Effects of Diffusion-limited Aggregation on Bacterial Systems

Miguel Garrido Zornoza

Supervised by Namiko Mitarai

Niels Bohr Institute

Faculty of Science

University of Copenhagen

May, 2021

A dissertation submitted in partial fulfilment of the requirements for the degree of M.Sc. in Physics.

To My Family

You know, for saving the world.

"... If the scientist had infinite time at her/his disposal, it would be sufficient to say to her/him, "Look, and look carefully" . But, since she/he has no time to look at everything, and above all to look carefully, and since it is better not to look at all than to look carelessly, she/he is forced to make a selection. The first question, then, is to know how to make this selection. This question confronts the physicist as well as the historian ... Trying to make science contain nature is like trying to make the part contain the whole, but scientists believe that there is a hierarchy of facts, and that a judicious selection can be made ... the more general a law is, the greater is its value. This shows us how our selection should be made. The most interesting facts are those which can be used several times, those which have a chance of recurring. We have been fortunate enough to be born in a world where there are such facts ... Which, then, are the facts that have a chance of recurring? In the first place, simple facts. It is evident that in a complex fact many circumstances are united by chance, and that only a still more improbable chance could ever so unite them again. But are there such things as simple facts? ..."

***Henry Poincaré** , Science and Method*

Acknowledgements

Special acknowledgements go to Namiko Mitarai, the creator and guide of the project, to whom I am grateful for letting me be a (small) part of the Biocomplexity Group throughout the duration of this dissertation. Another special acknowledgement goes to my *tío* Jose.

Abstract

When macromolecules are suspended in a fluid, they perform a random walk on account of the molecular movement of the fluid constituents. On average, these molecules move towards low particle concentration regions. This phenomenon is classically know as diffusion. In this work, we addressed the role this type of mass transport can play in the formation of bacterial colonies to find that, within a given density of bacteria, it can be a player of the same magnitude than exponential growth, and provokes the acceleration of bacterial clustering when overlapped to it. We, additionally, addressed the ecological implications to bacteria caused by the diffusive formation of colonies when these are exposed to predation by phages. We quantified a delay in the population extinction based on the increased viral searching times associated to bacterial clustering. Although this discards diffusive-limited aggregation as a solid passive surviving mechanism, it opens the possibility of colony survival upon a transient phage attack.

Contents

List of Figures

List of Tables

List of Abbreviations

SM Simple Model
BAM Bacteria Aggregation Model
PPAM Predator-Prey Aggregation Model

1

Introduction

Bacteria are prokaryotic [1] microorganisms with a typical lenght of $\sim 1\mu m$, mostly sphere or rod-shaped. Prokaryotes are ubiquitous in nature (1), and so are bacteria. Even in the human body, they are estimated to hold a $1:1$ ratio with respect to human cells (2). An important feature in order to understand how bacteria evolve is via their growth dynamics. They are usually understood, or at least explained, from the perspective of the growth patterns from cultured groups in laboratory conditions. For the interested reader we refer to (3). Since bacteria are the cornerstone of this work, it is of importance to briefly comment on these ways of bacterial growth, since the discussion involves concepts we will return to recursively.

Imagine we are in our laboratory. If we consider a low number of *E.coli* cells and put them into a transparent liquid with the proper growth conditions (this is, the necessary nutrients for the system to undergo cell division), the medium would gradually become turbid. This change in the liquid's optical properties is caused by cellular growth, and quantified by measuring the optical density[2] (OD) of the medium. If we plot against time in a semi-logarithmic scale the optical density (which is assumed to be proportional to the total cell mass), it will show a sigmoidal growth curve (see Figure 1.1). The first plateau is known as *lag phase*, and it corresponds to the time it takes our cells to adapt to their new environment (more precisely, to synthesise the necessary proteins to undergo cell division). Following the lag phase one would encounter a linear increase in OD, which in a semi-log scale translates to exponential growth, hence the received name

[1] Single-celled organism without a membrane-enclosed nucleus

[2] The optical density is a rescaled transmitance, defined as $OD_{600} = -log_{10}T(\lambda)$. The subscript indicates that the sample has been measured at a $\lambda = 600nm$. This wave length is usually used in spectrophotometry to estimate the cell concentration in a culture, since does little or no damage and has no interference with the growth pattern.

of *exponential phase*. Here the culture's rate of growth is proportional to the population number, as we shall see in the subsequent subsection. Since these types of growing experiments are usually performed on batch cultures [3], once growing cells have consumed most nutrients they enter the *stationary phase*, where the curve reaches a second plateau. Consequently, the death and growth rates are equal and the population number remains unchanged.

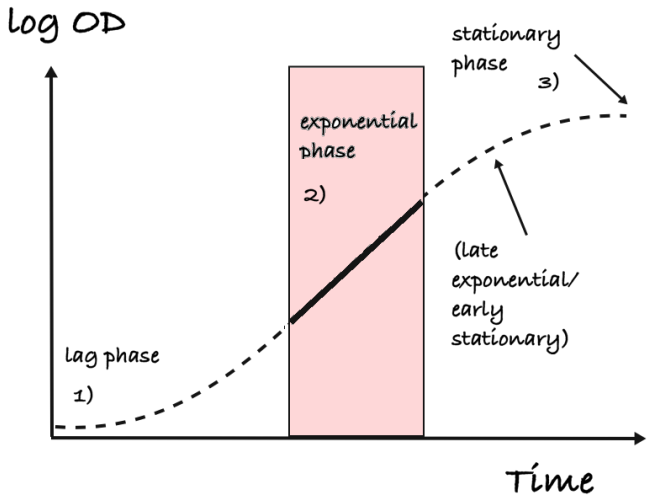


Figure 1.1: Growth curve, displaying the different growth phases of the cell culture.

As we said, bacteria are usually studied in liquid batch cultures. A recent paper by K. N. Kragh *et. al.* (4) investigates how different methods used to inoculate bacteria into these cultures could influence the development of biofilm [4] aggregates. In their research, they discover an impact of the inoculation method on the fraction of aggregated biomass in a liquid bacterial culture. Furthermore, cultures with more aggregated biomass show an increase of survival rates when exposed to the antibiotic tobramycin, a trait in accordance with biofilm-forming bacteria (5). The suggested mechanism of aggregate formation, based on observation, is the recruitment of single cells from the

[3] Technique to grow bacteria where a finite amount of nutrients is supplied. When these are used up cells stop growing in an exponential manner and the culture is then ready to be harvested.

[4] A biofilm is a way to label microbial aggregates wrapped in a matrix of extrecellular substances. These substances form a mechanical scaffold by providing adhesion to surfaces. This type of consortia develop a variety of properties, such as protection against some antibiotics (5).

surroundings of the aggregate.

From this investigation starts our project. More precisely, from this last thought. Can we explain the formation of bacterial aggregates via basic physical principles? If so, what is the impact of the different colony-forming mechanisms in the overall aggregate distribution? We based our approach on a basic principle know as *diffusion*, which we will briefly describe, and proceeded to study its impact and outreach when it comes to the formation of aggregates.

Since we can not characterize every layer of a system in a single framework or model, we made a selection of the length scale where the system would be studied. We were not interested in the specific bacteria-bacteria mechanical interactions, or the emergent colony properties that can fundamentally change bacterial growth via the presence of chemical gradients of nutrients (food) within the aggregate (6). Our aim turned out to be at a different length scale, where a coarse-grained description of bacteria as spherical colloids was acceptable (see Figure 1.2).

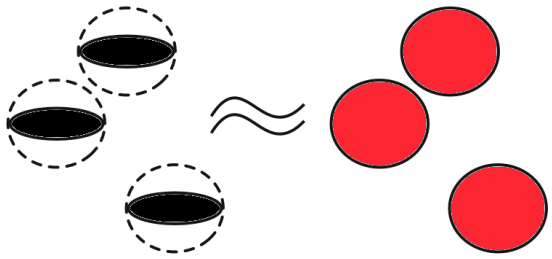


Figure 1.2: Spherical approximation. Bacteria, which display different morphology and size, are here approximated as spherical colloids of fixed radius and a movement ruled by diffusion.

At this scale, where the detailed interactions are not described, we focus on zoomed out events, described by rates. A particular rate encompasses an average behavior of the system in relation to a specific event. This can be, for example, how often do aggregates stick to each other upon collision, or the frequency of cell division, without diving into a molecular description of the different mechanisms at play. Consequently, we built a mass-action kinetics model [5] where colloidal particles (representing bacteria) could collide and merge to form bigger aggregates, as well as grow. We will address this model

[5] A mass action kinetics model is a kinetic model where events are proportional to the densities of the involved reactants.

as the Bacteria Aggregation Model (BAM). This was aimed to describe bacterial clustering, accomplished by both the growth in size of the colony via aggregation events, ruled by diffusion, and the growth and division of single cells embedded in the aggregate.

Environmental bacteria, however, might present themselves quite differently than in laboratory conditions. A concern raised given the impossibility to study most microbes in laboratory conditions (7) as well as the fact that bacteria are unlikely to encounter reasonable growth conditions outside batch cultures found in laboratories (8). This becomes evident by noticing that a single *E.coli* cell, if growing in the exponential phase, would grow to a population equivalent to the mass of the Earth within 2 days (9), a statement that obviously disagrees with both common sense, and the estimated $(4-6)\cdot 10^{30}$ (1) prokaryotes on Earth. This discrepancy between potential and actual behavior underscores the existence of energy-limited ways for bacteria to survive (9).

However, bacterial population levels are not fully explained via energy-limited mechanisms. They are part of, at least, one predator-prey system. So, then, we ask, who's eating them? In 1989 viruses came along as a plausible answer to this question, with the discovery that there are $\sim$ 10 million virus-like particles (VLPs) in every mL of sea water (10). Most of these particles are assumed to be phages (virus that infect bacteria), since viral abundance correlates with microbial presence with a ratio of $5-10$ VLPs per bacteria (11). There is an estimate of 10^{31} viruses on Earth (12), and they are considered to kill $4-50\%$ of the bacteria produced every day (13; 14).

With this in mind, and as we built the former bacterial aggregation framework, we raised the second question of this work. What are the ecological implications to bacteria associated with the formation of structured environments when exposed to predation by bacteriophages?

It has been shown in the past that bacteria-phage dynamics are fundamentally different in spatially structured environments (15; 16; 17), so, has our question already been addressed?

Former studies on the role of spacial structures for bacteria-phage fate have suggested microcolonies as a survival mechanism (16). However, these studies have focused on semisolid medium where bacteria are immobile and do not diffuse. This allows for a static description where bacteria grow into spherical aggregates, keeping phage infection on the surface. Prokaryotes are, on the other hand, ubiquitous in liquid medium

such as marine and fresh water (lakes, rivers, open ocean) (1). This represents a different paradigm and thus the need for a different mathematical description that includes bacterial diffusion and, therefore, aggregation events.

In order to address this question we used the model of bacteria aggregation (BAM) as a basis to incorporate bacteria-phage interactions. This expanded model, which we will label as Predator-Prey Aggregation Model (PPAM), considers phages as smaller (than bacteria) spherical colloidal particles that perform a diffusive search and encounter with bacteria. The concrete bacterium-phage interactions and the way to quantify them shall be left to future sections.

Document Structure

The project can be divided into different stages.

First - Selection and development of the tools to be used in order to fulfill the proposed objective. In **Chapter 2** we review the conceptual and mathematical background necessary to understand the following section. We first establish the mathematical grounds of mass aggregation, followed by the estimation of the aggregation rate between colloids assuming a diffusive transport mechanisms, labeled as *brownian kernel*. We then refresh the concept of *diffusion*, which will be used throughout the whole dissertation. We continue with a brief comment on the biological factors at play, *i.e.,* viruses and bacteria, in order to understand how to model their interplay. To properly identify aggregation traits on phage-bacteria dynamics we must first conform a background aggregation-less model "template" to be used as comparison. Here is where we go over a simple predator-prey model that will be used as a comparison to study aggregation effects on virus-bacteria interactions. In **Chapter 3** we make use of the reviewed material to build the models from which we shall address our research questions.

Second - Computer simulations, analysis and comment on the results. The models built in Chapter 3 were simulated on a specific range of parameters. In **Chapter 4** we explain step by step the procedure we followed to launch those simulations and the posterior analysis of the results.

Third - Conclusion. In **Chapter 5** an overall review of the project is made, followed by the accomplished goals and a critique of our models. Speculations about future work follow to end the project.

2

Background

Before getting to the development of the framework that we used to address our research questions, it is of importance to explain the theory required to understand its conceptual and analytical basis. In this sense, we review a kinetic way of describing mass aggregation, followed by a refreshment of the concept of diffusion, the principle that will determine the transport mechanism of bacteria and phages. From the conceptual definition and the equations describing this type of mass transport, we shall recover an important result, derived by Smoluchowski, which quantifies the rate at which two colloidal particles of different sizes coalesce. This result shall then be coupled, in the next chapter, to the kinetic aggregation framework and bacterial growth to establish the BAM.

We, then, will briefly stop and explain the interplay of the biological factors that conform the system of study in order to lay the grounds for the BAM expansion into the PPAM.

2.1 | Aggregation: Mathematical Description

We now start by finding a way to describe bacterial aggregation. We shall make a small abstraction and think of bacteria as monomers of some undetermined unit mass. These monomers will bump into each other to form a bigger aggregate, which shall be characterized by an integer number, representing the number of its constituents. With this in mind, we now describe a more general framework, where bacteria are not mentioned at all. This is a fairly well-established framework that can be found in literature such as (18).

In a kinetic description, the dynamics of aggregation involves a sequence of coalescence events represented as

$$A_i + A_j \xrightarrow{K_{ij}} A_{i+j},$$

where two active aggregates of mass i and j merge into one cluster of mass $i + j$ (Fig. 2.1), irreversibly losing their identity, as opposed to a scattering event. The process is regulated by an intrinsic aggregation rate, K_{ij}, know as the kernel of the reaction $(i + j)$. A reaction or event rate in a kinetic description is the probability of that event happening per unit time. The concise value of a rate is determined by each of the underlying mechanisms that make up the event associated to it. For example, in a chemical reaction we would find two time-scales influencing the reaction rate between reactants. The first one being the time of the (typically diffusive) searching process, and the second the actual reaction-upon-encounter time. The dominating time-scale (if any) would then define the value of the rate. We shall think of aggregation as an instantaneous event, thus the aggregation rate will be purely determined by the transport mechanism of the aggregates, *i.e.,* diffusion or ballistic movement.

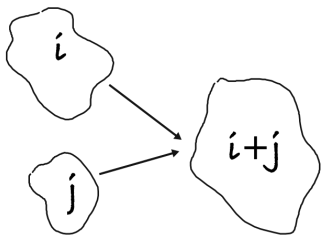


Figure 2.1: Irreversible merging of two aggregates of size i and j into one of mass $i + j$ with $(i, j) \in \mathbb{N}$.

In this description the total mass is measured in positive integer units, *i.e.*, everything is expressed in terms of an undetermined minimal monomer mass. Hence, a cluster of mass k, or a *k-mer*, is made out of k minimal mass units. The aggregates are referred to as *clusters* and it's density denoted as c_k. As in any dynamical study, the aspiration is to calculate the time evolution of the key observables. In this description, one aims to determine the time evolution of the distribution of cluster sizes, $c_k(\vec{r}, t)\ \forall\ k \in [1, \infty)$, for a given aggregation kernel, which is assumed to be known. The time evolution of the system is dictated by an infinite set of coupled nonlinear differential equations

analytically soluble for a few simple cases such as $K_{ij} = \text{constant}$, $K_{ij} = ij$ or $K_{ij} = i + j$. This set of master equations reads as

$$\frac{dc_k}{dt} = \frac{1}{2} \sum_{i+j=k} K_{ij} c_i c_j - c_k \sum_{i=1}^{\infty} K_{ik} c_i \qquad (k = 1, .., \infty). \tag{2.1}$$

The first term accounts for the gain in mass concentration from the coalescence of two clusters such that $i + j = k$. The second term is a loss term accounting for the reaction of clusters of mass k with any other cluster. These equations incorporate some underlying assumptions on the system's behaviour that we will now briefly comment.

Assumptions:

- *No inactive clusters:* all clusters are treated as active reactants. Hence, interaction probabilities, which dictate the rates, include total cluster size concentrations.
- *Geometry independence:* the only dynamical variable is the cluster mass. This assumption is not physically justifiable, with the exemption of systems with spherical symmetry, such as those where surface tension plays a key role on the aggregate's shape. Furthermore, this approach tacitly assumes a diffusive transport mechanism, since for ballistic aggregation each cluster is characterized by its mass and velocity, and this last one should, in such case, appear in the equivalent set of equations.
- *Well-mixed system:* the probability of two clusters interacting factorizes into the product of individual cluster size probabilities which, in a non-normalized way, corresponds to the concentrations of their respective cluster sizes. This interaction probability, as we can see, is weighted by the reaction kernel K_{ij}, which contains information on the intrinsic character of the system and the aggregates, beyond their actual concentrations. Since there is spatial homogeneity we get that $c_k(\vec{r}, t) = c_k(t)$, thus the rate of interaction between two clusters of sizes i and j takes the form $\sim c_i(t) c_j(t) K_{ij}$. We make here no assumptions regarding the functional dependence of the reaction kernel.
- *Dilute system:* aggregation events are second order cluster collisions. This is only asymptotically correct, as the number of aggregates decreases with time and merge into a single cluster. Hence, a system initially dense eventually becomes dilute and higher order cluster collisions become negligible. This approach would turn out

to be fundamentally wrong if higher order interactions played an essential role in the dynamics. An example would be catalyst-assisted reactions.

As analytic solutions are only available for a narrow palette of functional forms of the reaction kernel, it is worth looking at some general properties of aggregation by looking at the moments of the cluster size distribution,

$$M_n(t) = \sum_{k\geq 1} k^n \cdot c_k(t) \, . \tag{2.2}$$

Core information about the mass distribution can be extracted; for example, a typical definition for the average cluster mass is

$$\langle k \rangle = \sum k \cdot P_k = \frac{\sum k^2 \cdot c_k}{\sum k \cdot c_k} = \frac{M_2(t)}{M_1(t)} \, . \tag{2.3}$$

Where the probability of finding a cluster of size k, P_k, has been set equal to the number of monomers belonging to clusters of such size, normalized by the total mass $M = \sum k \cdot c_k$. A less intuitive way of defining this quantity is the ratio of the first and zero-th moments M_1/M_0. For our aggregation framework the equation for the moments reads as

$$\frac{dM_n}{dt} = \sum_{k=1}^{\infty} k^n \frac{dc_k}{dt} = \frac{1}{2} \sum_{i\geq 1} \sum_{j\geq 1} (i+j)^n K_{ij} c_i c_j - \sum_{k\geq 1} \sum_{i\geq 1} k^n K_{ik} c_k c_i. \tag{2.4}$$

The irreversibility of the process and the asymptotic disappearance of all clusters into one of infinite size can be predicted by the behaviour of the zero-th moment of the distribution

$$\frac{dM_0}{dt} = -\frac{1}{2} \sum_{i\geq 1} \sum_{j\geq 1} c_i c_j K_{ij} \tag{2.5}$$

which, as $(c_i,\ c_j,\ K_{ij}) \geq 0$, remains negative unless $c_k = 0 \ \forall k \in [1, \infty)$. Furthermore, the behavior of the first moment predicts mass conservation, since

$$\frac{dM_1}{dt} = 2 \cdot \frac{1}{2} \sum_{i\geq 1} \sum_{j\geq 1} i K_{ij} c_i c_j - \sum_{k\geq 1} \sum_{i\geq 1} k K_{ik} c_k c_i = 0. \tag{2.6}$$

Consequently, $M_1(t) = \sum k c_k(t) = M_1(0)$, and the total number of monomers is conserved. This is a good property for sanity checks in computational simulations involving aggregation events.

We now have a set of equations that conform a kinetic description of aggregation. This description relies on the aggregation kernel, K_{ij}, which, as we said, contains intrinsic information about the system. This information can, for example, contain idiosyncrasies in the way monomers react. Following the example illustrated at (18), imagine that monomers have f reactive parts (see Figure 2.2). This is, they have a limited number of reactive regions, and they are indistinguishable from each other. When they merge, the arising dimer will have $2f - 1$ reactive parts, a trimer $3f - 2$ and a k-mer will end up with $kf - 2(k-1) = (f-2) \cdot k + 2$. The merging of two clusters of sizes i and j has a reaction rate equal to the product of their reactive parts, which is

$$K_{ij} = [(f-2) \cdot i + 2][(f-2) \cdot j + 2] = (f-2)^2 ij + 2(f-2)(i+j) + 4\,.$$

This is a linear combination of the product, $K_{ij} = ij$, sum, $K_{ij} = i + j$, and constant kernels.

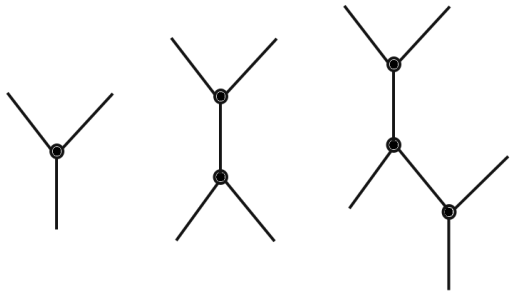


Figure 2.2: Aggregates of 3 functional units. From the left: monomer, dimer and trimer.

In the same way that the aggregation kernel can describe limiting ways of reacting between monomers, based perhaps on anisotropies, it can also incorporate information regarding the transport mechanism by which monomers collide, and it can do so as a function of the reactant sizes. The transport mechanism of our choice is diffusion, but before estimating the value of the aggregation kernel in a diffusion-limited paradigm, let's quickly review the concept, and the phenomenological equations used to describe it.

2.2 | Diffusion

Bodies that can be easily observed in a microscope, when suspended in a liquid, perform measurable displacements on account of the molecular movements of the fluid constituents, as suggested by Einstein and measured by Perrin (19; 20). These displacements follow a random walk, or *Brownian movement*, which, in the continuum limit, provokes the spread of particles from high to low particle concentration regions. This phenomenon is known as *diffusion*.

There are two differential equations that respond to this spatial and temporal variation of nonuniform particle distributions, $c(\vec{x}, t)$. These are Fick's first and second equations[1], and we directly introduce them both.

- Fick's first equation: without particle-particle interactions or external drift, the net particle flux per unit area and time is proportional to the particle concentration gradient

$$\vec{j}(\vec{x}, t) = -D\vec{\nabla}c(\vec{x}, t), \tag{2.7}$$

  with the proportionality constant equal to $-D$, where $D \geq 0$ is the so-called diffusion coefficient. This equation describes the particle movement from high to low concentration regions.

- Fick's second equation or Diffusion equation: obtained by imposing particle number conservation, tells us how a distribution of particles redistributes itself in time, and reads as

$$\frac{\partial c(\vec{x}, t)}{\partial t} = D\nabla^2 c(\vec{x}, t)\ . \tag{2.8}$$

It is worth mentioning that by accepting Fick's equations as a proper way to describe dynamics in the upcoming sections, we are neglecting the hydrodynamic effect (22), present at low Reynold's number, as well as the problem of sedimentation (23), which

[1] They can be derived on account of the random walk model in a very simple way (21).

might introduce a drift into the system [2]. However, for the purposes of this dissertation, these two equations shall be enough to work in a diffusive regime.

2.2.1 | Smoluchowski's Theory of Coagulation: the Brownian Kernel

We now know what diffusion is, and which is the way to describe it analytically. We are, therefore, in the position to dive into Smoluchowski's estimation of colloidal interactions, which we shall use in the subsequent chapter as the diffusion-limited aggregation rate between our colloids (bacteria).

A statistical description of coagulation (or sticky collisions) of two different diffusing particles was developed by Marian Smoluchowski (24; 25) as an application of the theory of Brownian motion. The development of this framework was based on the idea that each colloidal particle is surrounded by an *sphere of influence* of some diameter such that the particle follows a free Brownian motion unless another particle enters the sphere, in which case both colloids merge into a new single and bigger particle (23). Accepting this assumption, the Brownian coagulation process can be divided into two steps: encounter and reaction - if the reaction rate is high then the diffusive encounter dominates the time scale of the process. This type of reaction is known as *diffusion-limited*, and we shall consider coagulation as such. Prior to the general problem we first consider a limiting situation, where one of the two types of particles is big enough to be considered static. We then make the following question: how many colloids of the smaller type bump against one of these bigger static objects of radius R? For this we need to integrate the particle flux on the object's surface, which is related to the particle concentration by Fick's first law

$$j_1 = -D(\hat{n} \cdot \vec{\nabla} c(\vec{r}, t)). \tag{2.9}$$

[2] If, for example, the external field is gravity along the z direction, diffusion in the (x, y) plane would occur as in the field free case. However, the equation for the z direction would have to be modified resulting in

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} + c\frac{\partial c}{\partial z}$$

with $c = (1 - \rho_0/\rho) \cdot (g/\beta)$. Here $-(1 - \rho_0/\rho) \cdot g$ corresponds to the acceleration caused by the external force field (thus ρ_0 and ρ are the densities of the fluid and brownian particle, respectively), and β is the dynamical friction coefficient, a friction that experiences any moving particle by the surrounding medium and that is usually assumed to follow Stoke's law. Thus, for a particle of mass m and radius r, $\beta = 6\pi r\eta/m$, where η is the dynamic viscosity of the surrounding fluid.

This is the rate at which particles describing a Brownian motion coalesce with the static target, per unit area and unit time ; $\hat{n}$ is an unitary vector perpendicular to the object's surface. Note that we adopted a slightly different notation for the particle concentration, anticipating the spherical symmetry of the problem. The concentration of particles in turn obeys the diffusion equation

$$\frac{\partial c(\vec{r},t)}{\partial t} = D\nabla^2 c(\vec{r},t), \tag{2.10}$$

whose time-dependent solution can be expressed analytically assuming spherical symmetry, $c(\vec{r},t) = c(r,t)$, absorbing boundary conditions on the particle's surface, $c(r = R,t) = 0$, and a fixed concentration far away from where the reaction takes place, $c(r \to \infty,t) = c_0$, *i.e.,* in the thermodynamic limit. Equation 2.10 has a solution of the form (see A.2)

$$c(r,t) = c_0 \left[1 - \frac{R}{r} + \frac{2R}{r\sqrt{\pi}} \int_0^{(r-R)/2\sqrt{Dt}} e^{-x^2} dx\right], \tag{2.11}$$

with $c_0 = c(r, t = 0)$. The total rate on the static object's surface is computed by summing over the surface of a sphere of radius R

$$J_1 = \int_\Omega j_1 \, d\Omega \Big|_R = 4\pi D \left(r^2 \frac{\delta c}{\delta r}\Big|_{r=R}\right). \tag{2.12}$$

The derivative is calculated by remembering the first fundamental theorem of calculus

$$\frac{d}{dx}\int_{a(x)}^{b(x)} f(t)dt = \frac{d}{dx}\left(F(b(x)) - F(a(x))\right) = f(b(x)) \cdot b'(x) - f(a(x)) \cdot a'(x).$$

Hence,

$$r^2 \frac{\delta c}{\delta r}\Big|_{r=R} = c_0 \cdot \left(\frac{R}{r^2} \cdot r^2 - \frac{2R}{r^2\sqrt{\pi}} \cdot r^2 \int_0^{(r-R)/2\sqrt{Dt}} e^{-x^2} dx + \frac{2R}{r\sqrt{\pi}} \cdot r^2 \left[\frac{e^{(r-R)/2\sqrt{Dt}}}{2\sqrt{Dt}} - e^0 \cdot 0\right]\right)_{r=R}$$

$$= R \cdot c_0 \cdot \left(1 + \frac{R}{\sqrt{\pi Dt}}\right),$$

and

$$J_1\, dt = 4\pi \cdot D \cdot R \cdot c_0 \cdot \left(1 + \frac{R}{\sqrt{\pi D t}}\right)\, dt. \tag{2.13}$$

This is the reaction rate for one particle. If we consider all particles from the static type, in a well-mixed scenario and a concentration c_s, the total aggregation rate reads as

$$J\, dt = J_1 \cdot c_s\, dt. \tag{2.14}$$

The static target limit is a reasonable assumption when one of the particles is considerably larger than the other, making its movement negligible within the smaller colloid's diffusion time scale [3]. If both particle types (i, j) are similar in size the big particle that we have treated to be static describes now a Brownian motion, characterized by some diffusion constant, D_j. On the other hand, the former point-like smaller colloids shall now have their own radii of influence. Consequently, the problem has to be reformulated by replacing R by the sum $R_{ij} = (R_i + R_j)$ and D by the effective diffusive coefficient $D_{ij} = (D_i + D_j)$ in 2.13 (see A.3). Here the subscripts i and j have been used to label different types of colloids which, in an aggregation framework, correspond to the number of constituents of each aggregate, *i.e.*, a i-mer and a j-mer, respectively. The reaction rate equation for the coalescence events of all colloids of types (i, j) reads as

$$J_{i+j} dt = 4\pi D_{ij} R_{ij} c_i c_j \left(1 + \frac{R_{ij}}{\sqrt{\pi D_{ij} t}}\right) dt\ . \tag{2.15}$$

This reaction rate will be regarded as the *Brownian kernel*.

We now have the basis for our description for bacterial aggregation, but we are lacking the knowledge to incorporate growth into the scheme, as well as interaction with phages for our second research question. Let's, therefore, explain a bit more about how viruses interact with bacteria, and learn how bacterial growth can be modelled.

[3] A rough estimate can be computed with the ratio of variances of two colloids of different size, for the same displacement (see equation A.17 in A.1). This is $t_i/t_j \sim D_j/D_i \sim (j/i)^{1/3}$. Thus, a difference in 3 orders of magnitude in monomer number translates to a 10-fold increase of the time to get to the same place, on average.

2.3 | Biological Factors

2.3.1 | Phage-Bacteria Interplay

Viruses are obligate [4] intra-cellular parasites constituted by a protein shell (or capsid) filled with genetic material (DNA or RNA). These parasites reprogram the host cell metabolism, in a pathogen-specific manner, to increase the energy supply needed for the synthesis of viral components, such as nucleic acids and proteins, needed for the subsequent replication of its progeny.

We shall focus in a particular kind of virus known as *bacteriophage* (virus that infects bacteria), or *phage*, and proceed to describe the main interactions that said predator has with its host.

Given that viruses have no metabolism, they can't actively propel themselves and its movement is limited by diffusion. During their diffusive search of targets viruses are exposed to environmental agents that cause their death, at some rate δ. Upon encounter with the host, the infection (only) occurs when the virus finds a proper receptor situated on the bacterium's surface. The nucleic acid is then inserted into the host via a mechanism yet poorly understood. This process is usually kinetically quantified with a parameter called *absorption rate*, denoted by the letter η. If the virus is *virulent* the subsequent steps will result in the disruption (or lysis) of the cellular membrane of the host, conforming the so-called *lytic cycle* (26). Following these attachment and penetration steps comes the biosynthesis and maturation processes. Here the phage's nucleic acid is replicated and assembled with some deliberately made capsid proteins to form new viral entities that shall constitute the virus progeny. Once the viruses are morphologically completed and infectious, lysis is triggered, liberating a number of particles that can resume the searching process again to infect neighboring bacteria. The time it takes between the delivery of the genome into the host and the membrane desintegration receives the name of *latency time*, and is denoted by τ. The viral progeny number is called burst size, β, and follows a broad distribution rather than a constant value, reflecting perhaps heterogeneity in killing and predation strategies, as suggested in (27). In fact, all the above described events are of stochastic nature and we shall take, for simplicity, their average value.

[4] An *obligate parasite* is a parasitic organism that cannot complete its life-cycle without finding and exploiting an adequate host.

On the other hand, it is worth mentioning, as a parenthesis in this dissertation, that some phages, the so-called *temperate phages*, are capable to choose between two developmental regimes, the lytic and lisogenic cycles. In *lisogeny* (28), the phage can establish a stable relationship with the host cell, where it prevails indefinetely inside the bacterial host as a prophage [5], with the parasite genome either integrated into the the host's chromosome or remaining in an plasmid-like state in the lysogen [6]. In this state no phage particles are produced until the lysogenic state breaks down. This process is called prophage induction and represents a reversal to the lytic cycle. It is usually induced by environmental stresses such as DNA damage caused by UV radiation. Infected cells that enter the lysogenic cycle, continue to grow and are resistant to further infection, as opposed to sensitive bacteria.

In the present work we shall focus on virulent phages. These are phages which can't follow the lysogenic cycle.

2.3.2 | Mathematical and Physical Modelling

As we have seen, bacteria undergo different phases of growth depending on their environmental context. In our study we will focus on a particular phase, the exponential phase, thus we here introduce the mathematical formalism used to describe it. A simple predator-prey model that neglects aggregation is then introduced. It is our aim to use it as a way to identify aggregation traits in a predator-prey system by comparing it with our PPAM.

2.3.2.1 | Exponential growth

When a population experiences an instantaneous rate of change proportional to the population itself, the dynamics of the system is described, in Newton's notation, by a linear differential equation of the type

$$\dot{N}(t) = \pm g \cdot N(t). \tag{2.16}$$

[5] A prophage is a latent form of a phage, in which the viral genes are present in the bacterium without causing disruption of the bacterial cell. Pro means "before", so, prophage means the stage of a virus before being activated inside the host.

[6] A lysogen is a bacteria that carries at least one phage genome either integrated into its chromosome or as an independent replicating extrachromosomal system, such as a plasmid.

With $N(t)$ being the population number at time t, and g the growth $(+)$/decay $(-)$ rate per population unit. The exponential function

$$N(t) = N(0)e^{\pm gt} \tag{2.17}$$

satisfies equation 2.16. This phenomenological description is used to model a very wide set of situations, ranging from radioactive decay to bacterial growth. One could argue that, in the growth case, the description asymptotically tends to infinite size populations. That is the case indeed, but it is not the goal of this approach to predict each of the ways bacteria can grow . If we were interested, for example, in accounting for the system's capacity to sustain a population, we could include non-linearities. Such is the case of logistic equations of the type

$$\dot{B} = g \cdot B \cdot \left(1 - \frac{B}{K}\right), \tag{2.18}$$

where B stands now for bacteria density and K is the so-called carrying capacity, whose value depends on the system's properties. Note that the carrying capacity dictates the point where the growth rate equals the death rate, and that the growth rate slows down as $B \to K$. We, however, shall assume our system follows a homogeneous and constant exponential growth rate. This is, bacteria will grow in the exponential phase independently of their aggregate state, time or total density. This translates, in the context of aggregation, to volumic growth [7].

2.3.2.2 | Predator-prey Dynamics

Viruses and bacteria represent, as we have seen, a coupled system of the predator-prey kind. Dynamics are typically formulated in terms of the experimentally accessible life cycle parameters described in subsection 2.3.1. These parameters are tabulated in table 2.1 for some virulent phages infecting *E. coli*.

The description relies on rates, in a well-mixed scenario. The mean-field assumption makes the interaction probabilities to factorize into the product of individual probabili-

[7] It has been recently shown that bacteria can indeed grow exponentially even in structured environments for relevant periods of time. As an example we refer back to (15; 16). This might seem counter-intuitive, since nutrient availability might strongly depend on the structure of the aggregate, which could mechanically avoid or modify nutrient transport to the inner bacteria.

Name	Adsorption η (10^{-9} ml/h)	Latency τ (h)	Burst size β	Phage decay δ (h^{-1})
ϕX174	174	0.25	135	0.0083
MS2	39	0.67	400	0.0104
T2	24	0.38	135	0.0028
T3	96	0.28	200	0.0043
T4	30	0.38	150	0.0028
T5	12	0.73	290	0.0050
T7	180	0.22	260	0.0078

Table 2.1: Virulent phages infecting E. coli in rich medium at $37C^o$ (29). These values are the arithmetic mean of at least three independent experiments.

ties. In a description where, instead of probability functions, we work with population densities this translates to the product of densities, coupled with a parameter (which is, ideally, experimentally accessible). Consequently, events associated to a single actor are represented by its individual density and weighted by the corresponding event rate.

For our system we distinguish the following events

- Bacterial growth, g: bacteria grow in a possibly space and time-dependent manner. The nutrient availability as well as the structured environment in which they are embedded cause growth to be a function of time (via nutrient concentration) and space. This is, $r_g = f(\vec{r}, n(t))$. However, as discussed before, we here consider $r_g = const \equiv g$.

- Phage-bacteria encounter and absorption, η: viral diffusive search of targets and the subsequent infection process.

- Latency time, τ: starting with the penetration, it is the time it takes the virus to complete the biosynthesis and maturation processes, ending with the lysis of the host.

- Lysis: disruption of the host's membrane and release of the progeny.

- Phage decay, δ: phage death caused by external environmental agents. It is considered to be constant, for simplicity.

The dynamics of bacteria and a virulent phage with average lysis time τ are, therefore, described by the typical predator-prey model (30; 31)

$$\frac{dB}{dt} = g \cdot B - \eta \cdot B \cdot P \tag{2.19a}$$
$$\frac{dI}{dt} = \eta \cdot B \cdot P - \frac{1}{\tau} \cdot I \tag{2.19b}$$
$$\frac{dP}{dt} = \frac{\beta}{\tau} \cdot I - \delta \cdot P - \eta \cdot (B + I) \cdot P \tag{2.19c}$$

An intermediate state, I, has been introduced to account for infected bacteria, that lysis at a rate $1/\tau$. We here assumed that lysis follows a one-step Poisson process (see, for example, *B*.1). These set of euqations shall be addressed as the **Simple Model** (SM).

3

Models

3.1 | Bacteria Aggregation Model

We have now reached the point where we can start building the tools we shall use to quantify the impact of aggregation on the aggregate size distribution of exponentially growing bacteria and on the bacteria-phage interplay. We will start by obtaining an expression for the bacteria-bacteria and phage-bacteria rate of aggregation that explicitly contains the number of constituents of each reactant aggregate. This shall be done, of course, within our the diffusion-limited paradigm, thus via the use of Smoluchowski's rate of coagulation.

Smoluchowski's reaction rate (or brownian kernel), J_{ij}, developed to describe coagulation processes, can be used to treat bacteria aggregation. We start by simplifying the problem a bit. The second term in 2.15 can be ignored as long as $t \gg R^2/D$, which, for a micromiter-sized colloid immersed in water at $T = 30\,°\text{C}$, takes the value

$$\frac{R^2}{D} \sim \frac{1\ \mu m^2}{0.56\ \mu m^2/s} \approx 1.8\ s\ ,$$

where we used the Einsteins-Stokes relation, $D = D_{ES} = k_B T/(6\pi\eta R)$, with $\eta_{H_20}(30\,°\text{C}) = 0.798 \cdot 10^{-3}\ Pa \cdot s$ (32), to theoretically estimate the diffusion coefficient [1]. Hence, the

[1] For the micrometer-sized bacterium *E.coli*, which can be described best as a short cylinder with hemispherical caps at both ends, takes the values $D_X = 0.188 \mu m^2/s$ and $D_Y = 0.154 \mu m^2/s$ for a deflagellated strain (33), with X being the translational diffusion coefficient of the bacterium in the direction parallel to the cylinder's axis, and Y that perpendicular to it. The use of the Einsten-Stokes relation assumes spherical symmetry, which for the *E.coli* case is not fulfilled hence some discrepancy was expected and found. The discrepancy is $D_{ES} \sim 3 \cdot D_X$ and $D_{ES} \sim 3.6 \cdot D_X$. They are both within the same order of magnitude, and that, for this study, is acceptable.

coalescence rate can be written as

$$J_{i+j}dt \approx 4\pi D_{ij}R_{ij}c_i c_j dt \; . \tag{3.1}$$

Consequently, the intrinsic rate of aggregation from equations 2.1 takes the value

$$K_{ij} = 4\pi D_{ij}R_{ij} \; . \tag{3.2}$$

With this in mind equations 2.1 now read as

$$\frac{dc_k}{dt} = 4\pi \left(\frac{1}{2} \sum_{i+j=k} D_{ij}R_{ij}c_i c_j - c_k \sum_{i=1}^{\infty} D_{kj}R_{kj}c_j \right) \qquad (k = 1, ..., \infty). \tag{3.3}$$

This description is half way to where we need to get. Given that we are using a formalism where aggregates are characterized by the number of constituents, the kernel should be explicitly written in terms of the reactant masses (i, j). With this in mind we employ again the Einstein-Stokes relation, $D_i = k_B T/(6\pi\eta R_i) \propto i^{-1/3}$, to express it as

$$K_{ij} = \frac{2k_B T}{3\eta} \left[2 + \left(\frac{R_i}{R_j} \right) + \left(\frac{R_j}{R_i} \right) \right] \; . \tag{3.4}$$

Here k_B is the Boltzmann constant, T is the absolute temperature and η is the dynamic viscosity of the surrounding fluid. In order to obtain the prefactor to $R_i \propto i^{1/3}$ we should first reflect on the coagulation process. This process assumes spherical symmetry before and after merging events. This is a reasonable approximation if colloids lose their entity in the merging process to form a bigger unified aggregate in a context where surface tension is important. However, if the system is composed by rough colloids that keep their shape upon encounter, the aggregates might leave empty spaces in between, as well as form agglomerates far from being spherical. We must, therefore, proceed with some assumptions.

Assumptions:

- Coalescence events result in the clustering of monomers to give agglomerates of quasi-spherical symmetry.

- In each cluster, constitutive monomers keep their identity, shape and volume and, therefore, an empty space is left in between. This introduces a sphere packing problem.
- The sphere packing problem, the quasi-spherical shape and the radii of influence of aggregates is encompassed and quantified in a single parameter that we shall denote as f.

With this understanding, the effective radius of a cluster can be written in terms of the number of constitutive (spherical) monomers. We can picture this as if each aggregate of size i was occupying an effective volume $V_{TOT}^{(i)}$, a fraction of which is empty. If the monomer volume is v_i we have

$$i \cdot v_i = f_i \cdot V_{TOT}^{(i)} = f_i \cdot \frac{4\pi}{3} R_i^3 \,, \tag{3.5}$$

where the parameter is, in principle, a function of the cluster size, $f = f(i) \equiv f_i$. Thus,

$$R_i = \left(\frac{3\, v_i}{4\pi\, f_i} \right)^{1/3} \cdot i^{1/3} \,. \tag{3.6}$$

Plugging 3.6 into 3.4 we can now write the kernel in terms of the reactant masses (i, j)

$$K_{ij} = \frac{2k_B T}{3\eta} \left[2 + \left(\frac{v_i}{v_j} \frac{f_j}{f_i} \frac{i}{j} \right)^{1/3} + \left(\frac{v_j}{v_i} \frac{f_i}{f_j} \frac{j}{i} \right)^{1/3} \right] \tag{3.7}$$

This pre-factor correction to the radius plays an important role if monomer volumes are highly different, which is the case of phages and bacteria, for which the ratio roughly takes the value of $v_P/v_B \sim 10^{-3}$ (for bacteria we used a typical lenght of $l_B \sim 1\mu m$ and for phages, despite they range from $\sim 20 - 400\ nm$ (34) we took $l_P \sim 0.1\ \mu m$ [2].). Throughout this dissertation, we shall also assume that the scaling of the fraction of unoccupied space with cluster size is constant, *i.e.*, $f_i/f_j \sim 1\ \forall\ i, j$ [3]. Since the sub-unit

[2] If we take for phages the diffusion constant $D_P = 10^4\ \mu m^2/h = 2.76\ \mu m^2/s$ (35) the discrepancy with the theoretical estimate, for a particle radius of $0.05\mu m$ at $T = 30\,°C$, is $D_{ES} \approx 2 \cdot D_P$

[3] This assumption is not justifiable beyond the fact that f is not experimentally accessible. It becomes evidently wrong in the limit case $i = 1, j \neq 1$, where $f_1 = 1$ and $f_j \neq 1$. Notice that this approximation is conceptually different than saying $f_i \sim 1$, in which case the nutrient transport to inner bacteria becomes practically impossible, exposing the volumic growth assumption.

volume is considered independent of cluster size (second assumption), we get for bacteria agglomerates that $v_i/v_j \sim 1 \; \forall \; i, j$.

Under these assumptions 3.4 for bacteria-bacteria interaction reads as

$$K_{ij} = \frac{2k_B T}{3\eta} \left[2 + \left(\frac{i}{j} \right)^{1/3} + \left(\frac{j}{i} \right)^{1/3} \right] , \tag{3.8}$$

and the total reaction rate between a bacterial aggregate of size k and free phages of density P takes the form

$$K_{kP} \cdot c_k \cdot P = \frac{2k_B T c_k P}{3\eta} \left[2 + \frac{1}{10} \left(\frac{1}{k} \right)^{1/3} + 10k^{1/3} \right] . \tag{3.9}$$

The specific rate of bacteria-phage interaction is thus a function of both the size of the target and the number available aggregates. It is worth describing what equation 3.9 is tacitly telling us and to do that we focus on the particular case of single bacteria-phage interaction, *i.e.*,

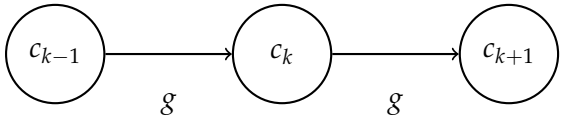
$$K_{1P} = \frac{2k_B T}{3\eta} \left[2 + \frac{1}{10} + 10 \right] . \tag{3.10}$$

If we look back to 3.4, the intrinsic aggregation rate mostly depends, for a fixed temperature, on the ratio of target radius R_i/R_j and R_j/R_i. Since diffusion is inversely proportional to particle size and target size is proportional to particle size, we see that phage diffusion contributes more than bacterial diffusion to the coalescence rate and so does bacterial target size compared to phage target size. This is all embedded in the ratio $v_P/v_B \sim 10^{-3}$. The diffusive target search is hence dominated by phage diffusion and bacterial target size, even in a context lacking spatially structured environments, as we can see from 3.10, where the $1/10$ term accounts for the phage size and bacterial diffusion contribution, whereas the 10 term accounts for that of phage diffusion and bacterial target size. However, even after these considerations, we don't limit ourselves to the static target limit approach of bacterial diffusion, since the total aggregation rate still depends on concentrations, and these can create regions in phase space where clustering by aggregation becomes a relevant trait in colony formation.

We are now in the position to couple aggregation to bacterial growth. This is straightforward in the framework we have been developing so far. The density of aggregates

formed by k individual bacterium will simply be represented as c_k. Hence, the rate of change of aggregate sizes caused by growth is

$$\frac{dc_k}{dt} = g \cdot (c_{k-1} \cdot (k-1) - c_k \cdot k) \qquad (k = 1, ..., \infty) \,. \tag{3.11}$$



It is sometimes helpful the use of flow diagrams to summarize the range of possible transitions within a dynamic description. With that in mind we present the above drawing. The positive contribution to the rate of change in the density of clusters of size k is proportional to the density of monomers belonging to clusters of one size below, $c_{k-1} \cdot (k-1)$, which, when hosting a growth event, become of size c_k. Consequently, the outgoing term accounts for fision events happening to bacteria that belong to clusters of size c_k, as we can see in the flow diagram.

It is easy to see that individual exponential growth on the volume translates to exponential growth for the typical cluster size

$$s(t) \equiv \frac{M_1(t)}{M_0(t)}. \tag{3.12}$$

The equation for the moments takes a particularly easy form in this case

$$\frac{dM_n}{dt} = \sum_{k=1}^{\infty} k^n \cdot g \cdot (c_{k-1} \cdot (k-1) - c_k \cdot k) = g \left(\sum_{k=1}^{\infty} k^n \cdot (k-1) \cdot c_{k-1} - \sum_{k=1}^{\infty} k^{n+1} \cdot c_k \right) \tag{3.13}$$

From which

$$\frac{dM_0}{dt} = g \left(\sum_{k=1}^{\infty} (k-1) c_{k-1} - \sum_{k=1}^{\infty} k c_k \right) \quad = 0$$

$$\frac{dM_1}{dt} = g \left(\sum_{k=1}^{\infty} k(k-1) c_{k-1} - \sum_{k=1}^{\infty} k^2 c_k \right) = g M_1$$

Thus,

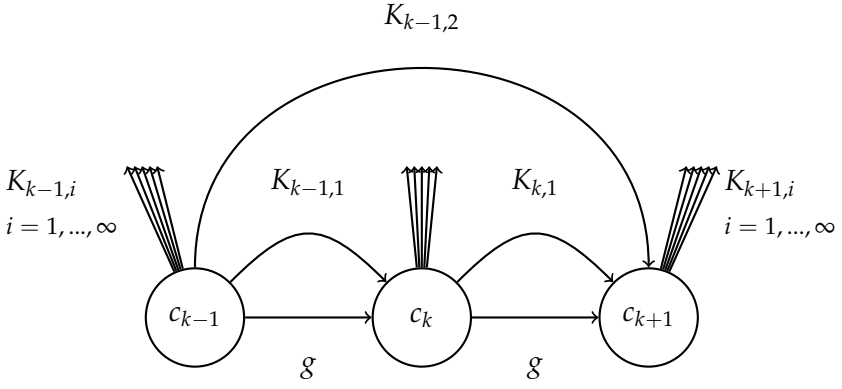
$$s(t) = \frac{M_1(0)}{M_0(0)} \cdot e^{gt} \tag{3.14}$$

On top of growing, if we ignore chemotaxis [4], aggregates describe brownian movements and react in a diffusion-limited manner. In a compact notation the set of master equations, therefore, reads as

$$\frac{dc_k}{dt} = 4\pi \left(\frac{1}{2} \sum_{i+j=k} D_{ij} R_{ij} c_i c_j - c_k \sum_{i=1}^{\infty} D_{kj} R_{kj} c_j \right) + g \cdot (c_{k-1} \cdot (k-1) - c_k \cdot k) \qquad (k = 1, ..., \infty). \tag{3.15}$$

These are the equations for the **Bacteria Aggregation Model** (BAM) in a well-mixed scenario.

In a continuous description (as opposed to a discrete one, proper of stochastic simulations, where there is a finite number of monomers and aggregate sizes) all cluster sizes are connected and susceptible to become any other size, via aggregation events and growth. The flow diagram for the system 3.15 looks like



[4] Chemotaxis is a random walk biased by a spatial gradient of a chemical attractant or repellent. In the presence of such gradient bacteria intercalate short periods of ballistic movement by random changes in direction. This behavior amounts to a diffusive movement with drift, not caused by an externally applied force but by the ability of bacteria to measure an external cue. The bacterium internally analyzes this input and generates the bias by changing the way it rotates the flagella. This transport mechanism is important to find food or to flee from poison. It is not the aim of a diffusion-limited aggregation formalism to account for this situation.

A more explicit way of expressing the coupled system of ODE's 3.15 is

$$\frac{dc_k^A(t)}{dt} = \frac{2k_BT}{3\eta}\left(\frac{1}{2}\sum_{j=1}^{k-1}\left[2+\left(\frac{k-j}{j}\right)^{1/3}+\left(\frac{j}{k-j}\right)^{1/3}\right]c_{k-j}c_j - c_k\sum_{i=1}^{\infty}\left[2+\left(\frac{i}{k}\right)^{1/3}+\left(\frac{k}{i}\right)^{1/3}\right]c_i\right) \tag{3.16}$$

for the diffusion-limited aggregation contribution and

$$\frac{dc_k^G(t)}{dt} = g\cdot(c_{k-1}\cdot(k-1) - c_k\cdot k) \tag{3.17}$$

for the volume-scaling growth terms, with the special case

$$\frac{dc_1(t)}{dt} = -g\cdot c_1\cdot 1 - \frac{2k_BT}{3\eta}\cdot c_1\sum_{i=1}^{\infty}\left[2+\left(\frac{i}{1}\right)^{1/3}+\left(\frac{1}{i}\right)^{1/3}\right]c_i\,. \tag{3.18}$$

We shall use these equations as the basis of the predator-prey model in order to incorporate aggregation events in the dynamics. We are now in the position to implement, on top of the model for bacterial growth and aggregation, phage-bacteria interactions.

It is useful, in order to make future comparisons between aggregation-free cluster size distributions and those with aggregation events to simulate only the necessary. In this sense we analytically solved the particular case where the kernel K_{ij} is constant and equal to zero.

- $K_{ij} = 0$

We will employ the monomer-only initial condition. Thus, we have the system

$$\frac{dc_k}{dt} = g\cdot c_{k-1}\cdot(k-1) - g\cdot c_k\cdot k \qquad ; \qquad c_k(0) = \delta_{k,1}\,. \tag{3.19}$$

By visual inspection we see that the following equation fulfills the system 3.19 and the initial condition

$$c_k(t) = \frac{(1-e^{-gt})^{k-1}}{e^{gkt}}\,. \tag{3.20}$$

We can easily see this by calculating the derivative

$$\begin{aligned}\frac{dc_k}{dt} &= \frac{(k-1)(1-e^{-gt})^{k-2}\, ge^{-gt}e^{gkt} - gke^{gkt}(1-e^{-gt})^{k-1}}{e^{2gkt}} = \\ &= \frac{g(k-1)(1-e^{-gt})^{k-2}}{e^{gt(k-1)}} - \frac{gk(1-e^{-gt})^{k-1}}{e^{gkt}} = \\ &= g\cdot(k-1)\cdot c_{k-1} - g\cdot k\cdot c_k\end{aligned} \tag{3.21}$$

3.2 | Predator-prey Aggregation Model

We have now the model to answer our first question and the basis to implement phage interactions. This new look aims to encompass growth, diffusion, aggregation, infection, latency and lysis. However, bacterial clustering changes phage-bacteria dynamics in various ways, hence we will have to carefully build up a description that properly accounts for it. The infected states are introduced through the use of an extra index in the cluster size notation, i.e.,

$$c_k(t) \longrightarrow c_k^n(t)\ ,$$

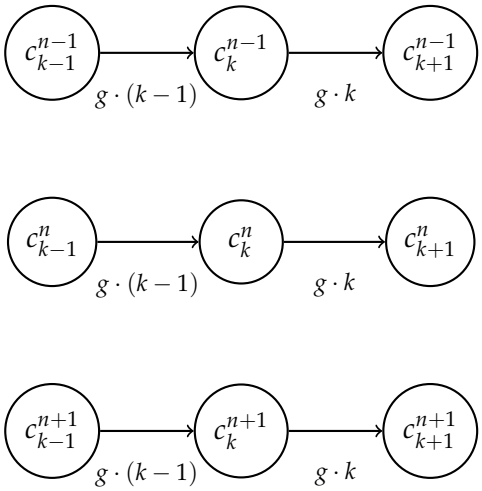
where k and n account for the number of sensitive (not infected) and infected bacteria, respectively. Let's write down the model equations.

- Growth

If we assume that infected bacteria don't grow, the volume-scaling growth equation 3.17 will be hamped by the number n. This translates to the following growth contribution to the total rate of change

$$\frac{dc_k^n(G)}{dt} = g\left(c_{k-1}^n\cdot(k-1) - c_k^n\cdot k\right)\ , \tag{3.22}$$

where it now only depends on the number of sensitive bacteria, as pictured in the following diagram



- Aggregation

As stated in the previous section, chemotaxis is ignored, and the only transport mechanism for bacteria is diffusion. Accordingly, the aggregation contribution to the total concentration change of cluster sizes over time is

$$\frac{dc_k^n(A)}{dt} = \frac{1}{2}\sum_{i+j=k}\left(\sum_{l+m=n} c_i^l c_j^m K_{i+l,j+m}\right) - c_k^n \sum_{i=0}^{\infty}\sum_{l=0}^{\infty} c_i^l K_{k+n,i+l}\,, \tag{3.23}$$

with $K_{0,i} = K_{i,0} = 0$. This is conceptually identical to the previous aggregation equations 3.16, with the proper adaptation to the new notation.

- Diffusion and Infection

The first step in phage-bacteria interactions is to implement the transport mechanism of both factors. We remember that, since viruses can't actively propel themselves, their movement is also limited by diffusion. This is modeled by the use of the same brownian kernel introduced to describe colloidal aggregation. In this case viruses are treated as single monomer clusters that react with bacteria in accordance with their lower monomer volume, as calculated in 3.9. Denoting the free-phage density as P, the corresponding contribution reads as

$$\frac{dc_k^n(P)}{dt} = f \cdot \left(K_{k+n,P} \cdot c_{k+1}^{n-1} \cdot S_{k+1,n-1} - K_{k+n,P} \cdot c_k^n \cdot S_{k,n}\right) \cdot P\,, \tag{3.24}$$

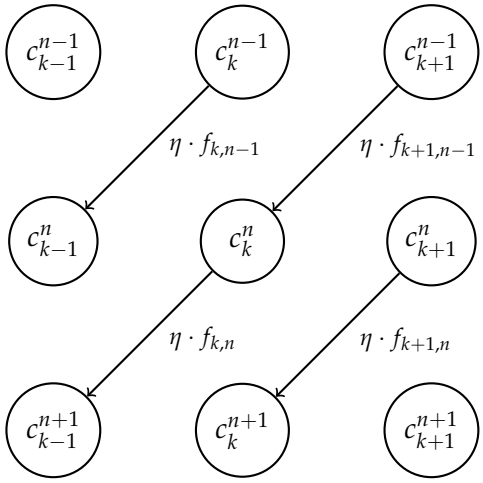
where the function $S(k,n)$, called *shielding function*, accounts for the possibility of absorption to an already infected bacteria. As we can see, the aggregation between a cluster of size (k,n), and free phages, *i.e.*, $K_{k+n} \cdot c_k^n \cdot P$, is weighted by the parameter f. The aim here is to account for the possibility of infection failure upon phage-bacteria encounter. The parameter f represents the fraction of those that not only collide, but bind to the proper receptor and manage to penetrate the bacterium's membrane. This is not experimentally accessible as it is. We therefore re-express it in terms of the absorption rate η. If we calibrate the absorption rate, η, as the rate of absorption to single-monomer clusters, $f \cdot K_{1,P} = \eta$, we can use equation 3.9 to get

$$K_{k,P} \cdot f \sim K_{1,P} \cdot f \cdot \frac{\left(2 + 10k^{1/3} + 0.1k^{-1/3}\right)}{12.1} = \eta \cdot h_k \,, \tag{3.25}$$

where $h_k = \frac{1}{12.1}\left(2 + 10k^{1/3} + 0.1k^{-1/3}\right)$. Equation 3.24 now reads as

$$\frac{dc_k^n(P)}{dt} = \eta \cdot h_{k+n} \cdot \left(c_{k+1}^{n-1} \cdot S_{k+1,n-1} - c_k^n \cdot S_{k,n}\right) \cdot P \,. \tag{3.26}$$

Free-phage-bacteria encounters imply the following flux in the distribution of cluster sizes



- Latency and Lysis

The infected population in a cluster of the type c_i^m will undergo lysis with some rate $r(m)$. These events provoke clusters (i,m) to change size, $(i,m) \to (k,n)$. If this is a

one-to-one transition (this is, (i, m) is not degenerate), the lysis contribution to the total rate of change should have the form

$$\frac{dc_k^n(L)}{dt} = -r(n) \cdot c_k^n(t) + r(m) \cdot c_i^m(t) \ , \tag{3.27}$$

The concrete relationship between cluster size indexes before and after lysis depends on several factors. First of all, and opposed to the SM, bacterial clustering introduces a local character to infection events. This means that, upon lysis, when the viral progeny is released into the system, they will have a higher number of bacteria from the same cluster already available for infection. This introduces a reabsorption effect that we shall quantify with a parameter α, representing the fraction of viral progeny that is reabsorbed into the same cluster. On the other hand, the fraction $1 - \alpha$ escapes the aggregate and becomes instantly available to interact with any other cluster.

These considerations establish the following relationship between the indexes (i, m) and (k, n)

$$n = m - 1 + \beta\alpha S_{i,m-1} \tag{3.28a}$$
$$k = i - \beta\alpha S_{i,m-1} \tag{3.28b}$$

where the viral progeny number, β, is weighted by the reabsorbed fraction of the phage's progeny, α, and the shielding function, given that from those reabsorbed, some might bump into an already infected bacterium. The solving of the system raises some special cases, given that the validity of the system fully depends on the (α, β) values. A careful description of the procedure to get the right equations is given in $(B.2)$, consequently, and without further delay, we directly write them down as

$$\frac{dc_k^n(L)}{dt} = -\frac{n}{\tau} \cdot c_k^n + \frac{m}{\tau} \cdot c_i^m \ , \qquad i + m \geq \beta\alpha \tag{3.29a}$$
$$\frac{dc_0^{\tilde{n}}(L)}{dt} = -\frac{\tilde{n}}{\tau} \cdot c_0^{\tilde{n}} + \sum_{i=0}^{min(\beta\alpha, \tilde{n})} \frac{\tilde{n} + 1 - i}{\tau} \cdot c_i^{\tilde{n}+1-i} \ . \qquad i + m \leq \beta\alpha \tag{3.29b}$$

Here lysis rate is treated as a one-step Poisson process (see $B.1$). These are the last of all contributions to the cluster size distribution change over time. The final expression is

made by the sum of the aggregation, growth, phage encounter and lysis contributions, *i.e.*,

$$\frac{dc_k^n(t)}{dt} = \frac{dc_k^n(A)}{dt} + \frac{dc_k^n(G)}{dt} + \frac{dc_k^n(P)}{dt} + \frac{dc_k^n(L)}{dt} \tag{3.30}$$

- Equation for Phages

The last equation is that of the phage population density and is completely restrained based on the former considerations. Phages will diffusively bump into clusters, and a fraction of them will get absorbed. This introduces a negative contribution to the total density, proportional the number of bacterial clusters as well as the phage density itself, of the type $\sim f \cdot K_{k+n} \cdot c_k^n \cdot P$, summed over all possible (k, n) combinations. Upon lysis within some cluster c_k^n, $(1-\alpha)\cdot\beta$ new viruses fully escape from the colony, positively contributing to the total density at a rate $r(n)$. The total lysis contribution to the free-phage density comes from considering all cluster sizes. Finally, there is a term accounting for phage death caused by external environmental factors, at some rate δ. With this in mind the free-phage equation reads as

$$\frac{dP}{dt} = (1-\alpha)\cdot\beta\cdot\sum_k\sum_n r(n)\cdot c_k^n - f\cdot\sum_k\sum_n \cdot K_{k+n,P}\cdot c_k^n\cdot P - \delta\cdot P\,, \tag{3.31}$$

and in terms of the absorption rate, η, and the one-step lysis rate per bacterium, τ^{-1}, reads as

$$\frac{dP}{dt} = (1-\alpha)\cdot\frac{\beta}{\tau}\cdot\sum_k\sum_n n\cdot c_k^n - \eta\cdot\sum_k\sum_n h_{k+n}\cdot c_k^n\cdot P - \delta\cdot P \tag{3.32}$$

As we can see the phage aggregation term is not weighted by the shielding function, since it doesn't matter whether bacteria are infected or not for the virus to be absorbed.

Equation 3.32 can be writen more explicitly to account for lysis events of clusters of the type $(k, n) = (0, 1)$, cases where, upon lysis, there is no re-absorption event. With this in mind we write

$$\frac{dP}{dt} = (1-\alpha)\cdot\frac{\beta}{\tau}\cdot\sum_{(k,n)\neq(0,1)} n\cdot c_k^n + \frac{\beta}{\tau}\cdot c_0^1 - \eta\cdot\sum_k\sum_n h_{k+n}\cdot c_k^n\cdot P - \delta\cdot P\,. \tag{3.33}$$

To summarize, the final equations for the **Predator-Prey Aggregation Model** (PPAM) are

$$\frac{dc_k^n(A)}{dt} = \frac{1}{2}\sum_{i+j=k}\left(\sum_{l+m=n} c_i^l c_j^m K_{i+l,j+m}\right) - c_k^n \sum_{i=0}^{\infty}\sum_{l=0}^{\infty} c_i^l K_{k+n,i+l} \tag{3.34a}$$

$$\frac{dc_k^n(G)}{dt} = g\left(c_{k-1}^n \cdot (k-1) - c_k^n \cdot k\right) \tag{3.34b}$$

$$\frac{dc_k^n(P)}{dt} = \eta \cdot h_{k+n} \cdot \left(c_{k+1}^{n-1} \cdot S_{k+1,n-1} - c_k^n \cdot S_{k,n}\right) \cdot P \tag{3.34c}$$

$$\frac{dc_k^n(L)}{dt} = \frac{m}{\tau} \cdot c_i^m - \frac{n}{\tau} \cdot c_k^n\,, \qquad k \neq 0,\ i+m \geq \beta\alpha \tag{3.34d}$$

$$\frac{dc_0^{\tilde{n}}(L)}{dt} = \sum_{i=0}^{min(\beta\alpha,\tilde{n})} \frac{\tilde{n}+1-i}{\tau} \cdot c_i^{\tilde{n}+1-i} - \frac{\tilde{n}}{\tau} \cdot c_0^{\tilde{n}}\,, \qquad k = 0,\ i+m \leq \beta\alpha \tag{3.34e}$$

for the rate of change in bacterial aggregates and

$$\frac{dP}{dt} = (1-\alpha) \cdot \frac{\beta}{\tau} \cdot \sum_{(k,n)\neq(0,1)} n \cdot c_k^n + \frac{\beta}{\tau} \cdot c_0^1 - \eta \cdot \sum_k \sum_n \cdot h_{k+n} \cdot c_k^n \cdot P - \delta \cdot P \tag{3.35}$$

for the phage population density.

4

Simulations, Data Analysis and Results

Now that the mathematical modelling has been established, we are in the position to answer our research questions. We do that by computationally simulating our models.

All simulations involve the numerical integration of a set of ODEs (either those from the BAM, PPAM or SM). For this we used Python's *Scipy* library, which contains an adequate integrator for initial value problems [1]. To speed up calculations the *Numba* library was also implemented in the code [2].

4.1 | Aggregation Impact on the Aggregate Size Distribution

Let's review the first question

- **Aggregation impact in an exponentially growing, and well-mixed, bacterial culture** Can diffusion-limited aggregation display a relevant role in the distribution of bacterial aggregates? For this, we built a model where bacterial growth was coupled to a diffusion-limited aggregation kinetic framework.

Aggregation displays a squared dependence on cluster size concentrations, remember that the rate of interaction of two clusters of size i and j is $\sim K_{ij}c_ic_j$. This means diffusion-limited cluster formation is susceptible to be negligible if we don't focus on

[1] The specific integrator can be found here `https://docs.scipy.org/doc/scipy/reference/generated/scipy.integrate.solve_ivp.html`

[2] The corresponding source, when last used, could be found here `http://numba.pydata.org`.

the "right" density region. In order to identify this region we simulated the dynamics of the phage-free model (BAM) with different initial monomer-only concentrations, $c_k(0) = \delta_{1,k}$.

- Simulation set #1

Here we chose $\mathbf{100\,\mu m \equiv 1}$ as unit length and $\mathbf{1h \equiv 1}$ as unit time. All cluster densities are therefore expressed in $cells/(100\mu m)^3$ and rates in h^{-1}. However, we will recover units typical of microbiology when plotting in order to make delivered information more accessible, *i.e.*, $cells/mL$.

The number of launched simulations was 5, and we used the monomer-only initial concentrations $\boldsymbol{c_k(0) = \delta_{1,k} = 10^{\alpha}}$, with $\alpha = (-1, 0, 1, 2, 3)$ [3]. The growth rate was set to zero, $\boldsymbol{g = 0}$. The cut-off cluster size was set to $\boldsymbol{k_{cut} = 2000}$ and each simulation stopped when the mass percentage of the cut-off size was bigger than 10^{-10} % of the total mass, $k_{cut} \cdot c_{k_{cut}} \geq 10^{-10} \cdot M = 10^{-10} \cdot \sum k \cdot c_k$. This is a safety measure to avoid the divergence of the integrator.

The physical parameters were chosen based on the optimal growing temperature on a batch culture, $\boldsymbol{T = 37\,°C}$, and the dynamic viscosity of water, $\boldsymbol{\eta_{H_20}(37\,°C) = 0.0006922}\,[Pa \cdot s]$. These two last parameters will be the same for all future simulations.

The unit length and unit time have an impact on the numerical values for the brownian kernel pre-factor, as well as the values for cluster size concentrations. For example, the brownian kernel pre-factor

$$B(T, \eta(T)) \equiv \frac{2k_B T}{3\eta} = \frac{2 \cdot 1.38 \cdot 10^{-23}[\frac{kg\ m^2}{s^2\ K}] \cdot (T(°C) + 273.15)[K]}{3 \cdot \eta(T)[\frac{kg}{m\ s}]}$$

$$\rightarrow B(T, \eta(T)) = f(T) \left[\frac{m^3}{s}\right]$$

For water at 37 °C $\rightarrow \eta_{H_20}(T = 37\,°C) = 0.0006922 \left[\frac{kg}{m \cdot s}\right] \rightarrow f(T) \sim 4 \cdot 10^{-18}$

[3] These values were not chosen at random. We scanned around what in microbiology is considered to be a *diluted* system, *i.e.*, $\sim 10^6\ cells/mL$ or, in our units, $\sim 10^0\ cells/100\mu m^3$.

If the unit length and time

$$\mu m \equiv 1 \;\; ; \;\; h \equiv 1 \longrightarrow m^3 = 10^{18} \;\; ; \;\; s = \frac{1}{3600}$$

the pre-factor takes the value

$$B(T, \eta(T)) = 4 \cdot 10^{-18} \cdot \frac{10^{18}}{1/3600} \left[\frac{\mu m^3}{h}\right] = 14839.88 \left[\frac{\mu m^3}{h}\right]$$

Conversely, if we instead choose the aforementioned units of $100\mu m^3 \equiv 1$; $h \equiv 1$, the prefactor takes a value closer to unity, $B = 1.483 \cdot 10^{-2} \left[\frac{(100\mu m)^3}{h}\right]$, which is computationally less expensive.

As sanity check for the reader, the set of initial concentrations was given in simulation units. If we adopt those typical from microbiology these translate to

$$c_k(0) = 10^{\alpha} \left[\frac{cells}{(100\mu m)^3}\right] = 10^{\alpha+6} \left[\frac{cells}{mL}\right] \qquad (\alpha = -1, 0, 1, 2, 3) \; .$$

- Analysis set #1

To make a visual comparison of the impact of this concentration change we plot in Figure 4.1 the cluster size probability distribution [4] for this set of simulations at 4 different times, this is

$$P_k(t) = \frac{k \cdot c_k(t)}{\sum k \cdot c_k(t)} \; , \tag{4.1}$$

at $t = [0.0, 0.5, 1.0, 1.5, 2.0]$.

Aggregation displays a great dependence on monomer concentration, as Figure 4.1 illustrates. Low density regions show no aggregation effect on the studied time interval, such as $c_k(0) = 10^5 \; cells/mL$, where no effect is observed. Conversely, higher order initial monomer concentrations, such as $c_k(0) = 10^8 \; cells/mL$, reveal the existence of a

[4] As a reminder, we consider the probability of finding a cluster of size k, P_k, as the number of monomers belonging to clusters of such size.

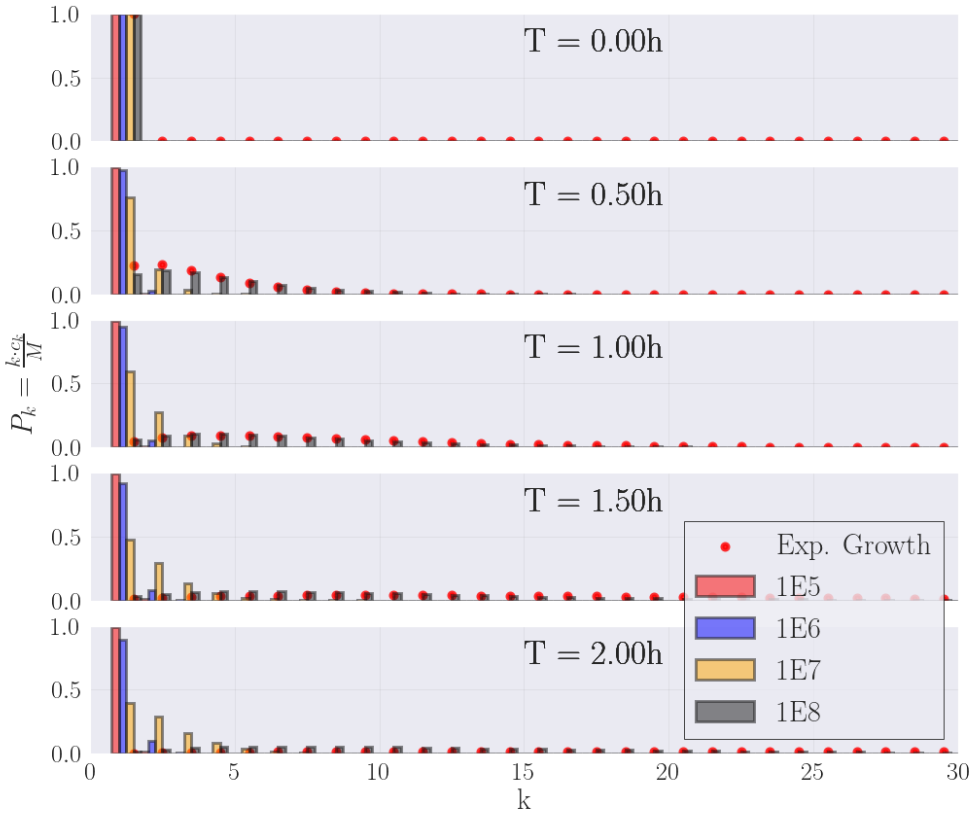


Figure 4.1: Distribution of cluster sizes, P_k, for different initial densities $c_k = \delta_{1,k}$. Aggregation kinetics greatly depends on initial concentrations, as one could intuitively have seen from its squared dependence in the aggregation kernel, $\sim K_{ij} c_i c_j$. The unit length is expressed back in units typical of microbiology, thus concentrations read in $cells/mL$. It has been plotted an exponentially growing system, see equation 3.20, as a visual comparison to our simulations.

density region where aggregation turns out to have a large impact upon the cluster size distribution. These qualitative statements can be translated by computing the system's typical mass, $\langle k \rangle$, (see Eq. 2.3), over time, as shown in Figure 4.2. This provides a way to compute the aggregation time-scale (we compute an event's time scale as the inverse of the rate of increase caused by that particular type of event, either aggregation or growth, in our case).

As we can see, the functional dependence turns out to be linear, a feature that, in prin-

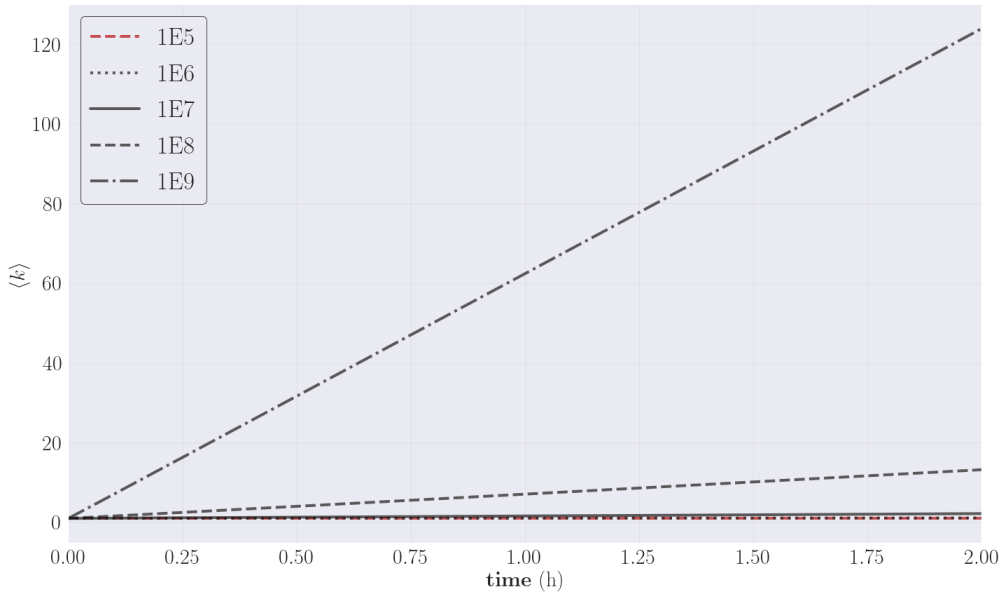


Figure 4.2: Aggregation time-scale experiment. The individual monomer growth rate, g, is set to zero. We plot the evolution in time of the typical mass of the cluster size distribution, $\langle k \rangle$, for different monomer-only initial concentrations. Linearity is not trivially explained, however, it provides an easy and direct way of estimating the aggregation time-scale as the inverse of the slope $\tau_a \sim 1/m$.

ciple, is not trivially explained but that, as we will see, generates a very interesting hypothesis regarding the nature of the brownian kernel. Linearity provides an easy way for a rough estimate of the aggregation time-scale as the inverse of the slope [5]. If

$$\frac{M_2}{M_1} = \frac{\sum k^2 \cdot c_k}{\sum k \cdot c_k} \equiv s(t) = m \cdot t + s(0) \tag{4.2}$$

then

$$m = \frac{\Delta y}{\Delta x} = \frac{s(t_{final}) - s(0)}{t_{final} - 0} = \frac{1}{\tau_a} \tag{4.3}$$

with $s(0) = M_2(0)/M_1(0) = 1$, since, given the monomer-only initial condition, $M_n(0) = 1 \; \forall n \geq 0$. The results from this analysis have been summarized in Table 4.1.

[5] If the functional dependency was different a χ^2 test would have been a better procedure. Given the simplicity of the results it was considered unnecessary. Furthermore, it was our aim to only make a rough estimation of the time-scale.

$c_k(0) = \delta_{1,k}$					
$c_k(0)$	$1E5$	$1E6$	$1E7$	$1E8$	$1E9$
m	0.006	0.059	0.600	6.110	61.473
τ_a	168.438	16.820	1.666	0.163	0.016

Table 4.1: Values of the aggregation time-scale for a set of initial monomer-only concentrations. The linear dependence of the ratio of moments M_2/M_1 provided a direct way of evaluating it as the inverse of the slope, *i.e.*, $\tau_a = 1/m$.

The squared dependency on cluster concentration, $\sim K_{ij}c_ic_j$, approximately translates to 10-fold decreases of the aggregation time-scale upon 10^3-fold increases in initial monomer concentration. It is interesting to see that the rate of aggregation remains constant no matter the cluster size distribution, as the linear behavior suggests, depending only on the initial mass $M = \sum k \cdot c_k(0)$. This indicates a balancing behavior between the earlier narrow free monomer distribution, where aggregation events may occur more often and the later and wider, where diffusion-limited reactions are less frequent, but translate to a higher increase in cluster size. This is an ideosyncrasy of the thermodynamic limit, where there is no finite number of monomer units.

As a last comment on the linearity of the typical mass estimate - It is suggested in literature (18) that the brownian (yet unsolved) and constant kernels might be related, being the second an uncontrolled approximation of the first. Suggestions are based on the fact that both kernels remain invariant under the transformation $(i, j) \to (ia, ij)$, *i.e.*, $K_{ai,aj} = K_{i,j}$. We here see that the brownian kernel's typical mass follows a linear behavior, and so it does the constant kernel's, whose first moments are $M_0 = 1/(1+t)$, $M_1 = 1$ and $M_2 = 1 + 2t$ [6], supporting the aforementioned hypothesis.

Since we want to see whether diffusion-limited aggregate formation has an impact on cluster size dynamics, calculating aggregation time-scales without comparing them to that for growth is completely useless. Luckily enough, the growth's time scale is straight

[6] To see this set $K_{i,j} = 2$ and solve for particular values of n in the moments equation 2.2 employing the monomer-only initial condition, $M_n(0) = 1 \; \forall n \geq 0$.

$$\frac{dM_n}{dt} = \sum_{i,j}^{\infty} (i+j)^n c_i c_j - 2M_n M_0$$

forward to compute. The system's typical mass for exponential growth on the volume can be calculated from equation 3.14, resulting in a growth rate of

$$\tau_g = \frac{1}{g} \,. \tag{4.4}$$

If we assume that typical values for the growth rate are $[1-3]\ h^{-1}$, we then have a growth time-scale of $\tau_g = [0.33, 1]\ h$. Thus, we find a concentration interval where aggregation and growth time-scales are of the same order or magnitude. This is, around $c(0) = 10^7 - 10^8\ cells/mL$, as visually suggested in Figure 4.1.

This argument revolves around a fixed total mass, since monomer number is conserved in aggregation. Conversely, how does the system evolve if we, on top of aggregation, allow monomer growth? The monomer-only initial condition might represent a kinetic bottle neck in the formation of bigger aggregates, but with a growth contribution aggregation might play a bigger role in colony formation for initial concentrations where it seemed negligible. If so, an initially diluted medium could evolve to become an undesirable set of aggregates with negative effects on clinical studies (4).

With this in mind, as a reinforcement to our argument that there exists a density region where aggregation plays an important role upon cluster size distribution dynamics (within the density region typical of microbiology studies), we launched another set of simulations where now the BAM has a finite value for the growth rate.

- Simulation set #2

We kept $\mathbf{100\ \mu m \equiv 1}$ as unit length and $\mathbf{1h \equiv 1}$ as unit time.

The number of launched simulations was 5, and we used the monomer-only initial concentrations $\boldsymbol{c_k(0)} = \delta_{1,k} = 10^{\alpha}$, with $\alpha = (-1, 0, 1, 2, 3)$. The growth rate now has a finite value, $\boldsymbol{g = 1.5}$. Since the growth contribution implied a faster increase in cluster sizes the cut-off cluster size was set to $\boldsymbol{k_{cut} = 5000}$ and each simulation was again stopped when the mass percentage of the cut-off size was bigger than 10^{-10} % of the total mass, $k_{cut} \cdot c_{k_{cut}} \geq 10^{-10} \cdot M = 10^{-10} \cdot \sum k \cdot c_k$.

- Analysis set #2

We start again with a visual comparison where we plot the cluster size distributions for all simulated initial densities against that related to exponential growth only (see equation 3.20), at time $t = 0.35h$.

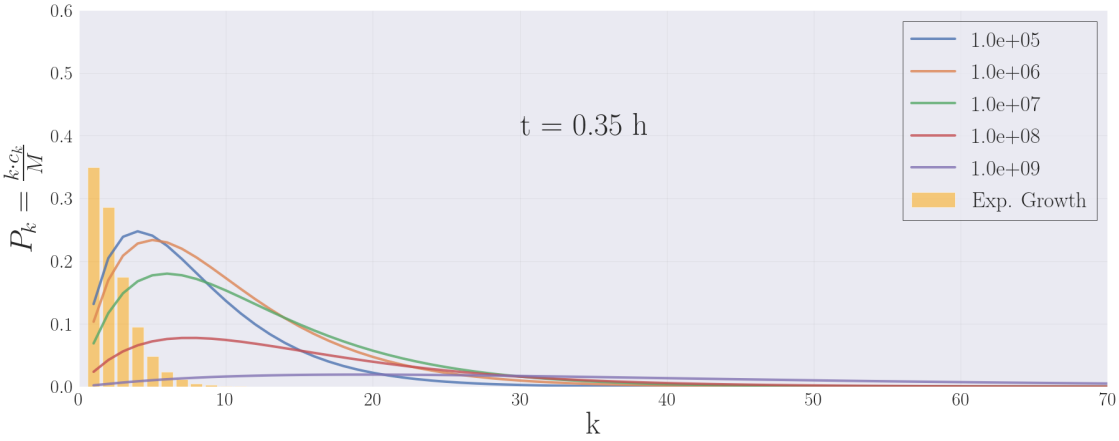


Figure 4.3: The distribution of cluster sizes in an exponentially growing on the volume set of bacteria is compared to that generated by a framework that also includes aggregation events. The integer-valued BAM cluster size probabilities are plotted as continuous lines for visual simplicity, in order to avoid overlapping. Even low density regions such as 10^5 $cells/mL$ represent now density regions where aggregate formation is more than plausible.

We now see that, when overlapped to exponential growth, even density regions that were absolutely negligible before, aggregation-wise , now turn out to be relevant in the diffusive-limited formation of aggregates.

We have thus seen that diffusion-limited aggregation can match colony formation associated to only growth environments and clearly snowballs when overlapped to exponential growth. A clear picture of it can be seen in Figure 4.4.

4.2 | Ecological Implications of Aggregation in Phage-Bacteria Dynamics

We now change our focus to address our second research question. In order to do that, as we have commented before, we simulated the PPAM (equations 3.34) against the SM (equations 2.19). Let's remember the question we are addressing.

- **Ecological implications to bacteria, when exposed to predation by bacteriophages, that are associated with the formation of structured environments** - From a kinetic

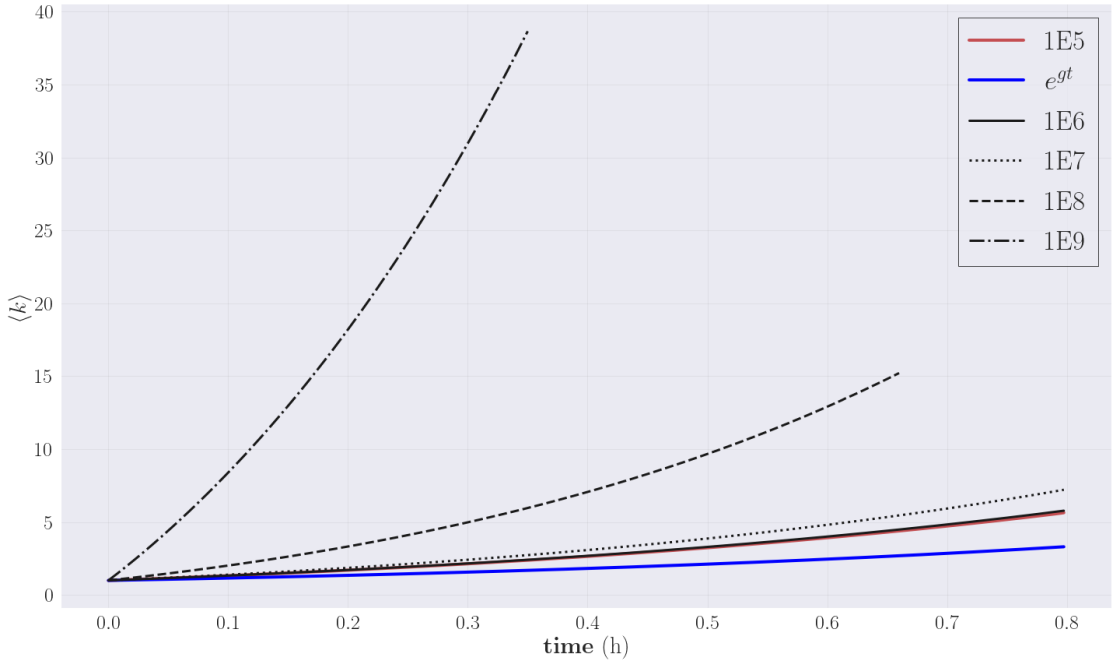


Figure 4.4: Typical mass of the BAM for different initial monomer-only concentrations. The time evolution of the typical mass in a system with no aggregation growing exponentially is plotted as a scale comparison. We here see how aggregation is accelerated when monomers are allowed to grow.

stand point we address the impact of aggregate formation upon predation by phages. For this, we built a model that encompasses bacteria aggregation, exponential volumic growth, diffusion-limited search and infection by phages, latency and lysis. This model, labeled as PPAM, is meant to be compared with a more simple predator-prey model, the SM, which neglects clustering formation between bacteria. Both models assume a well-mixed system and they also include latency with an intermediate infected state that lysis as a one-step Poisson process. The difference lies in the clustering and hence in aggregate-related characteristics such as the re-absorption weights $(\alpha, S_{k,n})$, accounting for the locality of an infection, or the different viral absorption rate $\sim f \cdot K_{k,P}$, cluster size dependent.

- Simulation set #3

We kept $\mathbf{100\ \mu m \equiv 1}$ as unit length and $\mathbf{1h \equiv 1}$ as unit time. The fraction of reabsorbed viral progeny upon lysis was set to $\boldsymbol{\alpha = 0.5}$.

Since the PPAM involves infected states that act as a damping mechanism on growth, the mentioned growth time-scale, $\tau_g = 1/g$, represents an upper estimation. We expect

the effective growth to be lower thus, the set of simulations were performed over the density grid $\boldsymbol{c_k^n(0)} = \delta_{1,k}^{0,n} = (5 \cdot 10^{-1} - 10^1)\ cells/(100\mu m)^3 = (5 \cdot 10^5 - 10^7)\ cells/mL$ and $P(0) = (10^0 - 10^3) PFU/(100\mu m)^3 = (10^6 - 10^9) PFU/mL$. The range of the phage density grid over which we shall perform our study is related to computational limitations, rather than physical or biological arguments.

The phage's parameters were those from ***T7***'s *E.coli* bacteriophage (see table 2.1).

The cut-off cluster size was set to $\mathbf{k_{cut} = 214}$ [7] and each simulation stopped when the mass percentage of one of the the cut-off sizes [8] was bigger than 10^{-10} % of the total mass, $(k+n) \cdot c_k^n = k_{cut} \cdot c_k^n \geq 10^{-10} \cdot M = 10^{-10} \cdot \sum_k \sum_n (k+n) \cdot c_k^n$.

- Analysis set #3

The PPAM's resulting trajectories were compared to those of the SM. The comparison was made at two different points. First, at the SM's half-population time - when the sensitive (not infected) bacteria population levels of this model without aggregation events reached $B(0)/2$, we looked at the population levels of the PPAM. Second, at the SM's extinction time, defined as the time when the population of sensitive bacteria reaches $B = 1\ cell/mL$ [9] [10].

Figures 4.5 and 4.6 are examples of this analysis for two particular simulated trajectories. As we can already see, they show a delaying in the crashing point. The typical mass provides information on the clustering dynamics - in Fig. 4.5, where the population descent takes a longer time, the typical mass shows some curvature, reflecting relevant growth contributions to the sensitive bacteria population. Conversely, in Fig. 4.6, the phage invasion and posterior lysis paradoxically [11] happens much faster thus, the typical mass is determined by aggregation, and this is reflected in the linearity of the

[7] The cut-off size is lower than before since we now solve for k_{cut}^2 differential equations, instead of just k_{cut}.

[8] Sizes, in plural. Remember that we now have sensitive plus infected bacteria, (k, n), thus the cut-off size has a degeneracy of pair combinations such that $k + n = k_{cut}$ of $k_{cut} + 1$.

[9] Stochastic simulations have the advantage of having well-defined extinctions ($M = 0\ cells/mL$) , whereas continuous descriptions involve an artificial definition of when the system is considered to be extinct, since the absolute 0 is only asymptotically achieved - think, for example, of the function e^{-x}.

[10] Within our range of densities and the parameters we used, all trajectories are crashing trajectories, *i.e.*, the system asymptotically moves towards the fixed point $(B, I, P) = (0, 0, 0)$.

[11] This is just a consequence of the number of sensitive bacteria infected by phage being proportional to both, phage and sensitive bacteria densities.

typical mass.

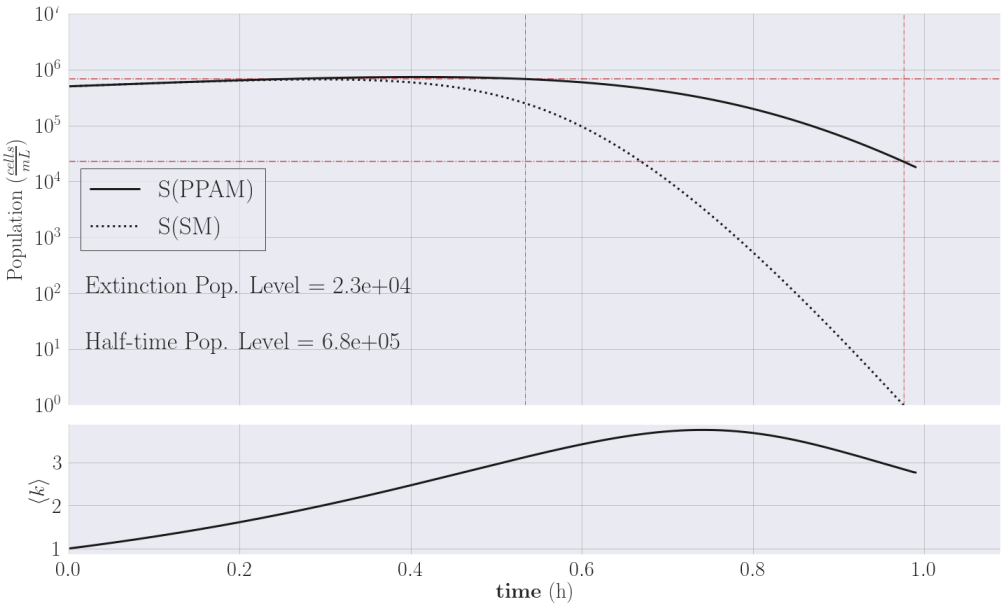


Figure 4.5: Analysis of the difference in trajectories between the two phage-bacteria models. We show an example of the BAM sensitive population trajectory, $S(BAM)$, against it's SM's equivalent, $S(SM)$. The example is a case with $c_k^n = \delta_{1,k}^{0,n} = 5 \cdot 10^5\ cells/mL$ and $P(0) = 10^6\ PFU/mL$.

Via the above procedure a set of 2 PPAM population values was obtained for each initial bacteria and phage densities. To summarize our results we decided to construct 2 heat maps, were we show the differences provoked by aggregation. With those associated to the half-population time of the SM we built the heat map from Figure 4.7, where we plot them as the fraction with respect to their initial population value, *i.e.*,

$$f = \frac{B^{PPAM}(t^{SM}_{half-pop.})}{B(0)} \,. \tag{4.5}$$

PPAM values associated to the extinction events of the SM's trajectories were used to build the heat map from Figure 4.8, where we directly provide the absolute value of the sensitive population level.

Trajectories displayed, over the studied phase space, a displacement of the crashing point. However, no trajectory has any signs of representing a survival situation for bacteria. Hence, from a purely kinetic stand point, aggregation events on a non-clustered initial population might help survive a transient virus invasion, whereas if permanent,

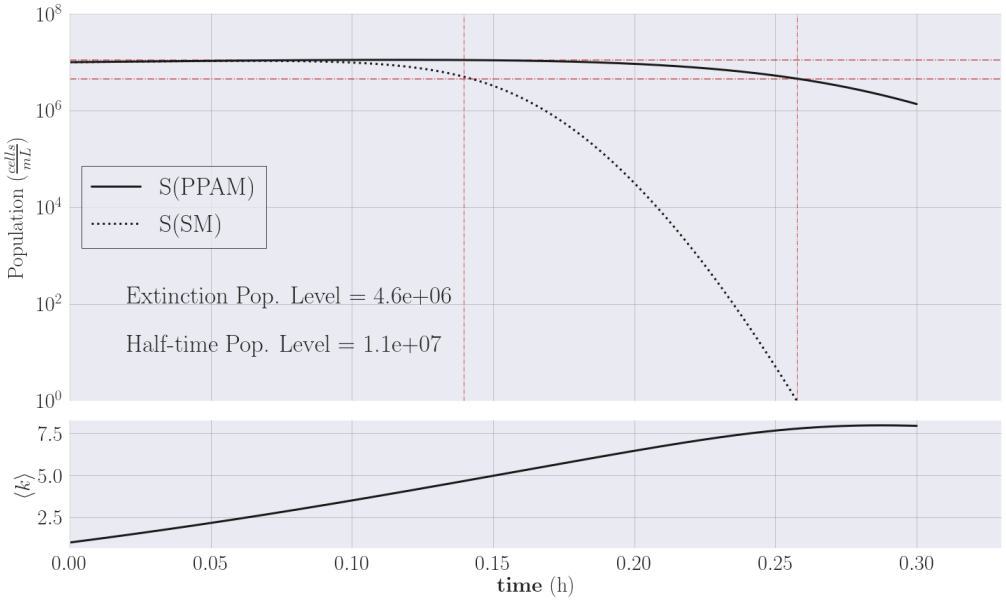


Figure 4.6: Analysis of the difference in trajectories between the two phage-bacteria models. We show an example of the BAM sensitive population trajectory, $S(BAM)$, against it's SM's equivalent, $S(SM)$. The example is a case with $c_k^n = \delta_{1,k}^{0,n} = 10^7\ cells/mL$ and $P(0) = 10^6\ PFU/mL$.

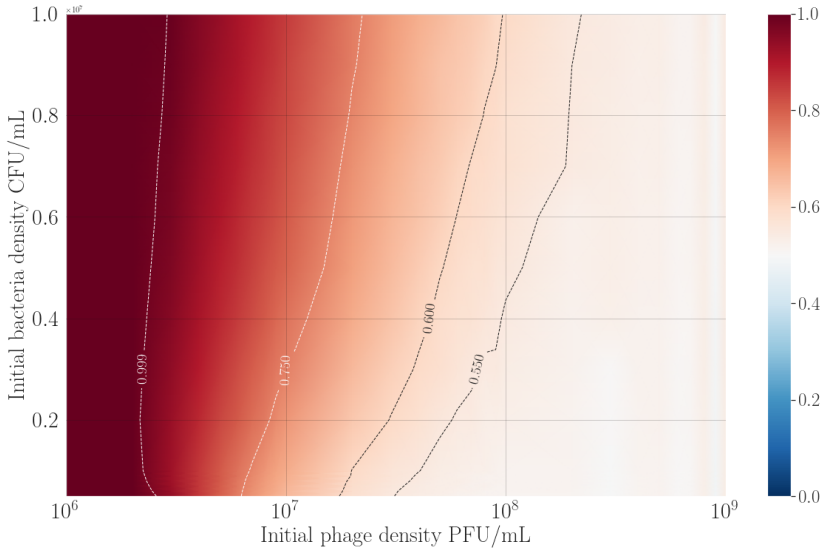


Figure 4.7: Fraction of the initial population at the SM's half-population time, *i.e.*, $S = S(0)/2 = M(0)/2$.

the system is bound to the same fate as that without aggregation.

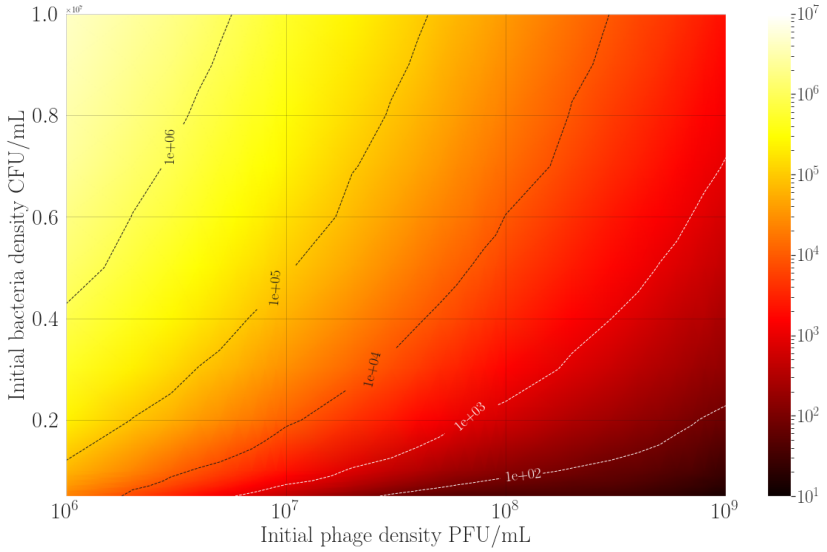


Figure 4.8: Density of bacteria at the SM's extinction time, *i.e.*, $S = 1\ cells/mL$.

Given that the PPAM has the free parameter α (we don't know how it might be distributed), the previous set of simulations were repeated for different α values. Since the value or distribution of α is an unknown for us, we must study how our results change over the possible range of α values, *i.e.*, $\alpha \in [0.0 - 1.0]$.

- Simulation set #4

To study the dependence on α, a new set of simulations was launched with the same conditions than Simulation set #3, we thus refer back to it. We studied the system's behavior over the range $\boldsymbol{\alpha = [0.0 - 1.0]}$ in steps of 0.1.

- Analysis set #4

A summary of the α effect over trajectories is provided in Figures 4.9 and 4.10, with the grey area being the region where all $\alpha \in [0.1 - 1.0]$ trajectories are contained.

$\alpha = 0.0$ trajectories represent the kinetic advantage of having a clustered population, where we don't find the trade-off of the re-absorption effect. This comes from the fact that phage absorption rates to a clustered population are lower than those for unstructured environments. To see this lets consider a fixed number of bacteria, and the special

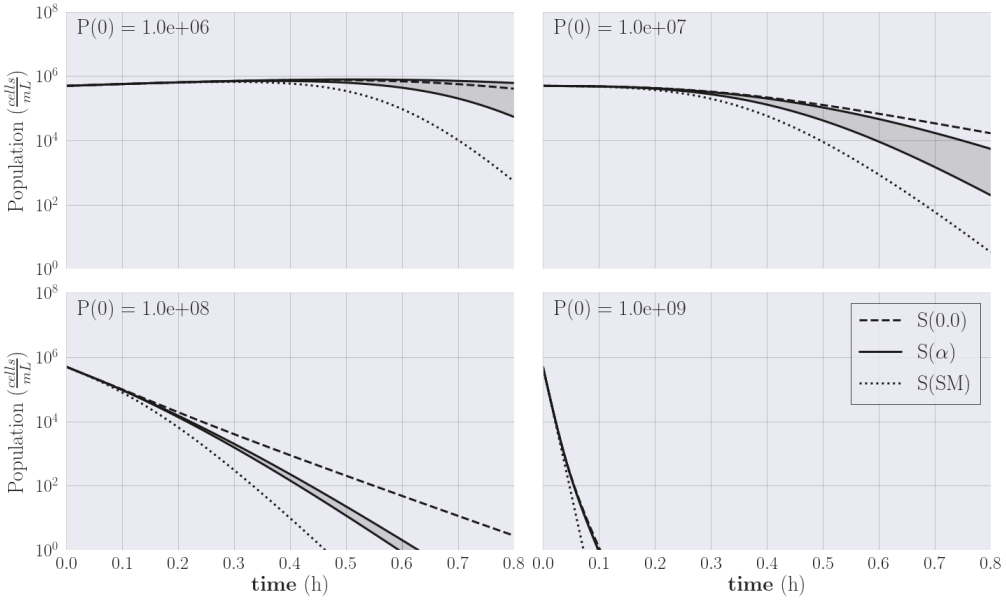


Figure 4.9: Example trajectories for a fixed number of initial sensitive bacteria density, $B(0) = 5 \cdot 10^5$ $cells/mL$, and a progressive 10^3-fold increase in viral initial concentration. The dotted line represents the population level of the SM, whereas the dashed and grey area represent that of the PPAM, for $\alpha = 0.0$ and the α values of $[0.1 - 1.0]$, respectively. The value $\alpha = 0.1$ is the black line on the lower edge of the grey area, whereas $\alpha = 1.0$ is the higher one.

case where all belong to clusters of the same size. Does the size of these clusters in which bacteria are embedded influence the absorption rate? To answer this we compute the searching time as a function of the cluster size to which these monomers belong. An estimate of this time is

$$\tau_k \sim (K_{k,P})^{-1} . \tag{4.6}$$

To get rid of the temperature dependent prefactor (it also depends on the properties of the surrounding fluid) we can instead compute the ratio

$$r(k) = \frac{\tau_k}{\tau_{k+1}} , \tag{4.7}$$

which already provides the information needed to answer the question. If the system is well-mixed, the equation $c_1 = c_k \cdot k$ holds, therefore

$$\frac{c_i}{c_j} = \frac{c_1/i}{c_1/j} = \frac{j}{i} , \tag{4.8}$$

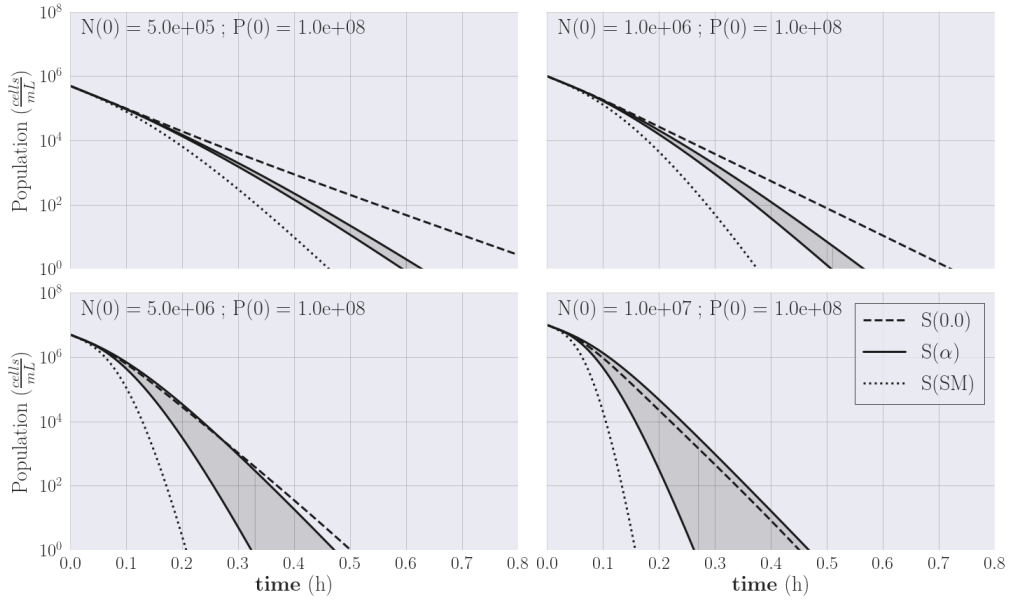


Figure 4.10: Example trajectories for a fixed viral initial concentration of $P(0) = 10^8\ PFU/mL$ and a progressive increase in initial bacteria concentration. The same comparison between SM and PPAM for $\alpha \in [0.0 - 1.0]$ is carried.

and the typical searching time ratio is (Figure 4.11- *a*)

$$r(k) \sim \left(\frac{K_{k,P}}{K_{k+1,P}}\right)^{-1} = \frac{c_{k+1}}{c_k} \cdot \frac{h_{k+1}}{h_k} = \frac{k}{k+1} \cdot \frac{h_{k+1}}{h_k}\,. \tag{4.9}$$

The ratio displays interesting information on the early clustering of monomers. Whereas big clusters tend to be equally likely to be found, for a fixed total mass, being distributed in dimers, trimers or any other low number implies a relevant difference in viral absorption rates. For example, the searching time for monomers is 60% of that for dimers, since $r(1) \approx 0.607$, and the searching time for dimers is 75% of that for trimers, since $r(2) \approx 0.75$. This quantity is useful if we want to compare clustering differences in neighboring sizes. With this in mind, if we want to see absolute differences, a complementary quantity to look at is the k-sized aggregate searching time, τ_k, expressed in units of τ_1 (Figure 4.11- *b*), *i.e.*,

$$\frac{\tau_k}{\tau_1} = \frac{h_1}{h_k} \cdot k = \frac{12.1 \cdot k}{2 + 10k^{1/3} + 0.1k^{-1/3}}\,, \tag{4.10}$$

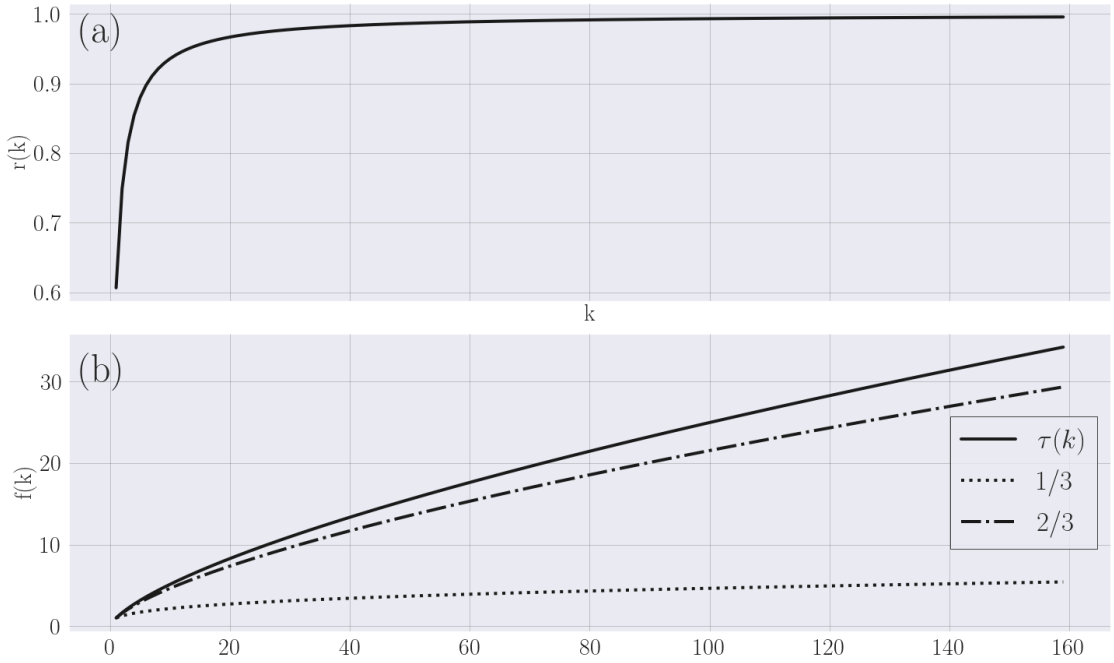


Figure 4.11: (a) Behavior of the integer-valued searching time ratio function, $r(k)$. As cluster size increases the difference becomes asymptotically negligible, since $r(k) \to 1$ as $k \to \infty$. On the other hand, small clustering differences at low k values represent a major trait in phage adsorption rates, as reflected by the value $r(1) = \tau_1/\tau_2 \approx 0.607$. This is, the searching for monomers is 60% of that for dimers, on average. (b) Target search time for different sizes of cluster densities. Time is expressed in units of τ_1. Along with $\tau(k)$, as a scale to compare, the functional forms of $f(k) = k^{1/3}$ as well as $f(k) = k^{2/3}$ are also plotted. Searching time scales as the surface of the clusters bacteria are embedded in, as one could expect.

given that $h(1) = 1$.

Thus, for $\alpha = 0.0$, once a phage finds a cluster after a longer searching time, its progeny will have to resume a new search after lysis without the locality reward of having the same cluster nearby.

As we can see, for the studied system, clustering takes place in such a way that the typical cluster mass remains low with respect the total amount of bacteria and phages in the system (see the typical mass time evolution in Figures 4.5 and 4.6). The typical cluster size is smaller than any $\alpha\beta$ combination (*T7*'s burst size is $\beta = 260$) for $\alpha \in [0.1 - 1.0]$. This means that clusters will tend to end up infected as a whole and die, with the difference settled in the fraction of the phage progeny entering the system. This trait makes clusters into phage sinks, of different "quality" depending on the fraction of re-absorbed viral progeny. Different fractions of re-absorbed progeny makes the free-phage popula-

tion slightly different, thus creating a continuum of sensitive bacteria trajectories, hence the grey area from Figures 4.9 and 4.10.

If the typical cluster size is greater than some $\alpha \cdot \beta$ values, the fate of the system becomes harder to elucidate. This could represent a reasonable future question to ask within our framework.

5

Discussion and Conclusions

5.1 | Achieved Aims and Objectives

In this work two questions were raised. Firstly, we addressed aggregate formation caused by diffusion, inspired by the observations of (4). From a mass-action kinetics aggregation framework, we constructed, with the help of the brownian kernel, a way to describe colloidal coalescence processes ruled by diffusion. From this basis, we then coupled bacterial exponential growth to obtain the BAM, a model that aims to imitate bacterial clustering caused by both, encounter and merging events between aggregates, and cell divisions within each cluster. With this model, we found the existence of a density region where aggregation displays a relevant role in the distribution of cluster sizes, c_k, or P_k. This density region is, furthermore, where typical microbiology experiments lay - bacterial overnight cultures reach a stationary phase of $\sim 10^9$ $cells/mL$. In a context where the growth rate was set to zero, $g = 0$, the aggregation time-scale for densities around $c \sim 10^7 - 10^8$ $cells/mL$ displayed a similar capacity to increase the system's typical mass than that associated with exponentially growing cells on the volume. Conversely, when monomers were let to grow, the model with aggregation showed an acceleration over the clustering dynamics that overcame exponential growth alone, even for density regions 3 orders of magnitude below, *i.e.*, $c \sim 10^5$ $cells/mL$. This indicates that diffusion plays a relevant role in colony formation in liquid environments, where bacteria are allowed to move.

Secondly, we addressed the ecological implications of bacterial aggregation when exposed to predation by phages in a well-mixed scenario. We used the BAM model as a basis to expand on, and implement phage-bacteria interactions. Phages entered the scene as colloidal particles, smaller in size than bacteria. Consequently, the brownian

kernel was implemented to regulate the phage-bacteria diffusive search and encounter rate. Simulated trajectories displayed a time delay in the crashing point. This time delay is mainly related to aggregation effects, given the quasi-linear behavior of the system's typical mass. At the simulated time-scales growth turned out to have a secondary role in the overall dynamics. This can be partially explained by the way we modelled lysis events. They were treated as *one-step Poisson processes* with constant rate $r = 1/\tau$. The downside of this assumption is that the inter-event distribution is exponential, and is not peaked around the average value, or latency time, as it would be with a 10-*step Poisson process* [1]. This means lysis events will start at time $t = 0$ (instead of being distributed around $\tau = 0.22\ h$), hence growth is hampered even more. Although this is a trait shared by both the PPAM and the SM it indirectly affects aggregation rates, via the size dependence of the brownian kernel. A future aim, to improve this approach, could be the development of single infected states that lysis following a distribution peaked around its average value, which is set to the experimentally obtained latency time. Although it might not be a trivial task.

Since all presented burst sizes (see table 2.1) are quite bigger than the displayed typical cluster mass, the above arguments become fairly general, and not pathogen-specific. Given the displacement of the crashing point, and from a kinetic stand point, aggregation events into small clusters could represent a plausible survival trait upon a transient viral invasion.

It is worth repeating here that this study discusses aggregation traits from a purely kinetic point of view. It completely neglects structures or idiosyncrasies associated with bacterial colonies, such as biofilm related properties (5), colony protection against phage attack (15; 16) or the existence of nutrient gradients within the colony, affecting bacterial growth, among many properties present in these complex structures.

- Future Work

The information loss in the coarse-grained approach, if found relevant, could be corrected in future work by the use of multi-scale simulations, which, for example, could help to obtain the attachment probability of two bacterium upon encounter, for a determined set of parameters such as the temperature, the dynamic viscosity of the surrounding fluid, or the bacterium's morphology. This shall be done, perhaps, from the

[1] However, this alternative and more suitable latency time treatment would involve a 10-fold increase in the number of differential equations to be integrated, making the analysis far heavier.

perspective of Molecular Dynamics Simulations, where the shear between two surfaces could be obtain, and could be introduced in this coarse-grained view via a multiplicative parameter in the terms of the type $\sim K_{ij}c_ic_j \to f \cdot K_{ij}c_ic_j$.

Finally, since the brownian kernel is physically important, and we lack of an analytic solution, it would be useful to study in which situations can be approximated to the constant kernel, since this last is computationally less expensive to simulate.

References

[1] William B. Whitman, David C. Coleman, and William J. Wiebe *Prokaryotes: The unseen majority* PNAS 95 (12) 6578-6583; https://doi.org/10.1073/pnas.95.12.6578 (June 9, **1998**).

[2] Sender R, Fuchs S, Milo R *Revised Estimates for the Number of Human and Bacteria Cells in the Body*. PLoS Biol 14(8): e1002533. https://doi.org/10.1371/journal.pbio.1002533 (**2016**).

[3] Suckjoon Jun et al Rep. Prog. Phys. 81 056601 (**2018**).

[4] Kragh KN, Alhede M, Rybtke M, Stavnsberg C, Jensen PO, Tolker-Nielsen T. *The inoculation method could impact the outcome of microbiological experiments*. Appl Environ Microbiol 84:e02264-17. https://doi.org/10.1128/AEM.02264-17 (**2018**).

[5] Flemming, HC., Wingender, J. *The biofilm matrix*. Nat Rev Microbiol 8, 623–633. https://doi.org/10.1038/nrmicro2415 (**2010**).

[6] Stewart, P., Franklin, M. *Physiological heterogeneity in biofilms*. Nat Rev Microbiol 6, 199–210. https://doi.org/10.1038/nrmicro1838 (**2008**).

[7] Rappé MS, Giovannoni SJ. *The uncultured microbial majority*. Annu Rev Microbiol. 57:369-94. https://doi-org.ep.fjernadgang.kb.dk/10.1146/annurev.micro.57.030502.090759 (**2003**).

[8] *Population dynamics*. Nat Rev Microbiol 14, 543. https://doi.org/10.1038/nrmicro.2016.123 (**2016**).

[9] Bergkessel, M., Basta, D. & Newman, D. *The physiology of growth arrest: uniting molecular and environmental microbiology*. Nat Rev Microbiol 14, 549–562. https://doi.org/10.1038/nrmicro.2016.107 (**2016**)

[10] Bergh, Ø., Børsheim, K., Bratbak, G. et al. *High abundance of viruses found in aquatic environments*. Nature 340, 467–468. https://doi-org.ep.fjernadgang.kb.dk/10.1038/340467a0 (**1989**).

[11] Wommack, K. and Colwell, R. Virioplankton: Viruses in aquatic ecosystems. Microbiol. Mol. Biol. Rev. 64, 69–114. https://doi-org.ep.fjernadgang.kb.dk/10.1128/MMBR.64.1.69-114.2000 (**2000**).

[12] Breitbart M, Rohwer F. *Here a virus, there a virus, everywhere the same virus?* Trends Microbiol ; 13(6):278-84. https://doi.org/10.1016/j.tim.2005.04.003 PMID: 15936660 (Jun , **2005**.

[13] Heldal, M. and Bratbak, G. *Production and decay of viruses in aquatic environments*. Mar. Ecol. Prog. Ser. 72, 205–212 (**1991**).

[14] Steward, G.F. et al. *Estimation of virus production in the sea: II. Field results*. Mar. Microb. Food Webs 6, 79–90 (**1992**).

[15] Eriksen, R. S., Svenningsen, S. L. L., Sneppen, K. & Mitarai, N. A *growing microcolony can survive and support persistent propagation of virulent phages*. Proc. Natl. Acad. Sci. **115**, 337–342, https://doi.org/10.1073/pnas.1708954115 (**2018**).

[16] Eriksen, R.S., Mitarai, N. & Sneppen, K. *Sustainability of spatially distributed bacteriaphage systems*. Sci Rep 10, 3154. https://doi.org/10.1038/s41598-020-59635-7 (**2020**).

[17] Emilia L. Simmons, Matthew C. Bond, Britt Koskella, Knut Drescher, Vanni Bucci, Carey D. Nadell *Biofilm Structure Promotes Coexistence of Phage-Resistant and Phage-Susceptible Bacteria* mSystems , 5 (3) e00877-19; DOI: 10.1128/mSystems.00877-19 https://doi.org/10.1128/mSystems.00877-19 Jun (**2020**).

[18] Krapivsky, P., Redner, S., & Ben-Naim, E. *A Kinetic View of Statistical Physics*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511780516 (**2010**).

[19] Einstein, A. *On the movement of small particles suspended in a stationary liquid demanded by the molecular-kinetic theory of heat*. Ann. Phys. **17**, 549-560 (**1905**).

[20] *Development and Application of the Theory of Brownian Motion* Coffey, William; Evans, Myron W Advances in Chemical Physics, **1985**-01-01, p.69-252

[21] Berg, H. C. *Random walks in biology*. Princeton, N.J: Princeton University Press. (**1983**).

[22] J. Chem. Phys. 59, 1669 (**1973**); https://doi.org/10.1063/1.1680247

[23] Chandrasekhar, S. Rev. Mod. Phys. 15, 1 (**1943**), p.59-63.

[24] Smoluchowski, M. (**1906**) *Ann. Phys.* 21 756

[25] Smoluchowski, M. *Versuch einer mathematischen Theorie der Koagulationkinetic kolloider Loesungen*. Z. Phys. Chem. (**1917**), 92, 129–168.

[26] McAllister, W. & Hinton, D. M. *Lytic Transcription*, Encyclopedia of Virology, 4th Edition, Volume 4. https://doi.org/10.1016/B978-0-12-809633-8.20951-5 **2021**

[27] Sneppen, K. (2014). *Models of Life: Dynamics and Regulation in Biological Systems*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107449442 - Chapter 12

[28] Shearwin, K. E. & Truong, J. Q. *Lysogeny*, Encyclopedia of Virology, 4th Edition, Volume 4. https://doi.org/10.1016/B978-0-12-809633-8.20963-1 **2021**

[29] De Paepe M, Taddei F. *Viruses' Life History: Towards a Mechanistic Basis of a Trade-Off between Survival and Reproduction among Phages*. PLoS Biol 4(7): e193. https://doi.org/10.1371/journal.pbio.0040193 (**2006**).

[30] Yin J, McCaskill J. *Replication of viruses in a growing plaque: a reaction diffusion model*. Biophys J 61:1540. http://dx.doi.org/10.1016/S0006-3495(92)81958-6 (**1992**).

[31] You L, Yin J. *Amplification and spread of viruses in a growing plaque*. J Theor Biol 200:365–373. https://doi.org/10.1006/jtbi.1999.1001 (**1999**).

[32] J. K. Venard and R. L. Street *Elementary Fluid Mechanics*, 5th ed., Wiley, New York (**1975**).

[33] Tavaddod, S., Charsooghi, M. A., Abdi, F., Khalesifard, H. R. & Golestanian, R. *Probing passive diffusion of flagellated and deflagellated Escherichia coli*. The Eur. Phys. J. E 34, 16, (**2011**).

[34] Information obtained from: https://www.britannica.com/science/virus/Size-and-shape

[35] R. Moldovan, E. Chapman-McQuiston, X.L. Wu, *On Kinetics of Phage Adsorption*, Biophysical Journal, Volume 93, Issue 1, **2007**, Pages 303-315, ISSN 0006-3495, https://doi.org/10.1529/biophysj.106.102962.

A

Diffusion

A.1 | Solution of the Diffusion Equation

The diffusion equation describes the redistribution of particles in space, as time goes by, caused by particle density gradients. We here present the solution for the displacement in time of a set of colloids initially at $\vec{x_0} = (x_0, y_0, z_0)$. All particles are considered identical and particle-particle interactions are neglected. In this particular case solving for the individual particle probability distribution, $P(\vec{x}, t)$, is equivalent than doing it for the particle concentration $c(\vec{x}, t)$, since $c(\vec{x}, t) = N \cdot P(\vec{x}, t)$, with N equal to the initial number of colloids.

The equation

$$\frac{\partial P(\vec{x}, t)}{\partial t} = D \frac{\partial^2 P(\vec{x}, t)}{\partial \vec{x}^2} \tag{A.1}$$

is solved under the initial condition $P(\vec{x}, t_0) = \delta(x - x_0)\delta(y - y_0)\delta(z - z_0)$ and the boundary conditions $P(\vec{x}, t) \to 0$ and $\vec{\nabla} P(\vec{x}, t) \to \vec{0}$ as $|\vec{x}| \to \infty$.

We take the Fourier Transform (FT)of $P(\vec{x}, t)$ in space, defined as

$$Q(k, t) = \int_{\mathbb{R}^3} P(\vec{x}, t) e^{i\vec{k}\cdot\vec{x}} d\vec{x} \tag{A.2}$$

to obtain a simpler equation to solve. The transformed diffusion equation reads as

$$\frac{\partial Q(\vec{k},t)}{\partial t} = \int_{\mathbb{R}^3} D\vec{\nabla}^2 P(\vec{x},t) e^{i\vec{k}\cdot\vec{x}} dxdydz \,. \tag{A.3}$$

The laplace operator divides the integration into three terms of the type

$$\int_{\mathbb{R}^3} D\frac{\partial^2 P(\vec{x},t)}{\partial x^2} e^{i\vec{k}\cdot\vec{x}} dx = D\int_{-\infty}^{\infty} e^{ik_y\cdot y} dy \int_{-\infty}^{\infty} e^{ik_z\cdot z} dz \cdot I_x \,, \tag{A.4}$$

with

$$I_x = \int_{-\infty}^{\infty} \frac{\partial^2 P(\vec{x},t)}{\partial x^2} e^{-k_x\cdot x} = -k_x^2 \int_{-\infty}^{\infty} P(\vec{x},t) e^{-ik_x\cdot x} dx \,. \tag{A.5}$$

Result one can arrive to after applying integration by parts twice and both boundary conditions. We therefore have

$$\frac{\partial Q(\vec{k},t)}{\partial t} = -D(k_x^2 + k_y^2 + k_z^2)\cdot Q(\vec{k},t) = -D\vec{k}\cdot Q(\vec{k},t) \tag{A.6}$$

This equation is solved with the initial condition

$$Q(\vec{k},t_0) = \int_{-\infty}^{\infty} \delta(x-x_0)\delta(y-y_0)\delta(z-z_0) e^{i\vec{k}\cdot\vec{x}} dxdydz = e^{i\vec{k}\cdot\vec{x_0}} \tag{A.7}$$

to give

$$Q(\vec{k},t) = e^{-D\vec{k}^2(t-t_0)+i\vec{k}\cdot\vec{x_0}} \,. \tag{A.8}$$

We perform now the inverse transform, defined as

$$P(\vec{x},t) = \frac{1}{(2\pi)^3}\int_{-\infty}^{\infty} Q(\vec{k},t) e^{i\vec{k}\cdot\vec{x}} \,. \tag{A.9}$$

The integral factorizes into three terms of the type

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-(Dk_x^2 t + ik_x x)} dx \,, \tag{A.10}$$

which are solved by rewriting the exponent in such a way that the integral becomes

$$\frac{e^{-ab^2}}{2\pi}\int_{-\infty}^{\infty} e^{-a(x+ib)^2}dk_x\,. \tag{A.11}$$

In this case one can easily show that $a = D(t-t_0)$ and $b = \frac{x-x_0}{2a}$. The above integral is equivalent to the gaussian integral [1]

$$\frac{e^{-ab^2}}{2\pi}\int_{-\infty}^{\infty} e^{-ax^2}dx\,. \tag{A.12}$$

Hence, we finally get to the single-particle probability function

$$\begin{aligned} P(\vec{x},t) =& \frac{1}{(2\pi)^3}\cdot e^{-((x-x_0)^2+(y-y_0)^2+(z-z_0)^2)/4Dt-t_0}\cdot\sqrt{\frac{\pi}{Dt}}^3 \\ =& \frac{1}{\sqrt{4\pi D(t-t_0)}^3}\cdot e^{-(\vec{x}-\vec{x_0})/4D(t-t_0)} \end{aligned} \tag{A.13}$$

The quantity $P(\vec{x},t)d\vec{x}$ gives the probability of finding the particle at the point $[\vec{x},\vec{x}+d\vec{x}]$ at time t. The particle concentration for the original problem with N particles is obtained with the change $c(\vec{x},t) = N\cdot P(\vec{x},t)$, *i.e.*,

$$c(\vec{x},t) = \frac{N}{\sqrt{4\pi D(t-t_0)}^3}\cdot e^{-(\vec{x}-\vec{x_0})/4D(t-t_0)} \tag{A.14}$$

[1] Note that this type of integral can be calculated by integrating twice and changing to polar coordinates. If

$$\int_{-\infty}^{\infty} e^{-ax^2}dx \equiv J,$$

then

$$J^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-a(x^2+y^2)}dxdy = \int_0^{\infty}\int_0^{2\pi} e^{-ar^2}rdrd\theta = 2\pi\left(\frac{e^{-ar^2}}{-2a}\bigg|_0^{\infty}\right) = \frac{\pi}{a} \longrightarrow J = \sqrt{\frac{\pi}{a}}$$

Since diffusion is a non-directional transport mechanism the average displacement is zero, as we can easily compute

$$\langle \vec{X}(t) \rangle = \int_{\mathbb{R}^3} \vec{x} \cdot P(\vec{x}, t) dV = 0 \tag{A.15}$$

since the integration is performed over an odd function in a symmetric interval. An alternative magnitude to make a rough estimate of the average displacement of a particle comes from the square root of the variance,

$$Var(\vec{X}(t)) = \langle \vec{X}(t)^2 \rangle - \langle \vec{X}(t) \rangle^2 = \langle X(t)^2 + Y(t)^2 + Z(t)^2 \rangle = \langle X(t)^2 \rangle + \langle Y(t)^2 \rangle + \langle Z(t)^2 \rangle \tag{A.16}$$

where each of the last three terms has a value of

$$\langle X(t)^2 \rangle = \int_{\mathbb{R}^3} \frac{1}{\sqrt{4\pi D(t - t_0)}} \cdot x^2 \cdot e^{-(\vec{x} - \vec{x_0})^2 / 4D(t - t_0)} dV = 2D(t - t_0) \,, \tag{A.17}$$

hence

$$\Delta |\vec{x}| \sim \sqrt{6D(t - t_0)} \,. \tag{A.18}$$

A.2 | Diffusion to a Spherical Adsorber

What is the rate at which small particles, perhaps nutrients or phages, collide against a bigger and approximately static system? This is an interesting question that can be answered within the scope of diffusion.

In this static target limit, the coalescence rate is calculated as the rate of adsorption to a sphere of some radius, R, by some smaller colloids characterized by some diffusion constant, D. Each particle reaching the surface is instantly swallowed up by the bigger target, and the system is big enough to not notice about local coalescence events to it. These two properties translate to the boundary conditions $c(r = R, t) = 0$ and

$c(r \to \infty, t) = c(r, t = 0) \equiv c_0$, respectively.

Since the system is assumed to be well-mixed there are no spatial preferred direction and we can impose spherical symmetry for the solution of the diffusion equation, $c(\vec{r}, t) = c(r, t)$. Luckily enough, this assumption helps us simplify the solving by relating the three-dimensional Laplacian operator to that in one dimension:

$$\frac{\partial c(r,t)}{\partial t} = \nabla^2_{3d}\, c(r,t) = \frac{1}{r}\frac{\partial^2}{\partial r^2}[r\, c(r,t)] = \frac{1}{r}\nabla^2_{1d}\,[r\, c(r,t)] \tag{A.19}$$

Thus, if we define $u(r,t) \equiv r\ c(r,t)$ we can solve for the equivalent one-dimensional system

$$\frac{\partial u(r,t)}{\partial t} = D\nabla^2_{1d} u(r,t). \tag{A.20}$$

We shall first see, for clarity, the general lines for a "purely" 1-dimensional system with its corresponding boundary conditions. Once solved we can mimic the procedure for the equivalent A.20, with its corresponding, and slightly different, boundary conditions. We adopt the following notation to underline the small parenthesis we are taking

$$\frac{\partial u(x,t)}{\partial t} = D\frac{\partial^2 u(x,t)}{\partial x^2}. \tag{A.21}$$

Boundary conditions:

1. Constant initial concentration $\qquad u(x, t = 0) = u_0$
2. Absorbing boundary conditions on a point-like particle $\qquad u(x = 0, t) = 0$
3. Thermodynamic limit $\qquad u(x \to \infty, \forall t) = u(x, 0) = u_0$

Solution:

We define the Laplace transform with respect to t as:

$$\mathfrak{L}\{f(x,t)\} \equiv \int_0^\infty f(x,t)\, e^{-st} dt = F(x,s) \tag{A.22}$$

With the following properties in mind: $\mathfrak{L}\{f_t(x,t)\} = s\mathfrak{L}\{f(x,t)\} - f(x,0)$; $\mathfrak{L}\{f_{xx}(x,t)\} = \frac{\delta^2}{\delta x^2}\mathfrak{L}\{f(x,t)\}$ we transform the PDE we want to solve to get

$$s\ \mathfrak{L}\{u(x,t)\} - u(x,0) = D\frac{\delta^2}{\delta x^2}\mathfrak{L}\{u(x,t)\} \tag{A.23}$$

which, in typical notation, reads as

$$s\ U(x,s) - u_0 = DU''(x,s) \tag{A.24}$$

The particular solution is $U_p(x,s) = \frac{u_0}{s}$. The solution to the homogeneous equation is of the form

$$U(x,s) = A(s)e^{-x\sqrt{s/D}} + B(s)e^{x\sqrt{s/D}} \tag{A.25}$$

We now apply (3) to set $B(s) = 0$ and (2) to determine $A(s)$:

$$U(0,s) = \mathfrak{L}\{u(0,s)\} = 0 = A(s) + \frac{u_0}{s} \longrightarrow A(s) = -\frac{u_0}{s}$$

Thus,

$$U(x,s) = \frac{u_0}{s}\left(1 - e^{-x\sqrt{s/D}}\right), \tag{A.26}$$

which has an inverse transform of the form

$$\mathfrak{L}^{-1}\left(u_0\ \frac{1 - e^{-x\sqrt{s/D}}}{s}\right) = u_0\ erf\left(\frac{x}{\sqrt{4Dt}}\right) \tag{A.27}$$

Hence,

$$u(x,t) = \frac{2\ u_0}{\sqrt{\pi}}\int_0^{x/\sqrt{4Dt}} e^{-z^2}dz \tag{A.28}$$

Once we have the solution for the 1-dimensional case we can extrapolate to the 3-dimensional system by adapting the boundary conditions.

Boundary conditions:

1. Constant initial concentration

$$u(r, t = 0) = r\, c(r, t = 0) = r\, c_0$$

2. Absorbing boundary conditions on a finite-sized particle of radius a

$$u(r = a, t) = a \cdot c(r = a, t) = 0$$

Solution:

The solution follows the general lines of the former development with the following modifications

Lagrange transform: $sU(r,s) - r\, c_0 = DU''(r,s)$

Particular solution: $U(r,s) = \frac{r\, c_0}{s}$

Homogeneous solution: $A(s) = -\frac{c_0}{s} e^{-a\sqrt{s/D}}$

Thus,

$$U(r,s) = \frac{c_0}{s}\left(r - ae^{-(r-a)\sqrt{s/D}}\right), \tag{A.29}$$

expression from which we get our desired solution to the time-dependent diffusion equation

$$u(r,t) = \mathfrak{L}^{-1}\{U(r,s)\} = c_0\left[r - a - a\, erf\left(\frac{r-a}{\sqrt{4Dt}}\right)\right] = r\, c(r,t), \tag{A.30}$$

or, inverting the change,

$$c(r,t) = c_0\left[1 - \frac{a}{r} + \frac{a}{r}\, erf\left(\frac{r-a}{\sqrt{4Dt}}\right)\right] = c_0\left[1 - \frac{a}{r} + \frac{2a}{r\sqrt{\pi}}\int_0^{(r-a)/2\sqrt{Dt}} e^{-z^2} dz\right]. \tag{A.31}$$

A.3 | Out of the Static Target Limit

For the case where both particles are similar in size the problem has to be reformulated, *i.e.*, the merging event will now happen when the *relative* displacement between colloids reaches the sum of their radii of influence. When the separation between both particles first reaches $R_{ij} = R_i + R_j$ the reaction occurs.

Luckily enough, the equation

$$J_0\, dt = 4\pi D R c_0 \left(1 + \frac{R}{\sqrt{\pi D t}}\right) dt. \tag{A.32}$$

can be recycled by demonstrating that this relative displacement also follows the laws of Brownian motion with an effective diffusion coefficient $D_{ij} = D_i + D_i$. The problem then translates to the computation of the flux to an absorbing sphere of radius R_{ij} by an effective particle with diffusive constant D_{ij}.

The $D_{ij} = D_i + D_i$ equivalence is established by showing the probability of relative displacement lying within $[\vec{r}, \vec{r} + d\vec{r}]$ is

$$P(\vec{r})d\vec{r} = \frac{d\vec{r}}{\sqrt{4\pi(D_i + D_j)t}^3}\, e^{-r^2/4(D_i+D_j)t}, \tag{A.33}$$

and comparing it with the corresponding result for the displacement of an individual colloid A.13. In A.33 we took, for simplicity $\vec{r_0} = \vec{0}$ at $t_0 = 0$, thus

$$P(\vec{r})d\vec{r} = \frac{d\vec{r}}{\sqrt{4\pi D t}^3} e^{r^2/4Dt}. \tag{A.34}$$

Proof:

$$P(\vec{r},t)d\vec{r} = d\vec{r}\int_{\mathbb{R}^3} P_1(\vec{r_1},t)P_2(\vec{r_1}+\vec{r},t)d\vec{r_1} = \frac{d\vec{r}}{(4\pi D_1 t)^{3/2}(4\pi D_2 t)^{3/2}}\int_{\mathbb{R}^3} e^{-|\vec{r_1}|^2/4D_1 t} e^{-|\vec{r_1}+\vec{r}|^2/4D_2 t} d\vec{r_1} \tag{A.35}$$

The integral factorizes in three independent terms of the form

$$I(x) = \int_{-\infty}^{\infty} e^{-(x_1^2+2x_1x+x^2)/4D_2t} \cdot e^{-x_1^2/4D_1t} dx_1 \, . \tag{A.36}$$

Thus, the problem boils down to solving

$$I(x) = \int_{-\infty}^{\infty} e^{-\frac{1}{4D_1D_2t}[x_1^2(D_2+D_1)+x_1(2xD_1)+x^2(D_1)]} dx_1 \equiv \int_{-\infty}^{\infty} e^{-(ax'^2+bx'+c)} dx' \, . \tag{A.37}$$

By re-writing the second degree polynomial we get a gaussian integral

$$I = e^{(b^2-4ac)/4a} \int_{-\infty}^{\infty} e^{-(2ax'+b)^2/4a} dx' = \frac{e^{(b^2-4ac)/4a}}{2a} \int_{-\infty}^{\infty} e^{-\frac{1}{4a}y^2} dy$$

$$= \frac{e^{(b^2-4ac)/4a}}{2a} \cdot \sqrt{\frac{\pi}{1/4a}} = e^{(b^2-4ac)/4a} \sqrt{\frac{\pi}{a}}$$

Note that this type of integrals can be calculated by integrating twice and changing to polar coordinates. If

$$\int_{-\infty}^{\infty} e^{-ax^2} dx \equiv J,$$

then

$$J^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-a(x^2+y^2)} dxdy = \int_0^{\infty} \int_0^{2\pi} e^{-ar^2} rdrd\theta = 2\pi \left(\frac{e^{-ar^2}}{-2a} \Big|_0^{\infty} \right) = \frac{\pi}{a} \longrightarrow J = \sqrt{\frac{\pi}{a}}$$

We now subtitute back to get

$$I = \sqrt{\frac{4D_iD_j\pi t}{D_i+D_j}} \cdot exp\left[\frac{4x^2D_i^2}{16D_i^2D_j^2t} \cdot \frac{4D_iD_jt}{4(D_i+D_j)} - \frac{x^2D_i}{4D_iD_jt}\right] =$$

$$= \sqrt{\frac{4D_iD_j\pi t}{D_i+D_j}} \cdot exp\left[\frac{x^2D_i}{4D_j(D_i+D_j)t} - \frac{x^2(D_i+D_j)}{4D_j(D_i+D_j)t}\right] =$$

$$= \sqrt{\frac{4D_iD_j\pi t}{D_i+D_j}} \cdot exp\left[\frac{-x^2}{4(D_i+D_j)t}\right].$$

From which

$$P(\vec{r})d\vec{r} = \frac{d\vec{r}}{(4\pi t)^3(D_iD_j)^{3/2}} \cdot \sqrt{\frac{4D_iD_j\pi t}{D_i+D_j}}^3 e^{-r^2/4(D_i+D_j)t} = \frac{d\vec{r}}{\sqrt{4\pi(D_i+D_j)t}^3} e^{-r^2/4(D_i+D_j)t}. \tag{A.38}$$

Equation A.3 is then reformulated in terms of the sum of the raddi of influence $R_{ij} = R_i + R_j$, where the subscripts i and j denote a i-mer and a j-mer, respectively, and in terms of the effective diffusive coefficient $D_{ij} = D_i + D_j$

$$J_{i+j}dt = 4\pi D_{ij}R_{ij}c_ic_j\left(1 + \frac{R_{ij}}{\sqrt{\pi D_{ij}t}}\right)dt. \tag{A.39}$$

B

Predator-prey Aggregation Model

B.1 | Latency Time Distribution

A Poisson point process, or simply *Poisson process*, is a stochastic process where events happen at a constant rate. If a collection of points (representing events) in some space is split in finite size regions, then the number of points within each region is a random variable Poisson distributed, hence the name. If the model is made by discrete states where transitions are restricted to adjacent states the process is additionally labeled as *one-step* Poisson process. Phage-bacteria dynamics are modeled as such, with an intermediate infected state to delay lysis.

As a stochastic process, the time between events fluctuates. However, some useful information comes from how the time between events is distributed. The inter-event interval distribution for a one-step Poisson process is obtained by considering the probability of success at time $t = 0$ and $t = T + \Delta t$, as $\Delta t \to 0$. For this, the interval T is discretized into M segments, hence

$$P_{int}(r\Delta t)\Delta t = (1 - r\Delta t)^M \cdot (r\Delta t) \,, \tag{B.1}$$

with $\Delta t = T/M$, and then the limit $M \to \infty$ is computed. This is

$$P_{int}(T) = \lim_{M\to\infty} r\left(1 - r\frac{T}{M}\right)^M = re^{-rT} \,. \tag{B.2}$$

In both SM and PPAM latency is set to be equal to the average inter-event time of this distribution, which is

$$\langle T \rangle = \int_0^\infty T\, r\, e^{-rT} dT = r \cdot (-) \frac{d}{dr} \int_0^\infty e^{-rT} dT = -r \frac{d}{dr} \left(-\frac{1}{r} e^{-rT} \Big|_0^\infty \right) = -r \frac{d}{dr} \left(\frac{1}{r} \right) = \frac{1}{r} \,. \tag{B.3}$$

B.2 | Lysis

Given a cluster with i sensitive and m infected bacteria, c_i^m, the infected population will undergo lysis with a rate $r(m)$, yet undetermined. Lysis events provoke the cluster to change size, $(i, m) \rightarrow (k, n)$, thus, in principle, for every cluster size there should be incoming and outgoing terms of the type

$$\frac{dc_k^n(L)}{dt} = -r(n) \cdot c_k^n(t) + r(m) \cdot c_i^m(t) \,, \tag{B.4}$$

where the functional dependence, L, stands for the Lysis contribution to the total rate of change in cluster size density. It is understood that the actual explicit functional dependence is time.

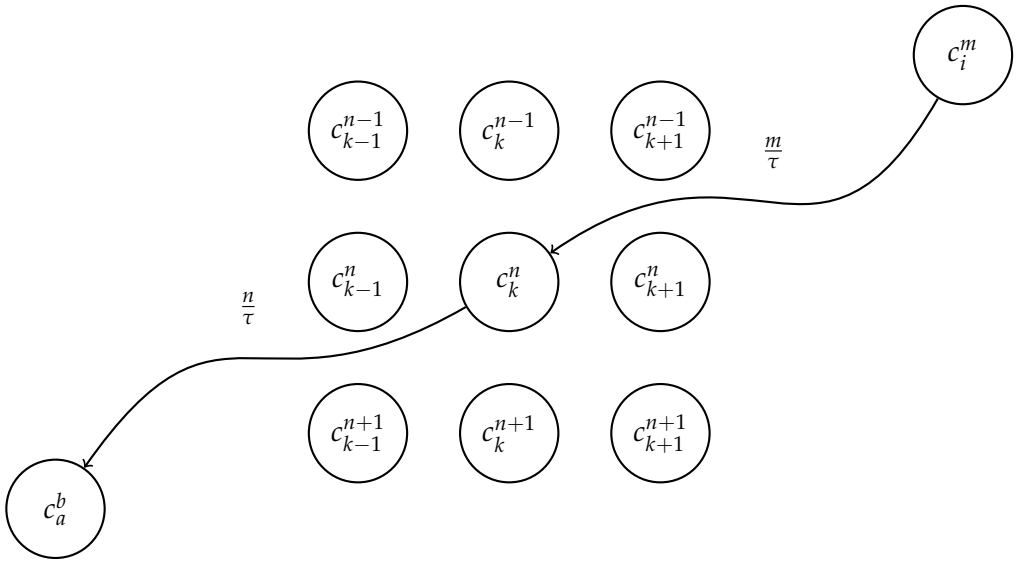
The latency time is set to be equal to the average inter-event time of this distribution, thus, the lysis rate per infected bacterium is

$$r_{lys} = \frac{1}{\langle T \rangle} \equiv \frac{1}{\tau} \,, \tag{B.5}$$

and equation B.4 now reads as

$$\frac{dc_k^n(L)}{dt} = -\frac{n}{\tau} \cdot c_k^n(t) + \frac{m}{\tau} \cdot c_i^m(t) \,. \tag{B.6}$$

A general lysis event is described in the following flow diagram



The relationship between indexes before and after lysis is determined by the burst size β, as well as the probability of the progeny to be reabsorbed by infected and sensitive cells of that same cluster. The reabsorption effect comes in as a weight, $\alpha \in [0, 1]$, to the burst size. To our knowledge, there is a lack of detailed information on the precise mechanisms ruling this process thus, we initially assumed $\alpha \sim 0.5$. Any statement derived from this model must, however, be independent of the specific value of α. Monomer conservation sets the restriction $i + m = n + k + 1$.

With this in mind, equation B.6 describes lysis events $(i, m) \to (k, n)$ such that

$$n = m - 1 + \beta\alpha S_{i,m-1} \tag{B.7a}$$

$$k = i - \beta\alpha S_{i,m-1} \tag{B.7b}$$

where the function $S_{i,m-1}$, known as shielding function, accounts for the absorption to an already infected bacteria. It is assumed that each virus that manages to get back to the cluster (α) and avoid infected cells ($S_{i,m-1}$) finds a bacteria which will not be infected by other viruses that suffer the same fate, this is, there is no infection overlapping. The validity of the pair (i, m) in the above system depends on the (α, β) values and whether $i \geq \beta\alpha S_{i,m-1}$. For a more explicit argument we first need to now the functional form of the shielding function.

Within this framework we don't have a precise way to describe the details of cluster aggregation. This forces us to assume a well-mixed infected-sensible distribution of bacteria within each cluster. The assumption has an impact on the functional form of the shielding function, which now takes a value equal to the probability of a virus being absorbed by an uninfected bacteria, *i.e.*,

$$S_{k,n} = \frac{k}{k+n} \,. \tag{B.8}$$

With an explicit expression for the shielding function we can now study the lysis equations more carefully. The system holds if

$$i \geq \beta \cdot \alpha \cdot S_{i,m-1} = \frac{i}{i+m-1} \cdot \beta \cdot \alpha \tag{B.9}$$

or

$$i + m \geq \beta\alpha + 1 \tag{B.10}$$

The system is easily solved

$$i = \left(\frac{n+k}{n+k-\beta\alpha} \right) \cdot k \tag{B.11a}$$

$$m = n + k + 1 - i \tag{B.11b}$$

From which we see again that $\beta\alpha < n + k = i + m - 1$.

If $i \leq \beta\alpha S_{i,m-1}$ then the system reads as

$$\tilde{n} = m - 1 + i \tag{B.12a}$$

$$k = i - \beta \cdot \alpha \cdot S_{i,m-1} = 0 \tag{B.12b}$$

and the reaction is now $(i, m) \to (0, \tilde{n})$. We effectively have 1 equation for 2 unknowns, this introduces a degeneracy over the pairs (i, m) that after a lysis event go to $(0, \tilde{n})$.

Pairs that fulfil $n + k = m + i - 1$ but are below the $\beta\alpha + 1$ threshold will not contribute to the lysis term for c_k^n. Instead, they will contribute to the pair $(0, \tilde{n})$. Thus, we raise a condition for equation B.6

$$\frac{dc_k^n(L)}{dt} = -\frac{n}{\tau} \cdot c_k^n + \frac{m}{\tau} \cdot c_i^m \,, \qquad i + m \geq \beta\alpha + 1 \tag{B.13}$$

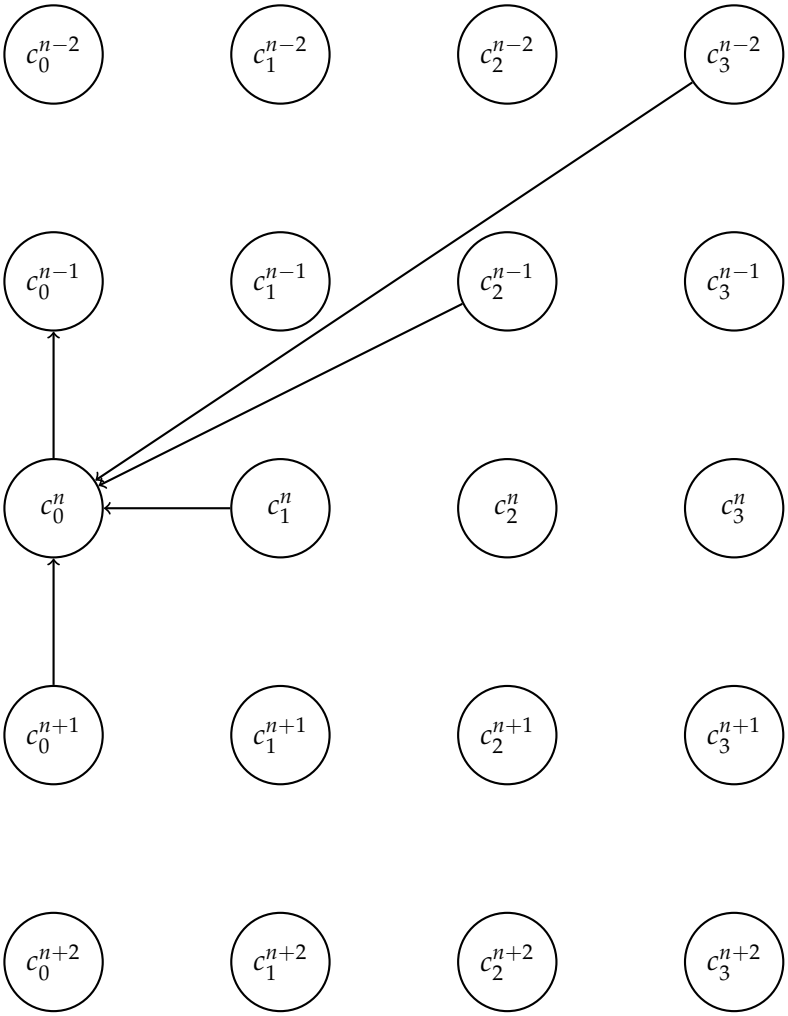
The reaction $(i, m) \rightarrow (0, \tilde{n})$ has a degeneracy an is restricted by

$$\tilde{n} + 1 = i + m \,, \qquad m \neq 0 \tag{B.14}$$

Therefore, there is an extra term for these cases,

$$\frac{dc_0^{\tilde{n}}(L)}{dt} = -\frac{\tilde{n}}{\tau} \cdot c_0^{\tilde{n}} + \sum_{i=0}^{min(\beta\alpha,\tilde{n})} \frac{\tilde{n} + 1 - i}{\tau} \cdot c_i^{\tilde{n}+1-i} \;\;, \;\; i + m \leq \beta\alpha + 1 \tag{B.15}$$

The following flow diagram might be helful to visualize the degeneracy



The no-overlapping-when-infecting assumption here breaks down, resulting in an overkill situation, where clusters act as phage sinks of higher "quality".