reasoning with a technique that is often used in the engineering community: the quasi-steady state approximation (QSSA; see [44, 77, 88]). As in the famous example from enzyme kinetics due to Segel and Slemrod [88], we show that QSSA correctly captures the leading term asymptotics, though, of course, there will be differences in higher order terms.

Chapter 5. Point Island Dynamics Under Fixed Rate Deposition

We restart with equations

$$\hat{c}_1' = \alpha - 2\hat{c}_1^2 + 2\hat{c}_2 - \hat{c}_1 \sum_{k=2}^{\infty} \hat{c}_k + \sum_{k=3}^{i} \hat{c}_k,$$

$$\hat{c}_j' = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j - \hat{c}_j + \hat{c}_{j+1}, \quad 1 < j < i,$$

$$\hat{c}_j' = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j - \hat{c}_j, \quad j = i,$$

$$\hat{c}_j' = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j, \quad j > i,$$

(5.62)

but now we immediately make the QSSA assumption that $\hat{c}_j' = 0$ for $1 < j \leq i$. We solve the i algebraic equations for $\hat{c}_j$, $1 < j \leq i$, in terms of $\hat{c}_1$. This clearly can be done consecutively, by starting with the $\hat{c}_i$ equation and solving it in terms of $\hat{c}_1$ and $\hat{c}_{i-1}$, substituting the expression we get for $\hat{c}_i$ into the $\hat{c}_{i-1}$ equation and continuing in this way, till $\hat{c}_2$ has been solved in terms of $\hat{c}_1$, after which we back-substitute.

For connivance, we will use Gauss's continued fraction notation throughout this section:

$$b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{\ddots + \cfrac{a_n}{b_n}}}} := b_0 + K_{k=1}^{n} \frac{a_k}{b_k}$$

Lemma 5.6.1. *If $\hat{c}_j' = 0$ for $1 < j \leq i$ in (5.62) then*

$$\hat{c}_j \sim \hat{c}_1^j + \sum_{k=1}^{\infty} -\hat{c}_1^{ki+1} + \hat{c}_1^{ki+j}, \quad 1 < j \leq i.$$

(5.63)

Chapter 5. Point Island Dynamics Under Fixed Rate Deposition

Proof. We have

$$\hat{c}_1' = \alpha - 2\hat{c}_1^2 + 2\hat{c}_2 - \hat{c}_1 \sum_{k=2}^{\infty} \hat{c}_k + \sum_{k=3}^{i} \hat{c}_k,$$

$$0 = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j - \hat{c}_j + \hat{c}_{j+1}, \quad 1 < j < i,$$

$$0 = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j - \hat{c}_j, \quad j = i,$$

$$\hat{c}_j' = \hat{c}_1\hat{c}_{j-1} - \hat{c}_1\hat{c}_j, \quad j > i,$$

$$(5.64)$$

Solving the right-hand side of the $\hat{c}_i$ equation from (5.64) for $\hat{c}_i$ in terms of $\hat{c}_1$ and $\hat{c}_{i-1}$,

$$\hat{c}_i = \frac{\hat{c}_1\hat{c}_{i-1}}{\hat{c}_1 + 1}. \tag{5.65}$$

Substituting $\hat{c}_i$ from (5.65) into the rhs of the $\hat{c}_{i-1}$ equation from (5.64) and solving for $\hat{c}_{i-1}$ in terms of $\hat{c}_1$ and $\hat{c}_{i-2}$.

$$\hat{c}_{i-1} = \frac{\hat{c}_1\hat{c}_{i-2}}{\hat{c}_1 + 1 - \frac{\hat{c}_1}{\hat{c}_1+1}}. \tag{5.66}$$

Continuing this process until we reach the $\hat{c}_2$ equation we find, using Gauss's continued fraction notation,

$$\hat{c}_2 = \frac{\hat{c}_1^2}{\mathrm{K}_{u=1}^{i-1} \frac{\hat{c}_1+1}{-\hat{c}_1}}. \tag{5.67}$$

Hence by induction,

$$\hat{c}_2 = \frac{\hat{c}_1^2}{\mathrm{K}_{u=1}^{i-1} \frac{\hat{c}_1+1}{-\hat{c}_1}} = \frac{\sum_{k=1}^{i-1} \hat{c}_1^{k+1}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}. \tag{5.68}$$

Substituting the expression $\hat{c}_2$ from (5.68) into the equation for $\hat{c}_3$ from (5.64)

94

when $i > 3$,

$$
\begin{aligned}
\hat{c}_3 &= \frac{\sum_{k=1}^{i-1} \hat{c}_1^{k+1}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}(1 + \hat{c}_1) - \hat{c}_1^2 \\
&= \frac{\sum_{k=1}^{i-1} \hat{c}_1^{k+1}(1 + \hat{c}_1) - \sum_{k=1}^{i} \hat{c}_1^{k+1}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}} \\
&= \frac{\sum_{k=1}^{i-2} \hat{c}_1^{k+2}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}.
\end{aligned}
\tag{5.69}
$$

Similarly, substituting the expression $\hat{c}_2$ from (5.68) into the equation for $\hat{c}_3$ from (5.64) when $i = 3$

$$
\begin{aligned}
\hat{c}_3 &= \frac{\sum_{k=1}^{i-1} \hat{c}_1^{k+2}}{(1 + \hat{c}_1)\sum_{k=1}^{i} \hat{c}_1^{k-1}} \\
&= \frac{\hat{c}_1^3}{\sum_{k=1}^{i} \hat{c}_1^{k-1}} \\
&= \frac{\sum_{k=1}^{i-2} \hat{c}_1^{k+2}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}.
\end{aligned}
\tag{5.70}
$$

Continuing this process by substituting the expression $\hat{c}_3$ from (5.69) into the equation for $\hat{c}_4$ from (5.64) when $i > 4$,

$$
\begin{aligned}
\hat{c}_4 &= \frac{\sum_{k=1}^{i-2} \hat{c}_1^{k+2}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}(1 + \hat{c}_1) - \frac{\sum_{k=1}^{i-1} \hat{c}_1^{k+2}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}} \\
&= \frac{\sum_{k=1}^{i-3} \hat{c}_1^{k+3}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}.
\end{aligned}
\tag{5.71}
$$

Similarly, when $i = 4$

$$
\hat{c}_4 = \frac{\sum_{k=1}^{i-3} \hat{c}_1^{k+3}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}.
\tag{5.72}
$$

Continuing by substituting the expressions into the equations we find the general

95

form

$$\hat{c}_j = \frac{\sum_{k=1}^{i-j+1} \hat{c}_1^{k+j-1}}{\sum_{k=1}^{i} \hat{c}_1^{k-1}}, \quad 1 < j \le i. \tag{5.73}$$

Expanding equation (5.73) into a Maclaurin series gives

$$\hat{c}_j = \hat{c}_1^j + \sum_{k=1}^{\infty} -\hat{c}_1^{ki+1} + \hat{c}_1^{ki+j}, \quad 1 < j \le i. \tag{5.74}$$

$\square$

Lemma 5.6.2. *The QSSA leads to an approximation of the centre manifold* $g_w(c_1)$ *as*

$$g_w(\hat{c}_1) \sim 2\hat{c}_1^2 + \frac{1}{\alpha} \sum_{k=1}^{\infty} -\hat{c}_1^{ki+3} + \hat{c}_1^{ki+4}$$

which is consistent with the centre manifold computation from Theorem 5.4.1 for the first four non-zero terms.

Proof. From the recurrences from Proposition 5.4.2

$$\gamma_{i,1} = 0, \gamma_{j,1} = 0,$$

$$\gamma_{w,3} = \frac{1 - (\gamma_{3,2} - 2\gamma_{2,2})}{\alpha},$$

$$\gamma_{w,4} = \frac{1}{\alpha}\left[\gamma_{3,3} - 2\gamma_{2,3} + 4\gamma_{2,3} + 2\sum_{n=3}^{i}\gamma_{n,2} - \gamma_{i,2} + 2\gamma_{w,2} + \right.$$

$$\left. + \sum_{m=2}^{3}\gamma_{w,m-1}\left(-2\gamma_{2,4-m} - \sum_{n=3}^{i}\gamma_{n,4-m} + \gamma_{w,4-m}\right) - 1\right],$$

$$\gamma_{w,k} = \frac{1}{\alpha}\left[\gamma_{3,k-1} - 2\gamma_{2,k-1} + 4\gamma_{2,k-2} + 2\sum_{n=3}^{i}\gamma_{n,k-2} - \gamma_{i,k-2} + 2\gamma_{w,k-2} + \right.$$

$$\left. + \sum_{m=2}^{k-1}\gamma_{w,m-1}\left(-2\gamma_{2,k-m} - \sum_{n=3}^{i}\gamma_{n,k-m} + \gamma_{w,k-m}\right)\right], \quad k > 4,$$

and using the result from (5.74) as the approximation to the centre manifolds $g_j(c_1)$:

$$g_w(\hat{c}_1) \sim 2\hat{c}_1^2 + \frac{1}{\alpha} \sum_{k=1}^{\infty} -\hat{c}_1^{ki+3} + \hat{c}_1^{ki+4},$$

which, compared to (5.43), agrees for the first four non-zero terms.

$\square$

Theorem 5.6.3. *The QSSA leads to an approximation of the flow governing the asymptotics of c_1 as*

$$\hat{c}_1' \sim \frac{1}{\alpha} \sum_{k=1}^{\infty} -\hat{c}_1^{ki+3} + \hat{c}_1^{ki+4} = -\frac{\hat{c}_1^{i+3}}{\alpha \sum_{k=1}^{i} \hat{c}_1^{k-1}}.$$

Proof. We can see this directly from (5.30) using Lemma (5.6.3). $\square$

Note that the power series in the expression for c_1' has a radius of convergence of infinity if $|c_1| \leq 1$. This is in contradiction to the power series from (5.43) which has radius of convergence of 0.

5.7   Conclusions

In this chapter we complemented the analysis of Section 3.1.2 by considering a more realistic dynamics of nucleating point islands with critical island size i by allowing subcritical islands of size $2 \leq j \leq i - 1$ to form and fragment. The mathematics of this new system of equations is more challenging than the fundamentally two dimensional system considered in Section 3.1.2 and we had to use both centre manifold techniques and a sophisticated globalisation argument using ideas from theories of compartmental systems and of asymptotically autonomous

dynamical systems; the globalisation methods used in this chapter are, in our opinion, more elegant than the "brute-force" asymptotics in Section 3.1.2. Furthermore, it appears that computations can be significantly simplified by making a sweeping assumption that all the clusters of size $1 < j \leq i$ are at a quasi-steady state.

Our asymptotic results in Section 5.5 are consistent with the leading term asymptotics for $c_1(t)$ of [70] (see our Theorem 5.5.1) and for $c_j(t)$ $(1 \leq j \leq i)$ of Section 3.1), as well as with the conjecture in [70] about the behaviour of $c_j(t)$, $j > i$ (see Theorem 5.5.1). In addition, our results are consistent with the work in Section 3.1.1 as $\beta = 0$ is equivalent to letting $i = 1$.

Chapter 6

The Dynamics of Erdős Numbers

6.1   Introduction

Paul Erdős (1913–1996) was a prolific mathematician who co-authored more than five hundred academic papers. During his lifetime, many began to wonder how their academic publications were connected to those of Erdős. This curiosity gave rise to the concept of an Erdős number which is defined by the Erdős Number Project [47] as follows: "Erdős's Erdős number is 0. Erdős's co-authors have Erdős number 1. People other than Erdős who have written a joint paper with someone with Erdős number 1 but not with Erdős, have Erdős number 2, and so on. If there is no chain of co-authorships connecting someone with Erdős, then that person's Erdős number is said to be infinite (undefined)".

Below we shall assume that a person's Erdős number can only change if that person collaborates with a person whose Erdős number is lower by two at least; in other words, it cannot change if the Erdős number of a collaborator changes independently. We shall also assume that all collaborations are binary; the case of multiple co-authors is of independent interest. We start by assuming that every

Chapter 6. The Dynamics of Erdős Numbers

collaborator is (intellectually speaking) immortal, and then, in Section 6.4, deal with the case when exiting the system is possible.

Let us specify, in the case of immortality, all the interactions that a collaborator can enter into. We denote the number of people with the Erdős number $k \in \mathbb{N}_0$ at time τ by $e_k(\tau)$ and the number of people with an undefined Erdős number at τ by $u(\tau)$. We will assume that the rate of arrival of new contributors with undefined Erdős number is a constant which we denote by $\hat{\alpha} > 0$. In addition, we denote the rate at which two people with assigned Erdős numbers' collaborate by $\beta \geq 0$, and the rate at which people without an assigned Erdős number collaborate by $\gamma \geq 0$. Clearly, the interesting case is when $\beta + \gamma > 0$.

Hence, the full list of possible collaborations and their outcomes in terms of Erdős numbers is

$$
\begin{aligned}
&\emptyset \xrightarrow{\hat{\alpha}} u, \\
&e_k + e_m \xrightarrow{\beta} e_k + e_m, && k \leq m \leq k+1, \\
&e_k + e_m \xrightarrow{\beta} e_k + e_{k+1}, && m > k+1, \\
&e_k + u \xrightarrow{\gamma} e_k + e_{k+1}, \\
&u + u \xrightarrow{\gamma} u + u,
\end{aligned}
\tag{6.1}
$$
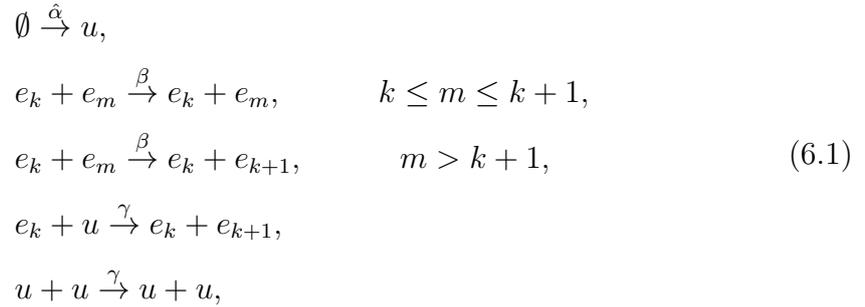
where $m, k \in \mathbb{N}_0$.

It is clear that (6.1) contains reactions that do not result in a net change of Erdős numbers:

$$
\begin{aligned}
&e_k + e_k \xrightarrow{\beta} e_k + e_k, \\
&e_k + e_{k+1} \xrightarrow{\beta} e_k + e_{k+1}, \\
&u + u \xrightarrow{\gamma} u + u.
\end{aligned}
\tag{6.2}
$$

100
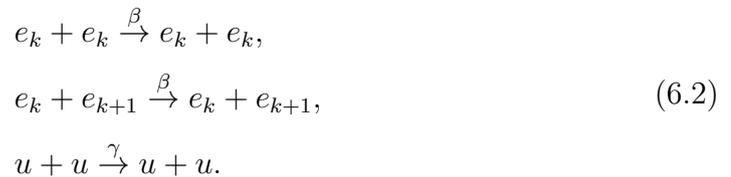
The reactions in (6.2) are often referred to as **futile cycles** or alternatively, substrate cycles. The effects futile cycles have on reaction systems are often unclear and, consequently, the effects of futile cycles on biological systems have been considered; for futile cycles in fatty acid and glucose growth see e.g. [23,74].

We can model the reactions in (6.1) as a system of rate equations by assuming that the futile cycles of (6.2) simply do not exist. Alternatively, we can model the reaction system (6.1) by considering a Gillespie type algorithm (GTA) which is described in detail later (see Section 6.3). The benefit of considering a GTA is all the futile cycles from (6.2) can be included in the process. For previous work using GTAs see, e.g, [34,35].

6.2 Governing Equations

Assuming the futile cycles as in (6.2) do not occur, the laws of mass kinetics give us the following infinite system of ODEs for the remaining reactions in (6.1):

$$\dot{e}_1 = \beta\left[e_0 \sum_{n=2}^{\infty} e_n\right] + \gamma e_0 u,$$

$$\dot{e}_k = \beta\left[e_{k-1} \sum_{n=k+1}^{\infty} e_n - e_k \sum_{n=0}^{k-2} e_n\right] + \gamma e_{k-1} u, \quad k = 2, 3, 4, \ldots,$$

$$\dot{u} = \hat{\alpha} - \gamma u \sum_{n=0}^{\infty} e_n.$$

We assume that $e_0 = 1$; dots denote differentiation with respect to τ.

It makes sense to scale the variables and the deposition rate $\hat{\alpha}$ in order to simplify the equations by combining β and γ. Thus, setting $t := c_1\tau$ for a constant c_1 to be chosen below, using primes for differentiation with respect to the new time scale t, setting $e_k(\tau) = E_k(t)$ and $u(\tau) = U(t)$, and choosing $c_1 = \beta+\gamma$, $\alpha = \hat{\alpha}/(\beta+\gamma)$,

Chapter 6. The Dynamics of Erdős Numbers

and $p := \beta/(\beta + \gamma) \in [0, 1]$, we obtain the system

$$
\begin{aligned}
E_0 &= 1, \\
E_1' &= p\left[E_0 \sum_{n=2}^{\infty} E_n\right] + (1-p)E_0 U, \\
E_k' &= p\left[E_{k-1} \sum_{n=k+1}^{\infty} E_n - E_k \sum_{n=0}^{k-2} E_n\right] + (1-p)E_{k-1}U, \quad k = 2, 3, 4, \ldots, \\
U' &= \alpha - (1-p)U \sum_{n=0}^{\infty} E_n.
\end{aligned}
\tag{6.3}
$$

Note, for reasons discussed in Section 1.3.1.4, we impose the restriction that solutions of (6.3) must have a finite mass, which implies that, for $t \geq 0$, a solution of (6.3) must be an element of the Banach space $X_1 \subset \ell^1$, where

$$
X_1 := \left\{E = (E_j) \in \mathbb{R}^{\mathbb{N}} \text{ such that } \|E\|_1 := \sum_{n=1}^{\infty} n|E_n| < \infty\right\}.
$$

As we are primarily interested in the average Erdős number we set

$$
T := \sum_{n=0}^{\infty} E_n, \ M := \sum_{n=0}^{\infty} nE_n, \text{ and } A := M/T.
\tag{6.4}
$$

To find the asymptotics for the average Erdős number A, we first find the asymptotics for the number of people with an undefined Erdős number U and the asymptotics for the total number T of collaborators with an assigned Erdős number. Once this is done, we discuss the asymptotic behaviour of M.

We set $T(0) =: T_0$, $U(0) =: U_0$ and $M(0) =: M_0$.

We begin by reducing the infinite system of ODEs (6.3) to a two dimensional system of ODEs in T and U by adding all equations for $k = 0, 1, 2, \ldots$ to obtain

Chapter 6. The Dynamics of Erdős Numbers

formally

$$\sum_{n=0}^{\infty} E_n' = (1-p)U\sum_{n=0}^{\infty} E_n,$$

$$U' = \alpha - (1-p)U\sum_{n=0}^{\infty} E_n.$$

Hence using the definitions of T from (6.4)

$$\begin{aligned} T' &= (1-p)UT, \\ U' &= \alpha - (1-p)UT. \end{aligned} \tag{6.5}$$

First of all, we have

Theorem 6.2.1. *If $\sum_{k=0}^{\infty} E_k(0) < \infty$, a solution of (6.3) for $k \geq 0$ is also a solution of (6.5).*

Proof. The argument of the proof is the same as in Theorem 5.3.1. $\qquad\square$

As a result of Theorem 6.2.1, we can use finite dimensional techniques to discuss the dynamics of T and U. We have

Lemma 6.2.1. *As $t \to \infty$, the asymptotics of U and T of (6.5) satisfies*

$$T = \begin{cases} \alpha t + O(1), & p \in [0,1), \\ T_0, & p = 1, \end{cases}$$

$$U = \begin{cases} \frac{1}{(1-p)t} + O\left(\frac{1}{t^2}\right), & p \in [0,1), \\ \alpha t + U_0, & p = 1. \end{cases}$$

Proof. We split the proof into two parts:

Chapter 6. The Dynamics of Erdős Numbers

Case 1: $p = 1$. The case of $p = 1$ is trivial as from (6.5) we have that $T' = 0$ and $U' = \alpha$, which implies $T = T_0$ and $U = \alpha t + U_0$.

Case 2: $p \in [0, 1)$. Set

$$P = 1 - p,$$
$$C_0 = T_0 + U_0 > 0. \tag{6.6}$$

From (6.5) we have that $(T + U)' = \alpha$, which implies that

$$U = \alpha t - T + C_0. \tag{6.7}$$

Substituting (6.7) into the first equation from (6.5) we have that

$$T' = P(\alpha t - T + C_0)T.$$

As $\alpha > 0$ and $p \in [0, 1)$ we can now solve explicitly for T as functions of t using MAPLE:

$$T = \frac{T_0\sqrt{2\alpha P}e^{P(\frac{\alpha t}{2} + C_0)t}}{T_0\sqrt{\pi}Pe^{\frac{-C_0^2 P}{2\alpha}}[\mathrm{erfi}(\nu_1(t) + \varphi) - \mathrm{erfi}(\varphi)] + \sqrt{2\alpha P}},$$

where,

$$\nu_1 = \frac{P\alpha t}{\sqrt{2\alpha P}}, \quad \varphi = \frac{C_0 P}{\sqrt{2\alpha P}}, \quad \mathrm{erfi}(x) := -i\,\mathrm{erf}(ix) = \frac{2}{\sqrt{\pi}}\int_0^x e^{r^2}\,dr.$$

Substituting T into (6.7):

$$U = \alpha t - \frac{T_0\sqrt{2\alpha P}e^{P(\frac{\alpha t}{2} + C_0)t}}{T_0\sqrt{\pi}Pe^{\frac{-C_0^2 P}{2\alpha}}[\mathrm{erfi}(\nu_1(t) + \varphi) - \mathrm{erfi}(\varphi)] + \sqrt{2\alpha P}} + C_0.$$

Chapter 6. The Dynamics of Erdős Numbers

As the exact expressions for T and U are cumbersome, we consider the asymptotics as $t \to \infty$:

$$T = \alpha t + C_0 + O\left(\frac{1}{t}\right),$$

$$U = \frac{1}{Pt} - \frac{C_0}{\alpha P t^2} + O\left(\frac{1}{t^3}\right).$$

Hence, in terms of p, using the definition of P from (6.6), as $t \to \infty$,

$$T = \alpha t + C_0 + O\left(\frac{1}{t}\right),$$

$$U = \frac{1}{(1-p)t} - \frac{C_0}{\alpha(1-p)t^2} + O\left(\frac{1}{t^3}\right).$$

Note that the leading terms of the asymptotic expansions for both T and U do not depend on the initial conditions. □

As we will use the asymptotics of U and T from Lemma 6.2.1 in our calculation of the asymptotics of A we compare our results with those obtained by numerical integration of the ODEs from (6.5).

To be able to directly compare the numerical integration of the ODEs from (6.5) with the results from Lemma 6.2.1, we linearise the equation for U from Lemma 6.2.1 when $p \in [0, 1)$ by multiplying U by $(1-p)t^2$ and taking the reciprocal of both sides:

$$\frac{1}{U(1-p)t^2} = \frac{1}{t + O(1)}, \quad \text{as } t \to \infty. \tag{6.8}$$

Chapter 6.   The Dynamics of Erdős Numbers

From the Taylor series, of the right-hand side of (6.8), at infinity:

$$\frac{1}{U(1-p)t^2} = \frac{1}{t} + O\left(\frac{1}{t^2}\right) \quad \text{as,} \quad t \to \infty,$$

which implies that

$$\frac{1}{U} = (1-p)t + O(1), \quad \text{as } t \to \infty.$$

We illustrate our numerical integration of (6.5) in Figure 6.1 where we let $\alpha = 2, p = 0.5, T_0 = 1$ and $U_0 = 1$.
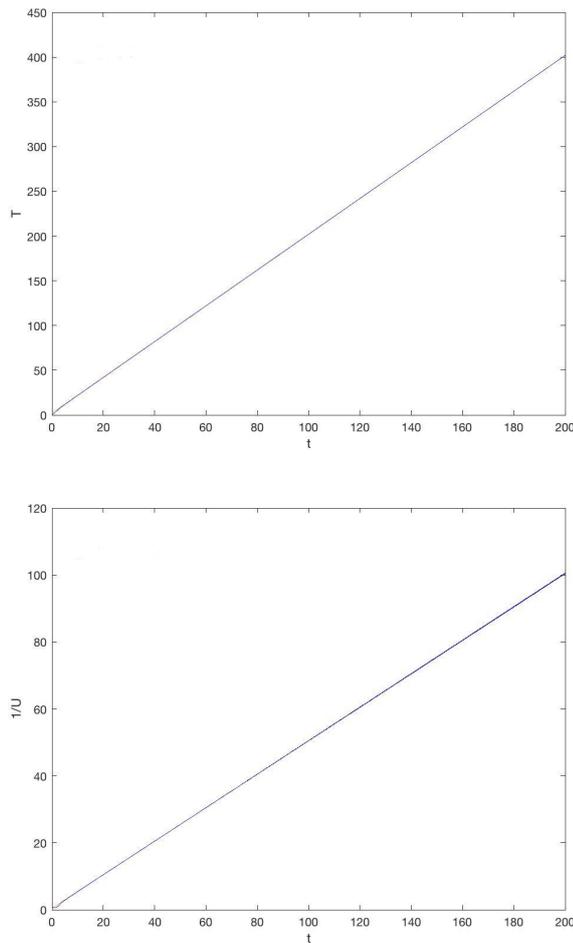
Figure 6.1: Numerical integration of system (6.5) where $\alpha = 2$, $p = 0.5, T_0 = 1$ and $U_0 = 1$.

It is clear the results from Figure 6.1 match the asymptotics of for T and U from Lemma 6.2.1.

We are now in a position to consider the asymptotics of A.

For convenience, we begin by considering the $p = 0$ case.

Lemma 6.2.2. *If $p = 0$, as $t \to \infty$, the asymptotics of A, as defined in (6.4), of*

Chapter 6.   The Dynamics of Erdős Numbers

(6.3) *satisfies*

$$A = \ln t + O(1),$$

independently of α.

Proof. We need to derive a differential equation for M, as once the asymptotics of M are found A falls out automatically as we know the asymptotics of T. We multiply all the E_k equations in (6.3) by k and add all the equations:

$$\sum_{n=0}^{\infty} nE_n' = \sum_{n=1}^{\infty} nE_{n-1}U = \left(\sum_{n=0}^{\infty} nE_n + \sum_{n=0}^{\infty} E_n \right)U,$$

which leads to

$$M' = (M + T)U. \tag{6.9}$$

Now we proceed as follows: we use the asymptotics of T and U to obtain the equation M satisfies for large times. We write $M \sim M_0 + M_1$, where $M_1 = o(M_0)$ as $t \to \infty$ and choose M_0 so that the right hand-side of the equation is $o(M_0')$ as $t \to \infty$.

From the asymptotics of U and T from Lemma 6.2.1, we have that:

$$M' = \left(M + \alpha t + O(1) \right)\left[\frac{1}{t} + O\left(\frac{1}{t^2} \right) \right]$$
$$= \alpha + \frac{M}{t} + O\left(\frac{1}{t} \right) + O\left(\frac{M}{t^2} \right),$$

as $t \to \infty$.

In terms of M_0 and M_1, we have

$$M_0' + M_1' = \alpha + \frac{M_0}{t} + \frac{M_1}{t} + O\left(\frac{1}{t} \right) + O\left(\frac{M}{t^2} \right), \text{ as } t \to \infty.$$

The only choice that satisfies the asymptoticity condition is

$$M_0' \sim \alpha + \frac{M_0}{t}, \text{ as } t \to \infty,$$

from which we obtain

$$M_0 \sim \alpha t \ln t + Ct, \text{ as } t \to \infty,$$

where C is a constant. Hence

$$M = \alpha t \ln t + O(t), \text{ as } t \to \infty,$$

Finally, from the definition of A from (6.4) and using the asymptotics of T from Lemma 6.2.1 and the result for M:

$$A = \frac{\alpha t \ln(t)}{\alpha t + O(1)} + \frac{O(t)}{\alpha t + O(1)}, \text{ as } t \to \infty.$$

From the Taylor series expansions of A at infinity we find that,

$$A = \ln t + O(1), \text{ as } t \to \infty.$$

$\square$

It is once again useful to compare the results from Lemma 6.2.2 with those indicated by numerical integration of the ODEs from (6.5) and (6.9) to find A.

For comparison, linearise A from Lemma 6.2.2. Hence, we have that

$$e^A = e^{\ln t + O(1)}, \text{ as } t \to \infty,$$

Chapter 6.   The Dynamics of Erdős Numbers

which implies that

$$e^A = O(t), \text{ as, } t \to \infty.$$

We illustrate our numerical integration in Figure 6.2 where we let $\alpha = 2, p = 0, T_0 = 1$ and $U_0 = 1$, $M_0 = 0$.
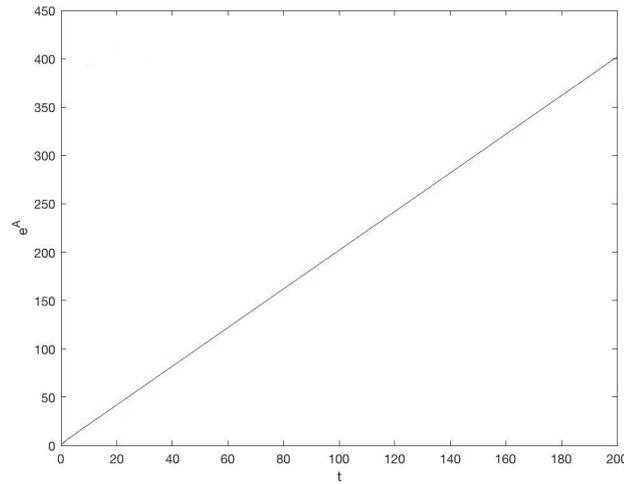
Figure 6.2: Numerical integration of system (6.5) and (6.9) where $\alpha = 2$, $p = 0, T_0 = 1$, $U_0 = 1$ and $M_0 = 0$ to find A.

It is clear that the results from Figure 6.2 indicate the same results as the asymptotics for A from Lemma 6.2.2.

We can now consider the asymptotics of A when $p \in (0, 1]$. As in the $p = 0$ case, we begin by considering the asymptotics of M. To that end, we again multiply all the E_k equations of (6.3) by k to obtain

$$E_1' = p\left[E_0 \sum_{n=2}^{\infty} E_n\right] + (1-p)E_0 U,$$

$$kE_k' = pk\left[E_{k-1} \sum_{n=k+1}^{\infty} E_n - E_k \sum_{n=0}^{k-2} E_n\right] + k(1-p)E_{k-1}U, \quad k = 2, 3, 4, \ldots.$$

110

Chapter 6. The Dynamics of Erdős Numbers

Adding all these equations and using the definitions of M and T from (6.4), we have

$$M' = p\psi + (1 - p)(M + T)U,$$

where

$$\psi = \sum_{m=0}^{\infty} \left[(m+1)E_m \sum_{n=m+2}^{\infty} E_n \right] - \sum_{m=2}^{\infty} \left[mE_m \sum_{n=0}^{m-2} E_n \right].$$

As it is not possible to express ψ in terms of M, T and U we will derive differential inequalities satisfied by M. First of all, we have a compact representation for ψ (note that ψ has to be negative as it encodes migration towards lower Erdős numbers).

Claim 6.2.3. *We have that*

$$\psi = -\sum_{m=0}^{\infty} E_m \sum_{n=0}^{\infty} nE_{n+m+1}.$$

Proof. We write

$$\psi = E_0(E_2 + E_3 + \dots) +$$
$$2((E_1(E_3 + E_4 + \dots) - E_2(E_0)) +$$
$$3((E_2(E_4 + E_5 + \dots) - E_3(E_0 + E_1)) +$$
$$4((E_3(E_5 + E_6 + \dots) - E_4(E_0 + E_1 + E_2)) +$$
$$5((E_4(E_6 + E_7 + \dots) - E_5(E_0 + E_1 + E_2 + E_3)) + \dots.$$

Rearranging, we have

$$
\begin{aligned}
\psi = & - E_0(E_2 + 2E_3 + 3E_4 + \ldots) + \\
& - E_1(E_3 + 2E_4 + 3E_5 + \ldots) + \\
& - E_2(E_4 + 2E_5 + 3E_6 + \ldots) + \\
& - E_3(E_5 + 2E_6 + 3E_7 + \ldots) + \\
& - E_4(E_6 + 2E_7 + 3E_8 + \ldots) + \ldots. \\
= & -\sum_{m=0}^{\infty} E_m \sum_{n=0}^{\infty} n E_{n+m+1}
\end{aligned}
$$

as required. □

Consider

$$
\psi = -\sum_{m=0}^{\infty} E_m \sum_{n=0}^{\infty} n E_{n+m+1} < 0.
$$

Lemma 6.2.2. *We have the following inequalities:*

$$
\psi \leq -M - 1 + T, \tag{6.10}
$$

and

$$
\psi \geq -MT - T + T^2. \tag{6.11}
$$

Proof. For the first inequality, we just note that

$$
\psi \leq -E_0(E_2 + 2E_3 + \ldots) = -E_0(E_0 + E_1 + 2E_2 + 3E_3 + \ldots - T) = -M - 1 + T,
$$

where we have used the fact that $E_0 = 1$. The second inequality, (6.11), requires

Chapter 6. The Dynamics of Erdős Numbers

more work. Consider,

$$\psi = -E_0(E_2 + 2E_3 + 3E_4 + \ldots)+$$
$$-E_1(E_3 + 2E_4 + 3E_5 + \ldots)+$$
$$-E_2(E_4 + 2E_5 + 3E_6 + \ldots)+$$
$$-E_3(E_5 + 2E_6 + 3E_7 + \ldots)+$$
$$-E_4(E_6 + 2E_7 + 3E_8 + \ldots)+$$
$$\ldots.$$

Rearranging and using the definition of M and T:

$$\psi = -E_0(1 + M) + E_0 T+$$
$$-E_1(1 + M) + E_1(E_0 + E_1 + 2E_2 + 2E_3 + 2E_4 + \ldots)+$$
$$-E_2(1 + M) + E_2(E_0 + E_1 + 2E_2 + 3E_3 + 3E_4 + \ldots)+$$
$$-E_3(1 + M) + E_3(E_0 + E_1 + 2E_2 + 3E_3 + 4E_4 + 4E_5 + \ldots)+$$
$$-E_4(1 + M) + E_4(E_0 + E_1 + 2E_2 + 3E_3 + 4E_4 + 5E_5 + 5E_6 + \ldots)+$$
$$\ldots.$$

Thus, only considering the first column and the definition of T :

$$\psi = - T(1 + M) + E_0 T+$$
$$+ E_1(E_0 + E_1 + 2E_2 + 2E_3 + 2E_4 + \ldots)+$$
$$+ E_2(E_0 + E_1 + 2E_2 + 3E_3 + 3E_4 + \ldots)+$$
$$+ E_3(E_0 + E_1 + 2E_2 + 3E_3 + 4E_4 + 4E_5 + \ldots)+$$
$$+ E_4(E_0 + E_1 + 2E_2 + 3E_3 + 4E_4 + 5E_5 + 5E_6 + \ldots)+$$
$$\ldots.$$

Factorising a T from each line:

$$\psi = -MT - T + E_0 T +$$
$$+E_1(T + E_2 + E_3 + E_4 + \ldots) +$$
$$+E_2(T + E_2 + 2E_3 + 2E_4 + \ldots) +$$
$$+E_3(T + E_2 + 2E_3 + 3E_4 + 3E_5 + \ldots) +$$
$$+E_4(T + E_2 + 2E_3 + 3E_4 + 4E_5 + 4E_6 + \ldots) +$$
$$\ldots .$$

Hence,

$$\psi = -MT - T + T^2 +$$
$$+E_1(E_2 + E_3 + E_4 + \ldots) +$$
$$+E_2(E_2 + 2E_3 + 2E_4 + \ldots) +$$
$$+E_3(E_2 + 2E_3 + 3E_4 + 3E_5 + \ldots) +$$
$$+E_4(E_2 + 2E_3 + 3E_4 + 4E_5 + 4E_6 + \ldots) +$$
$$\ldots ,$$

as required. $\qquad\square$

Given the result of Lemma 6.2.2, we can formulate the following result.

Lemma 6.2.3. *If $p \in (0, 1)$, as $t \to \infty$, the asymptotics of A, as defined in (6.4), of (6.3) satisfies*

$$A = 1 + O\left(\frac{1}{t}\right).$$

Proof. Consider functions M_a and M_b that satisfy the following ODEs:

$$M_a' = p(T - M_a - 1) + (1 - p)(M_a + T)U, \qquad (6.12)$$

and

$$M_b' = p(T^2 - M_b T - T) + (1 - p)(M_b + T)U. \tag{6.13}$$

Then, for all time, $M_a' \leq M' \leq M_b'$. Hence, if we find that the asymptotics of M_a and M_b are the same, we can conclude that we have found the asymptotics of M.

Let us consider M_a first, as $t \to \infty$, we have

$$M_a' = p\alpha t - pM_a + O(1) + O\left(\frac{M_a}{t}\right).$$

The only choice that satisfies the asymptoticity condition is

$$M_a' \sim p\alpha t - pM_a, \;\; as \; t \to \infty.$$

From this we conclude that

$$M_a = \alpha t + O(1), \;\; as \; t \to \infty.$$

Similarly, for M_b we have

$$M_b' = p\alpha^2 t^2 - \alpha p t M_b + O(t) + O(M_b), \; as \; t \to \infty,$$

and the only choice that satisfies the asymptoticity condition is

$$M_b' \sim p\alpha^2 t^2 - \alpha p t M_b, \; as \; t \to \infty,$$

which again leads to

$$M_b = \alpha t + O(1), \; as \; t \to \infty.$$

Thus, $M = \alpha t + O(1)$ as $t \to \infty$. Given the asymptotics of T as given in Lemma 6.2.1, we obtain the required result. $\qquad\square$

Additionally, the case $p = 1$ is trivial, as in this case $M = T_0 - 1 + o(1/t)$, as $t \to \infty$, and T remains constant, so $A = (T_0 - 1)/T_0 + o(1/t)$.

Combining our results we have

Theorem 6.2.4. *As $t \to \infty$, the asymptotics of A as defined in (6.4), of (6.3) satisfies*

$$
A = \begin{cases}
\ln t + O(1), & p = 0, \\
1 + O\left(\frac{1}{t}\right), & p \in (0,1), \\
\frac{T_0 - 1}{T_0} + o\left(\frac{1}{t}\right), & p = 1.
\end{cases}
$$

In the interest of completeness, it is once again useful to compare the results from Theorem 6.2.4 when $p \in (0,1)$ with those indicated by numerical integration of the ODEs from (6.5), (6.12) and (6.13) to find A.

We illustrate our numerical integration in Figure 6.3 where we let $\alpha = 2, p = 0, T_0 = 1$ and $U_0 = 1$, $M_0 = 0$.
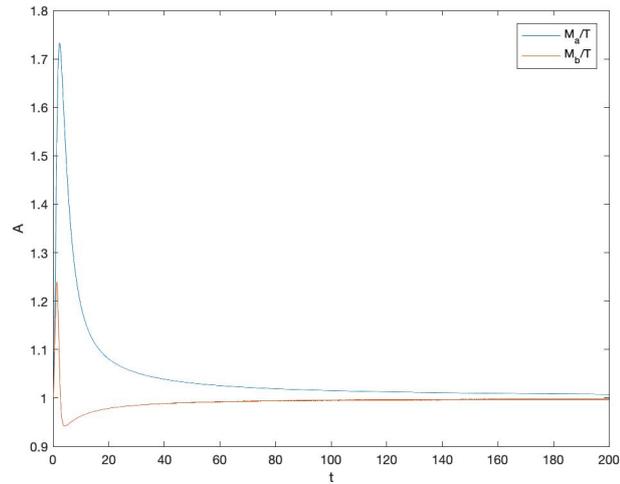
Figure 6.3: Numerical integration of system (6.5) (6.12) and (6.13) where $\alpha = 2$, $p = 0.5, T_0 = 1$, $U_0 = 1$ and $M_0 = 0$ to find A.

It is clear the results from Figure 6.3 indicate the same results as the asymptotics for A from Theorem 6.2.4.

6.3 Gillespie Type Algorithm

As we now have the asymptotics for A for system (6.1), from Theorem 6.2.4, in this section we will describe a GTA, related to the stochastic particle system (6.1) which includes all futile cycles. We will introduce the GTA using the methods of [34, 35] and [49, Chapter 4]. As our algorithm is similar to those used in the above papers and book, we indicate only the main ideas.

For comparison with Theorem 6.2.4 we use the scaling for $\tau, \hat{\alpha}, \beta, \gamma, e_k$ and u from (6.4).

Suppose we know the number of people with each Erdős number at time t, our aim is to describe how the number of people with each Erdős number evolves over

time. We therefore think of a **state vector**, $X(t) := X_t$ at t as

$$X_t = (E_0, E_1, \ldots, E_n, \ldots, U),$$

where we denote the maximum Erdős number at t by $n(t) := n_t$ and the elements
of the state vector X_t are positive integers that record the number of people
with each Erdős number at time t. The state vector X_t changes whenever a
collaboration (that is not futile) takes place. The effect of the collaborations on
the state vector is captured by the **stoichiometric vectors** $v_j(t) = v_{j_t}$ at t:

$$v_{j_t} = \begin{cases} [0, \ldots, 0], & 0 \leq j \leq n_t, \\ [0, \ldots, 0], & j \leq 2n_t, \\ [0, 1, -1, \ldots, 0], & j = 2n_t + 1, \\ \ldots \\ [0, 0, 1, \ldots, -1, 0], & j = (n^2 + 3n - 2)/2, \\ [0, 1, 0, \ldots, -1, 0], & j = n_t(n_t + 3)/2, \\ [0, \ldots, 1, 0, \cdots - 1], & \eta_1 \leq j \leq \eta_2, \\ [0, \ldots, 0], & j = (n_t + 1)(n_t + 4)/2, \\ [0, \ldots, 1], & j = (n_t + 3)(n_t + 2)/2, \end{cases}$$

where $\eta_1 = (n_t + 1)(n_t + 2)/2$ and $\eta_2 = (n_t^2 + 5n_t + 2)/2$.

The state vector at time t, X_t, changes when a collaboration (that is not futile)
occurs, we will work in terms of the probability of a collaboration taking place,
based on the current state of the system. The jth collaboration has an associated
stoichiometric vector, $v_j \in \mathbb{R}$, whose k^{th} component is the change in the number
people with the Erdős number E_k or U caused by the jth reaction. So one reaction
of type j has the effect of updating the state vector from X_t to $X_t + v_j$. The

probability of this reaction taking place in the infinitesimal time interval $[t, t + dt)$ is assumed to take the form $a_j(X_t)dt$, where $a_j(X_t)$ is called a **propensity function**:

$$a_j(X_t) = \begin{cases} \frac{1}{2}pE_j(E_j - 1), & 0 \le j \le n_t, \\[6pt] pE_{j-n_t-1}E_{j-n_t}, & n_t + 1 \le j \le 2n_t, \\[6pt] pE_0E_2, & j = 2n_t + 1, \\[6pt] \dots \\[6pt] pE_1E_{n_t} & j = (n^2 + 3n - 2)/2, \\[6pt] pE_0E_{n_t}, & j = n_t(n_t + 3)/2, \\[6pt] pE_0E_{n_t}, & j = n_t(n_t + 3)/2, \\[6pt] (1 - p)UE_{j-\eta_1}, & \eta_1 \le j \le \eta_2, \\[6pt] \frac{1}{2}U(U - 1), & j = (n_t + 1)(n_t + 4)/2, \\[6pt] \alpha, & j = (n_t + 3)(n_t + 2)/2. \end{cases}$$

Hence, using the state vector, the stoichiometric vectors and the propensity functions we can formulate the resulting GTA, where an initial state vector X_0 is given,

1. Set $t = 0$ and choose n_0 and t_f.

2. Set $Q = (n_t + 3)(n_t + 2)/2$.

3. Evaluate $\{a_j(X_t)\}^Q$ and $a_{sum} := \sum_{k=0}^{Q} a_k(X_t)$.

4. Pick two random numbers, R_1 and R_2 between $U[(0, 1)]$.

5. Set s to be the smallest integer that satisfies $\sum_{k=0}^{s} a_k(X_t) > R_1 a_{sum}$.

6. Set $\xi = (\ln(1/R_2))/a_{sum}$.

7. Set $X_{t+\xi} = X_t + v_{st}$.

8. Update t to $t + \xi$.

9. Go back to step 2 or terminate if $t \geq t_f$.

From our simulations it appears that the asymptotics of A for system (6.1) from the GTA are consistent with the asymptotics of A from the rate equations from (6.3). To illustrate the similarity we consider the cases when $n_0 = 20$, $\alpha = 2$, $p = 0$, $E_0(0) = 1$, $E_k(0) = 0$ where $1 \leq k \leq 20$ and $U = 1$ and $n_0 = 20$, $\alpha = 2$, $p = 0.5$, $E_0(0) = 1$, $E_k(0) = 0$ where $1 \leq k \leq 20$ and $U = 1$ as illustrated in Figure 6.4 – 6.5.
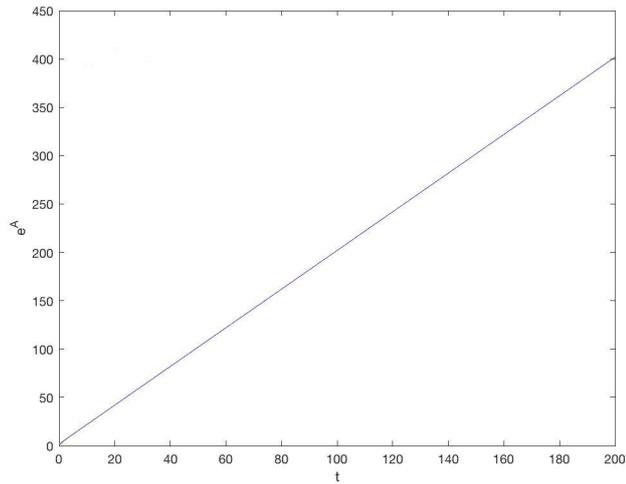
Figure 6.4: An example of a GTA for the reaction system (6.1) where $n_0 = 20$, $\alpha = 2$, $p = 0$, $E_0(0) = 1$, $E_k(0) = 100$ where $1 \leq k \leq 20$ and $U = 100$.
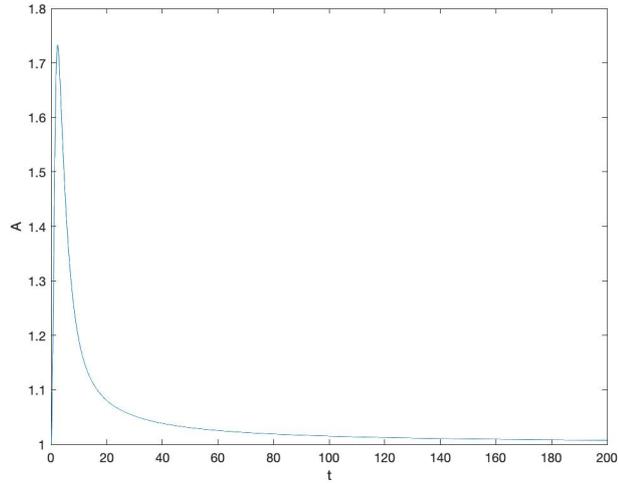
Figure 6.5: An example of a GTA for the reaction system (6.1) when $n_0 = 20$, $\alpha = 2$, $p = 0.5$, $E_0(0) = 1$, $E_k(0) = 100$ where $1 \leq k \leq 20$ and $U = 100$.

We find that changes in α, n_0 and the initial conditions have a negligible effect on the underlying distributions. From Figures 6.4 – 6.5, we can formulate

Proposition 6.3.1. *For the reaction system* (6.1) *the asymptotics of A from the GTA are given by*

$$
A = \begin{cases} \ln t + O(1), & p = 0, \\ 1, & p \in (0, 1], \end{cases}
$$

as $t \to \infty$.

It is clear the results from Proposition 6.3.1 indicate the same results as the asymptotics for A from Theorem 6.2.4; this result is not surprising as we assume the collaborations occur instantaneously. We illustrate this with a simple example:

Example 6.3.1. *Let $t = 0$, the maximum Erdős number at $t = 0$ be $n_0 = 2$ and the state vector $X_0 = (E_0, E_1 E_2, U) = (1, 1, 1, 0)$.*

Chapter 6. The Dynamics of Erdős Numbers

Let the probability density function (pdf) of the waiting time until the state vector
becomes $X_T = (1, 2, 0, 0)$ be $f(T)$.

In the case when no futile cycles occur the pdf is $f(T) = e^{-T}$.

In the case when futile cycles are allowed to occur

$$\int_0^T f(t)dt = \frac{1}{3}\sum_{k=1}^{\infty} P_k(t \leq T)\left(\frac{2}{3}\right)^{k-1},$$

where $P_k(t \leq T)$ is the probability that exactly k reactions happen before T. Hence,
using the Erlang distribution

$$P_k(t \leq T) = \int_0^T f_k(t)dt = \int_0^T r^k \frac{t^{k-1}}{(k-1)!}e^{-rt}dt.$$

In our case $r = 3$, so by differentiation with respect to T we get

$$f(T) = \frac{1}{3}\sum_{k=1}^{\infty} 3^k \frac{T^{k-1}}{(k-1)!}\left(\frac{2}{3}\right)^{k-1}e^{-3T}$$

$$= \sum_{k=1}^{\infty} \frac{(2T)^{k-1}}{(k-1)!}e^{-3T} = e^{2T}e^{-3T} = e^{-T}.$$

Hence, the system that contains futile cycles and the system that does not contain
the futile cycles produce the same pdf. So the futile cycles have no effect.

As the futile cycles of (6.2) have no effect we assume that the futile cycles simply
do not exist for the remainder of this chapter.

6.4   Accounting for Mortality

We next consider the effects of people being able to exit the system. The simplest way to do that is to put $E_0 = 0$, alas, and to denote the exit of people with the Erdős number k from the system by the function $D(E_k, t)$ and the exit of people with an undefined Erdős number from the system by $D(U, t)$. Neglecting temporal effects we assume that $D(E_k, t) = D(E_k)$ and $D(U, t) = D(U)$ and that the exit of people is proportional to the population size, i.e. $D(E_k) = \delta E_k$ and $D(U) = \delta U$ where $\delta > 0$.

We have

$$
\begin{aligned}
E_1' &= -\delta E_1, \\
E_k' &= p\left[E_{k-1} \sum_{n=k+1}^{\infty} E_n - E_k \sum_{n=0}^{k-2} E_n \right] \\
&\quad + (1-p)E_{k-1}U - \delta E_k, \, k = 2, 3, 4 \ldots, \\
U' &= \alpha - (1-p)U \sum_{n=0}^{\infty} E_n - \delta U.
\end{aligned}
\tag{6.14}
$$

As in Section 6.2 we can reduce system (6.14) to a system of ODEs for T and U:

$$
\begin{aligned}
T' &= (1-p)UT - \delta T, \\
U' &= \alpha - (1-p)UT - \delta U,
\end{aligned}
\tag{6.15}
$$

where

Theorem 6.4.1. *If $\sum_{k=0}^{\infty} E_k(0) < \infty$, a solution of (6.14) for $k \geq 0$ is also a solution of (6.15).*

Note that this is a predator-prey type system. We consider the stability of the fixed points for system (6.15):

Lemma 6.4.1. *The fixed points of system* (6.15) *are:*

$$(T, U) = \begin{cases} \left(0, \frac{\alpha}{\delta}\right), & \alpha < \frac{\delta^2}{1-p} \text{ which is a stable node,} \\[2mm] \left(0, \frac{\alpha}{\delta}\right), & \alpha > \frac{\delta^2}{1-p} \text{ which is a saddle,} \\[2mm] \left(\frac{\alpha(1-p)-\delta^2}{\delta(1-p)}, \frac{\delta}{1-p}\right), & \alpha < \frac{\delta^2}{1-p} \text{ which is a saddle,} \\[2mm] \left(\frac{\alpha(1-p)-\delta^2}{\delta(1-p)}, \frac{\delta}{1-p}\right), & \alpha > \frac{\delta^2}{1-p} \text{ which is a stable node.} \end{cases}$$

Proof. The fixed points of the system are

$$\left(0, \frac{\alpha}{\delta}\right) \text{ and } \left(\frac{\alpha(1-p)-\delta^2}{\delta(1-p)}, \frac{\delta}{1-p}\right).$$

The Jacobian matrix is

$$J = \begin{bmatrix} (1-p)U - \delta & (1-p)T \\ -(1-p)U & -(1-p)T - \delta \end{bmatrix}.$$

At $\left(0, \frac{\alpha}{\delta}\right)$ the Jacobian matrix becomes

$$J = \begin{bmatrix} \frac{\alpha(1-p)-\delta^2}{\delta} & 0 \\ -(1-p)\frac{\alpha}{\delta} & -\delta \end{bmatrix}.$$

So the eigenvalues of J are $\lambda_1 = \frac{\alpha(1-p)-\delta^2}{\delta}$ and $\lambda_2 = -\delta$. Hence, as $\delta > 0$, the fixed point $\left(0, \frac{\alpha}{\delta}\right)$ is stable if $\alpha < \frac{\delta^2}{1-p}$ and unstable otherwise.

At $\left(\frac{\alpha(1-p)-\delta^2}{\delta(1-p)}, \frac{\delta}{1-p}\right)$ the Jacobian matrix becomes

$$J = \begin{bmatrix} 0 & \frac{\alpha(1-p)-\delta^2}{\delta} \\ -\delta & -\frac{\alpha(1-p)}{\delta} \end{bmatrix}.$$

So the eigenvalues of J are $\lambda_1 = -\frac{\alpha(1-p)-\delta^2}{\delta}$ and $\lambda_2 = -\delta$. Hence, as $\delta > 0$, the

fixed point $\left(\frac{\alpha(1-p)-\delta^2}{\delta(1-p)}, \frac{\delta}{1-p}\right)$ is stable if $\alpha > \frac{\delta^2}{1-p}$ and unstable otherwise.

$\square$

As we are only interested in solutions where the total number of people in the system is positive, i.e. $T > 0$ we let $\alpha > \delta^2/(1-p)$ for the remainder of this section.

The case of $p = 1$ is trivial, as $A \sim 0$ as $t \to \infty$, so we assume $p \in [0, 1)$.

Following the same methods as in Section 2, we first find the asymptotics for the number of people with an undefined Erdős number U and the asymptotics for the total number of people T.

We have

Lemma 6.4.2. *If* $p \in [0, 1)$, $\delta \neq 0$ *and* $\alpha \geq \frac{\delta^2}{1-p}$, *as* $t \to \infty$, *the asymptotics of* U *and* T *of* (6.15) *satisfy*

$$T = \frac{\alpha(1-p) - \delta^2}{\delta(1-p)} + o\left(\frac{1}{t}\right), \qquad U = \frac{\delta}{1-p} + o\left(\frac{1}{t}\right).$$

Proof. From (6.15), we have that $(U + T)' = \alpha - \delta(U + T)$, hence

$$U = \frac{\alpha}{\delta} + e^{-\delta t}\left(C_0 - \frac{\alpha}{\delta}\right) - T. \tag{6.16}$$

Substituting (6.16) into the first equation of (6.15) we have that

$$T' = (1-p)\left[\frac{\alpha}{\delta} + e^{-\delta t}\left(C_0 - \frac{\alpha}{\delta}\right) - T\right]T - \delta T.$$

Chapter 6. The Dynamics of Erdős Numbers

Which implies that, as $t \to \infty$,

$$T' = (1-p)\left[\frac{\alpha}{\delta} + o\left(\frac{1}{t}\right) - T\right]T - \delta T.$$

The only choice that satisfies the asymptoticity condition is

$$T' \sim -(1-p)T^2 + \gamma T \text{ as } t \to \infty,$$

where
$$\gamma = \frac{(1-p)\alpha - \delta^2}{\delta}.$$

As $\gamma > 0$, we can now solve asymptotically for T and hence for U as functions of t. The required result follows. □

We are now in a position to consider the asymptotics of A. We begin by considering the $p = 0$ case.

Lemma 6.4.3. *If $p = 0$, $\delta \neq 0$ and $\alpha \geq \frac{\delta^2}{1-p}$, as $t \to \infty$, the asymptotics of A of (6.15) satisfies*
$$A = \delta t + o\left(\frac{1}{t}\right)$$
independently of α.

Proof. Following the same method as in Lemma 6.2.2 we need to derive a differential equation for M. We have

$$M' \sim (M + T)U - \delta M \text{ as } t \to \infty.$$

From the asymptotics of T and U from Lemma 6.4.2, we have that

$$M' = \left(M + \frac{\alpha - \delta^2}{\delta} + o\left(\frac{1}{t}\right)\right)\left[\delta + o\left(\frac{1}{t}\right)\right] - \delta M = \alpha - \delta^2 + o\left(\frac{1}{t}\right) + o\left(\frac{M}{t}\right)$$

126

Chapter 6.   The Dynamics of Erdős Numbers

as $t \to \infty$.

The only choice that satisfies the asymptoticity condition is

$$M' \sim \alpha - \delta^2 \text{ as } t \to \infty,$$

from which we obtain

$$M \sim (\alpha - \delta^2)t \text{ as } t \to \infty,$$

and therefore, as $p = 0$ we have that

$$A = \delta t + o\left(\frac{1}{t}\right).$$

$\square$

We can now consider the asymptotics of A when $p \in (0,1)$. As it is not possible to express M in terms of T and U, or derive differential inequalities satisfied by M, we consider a finite truncation of (6.14) where we assume that $E_n = 0$ when $n > 20$:
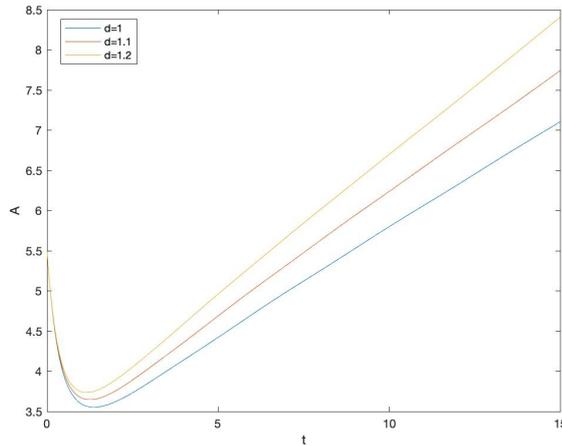
Figure 6.6: Numerical integration for the truncated system to calculate A when $p = 0.5$, $\alpha = 10$ and $\delta = 1, 1.1, 2$.

In addition, we find that changes in α, p, n and the initial conditions have a negligible effect on the underlying distributions.

From Figure 6.6, we can formulate

Conjecture 6.4.1. *If $p \in (0,1)$, $\delta \neq 0$ and $\alpha \geq \frac{\delta^2}{1-p}$, as $t \to \infty$, the asymptotics of A of* (6.15) *satisfies*

$$A \sim \delta t.$$

Additionally, it is worth noting the asymptotics of A from the GTA in the case of mortal collaborators are consistent with the asymptotics of A from the rate equations (6.14).

6.5   Conclusions

In this chapter we considered the asymptotics of the average Erdős number for both mortal and immortal collaborators using a rate equations approach assuming the futile cycles do not exist. To understand asymptotics of solutions, we used results on asymptotic power series. Subsequently, we considered a GTA where the futile cycles were included. Our asymptotic results of the average Erdős number for both mortal and immortal collaborators are consistent with the GTA, which may not be surprising as we assume the collaborations occur instantaneously.

An interesting question is how to extend this methodology to collaborations that do not occur instantaneously. In such a case we may expect the futile cycles to affect the average Erdős number, a possibility we will discuss this possibility in Section 7.

Chapter 7

Conclusions and Future Work

A full mathematical theory that explains coagulation and fragmentation processes in SD remains an open research question. Since the 1960s a considerable amount of work has been done with the aim of providing a theory that would explain the ISDs found both experimentally and in kMC simulations. A validated modelling framework with predictive capabilities would enable the development of new experiments and better control over industrial processes.

In Chapter 4, we have shown that the analysis of some of the properties of the VG generated from a kMC simulation allow us to determine the underlying nucleation mechanism. Both the degree distribution $(q(3))$ and the spectrum of the adjacency matrix reliably allow us to identify the value of i used in the kMC simulation. Therefore, we have created an effective characterisation process that can be applied to experimental data for SD in one dimension, such as island nucleation and growth on a stepped substrate [79].

The VG method has the potential to deal with more complicated mechanisms,including evaporation, mobile islands, the action of electric fields and any level of coverage within the scaling regime as discussed above. The generalisation of our work to

extended islands is also straightforward as we can create the vectors P, used in the construction of VG, by using the position of the centre of mass of an island and its mass as coordinates. We leave these versions of SD to future work.

An important question is how to extend this methodology to two and three space dimensions. In [59] a method is proposed to extend one dimensional VGs to higher dimensions which enables the construction of VGs of large-scale spatially-extended surfaces. The method uses one dimensional VGs along different straight lines in the multidimensional lattice to construct a single VG (only dependent on the number of lines one considers). An extension to two space dimensions is particularly important to industry.

In Chapter 5 we obtained the long term behaviour of clusters and monomers for rate equations with a critical island size i by allowing subcritical islands of size $2 \le j \le i-1$ to form and fragment, complementing the analysis of Section 3.1.2. Additionally, we have proved the ISD converges to a (discontinuous) similarity profile, which is discontinuous at $\frac{(i+1)\beta^{-\frac{i+1}{i+2}}}{i+2}$. The data from kMC simulations show no discontinuity exists for the ISD and hence the divergence discussed does not exist in reality. However, the asymptotics are only as good as the asymptotic sequence of functions used and so, the discontinuity can be smoothed if we use a better family, which may be investigated as future work [30]. Additionally, our asymptotic results in Section 5.5 are consistent with the leading term asymptotics for $c_1(t)$ of [70] (see our Theorem 5.5.1) and for $c_j(t)$ ($1 \le j \le i$ of Section 3.1), as well as with the conjecture in [70] about the behaviour of $c_j(t)$, $j > i$ (see Theorem 5.5.1). In addition, our results are consistent with the work in Section 3.1.1 as $\beta = 0$ is equivalent to letting $i = 1$.

In Chapter 5 we also showed that the asymptotics of solutions obtained in Section 5.5 based on the centre manifold analysis of Section 5.4 can be recovered more easily by making a sweeping assumption that all the clusters of size $1 < j \le i$ are

at a quasi-steady state. To justify the QSSA rigorously, in this case, one would need to write the system in terms of slow and fast variables and use the singular perturbation methods [43].

In Chapter 6, we considered the asymptotics of the average Erdős number for both mortal and immortal collaborators, using a rate equations approach assuming the futile cycles do not exist. To understand asymptotics of solutions, we used results on asymptotic power series. Later on in that chapter, we considered a GTA where the futile cycles were included. Our asymptotic results of the average Erdős number for both mortal and immortal collaborators are consistent with the GTA, which may not be surprising as we assume the collaborations occur instantaneously, as discussed in detail in Example 6.3.1.

This leads to an interesting question about how to extend this methodology to collaborations that do not occur instantaneously. In such a case we may expect the futile cycles to affect the average Erdős number. The simplest way to consider collaborations that do not occur instantly is to consider the system:

$$
\begin{aligned}
&\emptyset \xrightarrow{\hat{\alpha}} u, \\
&e_k + e_m \xrightarrow{\beta} e_k e_m \xrightarrow{b} e_k + e_m, && k \leq m \leq k+1, \\
&e_k + e_m \xrightarrow{\beta} e_k e_m \xrightarrow{b} e_k + e_{k+1}, && m > k+1, \\
&e_k + u \xrightarrow{\gamma} e_k u \xrightarrow{b} e_k + e_{k+1}, \\
&u + u \xrightarrow{\gamma} uu \xrightarrow{b} u + u,
\end{aligned}
$$
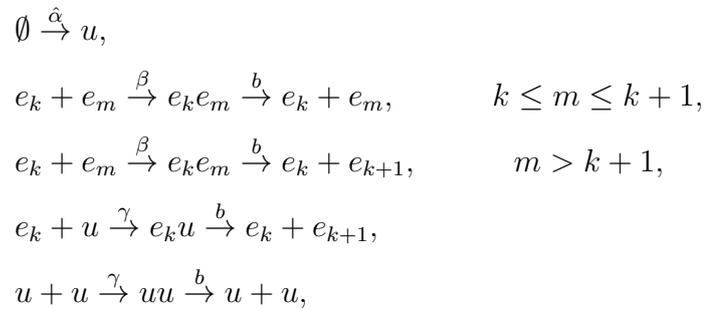
where $m, k \in \mathbb{N}_0$, and analyse the resulting GTA and rate equations. This approach has the potential to lead to useful insight into the effects of futile cycles on coagulation and fragmentation systems.

Additionally, there is work that is not in this thesis where we considered the formation of fouling phytoplankton patches on Navy vessels. This is a serious

Chapter 7. Conclusions and Future Work

practical problem as fouling allows vessels to be easily detected by radar and hence, they have to be docked to wash off unwanted materials. Our model for fouling patches formation is as follows: we consider a Becker-Döring model with deposition,

$$c_1' = \hat{\alpha} + D\left(-2c_1^2 - c_1 \sum_{k=2}^{\infty} c_k\right) + \beta\left(c_2 + \sum_{k=2}^{\infty} c_k\right),$$

$$c_j' = D(c_1 c_{j-1} - c_1 c_j) + \beta(c_{j+1} - c_j), \quad j > 1,$$

where $\hat{\alpha} > 0$ is the deposition rate, $c_j(t) := c_j$ is the density of phytoplankton clusters of size j at time t, $D > 0$ is the coagulation rate and $\beta > 0$ is the fragmentation rate.

Using the QSSA (see Section 2.4), we derived the following result:

$$c_N \sim \left(\frac{\beta}{D}\right)^{\frac{1}{2}} (\theta)^{\frac{1}{2}},$$

where $c_N := \sum_{k=2}^{\infty} c_k$ and $\theta = \hat{\alpha}t$, in agreement with the work of Wattis [96].

Although the QSSA is a largely heuristic technique, it may be said that if the monomer density c_1 is decaying over time as t^{-z}, then any processes that are higher order in c_1 must decay even faster, at the rate t^{-Nz}, so the QSSA can be justified in the long-time limit. That being said, both the paper by Wattis [96] and the QSSA method cannot be considered mathematically rigorous. As discussed in Section 2.4, it is generally accepted that the validity of the QSSA relies on the existence of some 'slow and fast variables', exploring this approach may be of potential interest in the future.

Appendix

Computations of Theorem 5.4.1

In this Appendix we supply the MAPLE code implementing the computations described in the proof of Theorem 5.4.1. We compute 15 terms in the expansion of the one dimensional centre manifold with components given by (5.32) of a system with $i = 5$.

```
n:=15:
```

First of all we set up the equations:

```
eqc1 := a-2*c1^2-c1*z+2*b*c2+b*c3+b*c4+b*c5:
eqc2 := c1^2-c1*c2-b*c2+b*c3:
eqc3 := c1*c2-c1*c3-b*c3+b*c4:
eqc4 := c1*c3-c1*c4-b*c4+b*c5:
eqc5 := c1*c4-c1*c5-b*c5:

eqz  := c1^2-b*c2:
```

```
z := (a-v)/c1:


eqv := -eqc1*z-c1*eqz:


eqc1s := eqc1*c1:

eqc2s := eqc2*c1:

eqc3s := eqc3*c1:

eqc4s := eqc4*c1:

eqc5s := eqc5*c1:


eqvs := simplify(eqv*c1):


v := w-2*b*c2-b*c3-b*c4-b*c5:


eqws := simplify(eqvs+2*b*eqc2s+b*eqc3s+b*eqc4s+b*eqc5s):


eqws := expand(eqws):
```

Now we use the expansions (5.32) and (5.31) and remove all the higher order terms that are not needed in the computation to save time:

```
c2 := sum('g2||j*c1^j','j'=2..n):

c3 := sum('g3||j*c1^j','j'=2..n):

c4:= sum('g4||j*c1^j','j'=2..n):

c5 := sum('g5||j*c1^j','j'=2..n):

w := sum('gw||j*c1^j','j'=2..n):
```

Chapter 7. Conclusions and Future Work

```
aw := collect(simplify(eqc1s*diff(w,c1)-eqws),c1):

ac2 := collect(simplify((w-2*c1^2)*diff(c2,c1)-eqc2),c1):

ac3 := collect(simplify((w-2*c1^2)*diff(c3,c1)-eqc3),c1):

ac4 := collect(simplify((w-2*c1^2)*diff(c4,c1)-eqc4),c1):

ac5 := collect(simplify((w-2*c1^2)*diff(c5,c1)-eqc5),c1):

aw:= convert(taylor(aw,c1=0,n+1),polynom):


ac2:= convert(taylor(ac2,c1=0,n+1),polynom):

ac3:= convert(taylor(ac3,c1=0,n+1),polynom):

ac4:= convert(taylor(ac4,c1=0,n+1),polynom):

ac5:= convert(taylor(ac5,c1=0,n+1),polynom):
```

Finally, we compute the coefficients of the expansion in the order indicated in the proof of Theorem 5.4.1.

```
for k from 2 to n do

    gw||k:= solve(coeff(aw,c1,k),gw||k):

    g5||k:= solve(coeff(ac5,c1,k),g5||k):

    g4||k:= solve(coeff(ac4,c1,k),g4||k):

    g3||k:= solve(coeff(ac3,c1,k),g3||k):

    g2||k:= solve(coeff(ac2,c1,k),g2||k):

od:
```

Now we print out the asymptotic ODE equation for c_1:

```
odec1 := w-2*c1^2;
```

135

Chapter 7. Conclusions and Future Work

The result is

$$c_1' \sim -\frac{c_1^8}{\alpha\beta^4} + \frac{c_1^9}{\alpha\beta^5} - \frac{c_1^{13}}{\alpha\beta^9} + \frac{(30\beta^2 + \alpha)c_1^{14}}{\alpha^2\beta^{10}} - \frac{80c_1^{15}}{\alpha\beta^9} + O(c_1^{16}).$$

Bibliography

[1] M. Aizenman and T. Bak, Convergence to equilibrium in a system of reacting polymers, Commun. Math. Phys. **65** (1979), 203–230.

[2] D. Allen, M. Grinfeld and P. Mulheran, Characterising submonolayer deposition via visibility graphs, PHYSICA A **532** (2019), 121872.

[3] D. Allen, M. Grinfeld and R. Sasportes, Point island dynamics under fixed rate deposition, J. Math. Anal. Appl. **472** (2019), 1716–1728.

[4] N. Alon, Eigenvalues and expanders, Combinatorica **6** (1986), 83–96.

[5] N. Alon and V. Milman, λ_1, isoperimetric inequalities for graphs and super-concentrators, J. Comb. Theory B **38** (1985), 73–88.

[6] J. Amar and F. Family, Critical cluster size: Island morphology and size distribution in submonolayer epitaxial growth, Phys. Rev. Lett. **74** (1995), 2066–2069.

[7] J. Amar, F. Family and P. Lam, Dynamic scaling of the island-size distribution and percolation in a model of submonolayer molecular-beam epitaxy, Phys. Rev. B **50** (1994), 8781–8800.

[8] N. Ardelyan, V. Bychkov, I. Kochetov and K. Kosmachevskii, Pre-breakdown air ionization in the atmosphere, in: *The Atmosphere and*

Bibliography

Ionosphere, V. Bychkov, G. Golubkov and A. Nikitin, Springer, Dordrecht/Heidelberg/London/New York 2010, 69–95.

[9] O. Arino and R. Rudnicki, Phytoplankton dynamics, C. R. Biol. **327** (2004), 961–969.

[10] J. Arthur, Molecular beam epitaxy, Surf. Sci. **500** (2002), 189–217.

[11] G. Bales and D. Chrzan, Dynamics of irreversible island growth during submonolayer epitaxy, Phys. Rev. B **50** (1994), 6057–6067.

[12] J. Ball and J. Carr, The discrete coagulation-fragmentation equations: existence, uniqueness and density conservation, J. Stat. Phys. **61** (1990), 203–234.

[13] M. Balluff and B. Wolf, Degradation of chain molecules. 1. Exact solution of the kinetic equations, Marcomolecules **14** (1981), 654–658.

[14] J. Banasiak, Kinetic models in natural sciences, in *Evolutionary Equations with Applications in Natural Sciences*, Springer, Cham 2010, 83–162.

[15] J. Barrow, Coagulation with fragmentation, J. Phys. A **14** (1981), 729–733.

[16] M. Bartelt and J. Evans, Exact island-size distributions for submonolayer deposition: influence of correlations between island size and separation, Phys. Rev. B **54** (1996), 17359–17362.

[17] C. Beisbart, R. Dahlke, K. Mecke and H. Wagner, *Morphology and Geometry of Spatially Complex Systems*, Springer, Berlin 2002, 238–260.

[18] K. Binder, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, Oxford University Press, New York 1995.

[19] J. Blackman and P. Mulheran, Scaling behaviour in submonolayer film growth: a one dimensional model, Phys. Rev. B **54** (1996), 11681–11692.

Bibliography

[20] J. Blackman and P. Mulheran, Growth of clusters on surfaces: Monte Carlo simulations and scaling properties, Comput. Phy. Commun. **137** (2001), 195–205.

[21] J. Blackman and A. Wilding, Scaling theory of island growth in thin films, Europhys. Lett. **16** (1991), 115–120.

[22] C. Bolton and J. Wattis, The Becker-Döring equations with monomer input, competition and inhibition, J. Phys. A **37** (2004), 1971–1986.

[23] B. Brooks, J. Arch and E. Newsholme, Effects of hormones on the rate of the triacylglycerol/fatty acid substrate cycle in adipocytes and epididymal fat pads, FEBS Lett. **146** (1982), 327-330.

[24] G. Carlsson, Topological pattern recognition for point cloud data, Acta Numerica **23** (2014), 289–368.

[25] J. Carr, *Applications of Centre Manifold Theory*, Springer, US 1981.

[26] G. Chaitin, Randomness and mathematical proof, Scientific American **232** (1975), 47–52.

[27] J. Collet, Some modelling issues in the theory of fragmentation-coagulation systems, Commun. Math. Sci. **1** (2004), 35–54.

[28] F. da Costa, Mathematical aspects of coagulation-fragmentation equations, in *Mathematics of Energy and Climate Change*, CIM Series in Mathematical Sciences, Springer, Cham 2015, 83–162.

[29] F. da Costa, J. Pinto and R. Sasportes, Rates of convergence to scaling profiles in a submonolayer deposition model and the preservation of memory of the initial condition, SIAM J. Math. Anal. **48** (2016), 1109–1127.

Bibliography

[30] F. da Costa, H. van Roessel and J. Wattis, Long-time behaviour and self-similarity in a coagulation equation with input of monomers, Markov Processes And Related Fields **12** (2006), 367–398.

[31] O. Costin, M. Grinfeld, K. O'Neill and H. Park, Long-time behaviour of point islands under fixed rate deposition, Comm. Inf. Syst. **13** (2013), 183–200.

[32] P. van Dongen and M. Ernst, Cluster size distribution in irreversible aggregation at large times, J. Phys. A **18** (1985), 2770–2793.

[33] R. Eckhardt, Stan Ulam and John von Neumann, the Monte Carlo method, Los Alamos Science **15** (1987), 131–137.

[34] A. Eibeck and W. Wagner, Approximative solution of the coagulation-fragmentation equation by stochastic particle systems, Stochastic Anal. Appl. **18** (2000), 921–948.

[35] A. Eibeck and W. Wagner, Stochastic particle approximations for Smoluchowski's coagulation equation, Ann. Appl. Probab. **11** (2001), 1137–1165.

[36] M. Einax, W. Dieterich and P. Maass, Colloquium: cluster growth on surfaces: densities, size distributions, and morphologies, Rev. Mod. Phys. **85** (2013), 921–939.

[37] M. Escobedo, S. Mischler and M. Ricard, On self-similarity and stationary problem for coagulation and fragmentation models, Ann. H. Poincarè Anal. Nonlin. **22** (2005), 99–125.

[38] E. Estrada and P. Knight, *A First Course in Network Theory*, OUP, UK 2015.

[39] J. Evans and M. Bartelt, Nucleation, adatom capture, and island size distributions: Unified scaling analysis for submonolayer deposition, Phys. Rev. B **63** (2001), 235408.

Bibliography

[40] J. Evans and M. Bartelt, Island sizes and capture zone areas in submonolayer deposition: Scaling and factorization of the joint probability distribution, Phys. Rev. B **66** (2002), 235410.

[41] J. Evans, P. Thiel and M. Bartelt, Morphological evolution during epitaxial thin film growth: Formation of 2D islands and 3D mounds, Surf. Sci. Rep. **61** (2006), 1–128.

[42] V. Galiev, A. Polupanov and I. Shparlinski, On the construction of solutions of systems of linear ordinary differential equations in the neighbourhood of a regular singularity, J. Comput. Appl. Math. **39** (1992), 151–163.

[43] A. Goeke, S. Walcher and E. Zerz, Classical quasi-steady state reduction - a mathematical characterization, Physica D **345** 2017, 11–26.

[44] D. Goussis, Quasi steady state and partial equilibrium approximations: their relation and their validity, Combustion Theory Model **16** (2012), 869–926.

[45] J. Gratton, Similarity and self similarity in fluid dynamics, Fundamentals of Cosmic Physics **15** (1991), 1–106.

[46] M. Groce, B. Conrad, W. Cullen, A. Pimpinelli, E. Williams and T. Einstein, Temperature-dependent nucleation and capture-zone scaling of C_{60} on silicon oxide, Surf. Sci. **606** (2012), 53–56.

[47] J. Grossman, The Erdős number project. `http://www.oakland.edu/enp/`.

[48] B. Hayes, Randomness as a Resource, American Scientist, **89** (2001), 300–304.

[49] D. Higham, *An Introduction to Financial Option Valuation: Mathematics, Stochastics and Computation*, Cambridge University Press, London 2004.

[50] R. Isaac, *The Pleasures of Probability*, Springer, New York 2013.

Bibliography

[51] J. Jacquez and C. Simon, Qualitative theory of compartmental systems, SIAM Review **35** (1993), 43–79.

[52] A. Jansen, *An introduction to kinetic Monte Carlo Simulations of Surface Reactions*, Springer, Netherlands 2002.

[53] M. Jerrum and A. Sinclair, Approximating the permanent, SIAM J. Comput. **18** (1989), 1149–1178.

[54] C. Joshi, Y. Shim, T. Bigioni and J. Amar, Critical island size, scaling, and ordering in colloidal nanoparticle self-assembly, Phys. Rev. E **90** (2014), 032406.

[55] I. Kevrekidis, C. Gear and G. Hummer, Equation-free: The computer-aided analysis of complex multiscale systems, AICHE J. **50** (2004), 1346–1355.

[56] M. Körner, M. Einax and P. Maass, Island size distributions in submonolayer growth: successful prediction by mean field theory with coverage dependent capture numbers, Phys. Rev. B **82** (2010), 201401.

[57] M. Körner, M. Einax and P. Maass, Capture numbers and island size distributions in models of submonolayer surface growth, Phys. Rev. B **86** (2012), 085403.

[58] H. Krcelic, M. Grinfeld and P. Mulheran, Distributional fixed-point equations for island nucleation in one dimension: The inverse problem, Phys. Rev. E **98** 2018, 052801.

[59] L. Lacasa and J. Iacovacci, Visibility graphs of random scalar fields and spatial data, Phys. Rev. E **96** (2017), 012318.

[60] L. Lacasa, B. Luque, F. Ballesteros, J. Luque and J. Nuño, From time series to complex networks: the visibility graph, Proc. Nat. Acad. Sci. **105** (2008), 4972–4975.

Bibliography

[61] W. Lamb, Existence and uniqueness results for the continuous coagulation and fragmentation equation, Math. Meth. Appl. Sci. **27** (2004), 703–721.

[62] F. Leyvraz, Scaling theory and exactly solved models in the kinetics of irreversible aggregation, Phys. Rep. **383** (2003), 95–212.

[63] H. Maeda, S. Kodama and Y. Ohta, Asymptotic behavior of nonlinear compartmental systems: Nonoscillation and stability, IEEE Trans. Circ. Systems **25** (1978), 372–378.

[64] G. Menon and R. Pego, Approach to self-similarity in Smoluchowski's coagulation equations, Comm. Pure Appl. Math. **57** (2004), 1197–1232.

[65] G. Menon and R. Pego, Dynamical scaling in Smoluchowski's coagulation equations: uniform convergence, SIAM J. Math. Anal. **36** (2005), 1629–1651.

[66] P. van Mieghem, *Graph Spectra for Complex Networks*, CUP, Cambridge 2011.

[67] S. Miyamoto, O. Moutanabbir, E. Haller and K. Itoh, Spatial correlation of self-assembled isotopically pure Ge/Si(001) nanoislands, Phy. Rev. B **79** (2009), 165415.

[68] P. Mulheran, The dynamics of island nucleation and growth-beyond mean field theory, EPL **65** (2004), 379–385.

[69] P. Mulheran, Theory of cluster growth on surfaces, in *Metallic Nanoparticles (Handbook of Metal Physics vol 5)*, Amsterdam: Elsevier 2009, Cham 2015, 73–111.

[70] P. Mulheran and M. Basham, Kinetic phase diagram for island nucleation and growth during homoepitaxy, Phys. Rev. B **77** (2008), 075427.

Bibliography

[71] P. Mulheran and J. Blackman, Capture zones and scaling in homogeneous thin-film growth, Phys. Rev. B **53** (1996), 10261.

[72] P. Mulheran, K. O'Neill, M. Grinfeld and W. Lamb, Distributional fixed-point equations for island nucleation in one dimension: A retrospective approach for capture-zone scaling, Phys. Rev. E **86** (2012), 051606.

[73] P. Mulheran and D. Robbie, Theory of the island and capture zone size distributions in thin film growth, Europhys. Lett. **49** (2000), 617–623.

[74] M. Navas, S. Cerdán and J. Gancedo, Futile cycles in Saccharomyces cerevisiae strains expressing the gluconeogenic enzymes during growth on glucose, PNAS **90** (1993), 1290-1294.

[75] H. Niwa, School size statistics of fish, J. Theor. Biol. **195** (1998), 351–361.

[76] K. O'Neill, *Island Nucleation and Growth during Submonolayer Deposition*, Ph. D. Thesis, University of Strathclyde, Glasgow, UK 2012.

[77] C. Pantea, A. Gupta, J. B. Rawlings and G. Craciun, *The QSSA in Chemical Kinetics: As Taught and as Practiced*, Springer, Berlin/Heidelberg/Berlin/Heidelberg 2014, 419–442.

[78] A. Pimpinelli and T. Einstein, Capture-zone scaling in island nucleation: universal fluctuation behaviour, Phys. Rev. Lett **99** (2007), 226102.

[79] C. Pownall and P. Mulheran, Simulation and theory of island nucleation, growth, and coalescence on stepped substrates, Phys. Rev. B **60** (1999), 9037.

[80] S. Pratsinis, Flame aerosol synthesis of ceramic powders, Prog. Energy Combust. Sci. **24** (1998), 197–219.

Bibliography

[81] H. Pruppacher and J. Klett, *Microphysics of Clouds and Precipitation*, D. Reidel Publishing Company, Dordrecht 1997.

[82] C. Ratsch and J. Venables, Nucleation theory and the early stages of thin film growth, J. Vac. Sci. Technol. **21** (2003), 96–109.

[83] D. Ropers, V. Baldazzi and H. Jong, Reduction of a kinetic model of the carbon starvation response in escherichia coli, IFAC Proc. **42** (2009), 27–32.

[84] R. Rubinstein, *Simulation and The Monte Carlo Method*, Wiley, New York 1981.

[85] T. Saaty and L. Varga, *Decision Making with the Analytic Network Process*, Springer, New York 2013.

[86] W. Scott, Analytic studies of cloud droplet coalescence, I. J. Atmos. Sci. **25** (1968), 54–65.

[87] P. Seba, Parking in the city, Acta Physica Polonica A **112** (2007), 681–690.

[88] L. Segel and M. Slemrod, The quasi-steady state assumption: A case study in perturbation, SIAM Review **31** (1989), 446–477.

[89] L. Smith, W. Lamb, M. Langer and A. McBride, Discrete fragmentation with mass loss, J. Evol. Equ. **12** (2012), 181–201.

[90] M. Smoluchowski, Drei Vorträge über Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteilchen, Physik. Zeitschr., **17** (1916), 557–571, 587–599.

[91] M. Smoluchowski, Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen, Zeitschrift f. physik. Chemie, **92** (1917), 129–168.

Bibliography

[92] H. Thieme, Convergence results and a Poincaré–Bendixson trichotomy for asymptotically autonomous differential equations, J. Math. Biol. **30** (1992), 755–763.

[93] C. Thompson, C. Palasantzas, Y Feng, S. Sinha and J. Krim, X-ray reflectivity study of the growth kinetics of vapor-deposited silver films, Phys. Rev. B **49** (1994), 4902–4907.

[94] J. Venables, G. Spiller and M. Hanbucken, Nucleation and growth of thin films, Rep. Prog. Phys. **47** (1984), 399–459.

[95] R. Vigil and R. Ziff, On the stability of coagulation-fragmentation population balances, J. Colloid Interface Sci. **133** (1989), 257–264.

[96] J. Wattis, Similarity solutions of a Becker-Döring system with time dependent monomer input, J. Phys. A **37** (2004), 7823–7841.

[97] J. Wattis, An introduction to mathematical models of coagulation-fragmentation processes: A discrete deterministic mean-field approach, Physica D **222** (2006), 1–20.

[98] W. Wasow, *Asymptotic Expansions for Ordinary Differential Equations*, Dover Publications, New York 1965.

[99] B. Wichmann and D. Hill, Algorithm AS 183: An efficient and portable pseudo-random number generator, J. Royal Stat. Soc. **31** (1982), 188–190.

[100] Y. Yang, J. Wang, H. Yang and J. Mang, Visibility graph approach to exchange rate series, Physica A **388** (2009), 4431–4437.

[101] R. Ziff, An explicit solution to a discrete fragmentation model, J. Phys. A **25** (1992), 2569–2576.

Bibliography

[102] R. Ziff and E. McGrady, Kinetics of polymer degradation, Macromolecules **19** (1986), 2513–2519.

[103] Y. Zou, M. Small, Z. H. Liu and J. Kurths, Complex network approach to characterize the statistical features of the sunspot series, New J. Phys. **16** (2014), 013051.

[104] A. Zsom and C. Dullemond, A representative particle approach to coagulation and fragmentation of dust aggregates and fluid droplets, Astronomy and Astrophysics **489** (2008), 931–941.