

## **ABSTRACT**

LAVIGNE, GARY MICHAEL. Modeling and Quantifying Spatial Strategies of the Innate Immune Response. (Under the direction of Kevin Flores.)

The innate immune response, particularly the interferon response, represents a first line of defense against viral infections. The interferon proteins produced from infected cells signal to nearby cells that an infection is present, causing them to enter a protected anti-viral state. Although the molecular processes involved in interferon signaling have been well studied, the collective effect of interferon signaling to stop the spread of viral infection through a tissue has yet to be rigorously characterized. Here, we use mathematical modeling in both continuous (partial differential equation) and discrete (cellular automata) frameworks to study the host interferon response as a strategy for controlling disease. Furthermore, we contrast the results from spatial and non-spatial models of this process to demonstrate the importance of localized response in the efficacy of interferon response. Specifically, we find that cell-to-cell signaling of interferon is only impactful in a spatially structured infection, whereas its effect is minimal in a well-mixed infection. However, same-cell signaling can be efficacious in a well-mixed environment but loses its effectiveness when infection is spatially structured. Furthermore, we argue that the interferon response can be seen as a parallel to population-level epidemic prevention strategies such as "contact tracing" or "ring vaccination". Thus, our results here may have implications for the outbreak control at the population scale more broadly. We lastly employ a variety of established and novel data-driven equation learning frameworks to draw conclusions about the impact of interferon signaling on the spread of disease. We study how high levels of interferon signaling can increase the stochasticity of the spread of infection.

© Copyright 2020 by Gary Michael Lavigne

All Rights Reserved

Modeling and Quantifying Spatial Strategies of  
the Innate Immune Response

by  
Gary Michael Lavigne

A dissertation submitted to the Graduate Faculty of  
North Carolina State University  
in partial fulfillment of the  
requirements for the Degree of  
Doctor of Philosophy

Applied Mathematics

Raleigh, North Carolina

2020

APPROVED BY:

---

Mansoor Haider

---

Alun Lloyd

---

Ralph Smith

---

Kevin Flores  
Chair of Advisory Committee

## **BIOGRAPHY**

Michael was raised in Mandeville, Louisiana, just outside of New Orleans. Michael attended Tulane University from 2011 to 2015 where he graduated *summa cum laude* with a B.S. in Mathematics and a B.A. in Spanish. His interests include cooking, foreign languages, and the piano.



## **ACKNOWLEDGEMENTS**

This work was only possible due to countless conversations and brainstorming sessions with my many colleagues of diverse styles, perspectives, and backgrounds. I would like to thank Ruian Ke for being an incredible mentor and friend. I would to thank the members of the Ke lab – Marco Hamins-Puertolas, Hayley Russell, and Alex Farrell – for their camaraderie and invaluable interdisciplinary perspectives on mathematical biology. I would like to thank Kevin Flores for deftly shepherding me through the many difficulties, mathematical and personal, of this process. I thank the Flores Lab – John Nardini, John Lagergren, and Erica Rutter – for always setting the bar so high. I thank Barbara Sherry of the NCSU School of Veterinary Medicine for her invaluable expertise in mammalian innate immune response. I thank Mallory McMahon for her hours of voluntary copy editing and general commiseration.

## TABLE OF CONTENTS

<b>LIST OF TABLES</b> .....	<b>vi</b>
<b>LIST OF FIGURES</b> .....	<b>vii</b>
<b>Chapter 1 Introduction</b> .....	<b>1</b>
1.1 Review of biology literature on Interferon response .....	4
1.2 Review of modeling literature on Interferon response .....	5
<b>Chapter 2 An ODE Model of Viral Infection with Interferon Response</b> .....	<b>7</b>
2.1 Model Development .....	7
2.2 Equilibrium Analysis of the ODE Model .....	11
2.3 Basic Reproduction Number $R_0$ .....	12
2.4 Approximating Peak Viral Load .....	16
2.5 Approximating Time to Peak Viremia in ODE Model .....	19
2.6 Results .....	19
2.7 Contributions .....	21
<b>Chapter 3 PDE Model of Viral infection with IFN response</b> .....	<b>24</b>
3.1 Introduction .....	24
3.2 Model Development .....	25
3.3 Reaction Diffusion Systems and Traveling Wave Solutions .....	26
3.3.1 A Classical Example: The Fisher Equation .....	27
3.3.2 Traveling Wave Analysis of our PDE Model .....	28
3.3.3 Discriminant Method for Determining Exact Expression for Wave Speed ....	30
3.4 Results .....	33
3.5 Contributions .....	38
<b>Chapter 4 A Cellular Automata Model of in-host Viral Infection</b> .....	<b>39</b>
4.1 Cellular Automata Models .....	40
4.2 Model Development .....	41
4.2.1 Simulating Diffusion by Random Walk .....	42
4.2.2 Simulating Diffusion with a Mean-Field Approximation .....	44
4.3 Results .....	48
4.3.1 Stochasticity of model outcomes .....	53
4.4 Contributions and Future Work .....	54
<b>Chapter 5 Equation Learning</b> .....	<b>58</b>
5.1 Equation Learning via linear regression .....	59
5.2 PDE-Find .....	60
5.3 Learning Governing Equations for Cellular Automata Model .....	62
5.3.1 Reconstruction of continuum density surface .....	64
5.3.2 Library structure .....	65
5.3.3 Results .....	66
5.4 A novel equation learning framework with trainable candidate terms .....	67
5.4.1 The Logistic Equation .....	70
5.4.2 Moment Equations .....	71

5.5 Contributions and Future Work . . . . .	75
<b>Chapter 6 Discussion . . . . .</b>	<b>77</b>
<b>Acronyms . . . . .</b>	<b>85</b>
<b>Glossary . . . . .</b>	<b>86</b>

## LIST OF TABLES

Table 2.1	Parameters of the ODE model. Parameter $\beta$ chosen to produce peak viremia near 2 DPI, $\phi$ taken to match $\beta$ so that $\pi$ represents ratio of efficacious IFN contacts to infection events. Autocrine transition rate $k$ taken to equal $\delta$ so that cell has equal probability of dying or entering autocrine-signaled state. .	9
Table 3.1	Parameters of the PDE model. Parameters $\beta$ and $\phi$ taken to produce reasonable wave speed given remaining parameter set. Diffusion coefficient $D_V$ taken to be 1 as a baseline, $D_F$ taken to be between 1 and 2 orders of magnitude greater than $D_V$ [31]. . . . .	26
Table 4.1	Parameters of the CA model. . . . .	48
Table 4.2	<b>Table of extinction probabilities across CA parameters.</b> The table above shows the percentage of 100 independent simulations that resulted in extinction of infection within the first 150 time steps across varying infectious lifespan and IFN production levels. . . . .	53
Table 5.1	<b>Table showing recovered equation for logistic data</b> . . . . .	71
Table 5.2	<b>Learned moment equations from CA distribution data.</b> Table shows the recovered equation from CA data across 14 levels of IFN production. Trend shows that one term is recovered at low levels of IFN production, while two terms are recovered at higher levels of IFN production. This suggests that a simple power law describes the data well at low levels of IFN production, but may break down at higher levels. . . . .	74

## LIST OF FIGURES

Figure 1.1	<p><b>Conceptual diagram.</b> Infected cells produce IFN in response to the detection of intracellular viral matter. The IFN then exits the infected host cell and binds to cell surface receptors. IFN signaling occurs via paracrine (cell-to-cell) and autocrine (same cell) pathways. Autocrine IFN signaling can slow virion production in the already-infected cell, or can create a completely anti-viral state in uninfected cells. These two signaling pathways are present in both spatially structured infection in epithelial tissues and in well-mixed infection such as in the blood. We will investigate the diverging impact of these two key signaling pathways in both spatial and non-spatial frameworks to understand how they work in concert to protect an immune naïve host from novel viral pathogens. . . . .</p>	3
Figure 2.1	<p><b>Schematic of ODE model of viral infection with interferon response.</b> Schematic diagram with parameters in the model. Solid arrows indicate transition of cells from one state to another; dashed arrows indicate the production or binding of viruses and IFNs from cells. . . . .</p>	10
Figure 2.2	<p><b>Representative simulations of the ODE model.</b> Shown are simulations of the ODE model with no IFN response (solid line), paracrine IFN signaling only (dashed line), and full IFN response (dotted line). The simulations show that paracrine signaling alone does not alter the initial exponential growth rate of the viral load, and can only moderately advance the time to peak viral load by accelerating the depletion of target cells. However, autocrine signaling is able to decrease the exponential growth rate, observable as the slope of the viral load curve prior to the peak at day 2. . . . .</p>	21
Figure 2.3	<p><b>Representative ODE model simulations with supraphysiological values of <math>\pi</math> show little impact of paracrine signaling.</b> These two simulations of the ODE model demonstrate the conclusion that paracrine signaling alone is unable to meaningfully impact the establishment of infection in the spatially homogeneous setting. Since the paracrine signaling acts by depleting the population of target cells, this impact is not sufficiently large during the initial exponential growth phase of the infection. . . . .</p>	23
Figure 3.1	<p><b>Characteristic polynomial <math>p(\lambda; \nu)</math>.</b> The minimum admissible wavespeed <math>\nu^*</math> will be the smallest such that admits non-oscillatory solutions, meaning that we must enforce that <math>p(\lambda; \nu)</math> have only real roots. Therefore, the critical wavespeed <math>\nu^*</math> will be the small that admits real roots, meaning that we look specifically for the value of <math>\nu</math> that produces a double root. Shown are three cases: (1) a subcritical case, where only one real root exists, (2) a supercritical case, with three distinct real roots, and (3) the critical case <math>\nu^*</math> where a double root is attained on <math>\mathbb{R}^+</math>. . . . .</p>	31

Figure 3.2	<p><b>Model simulations show that paracrine IFN signaling strongly interferes with the spatial spread of infection.</b> Shown is the solution of the model system at successive time points, with arrows indicating the direction of progression with time. (a) A representative simulation of the PDE model with <math>k = \pi = 0</math> (no IFN), exhibiting traveling wave behavior initiated from a single nexus of infected cells at position <math>x = 0</math>. The infection travels an equal distance between successive times, demonstrating constant speed of spread. (b) A representative simulation of the PDE model with cell protection included (<math>\pi \neq 0</math>) showing how IFN signaling can stop the spread of infection by depleting target cells. The distribution of virions and infected cells can be seen to remain localized to the far left of the domain, as the rapid depletion of susceptible cells in the domain prevent the infection from establishing a traveling wave. . . . .</p>	36
Figure 3.3	<p><b>Comparison of the impacts of paracrine signaling and autocrine signaling on viral spatial spread.</b> Shown is the position <math>x(t)</math> of the infection front over time for various values of the free IFN production parameter <math>\pi</math> while <math>k = 0</math> (a), and varying the autocrine-mediated transition rate <math>k</math> while <math>\pi = 0</math> (b). Here we define the front position <math>x(t)</math> to be such that <math>I(x(t), t) = 0.01 T_0</math>. The red line shows the predicted front position given the analytically derived wave speed, i.e., <math>x(t) = v^*t</math>. (a) Sufficiently large production of free IFN (<math>\pi = 0.4, 0.6</math>) leads to halting the spread of infectious front. (b) Autocrine parameter <math>k</math> has only small effect on wave speed of infection. The supra-physiologic values <math>k = 10</math> and <math>48 \text{ days}^{-1}</math> (corresponding respectively to 0.1 day and 30 minute waiting periods before infected cells transition to <math>I^*</math> state). . . . .</p>	37
Figure 4.1	<p><b>Schematic diagram of cellular automata model.</b> Model consists of a rectangular lattice of cells in one of five states: Healthy, Exposed, Infected, Protected, and Dead. Infected cells produce virion and IFN particles that diffuse to nearby lattice points, resulting in transitions to the Exposed and Protected states, respectively. After 1 lifespan time steps, Infected cells become Dead. . . . .</p>	43
Figure 4.2	<p><b>Schematic of recipient cell selection process.</b> (a) The particle is initially located at the producing cell, located at lattice position <math>(x_0, y_0)</math>. We interpret its position to be initially delta distributed, i.e., <math>V(x, y, 0) = \delta(x - x_0, y - y_0)</math>. (b) The probability density function of the particle's position is assumed to be given by solving the diffusion equation <math>V_t = D_V \nabla^2 V</math> forward <math>\Delta t</math> time units from the dirac delta initial condition shown in (a). This gaussian functional form of the p.d.f is known explicitly. (c) The particle's binding location <math>(x^*, y^*)</math> is drawn from the p.d.f shown in (b). Since in general <math>(x^*, y^*) \in \mathbb{R}^2</math>, we snap this binding location to a lattice grid point by rounding to determine the recipient cell. This process is identical for both virion and IFN particles, differing only in the value of the diffusion coefficient. . . . .</p>	46

Figure 4.3	<b>Representative simulation of CA model with and without IFN signaling with long-lived infection.</b> The left column shows the progression of infection in the absence of IFN signaling. The lattice is initially populated with Healthy cells (white) at each grid point, with a single infected cell (red) located in the center of the lattice. Each time step, an infected cell produces <code>virus_prod</code> virions, each of which is assigned to a recipient cell according to Algorithm 4.1. If the recipient cell is Healthy, it becomes Exposed (yellow). Otherwise the recipient cell is unchanged. The right column shows the progression of infection in the presence of IFN signaling. Each time step, in addition to virions, an infected cell produces <code>ifn_prod</code> IFNs, each of which is assigned to a recipient cell according to Algorithm 4.1. If the recipient cell is Healthy, it becomes Protected (blue). We see that the abundance of Protected cells limits the number of virion binding sites that can lead to further infections, thus impeding the spread of the infected area. . . . .	50
Figure 4.4	<b>Representative simulation of CA model with and without IFN signaling with short-lived infection.</b> The left column shows the progression of infection in the absence of IFN signaling with short lived infected cells. We see the spread of infection is virtually unimpeded compared to the long-lived infection case, as cell death occurs long after the front of infection has move (i.e., all Healthy cells have been exhausted), meaning that cell death is inconsequential. However, the right column shows the progression of short-lived infection in the presence of IFN signaling. Here, infected cells may become rapidly sequestered by a neighborhood of Protected cells. Some cells may die before creating any secondary infections, leading to the possibility of extinction. . . . .	51
Figure 4.5	<b>Log-log plots of lattice sum of infections over time for long-lived infection.</b> Along the top row are representative lattice states at $t = 100$ for each IFN level. In the bottom row are the total number of Infected cells over time (light blue lines) for 100 independent CA realizations at four different IFN levels graphed on the log-log scale. In red is shown the mean and one standard deviation. The linear trends in these plots suggests a power law of the form $I(t) \propto t^\gamma$ , where $\gamma$ decreases with <code>ifn_prod</code> . . . . .	52
Figure 4.6	<b>Log-log plots of lattice sum of infections over time for short-lived infection.</b> Along the top row are representative lattice states at $t = 100$ for each IFN level. In the bottom row are the total number of Infected cells over time (light blue lines) for 100 independent CA realizations at four different IFN levels graphed on the log-log scale. In red is shown the mean and one standard deviation. The linear trends in these plots suggests a power law of the form $I(t) \propto t^\gamma$ , where $\gamma$ decreases with <code>ifn_prod</code> . However, we see the linear trend begins to fail for <code>ifn_prod</code> = 20, as extinction of infection from the lattice becomes increasingly probable. . . . .	53

Figure 4.7	<b>Time evolution of distribution of total (lattice sum) number of infected cells for different levels of IFN production.</b> Shown are the histograms of the total number of infected cells (sum over entire lattice) over $10^5$ independent realizations of the CA model for three different values of <code>ifn_prod</code> at 4 time points. We observe that the mean number of infections decreases with <code>ifn_prod</code> at all times and that the variance of the distribution seems to grow more quickly for larger <code>ifn_prod</code> . This suggests that high levels of IFN production increases the stochasticity of model outcomes and leads to greater downstream uncertainty of global infection levels. . . . .	55
Figure 4.8	<b>Representative simulations of CA model with viral evolution and distributions of viral phenotype.</b> Shown are two independent simulations of the CA model with viral evolution in the absence ( <code>ifn_prod</code> =0, top row) and presence ( <code>ifn_prod</code> =5, bottom row) of IFN signaling. The CA simulations show infected cells continuously varying in color between black and red to represent infections of phenotype 0 to 1, respectively. Cells protected by IFN are shown in blue, while exposed cells are shown in green/yellow. Below each snapshot of the CA grid is the distribution of viral phenotypes on the lattice at that time. . . . .	57
Figure 5.1	<b>Schematic diagram of PDE-Find algorithm.</b> Solution surface data in the form of a matrix is loaded or constructed. Then, data matrices representing the desired derivative surfaces are constructed using either finite difference method, polynomial interpolation, or some other context-dependent derivative approximation scheme. Matrices are reshaped into columns and point-wise multiplication is used to construct columns representing all pre-supposed candidate terms. Columns representing candidate terms are collocated into library matrix and linear system $\mathbf{u}_t = \Phi \xi$ is formed. Sparse regression is used to recover coefficient vector $\xi$ . Non-zero entries of $\xi$ represent recovered terms. . . . .	62
Figure 5.2	Two dimensional visualization of regression problem under no regularization, $\ell_1$ regularization (lasso), and $\ell_2$ regularization (ridge). (a) The ordinary least squares solution $\xi_{\text{OLS}}^* = \Phi^\dagger \mathbf{u}_t$ is the global minimum of the convex OLS error function $\ \mathbf{u} - \Phi \xi\ _2^2$ . (b) Due to the cusps on the level sets of the $\ell_1$ regularization term $\lambda \ \xi\ _1$ , the lasso solution $\xi_{\text{lasso}}^*$ is more likely to occur on the coordinate axes, thus encouraging sparsity of the solution vector. (c) The ridge solution $\xi_{\text{ridge}}^*$ is pushed towards the origin by the $\ell_2$ regularization term $\lambda \ \xi\ _2^2$ . This approach, coupled with sequential thresholding of small entries of $\xi$ , can result in a sparse solution. . . . .	63
Figure 5.3	<b>Diagram of KDE solution surface reconstruction procedure.</b> a) Noisy surface of consisting of aggregated histograms for $10^5$ independent realizations of CA model showing # infected cells $x$ observed at time $t$ . b) For each time $t_j$ , kernel density estimation is used to reconstruct the PDF $u(x, t_j)$ . c) The individual PDFs are re-aggregated into denoised solution surface. . . . .	65



Figure 5.4	<b>Coefficients of recovered diffusion-advection equation vary with IFN production.</b> Coefficients $c$ and $D$ of diffusion-advection equation $u_t = -c u_x + D u_{xx}$ recovered for varying levels of IFN production. The advection rate $c$ can be interpreted as the average rate at which the number of infections increases. The diffusion coefficient $D$ can be interpreted at the rate at which the variance of the distribution increase. The trend suggests that increasing IFN prouction has the expected consequence of decreasing the average rate at which the number of infections increase, but furthermore increases the rate at which the variance of the distribution increases, suggesting that the stochasticity of the infection process increases with IFN production. . . . .	68
Figure 5.5	<b>Representative simulations of learned equations from logistic data and candidate term training.</b> The left column shows the ground truth data (black dash), solution of equation fit using untrained initial candidate terms (red circle), and solution of final learned equation (blue circle) for three representative logistic datasets. The right column shows the training of the candidate terms corresponding to the data on the left. Shown in blue are the exponents $\gamma_i$ and in red are the coefficients $\xi_i$ , with the true values shown in dotted lines of the same color. Terminated lines represent candidate terms that were removed from the library via merging or thresholding. . . . .	72
Figure 5.6	<b>Parameters of learned moment equations from CA distribution data.</b> . . .	74

## CHAPTER

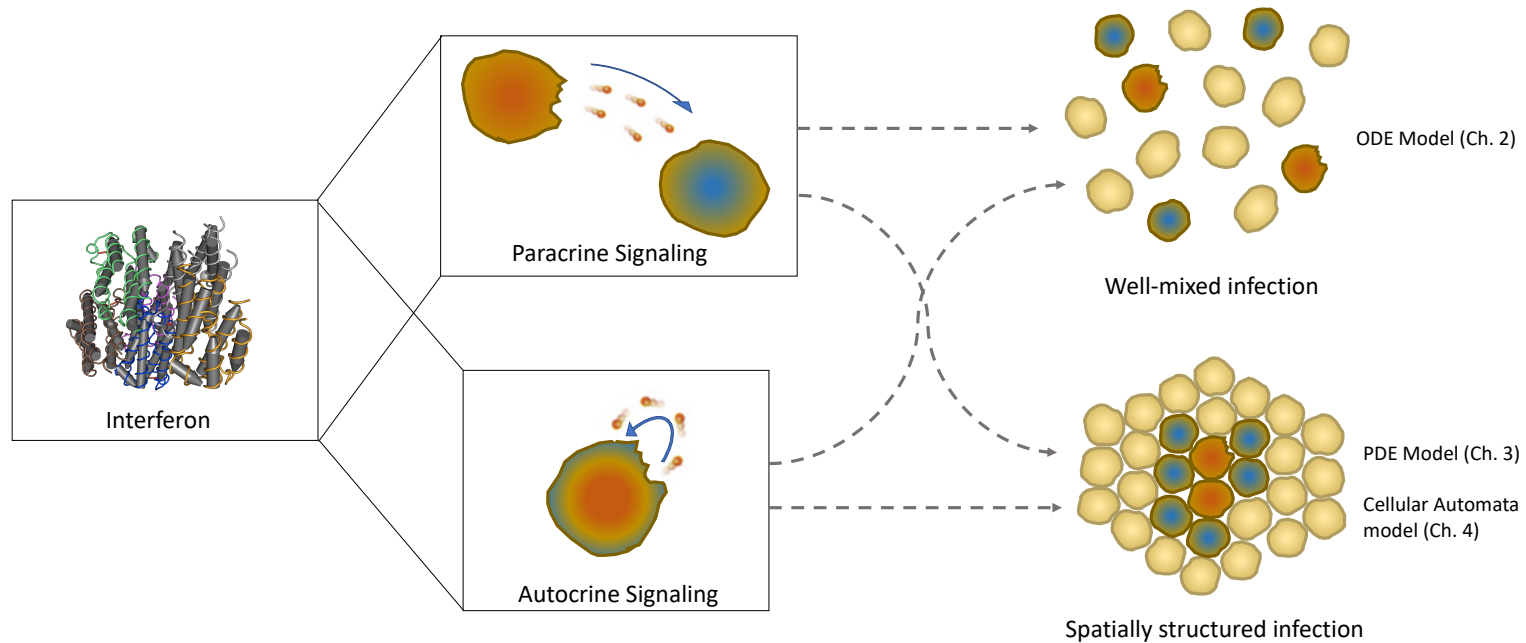
# 1

## INTRODUCTION

Mathematical models have a long history of use in describing biological phenomena across a wide array of temporal and spatial scales from natural selection and species migration to cell motility and insulinemic response. Biological phenomena invariably emerge as the result of local interaction, either of individuals on the population scale or of cells and molecules on the tissue scale. While many models can be of incredible utility without the explicit consideration of spatial arrangement and local interactions, certain phenomena are inherently spatial in nature. This is trivially true in the study of inherently spatial processes such as migration, motility, transport, and diffusion. But local interactions are equally important in the shaping processes such as cancer pathogenesis, embryonic development, and immune response. Especially at the cellular scale, it is the passive diffusion of nutrients, oxygen, and cytokines that is critical for shaping the way tissues organize themselves. Diffusion as a delivery mechanism is limited to this scale in the body, meaning the efficacy of diffusion as a delivery mechanism may act as a metric against which tissues measure themselves. Hypoxia in growing tumors due to insufficient diffusion of oxygen can lead to tumor angiogenesis, the first step in the cancer pathogenesis. In this way, diffusion, though a passive

process, can play a crucial role in the organization of tissues, especially when considered in concert with reaction processes. Alan Turing famously outlined the idea of diffusion-driven instability in his 1952 paper "A chemical basis of morphogenesis", wherein he describes how the passive diffusion of reactive species can result in the formation of patterns [1].

In this thesis, we will explore the importance of local interactions in the efficacy of the host immune response to viral infection. We will focus on the innate (naïve) response to a novel viral infection in the form of Interferon response. We aim to elucidate how this response is efficacious due to its ability to act locally within a tissue via the extension of previous work to spatially explicit modeling frameworks. Specifically, we hope to show how studying the Interferon signaling and viral infection process in consideration of explicit spatial arrangement of cells can lead to different conclusions than the study of the process under an assumption of well-mixedness. The Interferon signal received by a cell can be characterized as a "paracrine signal" if the signal is received from another cell or as an "autocrine signal" if the signal is received by the same cell that produced it, each resulting in different cell phenotypes during infection. We will study these two signaling pathways in both spatial and non-spatial frameworks to understand how they can work in concert to protect the host from viral infections (see Figure 1.1 for a conceptual diagram).



**Figure 1.1 Conceptual diagram.** Infected cells produce IFN in response to the detection of intracellular viral matter. The IFN then exits the infected host cell and binds to cell surface receptors. IFN signaling occurs via paracrine (cell-to-cell) and autocrine (same cell) pathways. Autocrine IFN signaling can slow virion production in the already-infected cell, or can create a completely anti-viral state in uninfected cells. These two signaling pathways are present in both spatially structured infection in epithelial tissues and in well-mixed infection such as in the blood. We will investigate the diverging impact of these two key signaling pathways in both spatial and non-spatial frameworks to understand how they work in concert to protect an immune naïve host from novel viral pathogens.

## **1.1 Review of biology literature on Interferon response**

The innate immune response provides critical protection against pathogen invasion of humans and other animals prior to establishment of adaptive immunity. It relies on multiple cytokines, chief among them being interferons (IFNs), a large, diverse family of signaling proteins that together induce a protective response [2]. The importance of IFN in the defense against viral infections is demonstrated by the fact that essentially all viral pathogens have developed mechanisms to interfere with the host IFN response [3, 4, 5]. Indeed, viral evasion of the IFN response strongly determines the rate of viral replication, the success of transmission and infection establishment in new hosts [6] and the range of species infection [7]. The capacity to inhibit the IFN response determines species tropism for human immunodeficiency virus [8], dengue virus [9], rotavirus [10], measles virus [11], and influenza virus [12].

The IFN response is commonly described by its two components: first, viral induction of IFN, and second, IFN induction of antiviral genes [13]. Upon infection, viral RNAs or DNAs are detected by the cell triggering a signaling cascade that results in the production of Type I IFNs [14, 15]. These IFN molecules are then secreted and bind to surface receptors located on the cell membrane. IFN binding to the surface of the cell from which it is produced is referred to as autocrine signaling, whereas binding to the surface of any other cell is referred to as paracrine signaling. This binding initiates a series of signaling events that ultimately result in the production of Interferon Stimulated Genes (ISGs), the expression of which repress viral replication in the cell at multiple steps ([16]. In an uninfected cell, binding of IFN to its receptor and subsequent IFN signaling renders the cell refractory to viral infection, while in an infected cell, this signaling can suppress viral replication and decrease release of viral progeny from the cell. An elegant analysis of the virus-induced IFN response at the single cell level demonstrated that paracrine signaling early in infection shapes the overall IFN response [16]. However, the inflammatory response elicited by IFN can have deleterious effects on the host if uncontrolled, with high levels of IFN being strongly correlated with host mortality [17].

## 1.2 Review of modeling literature on Interferon response

Although the molecular mechanisms of IFN signaling have been well characterized, the systems-level properties arising from the individual host cell response, particularly how the host cells collectively stop viral infection at the site of exposure before the infection becomes systematic, remain unclear. To address these questions, we use modeling approaches to understand how IFN signaling can stop early infection (i.e., at the site of initial entry) before adaptive immunity is developed. Previous modeling of virus infection and the IFN response has focused on the role of IFN response after the infection becomes systematic and used ordinary differential equations (ODEs) to model the dynamics of global target cell populations [18, 19, 20]. An ODE modeling study of equine IFN response in response to Influenza A Virus (H3N8) infection concluded that IFN response is important in modulating viral dynamics in a naïve host by accelerating the depletion of target cells [19]. A later study also using ODEs of the same experimental data was able to capture a bimodal trend in viral load which was explained by the temporary depletion of target cells due to IFN response followed by the bulk reentry of protected cells into the susceptible class, thus causing two peaks in measured viral titers [20]. A notable limitation of this dataset, however, was the collection method, wherein nasal wash from infected ponies was used as a surrogate for the global viral load of the host, which can be reasonably construed as the entire respiratory tract. ODE models necessarily include the implicit assumption that the host is treated as a single well-mixed compartment, and thus they neglect the spatial structure of infection [21]. ODE models of this type also typically model systemic infection of the host, and are therefore not useful for studying one of the key insights of the interferon literature: that IFN is a highly effective selection pressure and strongly shapes tropism of viral phenotypes. That is to say, ODE models of systemic infection can not capture the fundamental reasons for the efficacy of IFN in quashing incompetent viral infections at the initial site of transmission. Influenza infection, for example, starts at the epithelial lining of the upper respiratory tract, which is composed of a lattice monolayer of non-motile cells, and infection is initiated by a transmission bottleneck as small as 10-200 virions [22]. In HIV-1 infection, it has been shown that 80% of novel chronic infections are initiated by a single founder virus [23], suggesting that the vast majority of virions transmitted are unfit in the face selection pressure at the epithelial transmission barrier such as IFN response. This spatial structure ensures that infection – and the

ensuing interferon response – are inherently a spatial processes that depend of highly localized interactions. Therefore, for many viral pathogens, a spatially explicit model is most appropriate to investigate the interaction between virus and the IFN response early in infection.

In this thesis, we will develop various models with or without explicitly considering the impact of the spatial arrangement of cells to examine the roles of autocrine and paracrine IFN signaling. In Chapter 2 we will develop on ODE model of viral infection with IFN response based on previous work, with the novel inclusion of both autocrine and paracrine pathways for IFN signaling. We show that, in well-mixed ODE models, autocrine signaling can impact the course of infection by inhibiting virus production from already infected cells, thus directly decreasing the exponential growth rate during early infection. However, we will discover that paracrine signaling is less impactful during this early growth phase, as it can only act to decrease the availability of target cells. In Chapters 3 and 4, we will develop spatially explicit models of viral infection with IFN signaling and will show that IFN paracrine signaling can stop viral infection by segregating susceptible cells from areas of infection with an insulating layer of protected cells. In Chapter 3 we will develop a system of reaction-diffusion partial differential equations that allows for classical traveling wave analysis. In Chapter 4, we develop a novel Cellular Automata model that included the stochasticity inherent to the earliest stages of viral infection. Both of these models will show that the ability of IFN to locally target the cells most susceptible to infection allows it to strongly impede the spread of infection – a feature that is necessarily missed by the non-spatial ODE models. This strategy parallels the control strategies of "ring vaccination" and "contact tracing" in epidemiology and outbreak control which aim to stop spread of infection by targeting the most at-risk individuals [24, 25, 26].

## CHAPTER

# 2

# AN ODE MODEL OF VIRAL INFECTION WITH INTERFERON RESPONSE

## 2.1 Model Development

We first develop a model of viral infection with IFN signaling using ordinary differential equations (ODEs). ODE models of this type have been well established by previous work on in-vivo models of virus-immune interaction during systemic infection [18, 20]. These models consider all state variables to be functions of time  $t$  alone, and therefore neglect the spatial arrangement of individual cells, virions, and IFNs. Rather, in this model, the entire host is modeled as a single well-mixed compartment, with all interactions being governed by mass action kinetics. The equations of our model are as follows:



$$\begin{aligned}
\frac{dT}{dt} &= -\beta VT - \phi FT + \rho R \\
\frac{dI}{dt} &= \beta VT - \delta I - kI - \phi FI \\
\frac{dI^*}{dt} &= kI + \phi FI - \delta I^* \\
\frac{dR}{dt} &= \phi FT - \rho R \\
\frac{dV}{dt} &= pI + (1-f)pI^* - cV \\
\frac{dF}{dt} &= \pi p(I + I^*) - cF
\end{aligned} \tag{2.1}$$

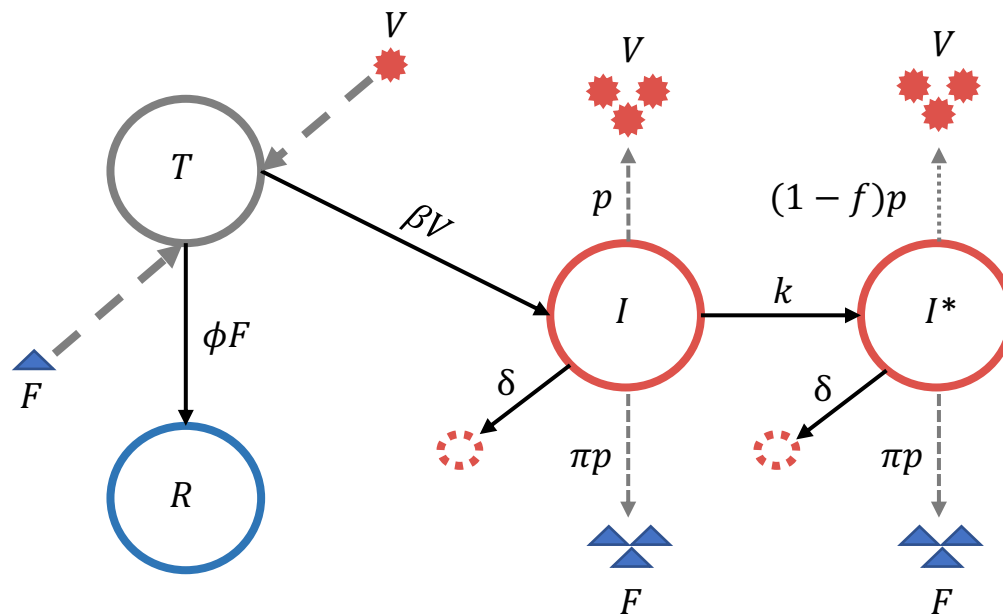
In this model (see Figure 2.1 or a schematic), cells are categorized into one of four states: uninfected target cells  $T$ , productively infected cells  $I$ , infected cells that are in an antiviral state  $I^*$  and refractory cells  $R$ . Uninfected cells are infected by virions  $V$  at rate  $\beta$  or become refractory to infection through paracrine signaling of IFN ( $F$ ) at rate  $\phi$ . Binding of IFN molecules to IFN receptors on infected cells ( $I$ ), including both autocrine and paracrine IFN signaling, may trigger an antiviral response in those cells, such that virus production is inhibited or reduced. We model the impacts of autocrine and paracrine signaling using two separate terms, i.e.  $kI$  and  $\phi FI$ . Note that  $F$  in our model represents the ambient concentration of unbound IFN (under the assumption of homogeneous concentration of IFN). We assume that autocrine signaling occurs independent of the ambient IFN concentration, because once produced from infected cells, IFNs preferentially bind to the producing cell due to proximity. The transition towards an antiviral state due to autocrine signaling is thus modeled by  $kI$ , i.e. independent of ambient IFN concentration. In contrast, the rate of transition due to the paracrine signaling is modeled to be dependent on the IFN concentration with the mass action term  $\phi FI$ .

We assume that infected cells (both  $I$  and  $I^*$ ) die at the same per capita rate  $\delta$ . Refractory cells remain protected for an average time of  $1/\rho$  before returning to the susceptible state, i.e. becoming target cells again. Infected cells  $I$  release viruses at rate  $p$ . Infected cells in an anti-viral state  $I^*$  release virions at a reduced rate of  $(1-f)p$ , where  $f$  is the fraction of reduction. For simplicity, we further assume that both  $I$  and  $I^*$  cells release IFNs at rate  $\pi p$  and that viruses and IFNs are cleared at per capita rate  $c$ . Note that since the time scale of the dynamics of IFNs is much faster than the time

scale of dynamics of the cells, we can make the quasi-equilibrium assumption for the concentration of IFN and then the level of IFNs are related to infected cells as  $F = \frac{\pi}{c} p(I + I^*)$ . Therefore, if IFN is cleared in the system at a rate different from  $c$ , the level of IFN can be compensated in the system by changing the value of  $\pi$ . In this way, when  $\beta$  and  $\phi$  are taken to be equal, the parameter  $\pi$  represents the ratio of successful IFN contact events to successful virion contact events on target cells  $T$ . The parameters used throughout are shown below in Table (2.1).

**Table 2.1** Parameters of the ODE model. Parameter  $\beta$  chosen to produce peak viremia near 2 DPI,  $\phi$  taken to match  $\beta$  so that  $\pi$  represents ratio of efficacious IFN contacts to infection events. Autocrine transition rate  $k$  taken to equal  $\delta$  so that cell has equal probability of dying or entering autocrine-signaled state.

Parameter	Description	Units	Default Value	Source
$T_0$	initial target cell population	cells	$3.5 \times 10^{11}$	[20]
$\beta$	virus contact rate	cm particles <sup>-1</sup> day <sup>-1</sup>	$5 \times 10^{-13}$	
$\phi$	IFN contact rate	cm particles <sup>-1</sup> day <sup>-1</sup>	$5 \times 10^{-13}$	
$\delta$	infected cell death rate	day <sup>-1</sup>	1	[20]
$k$	autocrine transition rate	day <sup>-1</sup>	1	
$f$	autocrine efficacy	unitless	0.9	
$p$	virion production	virion cell <sup>-1</sup> day <sup>-1</sup>	2400	[27]
$c$	virion clearance	day <sup>-1</sup>	14	[20]
$\rho$	reversion from $R$ to $T$	day <sup>-1</sup>	0.1	[20]



**Figure 2.1 Schematic of ODE model of viral infection with interferon response.** Schematic diagram with parameters in the model. Solid arrows indicate transition of cells from one state to another; dashed arrows indicate the production or binding of viruses and IFNs from cells.

## 2.2 Equilibrium Analysis of the ODE Model

The model presented in equations (2.1) is classified as a target cell limitation model. That is to say that there is no natural death or regeneration in the target cell compartment. We solve the steady state system:

$$\begin{aligned}
 -\beta VT - \phi FT + \rho R &= 0 \\
 \beta VT - \delta I - kI - \phi FI &= 0 \\
 kI + \phi FI - \delta I^* &= 0 \\
 \phi FT - \rho R &= 0 \\
 pI + (1-f)pI^* - cV &= 0 \\
 \pi p(I + I^*) - cF &= 0
 \end{aligned} \tag{2.2}$$

We find the equilibria of the model consist of the unstable manifold of “disease free” equilibria:

$$\begin{bmatrix} T, I, I^*, R, V, F \end{bmatrix}^T = \begin{bmatrix} T, 0, 0, 0, 0, 0 \end{bmatrix}^T, \quad T > 0$$

and the unique globally stable equilibrium

$$\begin{bmatrix} T, I, I^*, R, V, F \end{bmatrix}^T = \begin{bmatrix} 0, 0, 0, 0, 0, 0 \end{bmatrix}^T,$$

which is referred to as the “extinction” equilibrium. Our model is considered a *target cell limitation* model as the total number of cells inexorably decreases to 0. That is to say, the cell total  $N = T + I + I^* + R$  can only decrease due to cell death, as shown below.

$$\begin{aligned}
\frac{dN}{dt} &= \frac{d(T + I + I^* + R)}{dt} \\
&= \frac{dT}{dt} + \frac{dI}{dt} + \frac{dI^*}{dt} + \frac{dR}{dt} \\
&= -\beta VT - \phi FT + \rho R + \beta VT - \delta I - kI - \phi FI + kI + \phi FI - \delta I^* + \phi FT - \rho R \\
&= -\delta(I + I^*)
\end{aligned}$$

No “endemic” (i.e., non-trivially non-zero) equilibrium is admitted in this model since there is no proliferation of Target cells. Equations (??) show that  $\frac{dN}{dt}$  is always negative when any number of infected cells exists. In such target cell limitation models, the clearance of infection is only achieved by the extinction of the target cell compartment. These models have routinely been used to describe acute infections that do not establish persistent infection such as Influenza [18].

### 2.3 Basic Reproduction Number $R_0$

The *basic reproduction number*  $R_0$  of a disease is defined to be the expected value of the number of secondary infections produced by a single infected individual in a wholly susceptible population. For simple models with a single short-lived infectious class, a straightforward probabilistic argument can be made:

$$R_0 = [\text{\# successful contacts per unit time}] \times [\text{lifespan of infected cell}]$$

First, as a baseline, we will consider a reduced ODE model that neglects IFN signaling, which we will call the “TIV” model.

$$\begin{aligned}
\frac{dT}{dt} &= -\beta VT \\
\frac{dI}{dt} &= \beta VT - \delta I \\
\frac{dV}{dt} &= pI - cV
\end{aligned} \tag{2.3}$$

In this model, an infected cell lives for an average time of  $1/\delta$ , producing  $p$  virions per unit time.

Each virion, in turn, lives for an average time of  $1/c$ , producing  $\beta T_0$  infections per unit time. These two observations together yield the following expression for the basic reproduction number:

$$R_0 = p \cdot \frac{1}{\delta} \cdot \beta T_0 \frac{1}{c} = T_0 \frac{\beta p}{\delta c} \quad (2.4)$$

For our baseline parameter set shown in Table 2.1, the relation in 2.4 gives  $R_0 = 30$ . Note that this reproductive number would be very large in the context of the epidemiological spread of disease, but could be considered reasonable on the in-host scale in the absence of immune control. However, for models such as ours where there are several infectious classes with a variety of modalities of infection and state transition, this approach is rendered intractable, and we make use of a generalized Next Generation Matrix (NGM) approach, as outlined in [28].

### Deriving an expression for $R_0$ in ODE model

We will derive an expression for the *basic reproductive number*  $R_0$  for the ODE system below via construction of the Next Generation Matrix (NGM).

$$\frac{dT}{dt} = -\beta VT - \phi FT + \rho R \quad (2.5)$$

$$\frac{dI}{dt} = \beta VT - \delta I - kI - \phi FI \quad (2.6)$$

$$\frac{dI^*}{dt} = \phi FI + kI - \delta I^* \quad (2.7)$$

$$\frac{dR}{dt} = \beta FT - \rho R \quad (2.8)$$

$$\frac{dV}{dt} = pI + (1-f)pI^* - cV \quad (2.9)$$

$$\frac{dF}{dt} = \pi p(I + I^*) - cF \quad (2.10)$$

We consider the “infectious subsystem” consisting of consisting of equations (2.6), (2.7), and (2.9) above. We compute the Jacobian of this subsystem, evaluating it at the disease-free equilibrium  $[T, I, I^*, R, V, F]^T = [T_0, 0, 0, 0, 0, 0]^T$ , which yields the following:

$$\mathbf{J} = \begin{bmatrix} -\delta - k & 0 & \beta * T_0 \\ k & -\delta & 0 \\ p & (1-f)p & -c \end{bmatrix} \quad (2.11)$$

The Jacobian (2.11) is then partitioned as  $\mathbf{J} = \mathbf{T} + \Sigma$ , where the transmission matrix  $\mathbf{T}$  represents epidemiological birth events such as infections and virion production while  $\Sigma$  represents exponentially distributed transition events such as cell death.

$$\mathbf{T} = \begin{bmatrix} 0 & 0 & \beta T_0 \\ 0 & 0 & 0 \\ p & (1-f)p & 0 \end{bmatrix} \quad (2.12) \quad \Sigma = \begin{bmatrix} -\delta - k & 0 & 0 \\ k & -\delta & 0 \\ 0 & 0 & -c \end{bmatrix} \quad (2.13)$$

The NGM is then defined to be  $\mathbf{K} = -\mathbf{T}\Sigma^{-1}$ , and  $R_0$  is defined to be the spectral radius of the NGM. We note here that  $\Sigma$  is always non-singular, as the inevitable clearance of infected cells and virions is required for  $R_0$  to be finite. The following computations were performed with Maple.

$$\Sigma^{-1} = \begin{bmatrix} \frac{-1}{\delta+k} & 0 & 0 \\ \frac{-k}{(\delta+k)\delta} & \frac{-1}{\delta} & 0 \\ 0 & 0 & \frac{-1}{c} \end{bmatrix} \quad \text{and} \quad \mathbf{K} = -\mathbf{T}\Sigma^{-1} = \begin{bmatrix} 0 & 0 & \beta T_0/c \\ 0 & 0 & 0 \\ p/(\delta+k) + \frac{(1-f)pk}{(\delta+k)*\delta} & (1-f)p/\delta & 0 \end{bmatrix} \quad (2.14)$$

The spectral radius of  $\mathbf{K}$  as shown in (2.14) is found to be:

$$\rho(\mathbf{K}) = \sqrt{\frac{\beta p T_0 (-fk + \delta + k)}{\delta c (\delta + k)}} = \sqrt{\frac{\beta p T_0}{c \delta} \left(1 - \frac{fk}{\delta + k}\right)} \quad (2.15)$$

It is unclear if the expression for  $R_0$  given in (2.15) is properly interpretable, however, as the infectious system of equations contains state variables that are a mix of individuals (cells) and “vectors” (virions). We can produce a similar analysis for a system in which non-cell compartments are assume to immediately equilibrate.

$$\begin{aligned}
\frac{dV}{dt} &= pI + (1-f)pI^* - cV = 0 \implies V = \frac{p}{c}(I + (1-f)I^*) \\
\frac{dF}{dt} &= \pi p(I + I^*) - cF = 0 \implies F = \frac{\pi p}{c}(I + I^*)
\end{aligned} \tag{2.16}$$

By substituting the above expressions for  $V$  and  $F$  shown in (2.16) into equations (2.5) – (2.8), we obtain the *Quasi-equilibrium system*, shown below:

$$\begin{aligned}
\frac{dT}{dt} &= -\beta \frac{p}{c}(I + (1-f)I^*)T - \phi \frac{\pi p}{c}(I + I^*)T + \rho R \\
\frac{dI}{dt} &= \beta \frac{p}{c}(I + (1-f)I^*)T - \delta I - kI - \phi FI \\
\frac{dI^*}{dt} &= \phi \frac{\pi p}{c}(I + I^*)I + kI - \delta I^* \\
\frac{dR}{dt} &= \beta \frac{\pi p}{c}(I + I^*)T - \rho R
\end{aligned}$$

Reproducing the NGM analysis with the Quasi-equilibrium system, we obtain the following:

$$\begin{aligned}
\mathbf{T} &= \begin{bmatrix} \frac{\beta p T_0}{c} & \frac{\beta p(1-f)T_0}{c} \\ 0 & 0 \end{bmatrix} & \Sigma &= \begin{bmatrix} -\delta - k & 0 \\ k & -\delta \end{bmatrix} \\
\Sigma^{-1} &= \begin{bmatrix} \frac{-1}{\delta+k} & 0 \\ \frac{-k}{(\delta+k)\delta} & -1/\delta \end{bmatrix} \\
\mathbf{K}_{qe} &= \begin{bmatrix} \frac{\beta p T_0}{c(\delta+k)} + \frac{\beta p(1-f)k T_0}{c\delta(\delta+k)} & \frac{\beta p(1-f)T_0}{c\delta} \\ 0 & 0 \end{bmatrix} \\
\rho(\mathbf{K}_{qe}) &= \frac{\beta p T_0(-fk + \delta + k)}{\delta c(\delta + k)} = \frac{\beta p T_0}{c\delta} \left(1 - \frac{fk}{\delta + k}\right)
\end{aligned} \tag{2.17}$$

We note that  $\rho^2(\mathbf{K}) = \rho(\mathbf{K}_{qe})$ . Since the threshold for outbreak is  $R_0 > 1$ , this discrepancy does not affect the threshold for establishment of systemic infection. Furthermore, we note that the value



of  $R_0$  obtained from the quasi-equilibrium system in (2.17) is consistent with the probabilistic interpretation of the terms of the original model, as outlined below:

$$\begin{aligned}
R_0 &= \left[ \begin{array}{c} \# \text{ successful contacts} \\ \text{per time in class } I \end{array} \right] \times \left[ \begin{array}{c} \text{avg. time spent} \\ \text{in class } I \end{array} \right] \\
&\quad + \left[ \begin{array}{c} \# \text{ successful contacts} \\ \text{per time in class } I^* \end{array} \right] \times \left[ \begin{array}{c} \text{avg. time spent} \\ \text{in class } I^* \end{array} \right] \times \left[ \begin{array}{c} \text{prob. of entering} \\ I^* \text{ before cell death} \end{array} \right] \\
&= \frac{\beta p T_0}{c} \cdot \frac{1}{\delta + k} + (1-f) \frac{\beta p T_0}{c} \cdot \frac{1}{\delta} \cdot \frac{k}{\delta + k} \\
&= \frac{\beta p T_0}{c \delta} \cdot \left( \frac{\delta}{\delta + k} + \frac{(1-f)k}{\delta + k} \right) \\
&= \frac{\beta p T_0}{c \delta} \cdot \left( 1 - \frac{fk}{\delta + k} \right)
\end{aligned}$$

We can see now that the “baseline” value of  $R_0$  computed for the TIV model is exactly the value of  $R_0$  attained in the full model when IFN is switched off (i.e.,  $\delta = k = \pi = 0$ ,  $f = 1$ ). Furthermore, we note that the parameters governing the paracrine “free IFN” response do not appear at all in the expression for  $R_0$ , suggesting that in the spatially homogenous (i.e., well-mixed) setting, autocrine response is significantly more impactful. This is intuitive, as in a biologically realistic setting, paracrine signaling acts locally to decrease susceptibility in a targeted region around the infection, so in a non-spatial setting this effect is greatly diluted. We discuss this in greater depth later in section 2.6.

## 2.4 Approximating Peak Viral Load

We will approximate the peak viral load in consideration of an ODE model that considers only the protective effect of IFN signaling on susceptible cells and not the dampening of virion output from infected cells (i.e., no  $I^*$  class). Since paracrine IFN signaling is shown to be less influential than autocrine signaling in the ODE context, we consider this model as a “best-case scenario” for the efficacy of paracrine IFN signaling in a spatially-homogeneous setting. This reduces the model (2.1) to the following system of ODEs:

$$\begin{aligned}
\frac{dT}{dt} &= -\beta VT - \phi FT + \rho R \\
\frac{dI}{dt} &= \beta VT - \delta I \\
\frac{dR}{dt} &= \phi FT - \rho R \\
\frac{dV}{dt} &= pI - cV \\
\frac{dF}{dt} &= \pi pI - cF
\end{aligned} \tag{2.18}$$

We will once again assume that during initial infection the compartments  $V$  and  $F$  are near equilibrium, so we will take the following approximation in our analysis:

$$\begin{aligned}
0 &= pI - cV \implies V = \frac{p}{c}I \\
0 &= \pi pI - cF \implies F = \pi \frac{p}{c}I
\end{aligned} \tag{2.19}$$

We now consider an approximate model for the cell compartments alone that employs the rapid equilibration of the  $V$  and  $F$  compartments. We furthermore assume that in the initial stage of infection the target cell population remains close to the initial value  $T_0$  and that the transition from the refractory class back to the target population is negligible during initial infection (i.e.,  $\rho = 0$ ). Furthermore, we take the contact rates for both virions and IFNs to be  $\beta$ . Since there is mutual unidentifiability between  $p$  and  $\beta$  and between  $\pi p$  and  $\phi$ , we can let  $\beta = \phi$  as the parameter  $\pi$  can implicitly account for the difference in these contact rates as might be present in biology. Together these assumptions yield the following pair of ODEs:

$$\begin{aligned}
\frac{dT}{dt} &= -\beta(V + F)T = -\beta(1 + \pi)\frac{p}{c}IT \\
\frac{dI}{dt} &= \beta\frac{p}{c}IT - \delta I
\end{aligned}$$

We will now approximate the peak infected cell population. At peak infection, we have

$$\frac{dI}{dt} = 0 \implies T_{peak} = \frac{T_0}{R_0} \quad (2.20)$$

where  $R_0 = \frac{\beta p}{\delta c} T_0$ . We consider now a phase plane solution to the following:

$$\begin{aligned} \frac{dI}{dT} &= \frac{\beta \frac{p}{c} IT - \delta I}{-(1+\pi)\beta \frac{p}{c} IT} = -\frac{1}{1+\pi} \left(1 - \frac{T_0}{R_0} \frac{1}{T}\right) \\ \implies I(T) &= -\frac{1}{1+\pi} \left(T - \frac{T_0}{R_0} \ln(T) + C\right) \end{aligned} \quad (2.21)$$

Under the assumption that there is no initial infected cell population,

$$\begin{aligned} I(0) = 0 &= -\frac{1}{1+\pi} \left(T_0 - \frac{T_0}{R_0} \ln(T_0) + C\right) \implies C = -T_0 + \frac{T_0}{R_0} \ln(T_0) \\ I(T) &= \frac{1}{1+\pi} \left(T_0 - T + \frac{T_0}{R_0} \ln(T/T_0)\right) \end{aligned}$$

Since we now have  $I$  expressed explicitly as a function of  $T$ , we can substitute in the value of  $T_{peak}$  calculated in (2.20):

$$I_{peak} = I(T_{peak}) = \frac{1}{1+\pi} T_0 \left(1 - \frac{1 + \ln R_0}{R_0}\right)$$

We lastly return to our assumption that under rapid equilibration  $V = \frac{p}{c} I$ , so our expression for the peak viral load is

$$\begin{aligned} V_{peak} &= \frac{1}{1+\pi} \frac{p}{c} T_0 \left(1 - \frac{1 + \ln R_0}{R_0}\right) \\ &= \frac{1}{1+\pi} \frac{\delta}{\beta} \left(R_0 - 1 - \ln R_0\right) \end{aligned}$$

Since  $\pi = 0$  corresponds to the case of no paracrine IFN signaling, we see that paracrine IFN signaling has the effect of reducing peak viremia by a factor of  $f = \frac{1}{1+\pi}$  compared to the case of no paracrine IFN.

## 2.5 Approximating Time to Peak Viremia in ODE Model

We will find an approximate expression for the time to peak viremia. We again make the quasi-steady state assumption in  $V$  and  $F$ , take  $f = 1$  so that autocrine signaling completely shuts down virion output, meaning that we will work with the two-compartment model shown in (??). We see that peak infection ( $dI/dt = 0$ ) occurs at such a time that  $T(t_{peak}) = T_0/R_0$ , as shown in (2.20), where  $R_0 = \frac{\beta p}{c\delta} T_0$ . Approximating growth to be exponential until that time, we have:

$$I(t) = I_0 e^{(R_0-1)\delta t} \text{ while } t < t_{peak} \quad (2.22)$$

Substituting (2.22) into the equation for  $T$ , we can easily solve the single ODE, giving the following:

$$T(t) = e^{\left[ \frac{R_0}{R_0-1} \frac{I_0}{T_0} (1+\pi) (e^{(R_0-1)\delta t} - 1) \right]}.$$

Taking  $T(t_{peak}) = T_0/R_0$  in the above approximate solution for  $T(t)$ , we can easily solve for  $t_{peak}$  and find the following approximate expression.

$$t_{peak} = \frac{1}{(R_0-1)\delta} \ln \left( 1 + \frac{T_0}{I_0} \frac{R_0-1}{R_0} (1+\pi) \ln \frac{T_0}{R_0} \right) \quad (2.23)$$

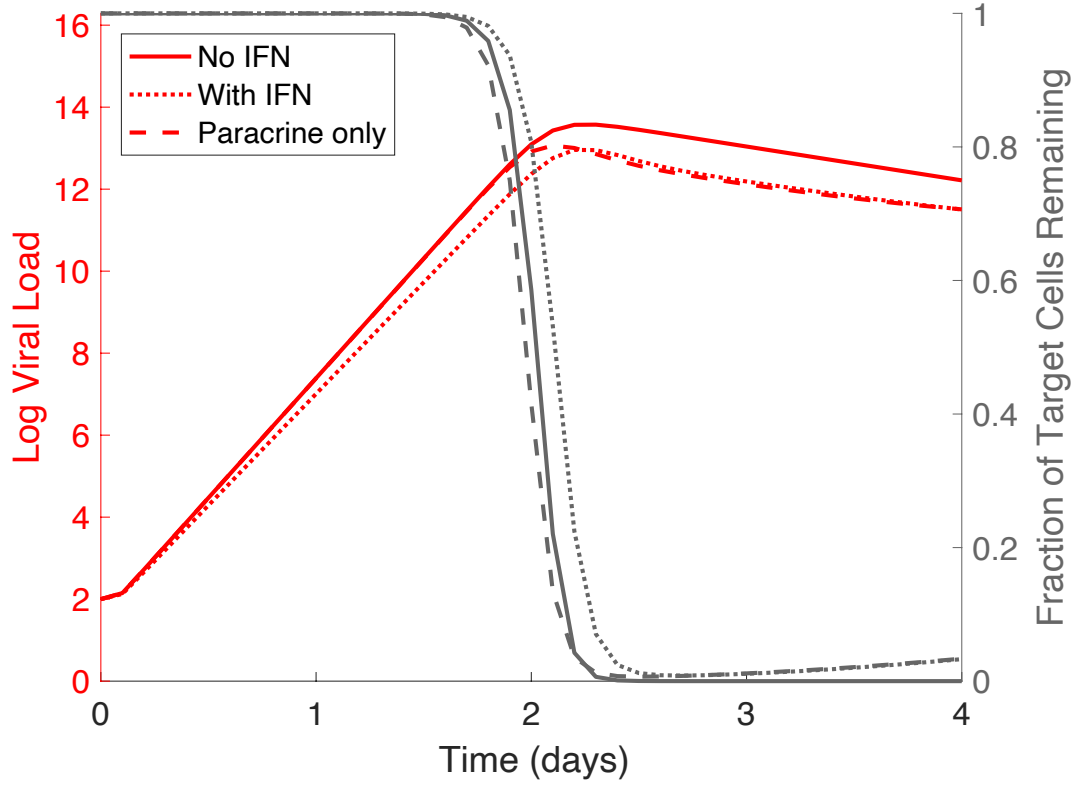
## 2.6 Results

We first constructed a model (see Figure 2.1 for a schematic) and analyzed the roles of autocrine and paracrine IFN signaling using ordinary differential equations (ODEs). To understand the impacts of autocrine and paracrine signaling on the virus dynamics after initial viral exposure, we calculated the Basic Reproductive Number  $R_0$  of the virus using the Next Generation Matrix technique [28]. Note that  $R_0 = 1$  is the threshold for establishment of infection, and viral population only grows when  $R_0 > 1$ . Thus, for an effective innate immune response to halt viral infection,  $R_0$  has to be less than 1. For the above model we find:

$$R_0 = T_0 \frac{\beta p}{c \delta} \cdot \left( 1 - \frac{f k}{\delta + k} \right)$$

This expression shows that the reduction of  $R_0$  due to autocrine signaling is  $\frac{f k}{\delta + k}$ , where  $f$  is the inhibition of virus production due to the cellular antiviral response and  $k/(\delta + k)$  is the probability that an infected cell becomes antiviral by the autocrine pathway before cell death occurs (see Table 2.1 for the parameter ranges considered for this model). For our baseline parameter set shown in table 2.1, this gives  $R_0 = 3$ . This is a reasonable value for in-host spread of disease as suggested in [29]. *In vivo* experiments suggest that the fraction of infected cells that successfully enter an antiviral state is in general low [30, 31, 32], i.e.  $k/(\delta + k)$  is much less than 1. If this observation is consistent with IFN response in vitro, then our results suggest that autocrine signaling alone has limited impact on arresting viral infection.

Importantly, we found that the parameters governing paracrine IFN signaling (i.e.  $\phi, \pi$ ) do not appear in the expression for  $R_0$ , i.e. paracrine signaling alone does not change the infection threshold. Therefore, the ODE model makes the surprising prediction that when cells, viruses and IFN are well mixed (as assumed in our ODE model and other models [20, 18], paracrine signaling has a negligible role in halting infection during early infection when the number of target cells are abundant). We further performed simulations of the model to compare the viral dynamics with and without paracrine IFN signaling (Figure 2.2). In agreement with the analytical derivation for  $R_0$ , we found that IFN paracrine signaling has negligible impact on the viral load during initial exponential growth period. This is true even for very large (biologically unrealistic) values of  $\pi$  (Figure 2.3). We found that IFN-mediated protection of target cells is only able to affect the course of infection after some period of viral growth once infected cell concentration, and thus IFN concentration, rises to a sufficiently high level that there is a notable impact on protecting target cells and infected cells. The peak viral load is decreased by approximately  $1/(1 + \pi)$ -fold and the time to peak viremia is relatively insensitive to changes in  $\pi$ . This nominal decrease in the time to peak viral load is a consequence of the accelerated target cell depletion due to IFN signaling to uninfected cells.

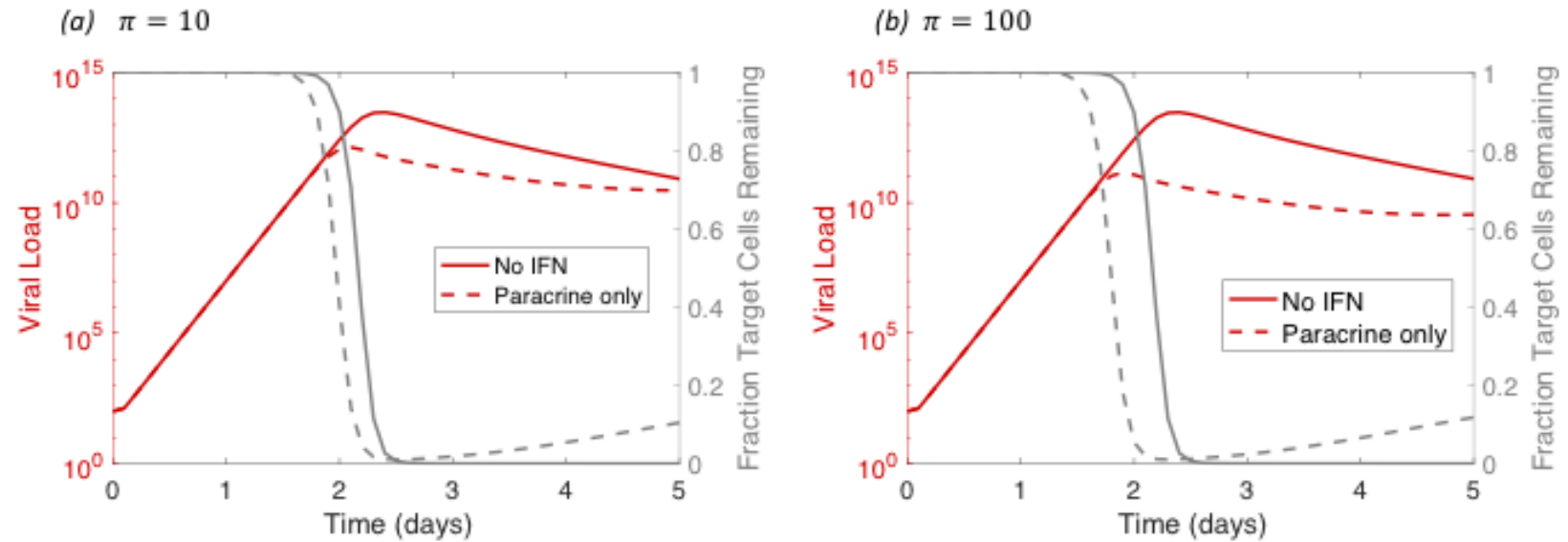


**Figure 2.2 Representative simulations of the ODE model.** Shown are simulations of the ODE model with no IFN response (solid line), paracrine IFN signaling only (dashed line), and full IFN response (dotted line). The simulations show that paracrine signaling alone does not alter the initial exponential growth rate of the viral load, and can only moderately advance the time to peak viral load by accelerating the depletion of target cells. However, autocrine signaling is able to decrease the exponential growth rate, observable as the slope of the viral load curve prior to the peak at day 2.

## 2.7 Contributions

In this Chapter we developed an ODE model of well-mixed viral infection with IFN signaling based on previous work [20]. Our novel contribution is the distinction between the paracrine (cell-to-cell) and autocrine (same-cell) signaling mechanisms. Whereas paracrine signaling is a density-dependent effect that requires the accumulation of extracellular IFN, autocrine signaling can be seen as density-independent and IFNs are highly likely to bind to receptor sites on the progenitor cell. Our results indicate that in the absence of spatial structure, where both virions and unbound IFN accumulate in the host environment homogeneously, autocrine signaling is considerably more

impactful than paracrine signaling. Whereas paracrine signaling is untargeted in a well-mixed infection, homogeneously decreasing the susceptibility of host cells, the autocrine signaling received by infected cells is inherently targeted and has the effect of reducing the viral load, thereby slowing the growth of the infection or even preventing it by reducing  $R_0$  below the threshold value of 1.



**Figure 2.3 Representative ODE model simulations with supraphysiological values of  $\pi$  show little impact of paracrine signaling.** These two simulations of the ODE model demonstrate the conclusion that paracrine signaling alone is unable to meaningfully impact the establishment of infection in the spatially homogeneous setting. Since the paracrine signaling acts by depleting the population of target cells, this impact is not sufficiently large during the initial exponential growth phase of the infection.



## CHAPTER

# 3

# PDE MODEL OF VIRAL INFECTION WITH IFN RESPONSE

### 3.1 Introduction

The analysis of the ODE model of viral infection with IFN signaling in developed in Chapter 2 revealed that, in a spatially homogeneous setting, the paracrine (cell-to-cell) signaling of interferon has little to no impact on the early stages of infection. In the ODE model, individual virions and IFNs, once produced, have an equal probability of contacting any cell in the entire host compartment. This means the protective effect of paracrine IFN signaling is diluted across the entire host compartment and that a virion's likelihood of contacting susceptible target cells is essentially unimpeded during the initial exponential growth phase. However, both *in vitro* and *in vivo* experimentation has revealed that IFN signaling is an extremely local process. In [33], *in vitro* infection of human tracheobrocheal epithelial monolayers with an H1N1 strain of Influenza A Virus (IAV) was performed at a very low multiplicity of infection. Immunostain was applied after several rounds of replication and the cells

that were found to contain viral RNA were concentrated in tight clusters. This suggests that infection tends to spread from infected cells only to a small number of neighboring epithelial cells. In [34], Immunostaining of flash-frozen liver biopsies from hepatitis C virus (HCV) positive individuals found that interferon stimulated gene (ISG) expression and HCV viral RNA were highly co-localized. That is to say, individual cells were the most likely to demonstrate ISG expression if the neighboring cells were infected or if they were infected themselves. Together, these works suggest that the ability of IFN to spread and act locally to the site of infection is an important aspect of the efficacy of cell-to-cell IFN signaling. That is to say, local IFN signaling to cells that are near to the site of infection (and therefore the most likely to be contacted by a virion) could greatly increase the impact of IFN signaling as compared to the ODE model.

## 3.2 Model Development

We next develop a partial differential equation (PDE) model of viral infection and IFN response. This model explicitly considers the spatial arrangement of cells, virions, and IFNs, thus more accurately representing the dynamics of infection in an epithelial tissue. We assume that susceptible cells  $T$  are arranged on a 1-dimensional lattice with spatial variable  $x \in [0, L]$  with a uniform initial density  $T_0$ . Viruses and IFNs can diffuse to nearby locations, in contrast to the ODE model where viruses and IFNs are assumed to instantaneously be evenly distributed once produced. Virions and IFNs diffuse across the spatial domain with diffusion coefficients  $D_V$  and  $D_F$ , where we take  $D_F \gg D_V$  since IFNs are much smaller than virions and therefore diffuse at a much greater rate [27, 35]. These diffusion parameters determine the characteristic length scales on which IFNs and virions will be active [36]. Furthermore, we assume that  $\rho = 0$  due to an emphasis on modeling the initial stages of the infection process where anti-viral cells are unlikely to return to the susceptible state. Otherwise, the dynamics from the non-spatial ODE setting remain the same. The initial conditions are taken to be such that the domain is populated only with Target cells at a constant density and a single infected cell at the position  $x = 0$ , which is achieved using a Dirac delta distribution  $\delta_0(x)$ . The boundary conditions are taken to be homogeneous Neumann at  $x = 0$  to represent reflective symmetry of the spread of infection, and homogeneous Dirichlet at far-field  $x = L$ . The parameters of the model can

be found in Table , and the equations of the model are as follows:

$$\begin{aligned}
\frac{\partial T}{\partial t} &= -\beta VT - \phi FT + \rho R \\
\frac{\partial I}{\partial t} &= \beta VT - \delta I - kI - \phi FI \\
\frac{\partial I^*}{\partial t} &= kI + \phi FI - \delta I^* \\
\frac{\partial R}{\partial t} &= \phi FT - \rho R \\
\frac{\partial V}{\partial t} &= pI + (1-f)pI^* - cV + D_V \frac{\partial^2 V}{\partial x^2} \\
\frac{\partial F}{\partial t} &= \pi p(I + I^*) - cF + D_F \frac{\partial^2 F}{\partial x^2}
\end{aligned} \tag{3.1}$$

Boundary Conditions:  $\frac{\partial F}{\partial x} = \frac{\partial V}{\partial x} = 0$  at  $x = 0$ ,  $F = V = 0$  at  $x = L$

Initial Conditions:  $T(x, 0) = T_0$ ,  $I(x, 0) = \delta_0(x)$

**Table 3.1** Parameters of the PDE model. Parameters  $\beta$  and  $\phi$  taken to produce reasonable wave speed given remaining parameter set. Diffusion coefficient  $D_V$  taken to be 1 as a baseline,  $D_F$  taken to be between 1 and 2 orders of magnitude greater than  $D_V$  [31].

Parameter	Description	Units	Default Value	Source
$T_0$	initial target cell density	cells $\text{cm}^{-1}$	$6.6 \times 10^5$	[37]
$\beta$	virus contact rate	$\text{cm particles}^{-1} \text{ day}^{-1}$	$8.3 \times 10^{-6}$	
$\phi$	IFN contact rate	$\text{cm particles}^{-1} \text{ day}^{-1}$	$8.3 \times 10^{-6}$	
$\delta$	infected cell death rate	$\text{day}^{-1}$	1	[20]
$k$	autocrine transition rate	$\text{day}^{-1}$	1	
$f$	autocrine efficacy	unitless	0.9	
$p$	virion production	$\text{virion cell}^{-1} \text{ day}^{-1}$	2400	[27]
$c$	virion clearance	$\text{day}^{-1}$	14	[20]
$D_V$	virion diffusion	$\text{cm}^2 \text{ day}^{-1}$	1	
$D_F$	IFN diffusion	$\text{cm}^2 \text{ day}^{-1}$	40	

### 3.3 Reaction Diffusion Systems and Traveling Wave Solutions

Our spatial model (3.1) can be analyzed as a *reaction diffusion system*, a system of PDEs containing only diffusion terms and reaction terms. In general, a reaction diffusion system has the form

$$\mathbf{u}_t = \nabla^2 \mathbf{u} + \mathbf{f}(\mathbf{u})$$

where the vector-valued state variable  $\mathbf{u}$  is a function of a temporal variable  $t$  and spatial variables  $\mathbf{x}$ ,  $\Delta$  represents the Laplacian operator in the spatial variables, and  $\mathbf{f}$  is a non-linear function of  $\mathbf{u}$  referred to as the “reaction term”. Reaction diffusion equations are widely used to model biological phenomena across many spatial scales. Reaction diffusion systems may exhibit traveling wave solutions, where a waveform translates through the spatial domain with time at a constant speed. For a system with a single spatial variable  $x$ , a traveling wave solution takes the form  $u(x, t) = U(x - v t)$ , where the single variable function  $U$  describes the traveling front and  $v$  is the traveling wave speed.

### 3.3.1 A Classical Example: The Fisher Equation

The Fisher Equation is the most classical of all reaction diffusion PDEs that exhibit traveling wave solutions. It is achieved by taking the non-linear reaction term to be a logistic growth term,  $f(u) = r u(1 - u)$ . Fisher first proposed and studied this equation in 1937 to study the propagation of an advantageous gene through a spatially structured population [38]. The equation takes the form

$$u_t = r u(1 - u) + D u_{xx}$$

where  $u$  represents the allele frequency,  $r$  represents the allele’s reproductive rate, and  $D$  is the diffusion coefficient of the allele. The PDE admits two spatially homogeneous equilibrium solutions: the unstable “extinction solution”  $u = 0$  and the stable “saturation solution”  $u = 1$ . We will look for the admissible traveling wave solutions and their corresponding wave speeds by substituting  $u(x, t) = U(x - v t)$ . This yields the 2<sup>nd</sup> order traveling wave ODE:

$$-v U' = r U(1 - U) + D U''.$$

The ODE can be solved compute the form of the traveling front  $U(\xi)$  for any given wavespeed  $v$ , where  $\xi = x - v t$ . However, we are particularly interested in wave solutions that preserve the interpretation of  $u$  as allelic frequency, meaning that we require that  $0 \leq u \leq 1$ . Note ,however, that integration forward in  $\xi$  is backwards in  $t$  (assuming  $v > 0$ ), so in the traveling wave frame the  $U = 0$  equilibrium is an attractor. This means we will require that the  $U = 0$  equilibrium be non-oscillatory.

We perform an equilibrium analysis of the traveling wave ODE as follows. Letting  $W = U'$ , we have

$$\begin{bmatrix} U \\ W \end{bmatrix}' = \begin{bmatrix} W \\ -\frac{r}{D}U(1-U) - \frac{\nu}{D}W \end{bmatrix}$$

We linearize the system about the extinction equilibrium  $U = W = 0$ , and find the Jacobian of the system to be the following:

$$\mathbf{J} = \begin{bmatrix} 0 & 1 \\ -\frac{r}{D} & -\frac{\nu}{D} \end{bmatrix}$$

Since we are interested in values of the wave speed  $\nu$  that admit non-oscillatory solutions, we argue that the eigenvalues of  $\mathbf{J}$  must be real valued. The characteristic polynomial of  $\mathbf{J}$  is found below.

$$p(\lambda; \nu) = |\lambda I - \mathbf{J}| = \begin{vmatrix} \lambda & -1 \\ \frac{r}{D} & \lambda + \frac{\nu}{D} \end{vmatrix} = \lambda^2 + \frac{\nu}{D}\lambda + \frac{r}{D}$$

We can make the following argument real-valued eigenvalues in terms of the discriminant.

$$\text{Disc}_\lambda(p(\lambda; \nu)) = \frac{\nu^2}{D^2} - 4\frac{r}{D} \geq 0 \implies \nu \geq 2\sqrt{rD}$$

We term the minimum admissible wavespeed  $\nu^* = 2\sqrt{rD}$  to be the *critical wavespeed*. Furthermore, it is able to be shown that for initial data with compact support the asymptotic speed of the traveling wave solution is precisely the critical wave speed  $\nu^*$ .

### 3.3.2 Traveling Wave Analysis of our PDE Model

We will perform a traveling wave analysis of the PDE model. Since we are interested in quantifying the impact of the IFN response (both via paracrine and autocrine signaling) on the ability of infection to spread, we will perform our analysis on a reduced PDE model without IFN response, shown

below.

$$\begin{aligned}
\frac{\partial T}{\partial t} &= -\beta V T \\
\frac{\partial I}{\partial t} &= \beta V T - \delta I \\
\frac{\partial V}{\partial t} &= p I - c V + D_v \frac{\partial^2 V}{\partial x^2}
\end{aligned} \tag{3.2}$$

We term this reduced model the TIV model, as it includes only target cells, infected cells, and virions. We will consider a traveling wave solution of the TIV system. For each compartment  $u_i(x, t)$  there exists a function  $U_i(z)$  where the solution can be expressed as  $u_i(x, t) = U_i(x + vt)$ . Now for each compartment  $u_i$  we make the following substitutions:

$$\frac{\partial u_i}{\partial t} = v U_i'(z), \quad \frac{\partial^2 u_i}{\partial x^2} = U_i''(z)$$

Making the above substitutions and re-writing the second-order equation in  $V$  as two first order equations, we arrive at the following 4 dimensional system system of first order equations:

$$\begin{aligned}
T' &= -\frac{\beta}{v} V T \\
I' &= \frac{\beta}{v} V T - \frac{\delta}{v} I \\
V' &= W \\
W' &= -\frac{p}{D} I + \frac{c}{D} V + \frac{v}{D} W
\end{aligned}$$

This system has disease-free steady state  $u_e = [T_0, 0, 0, 0]$ . Evaluating the Jacobian of the system at  $u_e$ , we find:

$$\mathcal{J} = \begin{bmatrix} 0 & 0 & -\frac{\beta}{v} T_0 & 0 \\ 0 & -\frac{\delta}{v} & \frac{\beta}{v} T_0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -p/D & c/D & v/D \end{bmatrix} \tag{3.3}$$

To exclude biologically irrelevant solutions to the system, we will enforce that the eigenvalues of the Jacobian be non-complex to avoid oscillation of the populations about zero. That is to say, the minimum admissible wave speed will occur for the critical value  $v^*$  at which the the characteristic

polynomial of the Jacobian takes a double real root. This of course is the threshold value before which the polynomial takes complex conjugate-pair roots. The block diagonal structure of the Jacobian reveals there to be a zero eigenvalue, so we can examine instead the characteristic polynomial of the full-rank  $3 \times 3$  block.

$$p(\lambda; v) = \det \left( \lambda I - \begin{bmatrix} -\frac{\delta}{v} & \frac{\beta}{v} T_0 & 0 \\ 0 & 0 & 1 \\ -p/D & c/D & v/D \end{bmatrix} \right) = D v \lambda^3 + (D \delta - v^2) \lambda^2 - (c + \delta) v \lambda + c \delta (R_0 - 1)$$

We note the following when  $R_0 > 1$ :

$$\begin{aligned} \lim_{\lambda \rightarrow \pm\infty} p(\lambda; v) &= \pm\infty \text{ if } v > 0, \quad \lim_{\lambda \rightarrow \pm\infty} p(\lambda; v) = \mp\infty \text{ if } v < 0 \\ p(0; v) &= c \delta (R_0 - 1) > 0 \\ p'(0; v) &= -(c + \delta) v \end{aligned}$$

These three observations together ensure that  $p(\lambda)$  always has one negative real root and attains an extremum for some  $\lambda > 0$ . We will now construct  $v^*$  such that there exists a positive real double root on  $\mathbb{R}^+$  (see Figure 3.1). For a positive double root to exist it must also be an extremum of the polynomial. This means we solve the following system of equations for  $\lambda$  and  $v$ :

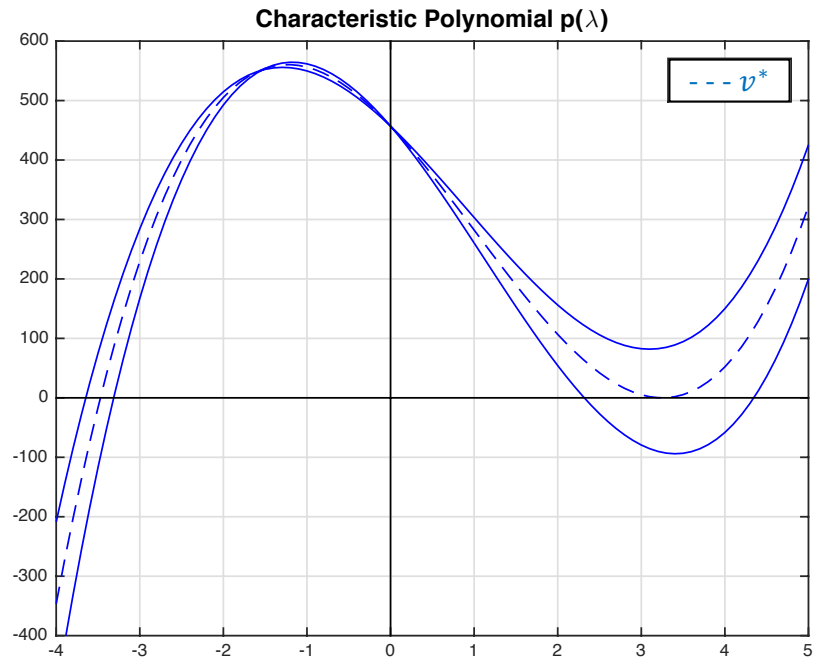
$$\begin{cases} p(\lambda; v) = 0 & \lambda > 0 \\ \frac{d}{d\lambda} p(\lambda; v) = 0 & \lambda > 0 \end{cases}$$

For a given parameter set, this system of two equations can be done numerically to find a prediction of the traveling wave speed  $v^*$ .

### 3.3.3 Discriminant Method for Determining Exact Expression for Wave Speed

With any polynomial  $p(\lambda) = W \lambda^3 + X \lambda^2 + Y \lambda + Z$  of degree 3 we can associate a discriminant  $\text{Disc}_\lambda p(\lambda) = X^2 Y^2 - 4 W Y^3 - 4 X^3 Z - 27 W^2 Z^2 + 18 W X Y Z$ .

The minimum admissible traveling wave speed of the system (3.2) is found to be the critical value  $v^*$  for which the characteristic polynomial of the Jacobian takes a double root, which is precisely



**Figure 3.1 Characteristic polynomial  $p(\lambda; \nu)$ .** The minimum admissible wavespeed  $\nu^*$  will be the smallest such that admits non-oscillatory solutions, meaning that we must enforce that  $p(\lambda; \nu)$  have only real roots. Therefore, the critical wavespeed  $\nu^*$  will be the small that admits real roots, meaning that we look specifically for the value of  $\nu$  that produces a double root. Shown are three cases: (1) a subcritical case, where only one real root exists, (2) a supercritical case, with three distinct real roots, and (3) the critical case  $\nu^*$  where a double root is attained on  $\mathbb{R}^+$ .



where  $\text{Disc}_\lambda p(\lambda; \nu^*) = 0$ . In terms of the discriminant, we frame the problem as:

$$\begin{aligned} \text{Disc}_\lambda p(\lambda; \nu) = & \nu^2(D\delta - \nu^2)(\delta + c)^2 + 4D\nu^4(\delta + c)^3 + 4(\nu^2 - D\delta)^3 c\delta(R_0 - 1)... \\ & ... - 27D^2\nu^2[c\delta(R_0 - 1)]^2 + 18D\nu^2(\nu^2 - D\delta)(c + \delta)[c\delta(R_0 - 1)] = 0 \end{aligned} \quad (3.4)$$

For the simpler problem where there is no death of infected cells and thus the total number of cells is conserved we can further reduce the dimensionality of the problem:

$$\begin{aligned} \frac{\partial I}{\partial t} &= \beta V(T_0 - I) \\ \frac{\partial V}{\partial t} &= pI - cV + D \frac{\partial^2 V}{\partial x^2} \end{aligned}$$

We find the corresponding 1<sup>st</sup> order traveling-wave ODE to have jacobian about the zero equilibrium.

$$\mathcal{J} = \begin{bmatrix} 0 & \beta T_0/\nu & 0 \\ 0 & 0 & 1 \\ -p/D & c/D & \nu/D \end{bmatrix}$$

We will consider the discriminant of the characteristic polynomial of the above jacobian matrix to be a function of the wavespeed  $\nu$  and collect terms accordingly:

$$\begin{aligned} p(\lambda; \nu) &= D\nu\lambda^3 - \nu^2\lambda^2 - q\nu\lambda + pT_0\beta \\ \implies \text{Disc}_\lambda(p(\lambda; \nu)) &= (4T_0\beta p + c^2)\nu^6 + (18T_0\beta cDp + 4c^3D)\nu^4 - (27T_0^2\beta^2D^2p^2)\nu^2 \end{aligned}$$

We see that the discriminant here is a sixth order polynomial in  $\nu$  with no odd order terms and no constant term. Since the trivial wavespeed  $\nu = 0$  is not of interest, we can neglect the double root of  $\nu = 0$  and look at the following:

$$\tilde{\Delta}_3(\nu) = (4T_0\beta p + c^2)\nu^4 + (18T_0\beta cDp + 4c^3D)\nu^2 - (27T_0^2\beta^2D^2p^2) := A\nu^4 + B\nu^2 + C$$

The minimal wavespeed  $\nu^*$  will be the largest value of  $\nu$  for which  $\tilde{\Delta}_3(\nu) = 0$ . Since the above is quadratic in  $\nu^2$ , it is easily solved by the quadratic formula. This furthermore implies that each

positive wave speed admitted also admits an equal and opposite wave speed, as the system is non-advective and thus has no directional bias. Letting  $R = T_0 \beta p$ , we have the following expression for the wave speed:

$$(v^*)^2 = \frac{-B + \sqrt{B^2 - 4AC}}{2A} = D_v \frac{-9Rc - 2c^3 + \sqrt{4c^6 + 36Rc^4 + 108R^2c^2 + 54R^3}}{R + c^2}$$

Recalling the parameter values used from Table 3.1, we can choose smaller order terms to neglect to arrive at a simpler expression that is sufficiently accurate for a neighborhood of our parameter set. Here, observing  $R \approx 10^4$  and  $c \approx 2 \times 10^1$ , we arrive at the following reduced expression:

$$v^* \approx \sqrt{\frac{D_v}{4} (\sqrt{108R} - 9c)}$$

This value is shown to be consistent with traveling speeds measured from numerical simulations (*relative error*  $\approx 5\%$ ). Since this method appears effective in producing an accurate closed form expression for the wave speed, we wish to recycle this logic as much as possible. For the more general case where  $\delta \neq 0$ , we would like to make small simplifications that will eliminate odd order terms and allow use of the quadratic formula. Specifically we will make the assumption that  $v^2 \gg D\delta$  and will substitute  $X = (D\delta - v^2) \rightarrow -v^2$ . Under this substitution, the non-zero positive root of the discriminant is approximated by quadratic formula and found be consistent with numerical simulation (*relative error*  $\approx 2 - 5\%$ ). The closed form is very complicated and is not included here.

### 3.4 Results

For almost all respiratory and enteric viral infections, the site of initial infection and viral replication is epithelial tissue, which is characterized by a monolayer structure [21]. Due to local diffusion of viral progeny over the epithelium, a virion is highly likely to infect one of a small number of neighboring cells rather than having an equal probability of infecting any target cell, as is the implicit assumption in an ODE model of viral infection. For this reason, we are primarily interested in comparing the behavior of our PDE model in the presence and absence of IFN response. Furthermore, we will acutely interested in the relative efficacy of paracrine signaling and autocrine signaling alone in stopping the spatial spread of viral infection.

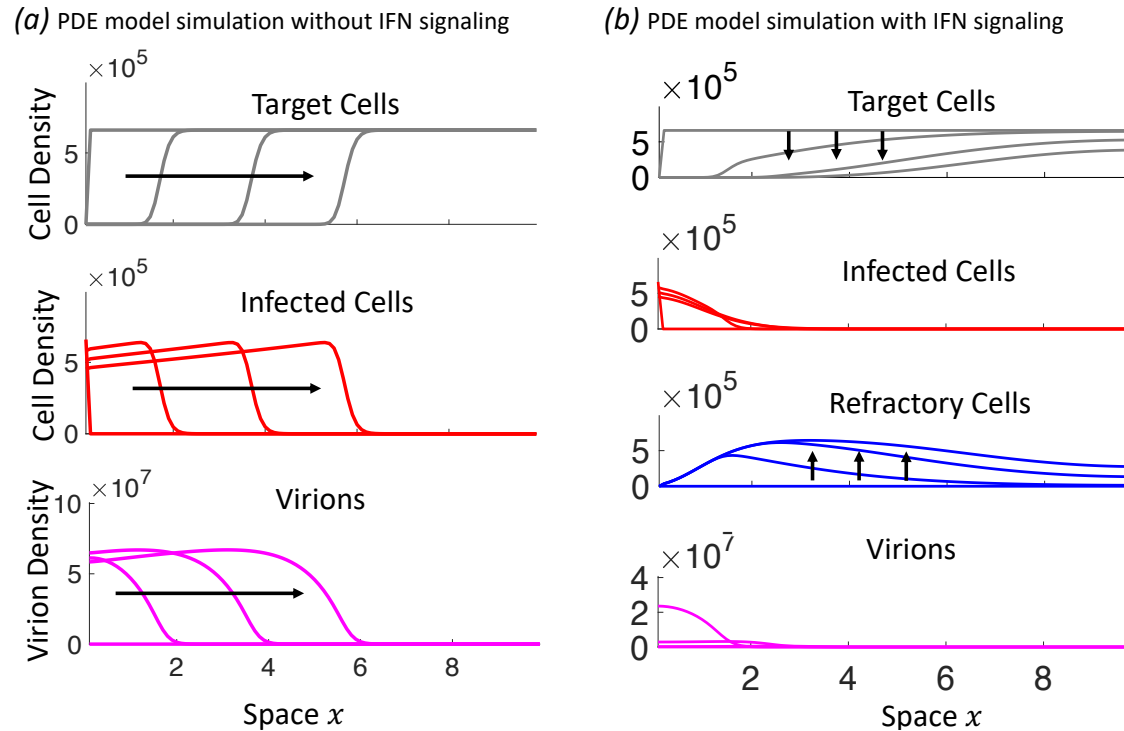
We begin by comparing simulations of the no-IFN TIV model and full PDE model, shown in Figure 3.2. All simulations are performed using finite difference method with Runge-Kutta numerical integration in time. The system of ODEs generated by the finite difference discretization is stiff, so a Runge-Kutta method designed for stiff equations was used. In the absence of IFN signaling ( $\pi = k = f = 0$ ), the solution of the PDE model exhibits a traveling wave solution, as revealed by the traveling wave analysis in the previous section. That analysis showed that, given that the initial infected region was compactly supported, a front of infection propagates through healthy epithelium with a constant velocity,  $v^*$ . Though a closed form expression is possible, it is very large and unwieldy. An approximate expression for  $v^*$  that ignores negligible terms given our parameter set is as follows:

$$v^* \approx \sqrt{\frac{D_V}{4} (\sqrt{108 T_0 \beta p} - 9c)} \quad (3.5)$$

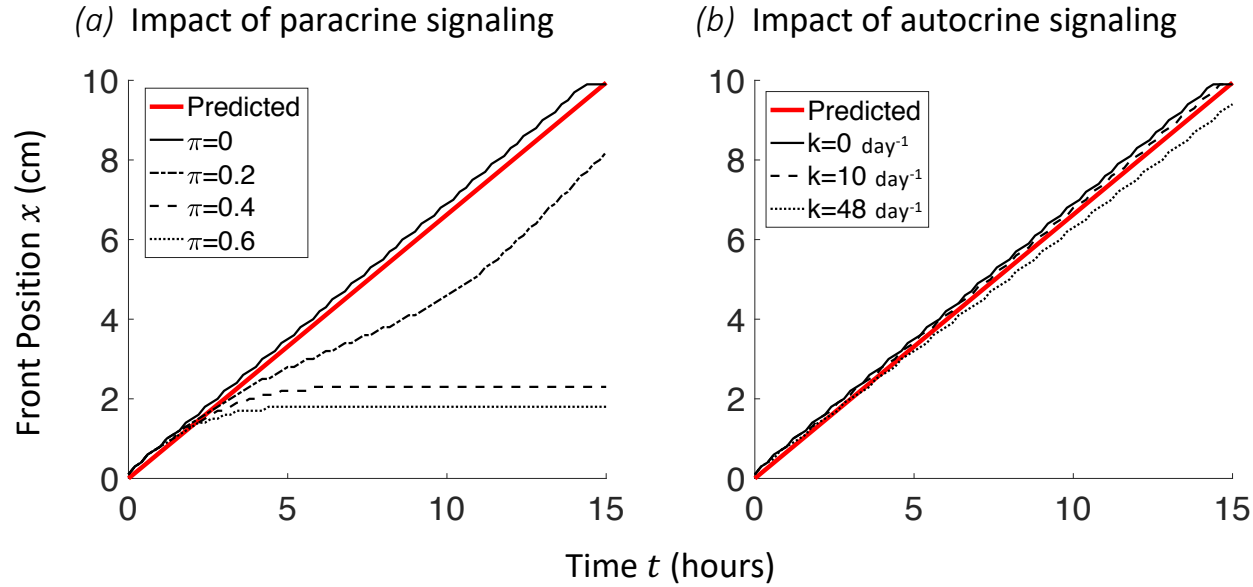
Recall that IFN signaling has the effect of both decreasing the production of virions ( $p$ ) and decreasing the number of cells susceptible to infection ( $T_0$ ), and thus it can in principal slow the spread of infection. The above expression shows that the spatial spread of infection is driven primarily by the production ( $p$ ) and diffusion ( $D_V$ ) of virions, the infection of target cells ( $\beta$ ) and the density of available target cells leading the front of the infection wave, i.e.  $T_0$ . Since autocrine signaling effectively decreases  $p$  for the portion of infected cells that are transitioned to the  $I^*$  class and paracrine signaling has the impact of decreasing the density of target cells  $T_0$ , we can expect the spatial spread of infection to be disrupted in the presence of IFN signaling. This is borne out in the simulation shown in Figure 3.2(b), where the rapid increase in refractory cells throughout the domain results in the infection remaining localized to the right boundary.

We furthermore find that, in the absence of paracrine IFN signaling, the observed traveling wave speed of the infection surprisingly does not depend strongly on the strength of autocrine signaling (i.e. the value of the autocrine parameter  $k$ ). The effect of  $k$  on the wave speed is small even for biologically unreasonable values of  $k$  (Fig. 3.3(b)). This is because the speed of spread is mostly driven by virus production from cells at the wave front. These infected cells are unlikely to be in an antiviral state, because of the waiting time (on average  $1/k$  days) for that to occur. Thus, the results from the PDE model is in a sharp contrast to the results from the ODE model, with respect to

the roles of the autocrine and the paracrine signaling on preventing the growth of infection than paracrine signaling. However, in Figure 3.3(a), we can clearly see that with increasing strength of paracrine signaling  $\pi$ , the position of the infectious front can be stalled.



**Figure 3.2 Model simulations show that paracrine IFN signaling strongly interferes with the spatial spread of infection.** Shown is the solution of the model system at successive time points, with arrows indicating the direction of progression with time. (a) A representative simulation of the PDE model with  $k = \pi = 0$  (no IFN), exhibiting traveling wave behavior initiated from a single nexus of infected cells at position  $x = 0$ . The infection travels an equal distance between successive times, demonstrating constant speed of spread. (b) A representative simulation of the PDE model with cell protection included ( $\pi \neq 0$ ) showing how IFN signaling can stop the spread of infection by depleting target cells. The distribution of virions and infected cells can be seen to remain localized to the far left of the domain, as the rapid depletion of susceptible cells in the domain prevent the infection from establishing a traveling wave.



**Figure 3.3 Comparison of the impacts of paracrine signaling and autocrine signaling on viral spatial spread.** Shown is the position  $x(t)$  of the infection front over time for various values of the free IFN production parameter  $\pi$  while  $k = 0$  (a), and varying the autocrine-mediated transition rate  $k$  while  $\pi = 0$  (b). Here we define the front position  $x(t)$  to be such that  $I(x(t), t) = 0.01 T_0$ . The red line shows the predicted front position given the analytically derived wave speed, i.e.,  $x(t) = v^* t$ . (a) Sufficiently large production of free IFN ( $\pi = 0.4, 0.6$ ) leads to halting the spread of infectious front. (b) Autocrine parameter  $k$  has only small effect on wave speed of infection. The supraphysiologic values  $k = 10$  and  $48 \text{ days}^{-1}$  (corresponding respectively to 0.1 day and 30 minute waiting periods before infected cells transition to  $I^*$  state).

### 3.5 Contributions

In this chapter, we developed a novel system of reaction-diffusion PDEs to model the spread of infection through a 1D simulated tissue in the presence of interferon signaling, with a particular interest in understanding the relative impact of the two IFN signaling pathways – autocrine and paracrine – on the progression of a spatially structured infection. We find that the model exhibits traveling wave solutions and perform a stability analysis in the traveling wave frame to determine the speed at which infection spreads. We found that the speed of the infection wave depends primarily on the virion diffusion coefficient, virion production rate, and the density of target cells at the front of infection, but that the infectious lifespan has negligible impact on the speed of spread. By extension, this implied that autocrine signaling can do little to reduce the speed of spread, since in the best-case scenario it is indistinguishable from cell death. However, since paracrine signaling can reduce the number of target cells susceptible to infection at the front of the infection wave, paracrine signaling can be highly impactful in interfering with the spatial propagation of infection. This stands in stark contrast to the well-mixed infection modeled by the ODE model from Chapter 2, where paracrine signaling had no impact on the initial growth phase of infection. In the spatially structured context, paracrine signaling targets the cells that are most susceptible to infection. And while autocrine signaling still inherently targets infected cells, the reduction of virion production mediated by the autocrine signal has little to no effect since the spread of infection is sustained almost entirely by the cells at the leading edge of the front. In the Chapter 4, we will develop a Cellular Automata framework for modeling this infection process with IFN signaling. This framework will both allow 2D simulations and will include the stochasticity that is inherent to this early stage of infection. We will then discuss the broader implications of our work for both cell-scale and epidemic disease.

## CHAPTER

# 4

# A CELLULAR AUTOMATA MODEL OF IN-HOST VIRAL INFECTION

Infectious disease is an inherently random process in the real world, with stochastic fluctuations such. This randomness mean that predictions of deterministic modeling frameworks often do not directly apply to the real world setting. For instance, in the deterministic PDE model explored in the previous chapter, the nature of parabolic PDEs demands that the virus density  $V(x, t)$  be non-zero on the entire domain for all  $t > 0$ . This means that fractional amounts of virus are present at parts of the domain that are unrealistically far from the initial site of infection given the size and diffusivity of virion particles. We can interpret this as representing the low but non-zero probability that virion particles exist far from the initial infection site. In the real-world, however, the implications for a virion reaching or failing to reach the far-field differ drastically, whereas the PDE framework can not distinguish between the two. These discrepancies between the deterministic models and the practical interpretations of their features and behaviors lead us to develop stochastic models that will better represent the inherent randomness in the infection process.



## 4.1 Cellular Automata Models

There is a wide array of agent-based modeling frameworks that would allow us to study the cell-virus-interferon system in a stochastic setting, including both on- and off-lattice models. However, since we are principally interested in studying the spread of infection through a static tissue, as in the PDE model, we can restrict ourselves to models with stationary, uniformly arranged agents. For this reason we chose to model this system in a Cellular Automata (CA) framework. A CA model consists of a regular 2D grid of cells  $\{s_{ij}\}$ , each in one of a finite number of states  $S = 1, 2, 3, \dots, m$ , with a predetermined ruleset for state transitions that occur between time steps. Throughout, we will take the mesh of the lattice to be  $\Delta x$  and the length of each time step to be  $\Delta t$ .

Realistic biological models that are continuous in both space and time such as those resulting from conservation laws often produce high-dimensional systems of non-linear partial (possibly integro- ) differential equations. There are many issues inherent to studying systems of this type. Firstly, the simulation of many coupled PDEs is a memory intensive process that is vulnerable to numerical instabilities. Furthermore, depending on the qualities of the model, vastly different numerical schemes may be necessary, thus limiting one's ability to freely explore the parameter space of the model. The next challenge is the interpretation of the model, where the deterministic forward-evolution of continuum densities of your state variables must be related back to the biological observables, which are often low resolution compared to the detail of the model and may contain noise. The cellular automata framework alleviates computational complexity of the model by reducing the model to a list of simple rules that can be executed in parallel.

In [39], a cellular automata framework was used to study the growth of tumors in an inhomogeneous environment. Specifically, they studied the impact of vascular profusion and the resulting oxygen landscape of the simulated tissue on the ability of tumors to grow. This problem would have been very difficult to study via a more traditional PDE model, as the vasculature and resulting inhomogeneously oxygenated environment would be difficult to describe in that context and would have complicated the numerical solution of such a system.

More recently, however, Cellular Automata models have come into their own for their ability to describe biological and ecological phenomena when prescribed stochastic rulesets. Specifically, in the field of in-vivo viral infection, several works implemented cellular automaton frameworks to

simulate intracellular infection processes while incorporating the inherent stochasticity at this scale that is crucial in shaping the long-term outcomes of disease [40, 27]. In particular, it is the ability to implement arbitrarily specific rulesets that include a great deal of biological phenomena that lend strength to this modeling approach. Whereas increasing model complexity leads to numerical and analytical difficulties for continuous and deterministic models, the ability to hard-code waiting times and other state transitions in a CA model allows for a more “kitchen sink” modeling approach.

In [41], a Cellular Automata model was developed to study the spread of Influenza A Virus (IAV) infection over a lattice of stationary epithelial cells. In this model, they include both epithelial cells as fixed agents to the lattice and patrolling immune cells that move about the lattice and kill infected cells. Virus particles were not modeled explicitly, but rather IAV infection was assumed to spread directly from cell-to-cell. This simple cellular automata model was able to reproduce the basic dynamical features of IAV infection. In a later work, [40], it was shown that the outcomes of infection are highly dependent on the initial distribution of infected cells. It was found that on average when initialized with tight clusters of infected cells there were fewer overall infections, whereas distributing infected cells more evenly over the lattice led to a greater number of total infections. They also found, however, that when infections were initialized in clusters there was greater downstream uncertainty in the severity of the infection, suggesting that when infection occurs more uniformly over the lattice the results are more deterministic in nature. These kinds of observations are completely inaccessible in deterministic modeling via ODE or PDE where the model outcomes are determined entirely by the initial condition.

## 4.2 Model Development

We are interested in developing a Cellular Automata (CA) model that will serve as a discrete, stochastic equivalent to the cell-infection-interferon process modeled by the PDE system in the previous chapter. Here, the rectangular lattice of cells  $\{s_{ij}\}$  will represent the population of stationary epithelial cells, with five possible states – Healthy, Exposed, Infected, Protected, and Dead. The state transitions will be governed by stochastic processes that are meant to simulate the intracellular diffusion and signaling of free infectious virion particles and IFN proteins. The basic rules governing the CA are listed below:

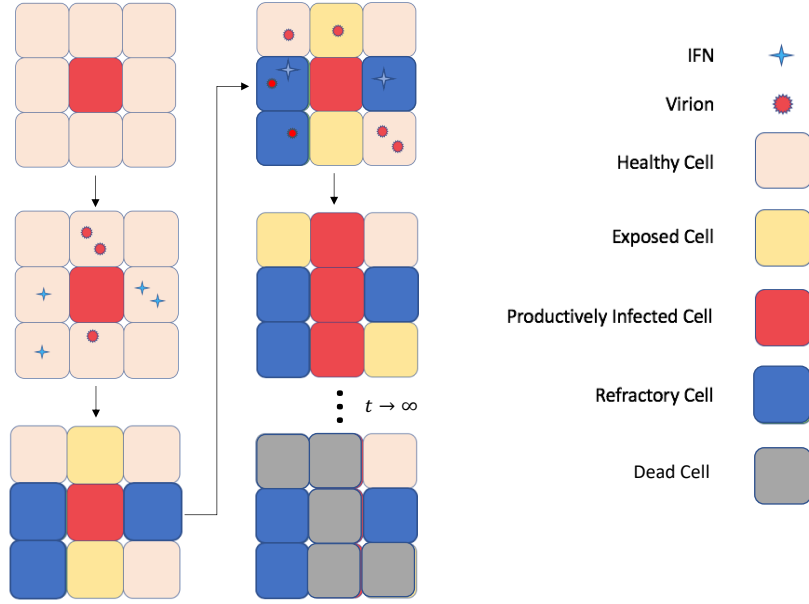
1. Each time step, each infected cell produces `virus_prod` Virions and `ifn_prod` IFNs.
2. Each IFN is assigned to a recipient cell by a simulated diffusion process.
3. A Healthy cell receiving an IFN becomes Protected.
4. Each Virion is assigned to a recipient cell by a simulated diffusion process.
5. A Healthy cell receiving a Virion becomes Exposed.
6. After `infect_delay` time steps, an Exposed cell becomes Infected.
7. After `lifespan` time steps, an Infected cell becomes Dead.
8. A cell in any state other than Healthy does not change state by receiving a Virion or IFN.

A schematic for the CA model can be seen in Figure 4.1. This model neglects the impact of autocrine IFN signaling and only includes the cell-to-cell protective effect of paracrine IFN signaling. Since the extreme case of complete inhibition of virion production in infected cells is indistinguishable from cell death, we can lump the autocrine effect into the `lifespan` parameter. This model simplification is supported by the results from the PDE model developed in Chapter 3 which suggested that the impact of autocrine signaling is diminished in the case of radial spatial spread.

#### 4.2.1 Simulating Diffusion by Random Walk

The most natural way to model the movement of virions and IFN particles over the cellular lattice is via a 2D random walk process in which each particle has a fixed probability `p_stay` of maintaining its current position on the lattice for the duration of the current time step and a corresponding probability  $(1 - p_{\text{stay}})$  of moving, in which case it would choose from the 4 possible directions to move in with equal probability. This approach is favorable because it is a discrete, stochastic process that is known analytically to upscale to the continuous fickian diffusion modeled in the PDE.

It will be valuable to be able to compare the probabilistic parameters of the 2D random walk with the upscaled diffusion coefficient  $D$ . To do so, we perform an analysis. Consider  $N(x_i, y_j, t)$  be the number of particles of a single species at the  $i, j$  position at time  $t$ . Recall that  $p_{\text{stay}}$  be the



**Figure 4.1 Schematic diagram of cellular automata model.** Model consists of a rectangular lattice of cells in one of five states: Healthy, Exposed, Infected, Protected, and Dead. Infected cells produce virion and IFN particles that diffuse to nearby lattice points, resulting in transitions to the Exposed and Protected states, respectively. After 1 lifespan time steps, Infected cells become Dead.

expected proportion of particles that do not move over any given time step. Then,

$$N(x_i, y_j, t + \Delta t) = p_{\text{stay}} N(x_i, y_j, t) + \frac{1}{4}(1 - p_{\text{stay}})(N(x_{i+1}, y_j, t) + N(x_{i-1}, y_j, t) + N(x_i, y_{j+1}, t) + N(x_i, y_{j-1}, t))$$

We assume that the grid spacing  $\Delta x$  is uniform and equal in the  $x$  and  $y$  directions. Now, by Taylor expansion, we see that

$$\begin{aligned} N(x_{i\pm 1}, y_j, t) &= N(x_i, y_j, t) + (\pm \Delta x) N_x(x_i, y_j, t) + \frac{1}{2}(\pm \Delta x)^2 N_{xx}(x_i, y_j, t) + \frac{1}{6}(\pm \Delta x)^3 N_{xxx}(x_i, y_j, t) + \mathcal{O}(\Delta x^4) \\ &= N(x_i, y_j, t) \pm \Delta x N_x(x_i, y_j, t) + \frac{1}{2}\Delta x^2 N_{xx}(x_i, y_j, t) \pm \frac{1}{6}\Delta x^3 N_{xxx}(x_i, y_j, t) + \mathcal{O}(\Delta x^4), \end{aligned}$$

and similarly in  $y$ . Substituting these series approximations in to the original equation, we note

that the odd-order terms annihilate, leaving:

$$\begin{aligned} N(x_i, y_j, t + \Delta t) &= p_{\text{stay}} N(x_i, y_j, t) + \frac{1}{4}(1 - p_{\text{stay}}) \left( 4N(x_i, y_j, t) + \Delta x^2 N_{xx}(x_i, y_j, t) + \Delta x^2 N_{yy}(x_i, y_{j+1}, t) + \mathcal{O}(\Delta x^4) \right) \\ &= N(x_i, y_j, t) + \frac{\Delta x^2}{4}(1 - p_{\text{stay}}) \left( N_{xx}(x_i, y_j, t) + N_{yy}(x_i, y_{j+1}, t) \right) + \mathcal{O}(\Delta x^4) \end{aligned}$$

We can similarly perform a Taylor expansion in  $t$  on the left-hand side and substitute  $N(x_i, y_j, t + \Delta t) = N(x_i, y_j, t) + \Delta t N_t(x_i, y_j, t) + \mathcal{O}(\Delta t^2)$ . Substituting above, we have

$$N_t(x_i, y_j, t + \Delta t) = \frac{\Delta x^2}{4\Delta t}(1 - p_{\text{stay}}) \left( N_{xx}(x_i, y_j, t) + N_{yy}(x_i, y_{j+1}, t) \right) + \mathcal{O}\left(\frac{\Delta x^4}{\Delta t}\right) + \mathcal{O}(\Delta t)$$

In the limit, as  $\Delta x, \Delta t \rightarrow 0$  while preserving the ratio  $\Delta x^2/\Delta t = \text{const.}$ , we see that the unbiased random walk process is equivalent to the continuous diffusion process governed by the classical heat equation  $N_t = D \nabla^2 N$  with diffusion coefficient  $D = \frac{\Delta x^2(1-p_{\text{stay}})}{4\Delta t}$ .

#### 4.2.2 Simulating Diffusion with a Mean-Field Approximation

In practice, we find that the random walk approach to simulating the diffusion of particles over the cellular lattice is ineffective due to the disparity between time scales relevant to particle movement and cellular state transitions. That is to say, the  $\Delta t$  sufficient to make the random walk process an acceptable approximation of the continuous time motion of small particles is much smaller than the time step that is appropriate for the transition of cell states. This is reflective of the inherent fast-slow nature of the ODE system developed in Chapter 2, where the parameters governing cell compartments are in general much smaller than the parameters governing the Virion and IFN compartments. In the ODE model, this allowed us to make the quasi-equilibrium assumption in the non-cell compartments, reducing the dimensionality of the system of equations and facilitating analyses that would be encumbered by a larger number of equations. Here, in the Cellular Automata setting, we can perform a similar reduction of state space dimension due to this disparity in time scales. Specifically, we can argue that the time scales for particle diffusion over the lattice are sufficiently small that the diffusion of particles away from the cell of origin and binding to a recipient cell occur within a single CA time step. This allows us to completely decouple the two

time scales – cellular and particulate. This is advantageous, as the diffusion of IFNs and virions can be approximated arbitrarily well by the diffusion equation  $V_t = D_V \nabla^2 V$ , as shown in the previous section. Below, we develop a framework for selecting a recipient cell for each virion by drawing the recipient cell location from a distribution that is consistent with a Wiener process.

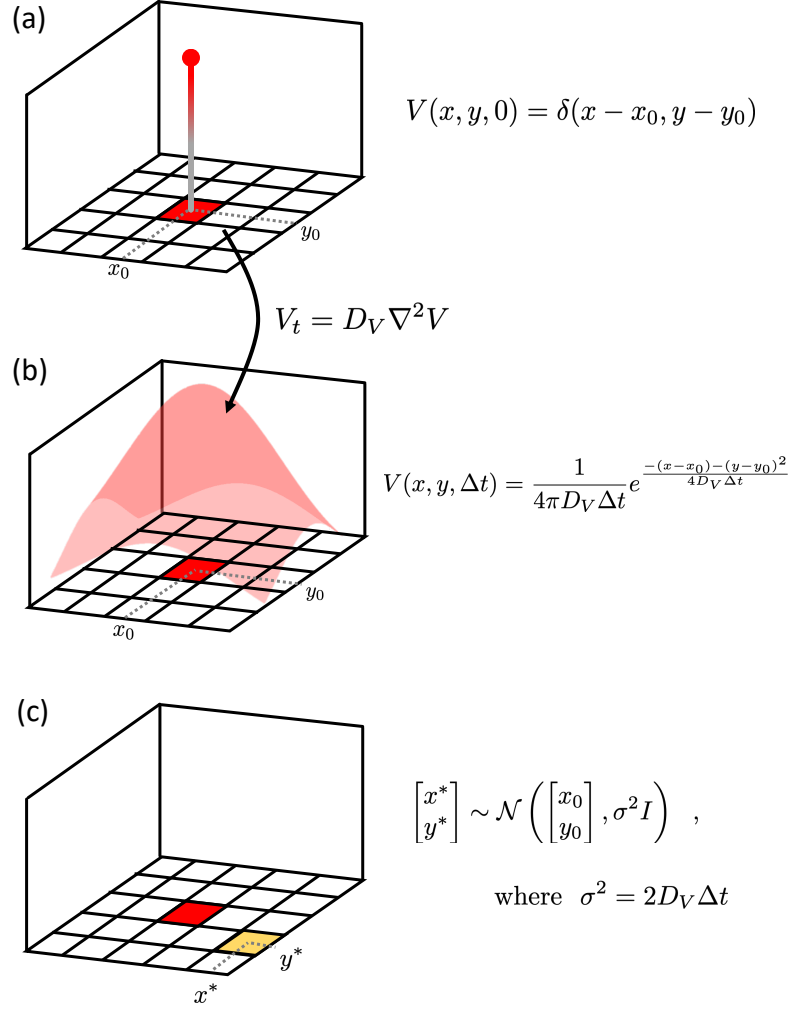
Consider a single virion originating from an infected cell with grid location  $(x_0, y_0)$ . Assuming that virions bind to cells on the lattice continuously and independently at a constant rate, then the time  $\Delta t$  between the production and binding of a single virion is an exponentially distributed random variable. Furthermore, we interpret the location of the virion to be a random variable with time-varying probability density function  $V(x, y, t)$  that is initially  $\delta$ -distributed, i.e.,  $V(x, y, 0) = \delta(x - x_0, y - y_0)$ . We can now interpret the diffusion equation

$$V_t = D_V \nabla^2 V \quad (4.1)$$

that resulted from the analysis of the random walk process to govern time-evolution of the probability density function  $V$  as the limiting distribution for a two-dimensional Wiener process. This is convenient, as when given dirac-delta initial conditions, an explicit solution can be found for (4.1):

$$V(x, y, t) = \frac{1}{4\pi D_V \Delta t} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{4D_V \Delta t}}$$

We recognize this Gaussian function as being precisely the probability density function of the bivariate normal distribution  $\mathcal{N}\left(\begin{bmatrix} x_0 & y_0 \end{bmatrix}^T, \sigma^2 I\right)$  where  $\sigma^2 = 2D_V \Delta t$ . This allows us to draw from this distribution to determine the coordinates of the binding site for the virion. Combining these insights, we develop algorithm 4.1 for the diffusion of virions and IFNs. We also show a schematic for this recipient cell selection procedure in Figure 4.2. We can now write down the algorithm for the complete Cellular Automata model, which is detailed in algorithm 4.2.



**Figure 4.2 Schematic of recipient cell selection process.** (a) The particle is initially located at the producing cell, located at lattice position  $(x_0, y_0)$ . We interpret its position to be initially delta distributed, i.e.,  $V(x, y, 0) = \delta(x - x_0, y - y_0)$ . (b) The probability density function of the particle's position is assumed to be given by solving the diffusion equation  $V_t = D_V \nabla^2 V$  forward  $\Delta t$  time units from the dirac delta initial condition shown in (a). This gaussian functional form of the p.d.f is known explicitly. (c) The particle's binding location  $(x^*, y^*)$  is drawn from the p.d.f shown in (b). Since in general  $(x^*, y^*) \in \mathbb{R}^2$ , we snap this binding location to a lattice grid point by rounding to determine the recipient cell. This process is identical for both virion and IFN particles, differing only in the value of the diffusion coefficient.

---

**Algorithm 4.1:** Recipient Cell Selection Algorithm

---

For each infected cell with lattice coordinates  $(x_0, y_0)$ :

1. sample  $\Delta t \sim \exp(\lambda)$
2. sample  $\begin{bmatrix} x^* \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \sigma^2 I\right)$  where  $\sigma^2 = 2D_F \Delta t$
3. snap binding site  $(x^*, y^*)$  to recipient cell grid location  $(x_i^*, y_j^*)$ :

$$(x_i^*, y_j^*) = \underset{(x,y) \in \mathbb{Z}^2}{\operatorname{argmin}} \left( \sqrt{(x - x^*)^2 + (y - y^*)^2} \right)$$

4. if  $(x_i^*, y_j^*)$  Healthy,  $(x_i^*, y_j^*)$  becomes Protected.
5. Repeat 1. – 4. IFN\_prod times.

For each infected cell with lattice coordinates  $(x_0, y_0)$ :

1. sample  $\Delta t \sim \exp(\lambda)$
2. sample  $\begin{bmatrix} x^* \\ y^* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} x_0 \\ y_0 \end{bmatrix}, \sigma^2 I\right)$  where  $\sigma^2 = 2D_V \Delta t$
3. snap binding site  $(x^*, y^*)$  to recipient cell grid location  $(x_i^*, y_j^*)$ :

$$(x_i^*, y_j^*) = \underset{(x,y) \in \mathbb{Z}^2}{\operatorname{argmin}} \left( \sqrt{(x - x^*)^2 + (y - y^*)^2} \right)$$

4. if  $(x_i^*, y_j^*)$  Healthy,  $(x_i^*, y_j^*)$  becomes Exposed.
  5. Repeat 1. – 4. virus\_prod times.
- 

---

**Algorithm 4.2:** Cellular automata model algorithm

---

**for**  $(i, j) \in \{1, 2, \dots, n\} \times \{1, 2, \dots, n\}$ :

1. **if**  $s_{ij} = \text{Exposed}$ 
    - (a) infection\_age++
    - (b) **if** infection\_age  $\geq$  infect\_delay,  $s_{ij} = \text{Infected}$
  2. **if**  $s_{ij} = \text{Infected}$ , **do** algorithm 4.1. Then:
    - (a) infection\_age++
    - (b) **if** infection\_age  $\geq$  lifespan,  $s_{ij} = \text{Dead}$ .
-



### 4.3 Results

We simulate the Cellular Automata model on a  $n \times n$  lattice, initialized with a single infected cell at the center of the grid, i.e.,  $(\lfloor \frac{n}{2} \rfloor, \lfloor \frac{n}{2} \rfloor)$ , taking  $n = 100$ . This initialization best simulates the initial conditions used in the PDE model of a single infected cell located at the left end of the domain. The parameters of the CA model, their units, and default values are enumerated in Table 4.1. We will be interested in studying the impact of IFN signaling on the outward spread of the infected region, so the value of `ifn_prod` will be varied over simulations. Note that since each virion and IFN in our formulation of the model successfully finds a recipient cell before being cleared, the `virus_prod` and `ifn_prod` parameters can not be directly interpreted as equivalent to the corresponding PDE parameters  $p$  and  $\pi p$ , as none of them are cleared before making contact.

**Table 4.1** Parameters of the CA model.

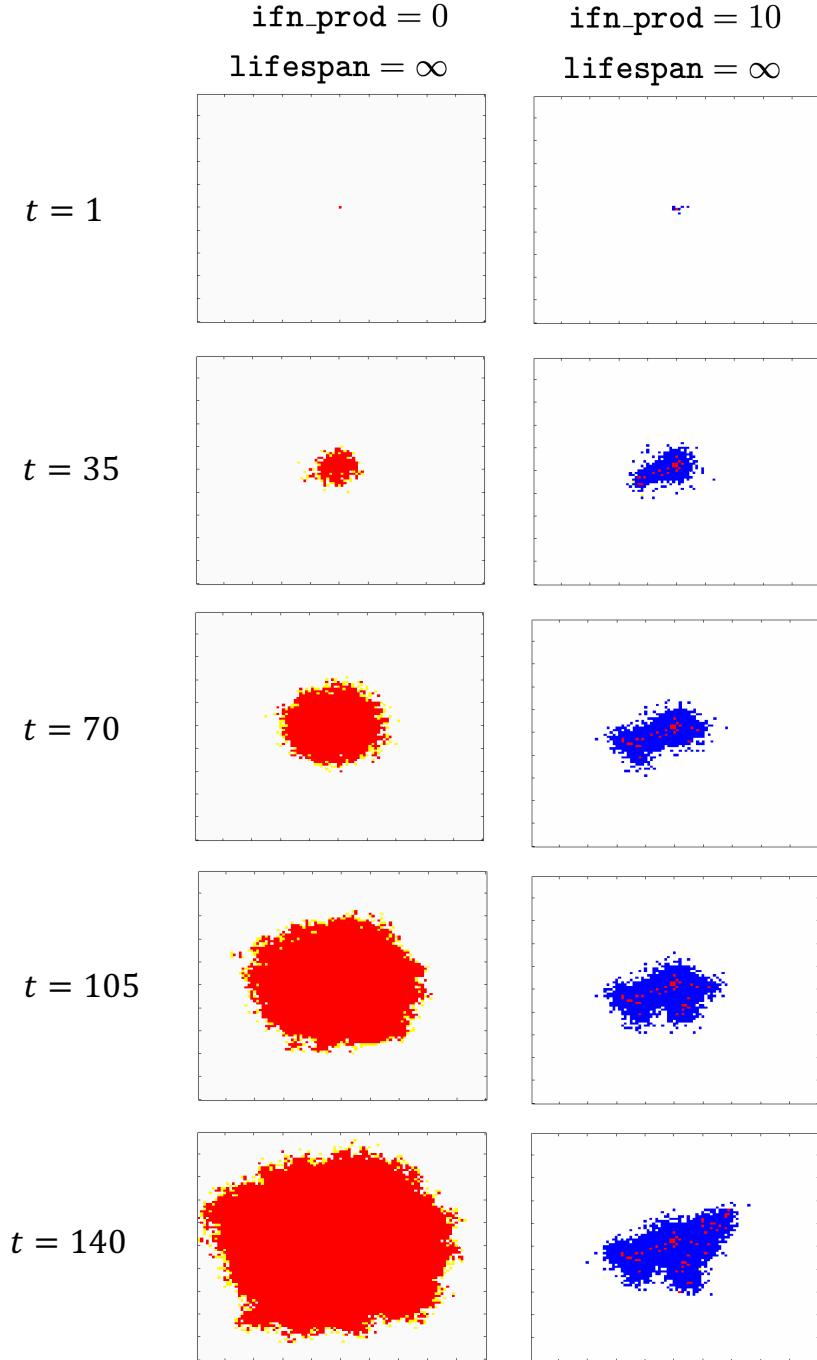
Parameter	Description	Units	Default Value
<code>virus_prod</code>	virus production rate	virus $\text{step}^{-1}$	1
<code>ifn_prod</code>	IFN production rate	IFN $\text{step}^{-1}$	NA
<code>virus_diff</code>	virus diffusion coeff.	cell-width <sup>2</sup> $\text{step}^{-1}$	0.1
<code>ifn_diff</code>	IFN diffusion coeff.	cell-width <sup>2</sup> $\text{step}^{-1}$	0.5
<code>infect_delay</code>	post-exposure wait time	steps	14
<code>lifespan</code>	post-exposure wait time	steps	48

We first simulate the CA model with long-lived infection (i.e., `lifespan` =  $\infty$ ) in the presence and absence of IFN response (see Figure 4.3). The simulations show that the radial spread of the infected region occurs relatively symmetrically and without impediment in the absence of IFN signaling, i.e., `ifn_prod` = 0 (see left column of Figure 4.3). This parallels the conclusion from the PDE model that the infection travels with a constant wave speed when no IFN is present. However, in the presence of IFN signaling, the local conversion of Healthy cells to Protected cells impedes the infection from spreading radially (see right column of Figure 4.3). We see in the figure that the shape of the affected region is highly amorphous. We can surmise that the limited availability of Healthy cells near to Infected cells leads to secondary infections being rare events, thus causing the spread of infection to be sporadic and asymmetric. This observation also reflects one of the key observations of viral phylogenetics: that the genetic diversity in viral inoculum is drastically reduced at the

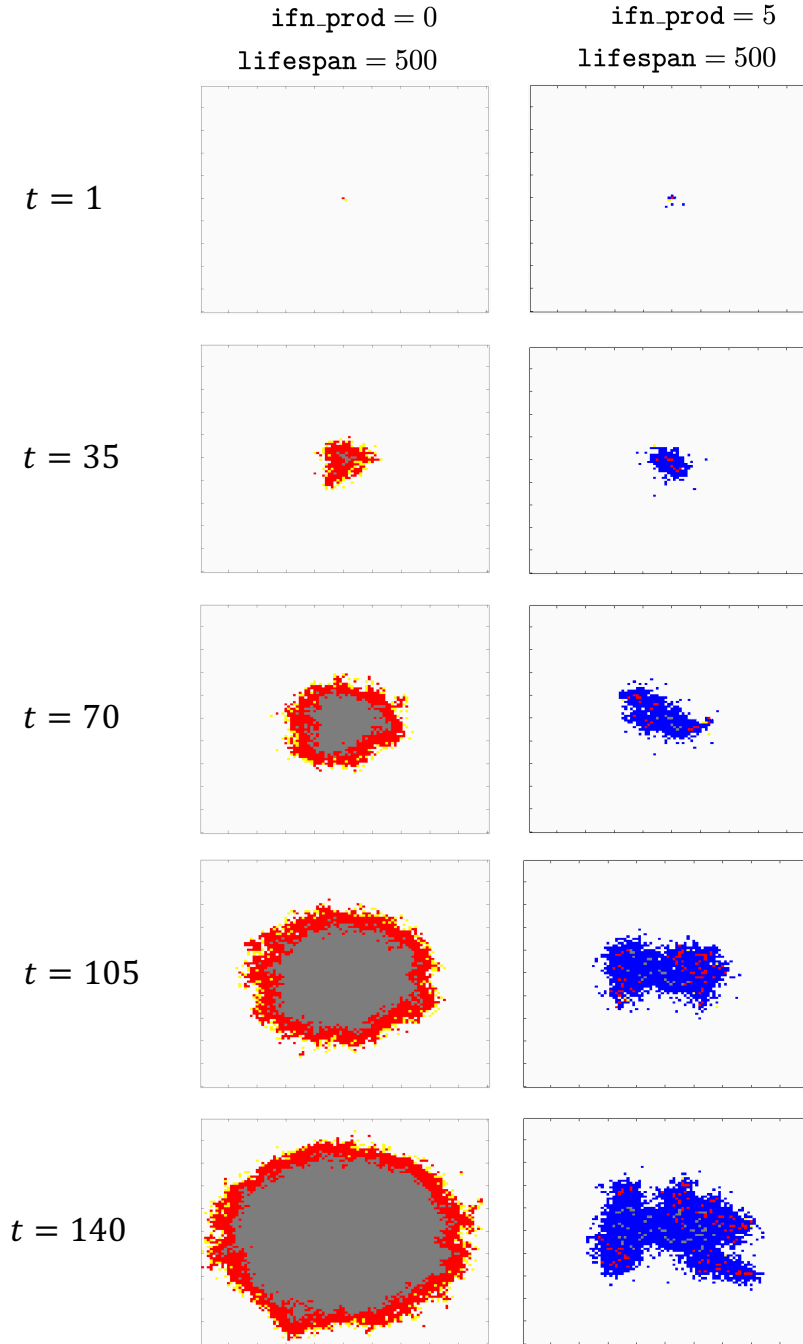
transmission bottleneck and that the “founder” virus that goes on to establish systemic infection represents a very small subset of the initial viral challenge. That is to say, as the effect of IFN signaling segregates infected cells behind a layer of protected anti-viral cells, the average number of productive secondary infections from each infected cell will decrease. If this number decreases below 1, then we can see the genetic diversity of the viral challenge is greatly reduced due to IFN signaling during this initial spatial phase of epithelial infection. Indeed, it has been found that the epithelial transmission barrier serves as a strong selection pressure for anti-IFN phenotypes ([19]).

We then simulate the CA model with short-lived infection (i.e., `lifespan` = 48) in the presence and absence of IFN response (see Figure 4.4). Visually comparing the left column of Figure 4.4 to that of Figure 4.3, we see that the spatial spread of the infected area is not hindered by introduction of cell death. The reason for this is that in the absence of IFN signaling, recipient cells for virions are mostly very close to the progenitor cell. That is to say, the new infections that cause the front of infection to propagate are being caused mostly by infected cells that are themselves located on or near the boundary of the infected region. Therefore, once an infected cell reaches an infectious age of `lifespan`, it is too far from the boundary to be responsible for new infections anyway, and its death is therefore inconsequential for the continued propagation of the infectious front. However, in the presence of IFN signaling, the segregation of infected cells from healthy ones can reduce the probability of an infected cell causing a secondary infection long enough for the cell to die before causing any new infections. This can lead to dead regions of the boundary where no infected cells remain, cutting off the continued spread of infection, and can potentially result in the extinction of infection from the lattice. This suggests that cell death is most impactful in a high-IFN regime.

We compare the effect of IFN on the total number of infections in both the long-lived and short-lived infection cases in Figures 4.5 and 4.6. In both figures, the top row shows a representative simulation of the CA model at  $t = 100$ . The bottom rows show the total number of infected cells over the entire lattice over time plotted on the log-log scale for 100 independent realizations of the CA model. We see that there is no possibility of extinction in the long-lived case, as the log-log plots in the bottom row of Figure 4.5 shows an uninterrupted linear trend. However, in the short-lived infection case (see Figure 4.6) where there is the possibility that the infection may die out entirely, we see that at the highest level of IFN production the log-log plot shows that the number of infections fails to continue increasing with time, suggesting that extinction is occurring in those simulations.

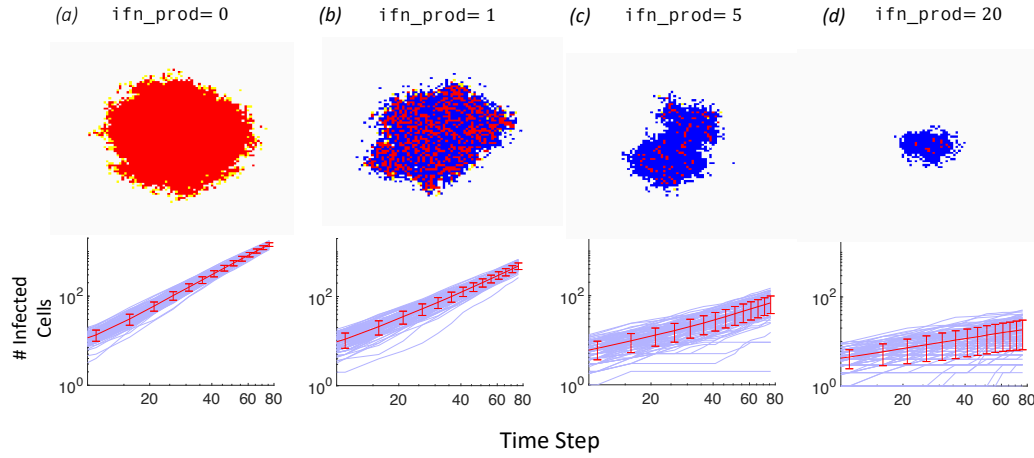


**Figure 4.3 Representative simulation of CA model with and without IFN signaling with long-lived infection.** The left column shows the progression of infection in the absence of IFN signaling. The lattice is initially populated with Healthy cells (white) at each grid point, with a single infected cell (red) located in the center of the lattice. Each time step, an infected cell produces `virus_prod` virions, each of which is assigned to a recipient cell according to Algorithm 4.1. If the recipient cell is Healthy, it becomes Exposed (yellow). Otherwise the recipient cell is unchanged. The right column shows the progression of infection in the presence of IFN signaling. Each time step, in addition to virions, an infected cell produces `ifn_prod` IFNs, each of which is assigned to a recipient cell according to Algorithm 4.1. If the recipient cell is Healthy, it becomes Protected (blue). We see that the abundance of Protected cells limits the number of virion binding sites that can lead to further infections, thus impeding the spread of the infected area.



**Figure 4.4 Representative simulation of CA model with and without IFN signaling with short-lived infection.** The left column shows the progression of infection in the absence of IFN signaling with short lived infected cells. We see the spread of infection is virtually unimpeded compared to the long-lived infection case, as cell death occurs long after the front of infection has move (i.e., all Healthy cells have been exhausted), meaning that cell death is inconsequential. However, the right column shows the progression of short-lived infection in the presence of IFN signaling. Here, infected cells may become rapidly sequestered by a neighborhood of Protected cells. Some cells may die before creating any secondary infections, leading to the possibility of extinction.

In Table 4.2 we show the probability of viral extinction across a range of infected cell lifespan and IFN strength. Furthermore, we note the linear trend in the log-log plots in the bottom row of Figure 4.5 suggest that the total number of infections roughly follows a power law with time, i.e.,  $I(t) \propto t^\gamma$ . The log-log plots in the bottom row of Figure 4.6 show that, short-lived infection case, this power law seems to hold for low levels of IFN production, fails to hold for higher levels of IFN production as the likelihood of extinction increases. For the case of low or no IFN production, we find that  $\gamma$  takes a value close to 2. This is consistent with a constant radial speed of spread, as the area of infection increases quadratically with the radius. That is, if  $r(t) = v^*t$  then  $I(t) \propto A(t) = \pi(r(t))^2 = \pi(v^*)^2 t^2$ .



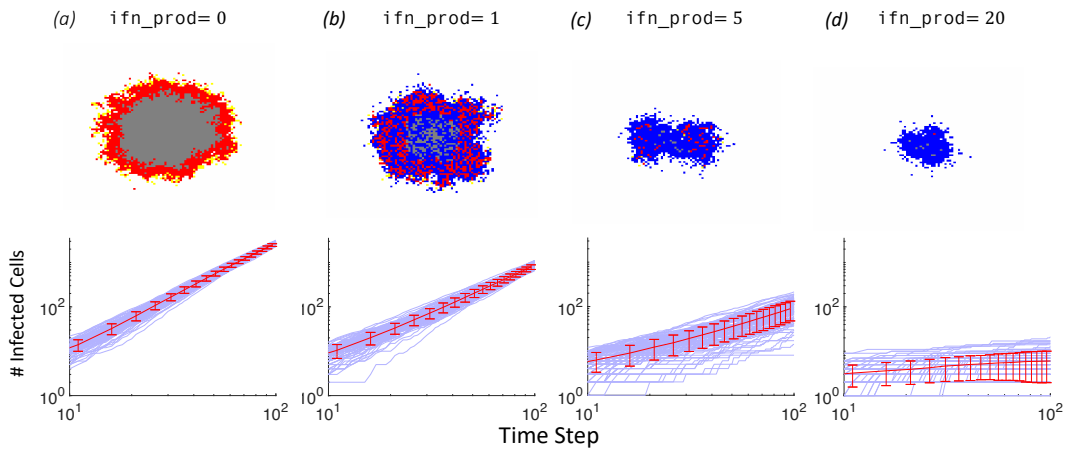
**Figure 4.5 Log-log plots of lattice sum of infections over time for long-lived infection.** Along the top row are representative lattice states at  $t = 100$  for each IFN level. In the bottom row are the total number of Infected cells over time (light blue lines) for 100 independent CA realizations at four different IFN levels graphed on the log-log scale. In red is shown the mean and one standard deviation. The linear trends in these plots suggests a power law of the form  $I(t) \propto t^\gamma$ , where  $\gamma$  decreases with `ifn_prod`.

**Table 4.2 Table of extinction probabilities across CA parameters.** The table above shows the percentage of 100 independent simulations that resulted in extinction of infection within the first 150 time steps across varying infectious lifespan and IFN production levels.

		ifn_prod				
		1	2	5	10	20
lifespan	6	80%	98%	100%	100%	100%
	12	0%	4%	77%	100%	100%
	24	0%	0%	12%	94%	100%
	48	0%	0%	3%	49%	98%

### 4.3.1 Stochasticity of model outcomes

One of the key features that distinguishes the cellular automata framework from the PDE model developed in Chapter 3 is the inherent stochasticity of the model outcomes. Above we showed that by aggregating many model realizations there is was a clear trend suggesting that the average number of infected cells over time follows a power law (see Figure 4.5 and 4.6). However, if we show the distribution of the number of infected cells at any given time, we see that the distribution depends heavily on the level of IFN production. In Figure 4.7 we show the histograms of the total number of

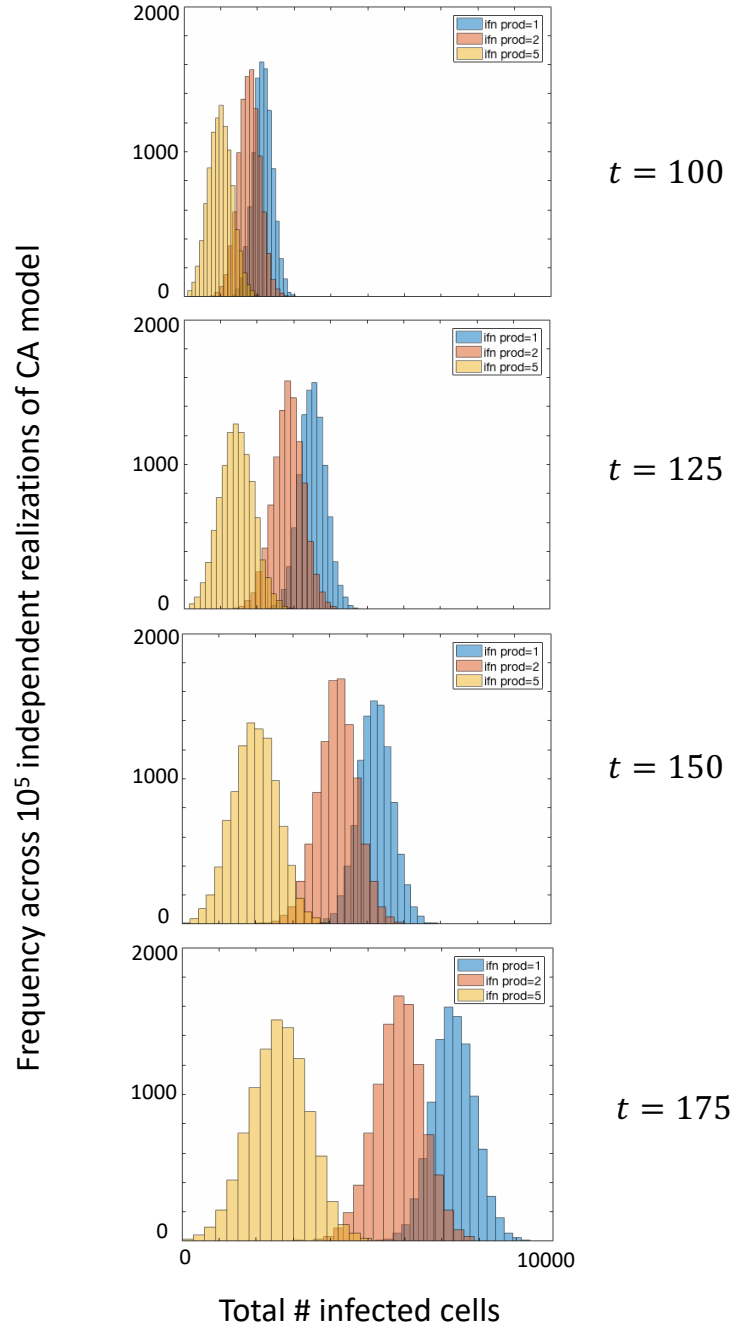


**Figure 4.6 Log-log plots of lattice sum of infections over time for short-lived infection.** Along the top row are representative lattice states at  $t = 100$  for each IFN level. In the bottom row are the total number of Infected cells over time (light blue lines) for 100 independent CA realizations at four different IFN levels graphed on the log-log scale. In red is shown the mean and one standard deviation. The linear trends in these plots suggests a power law of the form  $I(t) \propto t^\gamma$ , where  $\gamma$  decreases with `ifn_prod`. However, we see the linear trend begins to fail for `ifn_prod= 20`, as extinction of infection from the lattice becomes increasingly probable.

infected cells (sum over entire lattice) over  $10^5$  independent realizations of the CA model for three different values of `ifn_prod`. The panels show the histogram at 4 time points. We can clearly observe that the histogram corresponding to `ifn_prod`= 1 exhibits the largest mean number of infected cells at all 4 times. We furthermore see that the distance between the means of the three histograms increases with time, which is consistent with the observation that the mean numbers of infections follows a power law  $I(t) \propto t^\gamma$  where  $\gamma$  decreases with `ifn_prod`. Interestingly, we observe that the characteristic width of the histogram seems to increase with `ifn_prod`, suggesting that as IFN signaling increases the uncertainty of the model outcomes increases. For example, at  $t = 175$ , the tails of the yellow histogram show that some realizations of the CA with `ifn_prod`= 5 exhibit nearly no proliferation of infection where others show up to 5000 infected cells. This is consistent with our observation that successful virion contacts that produce new infections become increasingly rare as `ifn_prod` increases, implying a greater likelihood that the infection is unsuccessful over time.

#### 4.4 Contributions and Future Work

In this Chapter we have developed a cellular automata model that can be seen as the discrete stochastic 2D extension of the PDE model developed in Chapter 3. For the purpose of this model, we have developed a scheme for lattice-free diffusion in which virion and IFN particles are assigned to recipient cells in a way that mimics a continuous time diffusion process en lieu of a random-walk approach. We compared the simulation of the cellular automata in the presence and absence of IFN production for both long-lived and short-lived infection. We find, as with the PDE model, that the effect of IFN to decrease the number of healthy cells near to infected cells has the effect of reducing the rate at which the affected region grows. We also discovered that the lifespan of infected cells, as in the PDE model, does not strongly affect the ability of the infection to spread, as cell death on average occurs sufficiently far away from the active front of infection as to be inconsequential. This observation does not hold, however, for higher levels of IFN production, where the abundance of protected cells can sufficiently slow the spread of infection to such an extent where cell death can lead to extinction of infection from the lattice. This novel approach elucidates how paracrine IFN signaling can slow infection during the early, stochastic phase of epithelial spread and can stop IFN incompetent virus infection at the epithelial transmission barrier.

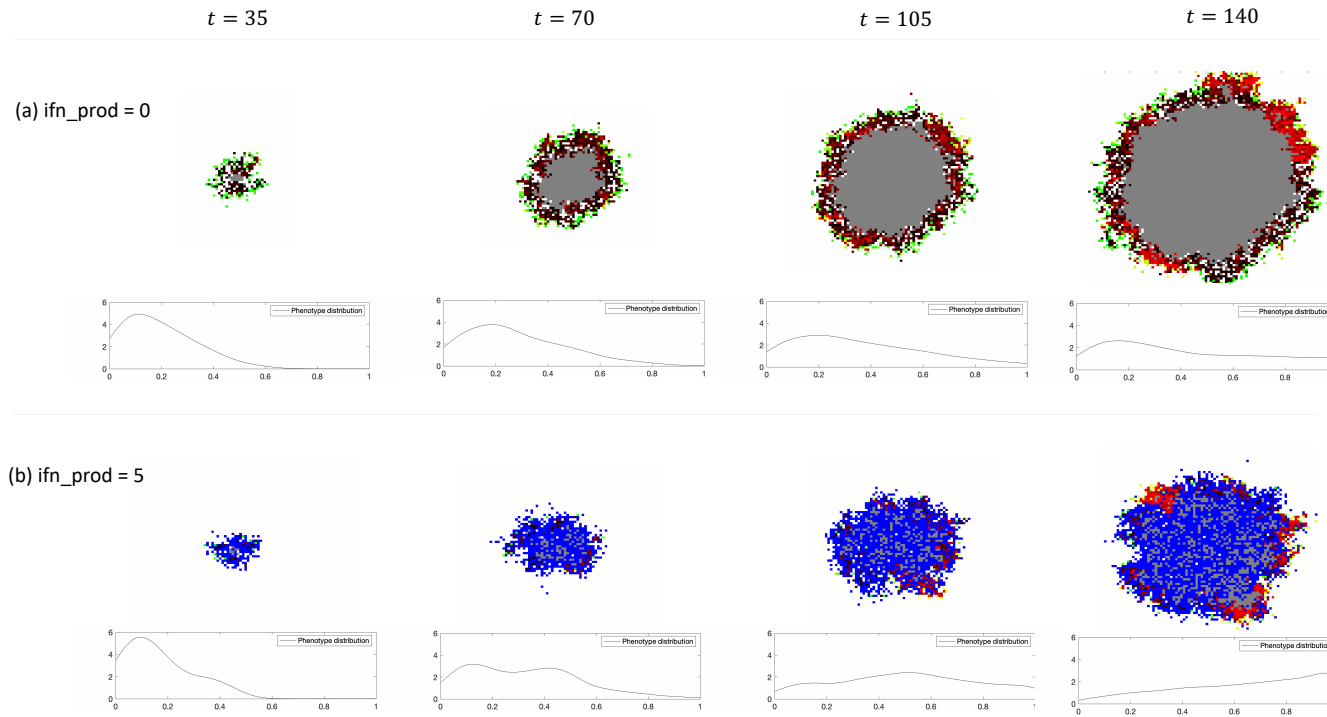


**Figure 4.7 Time evolution of distribution of total (lattice sum) number of infected cells for different levels of IFN production.** Shown are the histograms of the total number of infected cells (sum over entire lattice) over  $10^5$  independent realizations of the CA model for three different values of `ifn_prod` at 4 time points. We observe that the mean number of infections decreases with `ifn_prod` at all times and that the variance of the distribution seems to grow more quickly for larger `ifn_prod`. This suggests that high levels of IFN production increases the stochasticity of model outcomes and leads to greater downstream uncertainty of global infection levels.



As a future extension of this work, we are interested in studying more specifically the effect of paracrine IFN signaling as a selection pressure during early epithelial infection. As discussed in Chapter 1, IFN signaling is known to be an important determinant of viral species tropism, as many human viruses have non-structural proteins as well as specialized nucleic acid structures to that have adapted to interfere with the IFN signaling pathway or avoid detection by the host cell. We have begun development of an extended CA framework that includes viral evolution by incorporating a continuum of IFN-blocking viral phenotypes between 0 and 1. Phenotype 0 infections admit production of `ifn_prod` IFN particles per time step, while phenotype 1 infections completely prevent production of IFN Particles. The production of IFN varies continuously with phenotype between these two extreme cases, with an infection of phenotype  $p_{ij}$  producing an average of  $(1 - p_{ij}) \cdot \text{ifn\_prod}$  IFN particles per time step.

An infected cell at grid location  $(x_i, y_j)$  has phenotype  $p_{ij} \in [0, 1]$ . Daughter infections are assigned a phenotype that is close to but not equal to the phenotype of the parent infection. Specifically, the daughter phenotype  $p_{ij}^*$  is drawn from a uniform distribution  $\mathcal{U}[p_{ij} - \Delta p, p_{ij} + \Delta p]$ . Under this framework, phenotype 0 infected cells are unlikely to produce daughter infections since they will be producing large amounts of IFN and surrounding themselves in an insulting layer of protected cells. However, phenotype 1 infection will spread easily since it will produce no IFN and therefore have more ready access to healthy cells. Our preliminary observations show that, under this framework that allows for viral evolution, IFN signaling acts as a powerful selection pressure. That is, IFN signaling strongly impedes the spread of IFN incompetent ( $p \approx 0$ ) phenotypes, causing infections of phenotype near to 1 to emerge as the dominant “strain” of infection. In Figure 4.8 we compare simulations of this CA model at `ifn_prod=0` and `ifn_prod=5`, along with the distribution of viral phenotypes over the entire lattice. We see that in the presence of IFN signaling (`ifn_prod=5`, bottom row), there is strong species tropism as the phenotypic distribution becomes clustered around 1 as the simulation progresses. However, without IFN signaling (`ifn_prod=0`, top row) there is no species tropism and the distribution of phenotypes simply approaches  $\mathcal{U}[0, 1]$ . In the future, we are interested in studying this model in more depth to quantify the impact of IFN signaling on viral evolution.



**Figure 4.8 Representative simulations of CA model with viral evolution and distributions of viral phenotype.** Shown are two independent simulations of the CA model with viral evolution in the absence ( $\text{ifn\_prod}=0$ , top row) and presence ( $\text{ifn\_prod}=5$ , bottom row) of IFN signaling. The CA simulations show infected cells continuously varying in color between black and red to represent infections of phenotype 0 to 1, respectively. Cells protected by IFN are shown in blue, while exposed cells are shown in green/yellow. Below each snapshot of the CA grid is the distribution of viral phenotypes on the lattice at that time.

## CHAPTER

# 5

## EQUATION LEARNING

Classical mathematical dynamical models in biology and physics are most often direct expressions of physical conservation laws or biostatistical principals in the form of differential equations. These models have proven powerful for making insights into many systems. However, as the number of coupled phenomena becomes large, as is inherent to many biological problems, a first-principals approach to modeling often becomes intractable. Increasingly, with the development of new sensor technologies and machine learning frameworks that allow for high-throughput data collection, there has been a great deal of interest in data-driven modeling approaches. An iterative approach to model selection where several candidate models are proposed and then validated against experimental or observational data has become a staple of many applied math fields. This model selection approach may become unattractive, however, in the case of spatial systems, where the addition or removal of certain terms in a PDE model may fundamentally alter the numerical properties of the model and require the implementation of wildly different numerical solution schemes.

Many methods for data-driven modeling have emerged, including equation-free approaches [42], artificial neural networks [43], and natural selection of symbolic expressions [44]. Most recently,

great promise has been shown by a novel framework that learns non-linear differential equations by employing sparse regression techniques to a linear regression problem consisting of precomputed non-linear candidate terms [45, 46]. In this Chapter, we will adapt the framework developed in [45] and [46] to learn governing equations that describe the cellular automata model developed in Chapter 4, with the particular goal of studying the parametric response in the learned equations to IFN signaling strength.

## 5.1 Equation Learning via linear regression

We will discuss a general framework for using linear regression for fitting a differential equation model to data. Consider the 1-dimensional dynamical system of a single independent variable as follows:

$$y' = F(x, y)$$

Given a data set  $\{x_i, y_i, y'_i\}_{i=1}^N$  where  $y_i = y(x_i)$  and  $y'_i = y'(x_i)$ , we will determine the right-hand side dynamics  $F$ . Suppose that for some basis of functions  $\mathcal{B} = \{f_i\}_{i=1}^p$ , we have that  $F \in \text{span}(\mathcal{B})$ . That is, there exist coefficients  $\{w_i\}_{i=1}^p$  such that  $F = \sum_{i=1}^p f_i w_i$ . Then we can say

$$\begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_N \end{bmatrix} = \begin{bmatrix} f_1(x_1, y_1) & f_1(x_1, y_1) & \cdots & f_p(x_1, y_1) \\ f_1(x_2, y_2) & f_1(x_2, y_2) & \cdots & f_p(x_2, y_2) \\ \vdots & \vdots & & \vdots \\ f_1(x_N, y_N) & f_1(x_N, y_N) & \cdots & f_p(x_N, y_N) \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_p \end{bmatrix}, \quad \text{or} \quad \mathbf{y}' = \mathbf{F}\mathbf{w}$$

For instance, if we propose the basis of candidate functions  $\mathcal{B} = \{1, x, y, x^2, xy, y^2\}$ , then for the dynamical system  $y' = xy + 2y - 10$ , we expect to recover the coefficient vector  $\mathbf{w} = [-10, 0, 2, 0, 1, 0]^T$ .

In general, however, we expect the data will contain some combination of model error, measurement error, and numerical error. In these cases, we will develop approaches to solve this general linear regression problem in the presence of such error.

## 5.2 PDE-Find

The PDE-Find framework as developed by [46] is used to determine the right-hand side functional form of a time-dependent PDE of the form

$$u_t = F(t, x, u, u_x, u_{xx}, \dots).$$

Suppose that for some set of basis functions  $\mathcal{B} = \{f_i\}_{i=1}^p$  we have that  $F \in \text{span}(\mathcal{B})$ . That is, there exist some coefficients  $\{w_i\}_{i=1}^p$  such that  $F = \sum_{i=1}^p f_i w_i$ . Since the dynamics must be true across the entire spatio-temporal domain, if given some list of points  $\{(t_k, x_k)\}$ , we can argue that  $\mathbf{u}_t = \mathbf{F}\mathbf{w}$ , where

$$\mathbf{u}_t = \begin{bmatrix} \vdots \\ u_t(x_k, t_k) \\ \vdots \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ f_1(x_k, t_k, u(x_k, t_k), \dots) & f_2(x_k, t_k, u(x_k, t_k), \dots) & f_3(x_k, t_k, u(x_k, t_k), \dots) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

Now, suppose we are given experimental or artificial data  $\{u_{ij}\}$  representing the model solution at the point  $(x_i, t_j)$  plus measurement error  $\varepsilon_{ij}$ , i.e.,

$$u_{ij} = u(x_i, t_j) + \varepsilon_{ij}.$$

Our goal is to recover the right-hand side dynamics  $F$  that best describe the data. Since we can not be sure the true right hand side terms, we propose a library of candidate right-hand side terms  $\{\varphi_i\}_{i=1}^q$ . We choose the library to include sufficiently many terms as to be able to capture a wide variety of dynamics. Specifically, we hope that the library is sufficiently inclusive such that  $\text{span}\{f_i\} \subset \text{span}\{\varphi_i\}$ . Given our data, we approximate the derivative data necessary to evaluate our library of candidate terms  $-(u_t)_{ij}, (u_x)_{ij}, (u_{xx})_{ij}$ , etc. – using either finite difference method or polynomial interpolation. We then evaluate each candidate term at every domain point  $(x_i, t_j)$  represented by the data, reshape the data and candidate terms into columns, and collect them into

a matrix as follows:

$$\mathbf{u}_t = \begin{bmatrix} \vdots \\ (u_t)_{ij} \\ \vdots \end{bmatrix}, \quad \Phi = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ \varphi_1(x_i, t_j, u_{ij}, \dots) & \varphi_2(x_i, t_j, u_{ij}, \dots) & \varphi_3(x_i, t_j, u_{ij}, \dots) & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

We now will look to solve the linear system

$$\mathbf{u}_t = \Phi \xi \tag{5.1}$$

where the solution vector  $\xi$  is the vector containing the right-hand side coefficients of the dynamics  $F$  with respect to the library, i.e.,  $F = \sum_{i=1}^q \varphi_i \xi_i$ . Note that, even for the vector of true coefficients  $\xi_{\text{true}}$ , we have that  $\|\mathbf{u}_t - \Phi \xi_{\text{true}}\|_2^2 = \|\tilde{\epsilon}\|_2^2 \neq 0$ . Typically, the quantity of data vastly exceeds the number of candidate terms in the library, causing (5.1) to be a drastically overdetermined system. We can initially produce a least-squares solution of the system (see Figure 5.2(a)) by making use of the Moore-Penrose pseudo-inverse:

$$\xi_{\text{OLS}}^* = \Phi^\dagger \mathbf{u}_t = (\Phi^T \Phi)^{-1} \Phi^T \mathbf{u}_t$$

However, due to measurement error and error generated in the reconstruction of derivative data, the coefficient vector  $\xi$  will contain 0 with probability zero, meaning that every candidate term in the library will be present in the recovered right-hand side dynamics  $F$ . Rather, we will pursue a sparse solution using one of several methods to ensure sparsity. This will ensure that the terms recovered are highly informative of the data and are not fitting measurement error, model error, or numerical error from derivative reconstruction. The most elementary such method is to add  $\ell_1$  regularization to the ordinary least-squares (OLS) error.

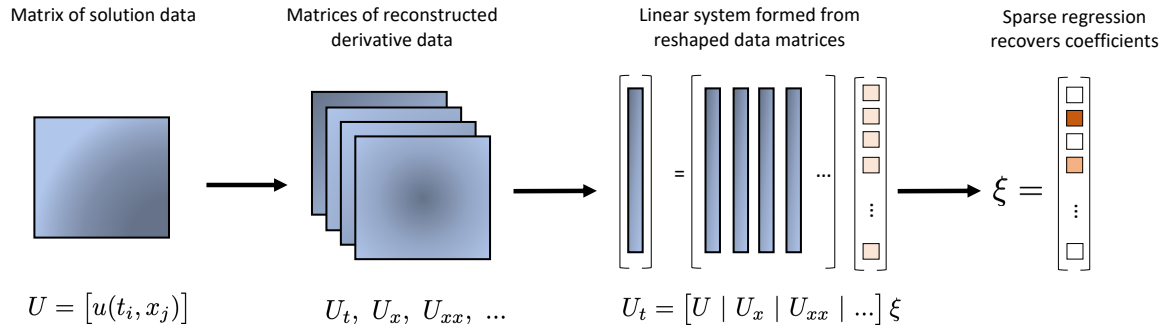
$$\underset{\xi \in \mathbb{R}^q}{\operatorname{argmin}} \|\mathbf{u}_t - \Phi \xi\|_2^2 + \lambda \|\xi\|_1$$

In addition to penalizing large coefficients, this so-called Lasso regression is naturally sparsifying since tangency between the level sets of the OLS error function and the  $\ell_1$  regularization term is more likely to occur on the coordinate axes of the parameter space (see Figure 5.2(b)). Another

option for generating sparse solutions is ridge regression with sequential thresholding, which uses  $\ell_2$  regularization, as follows:

$$\underset{\xi \in \mathbb{R}^q}{\operatorname{argmin}} \|\mathbf{u}_t - \Phi \xi\|_2^2 + \lambda \|\xi\|_2$$

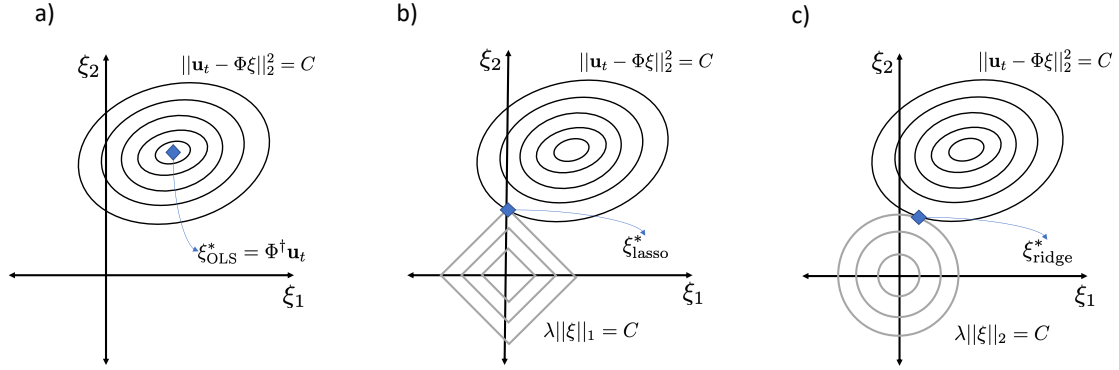
However, though ridge regression similarly discourages large coefficients, is not inherently sparsifying (see Figure 5.2(c)), meaning that in addition to regularization we include sequential thresholding step where small coefficients are clamped to 0. We refer to this as the Sequential Thresholding Ridge method (STRidge). Figure 5.1 shows a schematic of the PDE-Find algorithm.



**Figure 5.1 Schematic diagram of PDE-Find algorithm.** Solution surface data in the form of a matrix is loaded or constructed. Then, data matrices representing the desired derivative surfaces are constructed using either finite difference method, polynomial interpolation, or some other context-dependent derivative approximation scheme. Matrices are reshaped into columns and point-wise multiplication is used to construct columns representing all pre-supposed candidate terms. Columns representing candidate terms are collocated into library matrix and linear system  $\mathbf{u}_t = \Phi \xi$  is formed. Sparse regression is used to recover coefficient vector  $\xi$ . Non-zero entries of  $\xi$  represent recovered terms.

### 5.3 Learning Governing Equations for Cellular Automata Model

A key observation from our work on the cellular automata model developed in Chapter 4 was the increasing stochasticity of viral spread as IFN production increased. That is, when many independent realization of the model were aggregated, we found that the distribution of the total number of infected cells had a larger variance for higher levels of IFN production. We observed this in



**Figure 5.2** Two dimensional visualization of regression problem under no regularization,  $\ell_1$  regularization (lasso), and  $\ell_2$  regularization (ridge). (a) The ordinary least squares solution  $\xi_{OLS}^* = \Phi^\dagger \mathbf{u}_t$  is the global minimum of the convex OLS error function  $\|\mathbf{u} - \Phi\xi\|_2^2$ . (b) Due to the cusps on the level sets of the  $\ell_1$  regularization term  $\lambda\|\xi\|_1$ , the lasso solution  $\xi_{lasso}^*$  is more likely to occur on the coordinate axes, thus encouraging sparsity of the solution vector. (c) The ridge solution  $\xi_{ridge}^*$  is pushed towards the origin by the  $\ell_2$  regularization term  $\lambda\|\xi\|_2$ . This approach, coupled with sequential thresholding of small entries of  $\xi$ , can result in a sparse solution.

Figure 4.7 as the increasing characteristic width of the histograms. We argued that, as IFN signaling reduced the availability of Healthy cells, successful virion contacts become an increasingly rare event. So while the spread of infection is sustained by a high-probability event in the low-IFN case, the spread of infection is sustained by a low-probability event in the high-IFN case. We are interested in studying the relationship between IFN production and the variability of viral spread. That is, imagining the histograms of the model results shown in Figure 4.7 as approximations of a probability density function, we want to quantify the rate at which the variance of the distribution increases with time as a function of IFN production. This would quantify the effect of IFN in increasing the stochasticity of early infection.

In this section, we shall apply equation learning techniques to learn a partial differential equation that governs the time evolution of the probability distribution of the number of infected cells observed on the cellular automata lattice. Such equations are similar to the Focker-Plank equations of statistical physics which govern the time evolution of the probability distribution of a particle's momentum under Brownian motion. Our goal is to learn governing PDEs for each level of IFN response to describe the rate at which key statistical moments such as mean and variance increase



with time. Let  $u(x, t) = P(X = x | T = t)$  where  $X$  is a random variable representing the number of infected cells observed on the lattice and  $T$  is the number of time steps since the initial infected cell was born. Then we are interested in learning a PDE of the form

$$u_t = F(u, u_x, u_{xx}, \dots)$$

to describe the time evolution of the probability density function  $u$ .

### 5.3.1 Reconstruction of continuum density surface

We began by generating  $N = 10^5$  independent realization of the CA model for  $T$  time steps for all integer values of `ifn_prod` from 0 to 15. For each realization, the total number of infected cells across the entire lattice  $I_k(t)$  is computed at each time step  $t$ . We need to construct a matrix  $U$  representing the solution surface where  $U_{ij} = u(x_i, t_j)$ , and similarly for the derivative surfaces  $u_t$ ,  $u_x$ , and  $u_{xx}$ . At each time  $t_j$ , we can create a continuous reconstruction of  $u(x, t_j)$  using kernel density estimation. That is, by choosing a kernel function  $K(x) \in C^\infty$  with the properties:

1.  $K(-x) = \psi(x)$
2.  $\lim_{x \rightarrow \pm\infty} K(x) = 0$
3.  $\int_{-\infty}^{\infty} K(x) dx = 1$

we reconstruct  $u(x, t_j)$  as:

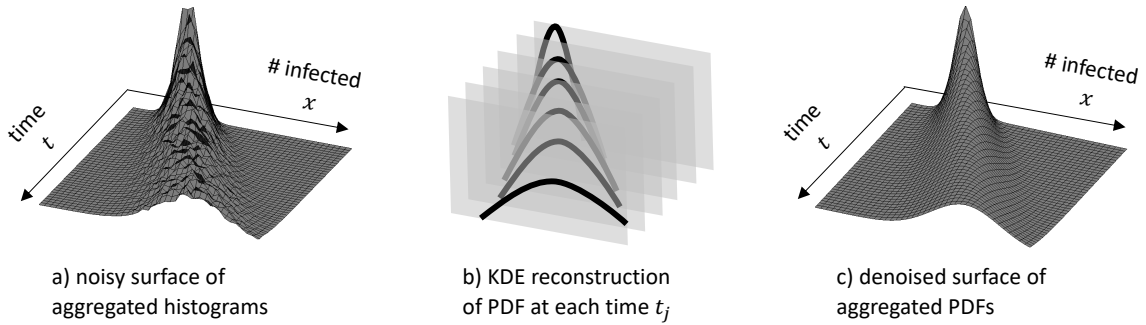
$$u(x, t_j) = \sum_{k=1}^N K(x - I_k(t_j))$$

where now the solution surface data matrix  $U$  is given as  $u(x_i, t_j)$  where  $x_i \in \left\{1, 2, \dots, \max_k I_k(T)\right\}$ . This procedure is outlined by the diagram in Figure 5.3. Furthermore, since the kernel  $K$  is differen-

tiable in  $x$ , we are able to reconstruct the derivative surfaces in the same way:

$$\begin{aligned} u_x(x, t_j) &= \sum_{k=1}^N K_x(x - I_k(t_j)) \\ u_{xx}(x, t_j) &= \sum_{k=1}^N K_{xx}(x - I_k(t_j)) \\ &\vdots \end{aligned}$$

Common choices for the kernel function  $K$  include uniform, linear, and gaussian. Here, since we are interested in recovering derivative information as well, we choose a gaussian kernel function  $K(x) = \frac{\sqrt{\pi}}{h} \exp(-(x/h)^2)$ , where the so-called “bandwidth”  $h$  controls the characteristic width of the kernel. This choice of kernel to allow for several orders of derivative information to be recovered while performing the KDE.



**Figure 5.3 Diagram of KDE solution surface reconstruction procedure.** a) Noisy surface of consisting of aggregated histograms for  $10^5$  independent realizations of CA model showing # infected cells  $x$  observed at time  $t$ . b) For each time  $t_j$ , kernel density estimation is used to reconstruct the PDF  $u(x, t_i)$ . c) The individual PDFs are re-aggregated into denoised solution surface.

### 5.3.2 Library structure

We next need to build the library of candidate terms  $\{\varphi_i\}_{i=1}^q$ , where the right hand side dynamics of the equation  $u_t = F(t, x, u, u_x, \dots)$  must be recoverable as a linear combination of our library, i.e.,  $F = \sum_{i=1}^q \xi_i \varphi_i$ . Our primary concerns are that the library be sufficiently inclusive such that many dynamics can be captured and that the recovered right hand side preserve desired mathematical properties of the solution. In our case, since we require that  $u(x, t_j)$  be a valid probability density

function at each time  $t_j$ , we must require that our equation be mass conserving:

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} u(x, t) dx = 0 \quad (5.2)$$

Consider the generic 1D conservation law

$$u_t + \phi_x = 0$$

where  $\phi = \sum_{i=1}^q \xi_i \phi_i$  is the net flux, and  $\phi_i$  are candidate fluxes. Then

$$\frac{\partial}{\partial t} \int_{-\infty}^{\infty} u(x, t) dx = \int_{-\infty}^{\infty} u_t(x, t) dx = \int_{-\infty}^{\infty} \phi_x dx = \lim_{a \rightarrow \infty} \phi(a) - \lim_{b \rightarrow -\infty} \phi(b)$$

Therefore, in order for 5.2 to hold, we must enforce that  $\phi$  vanish at  $x = \pm\infty$ . Since we want this property to hold for all possibly recoverable right hand sides, we extend this vanishing requirement to all  $\phi_i$ . Now, we may choose candidate fluxes  $\phi_i$  that meet this vanishing criterion and build our library  $\{\phi_i\}_{i=1}^q$  as  $\phi_i = (\phi_i)_x$ , therefore guaranteeing the recovered equation to conserve mass. Following this procedure, we design the following set of candidate fluxes and construct the corresponding library of mass conserving candidate terms as follows:

$$\begin{aligned} \{\phi_i\} &= \{u, u^2, uu_x, (u_x)^2, u_{xx}, uu_{xx}, u_x u_{xx}, (u_{xx})^2\} \\ \{\varphi_i\} &= \{u_x, 2uu_x, (u_x)^2 + uu_{xx}, 2u_x u_{xx}, u_{xxx}, u_x u_{xx} + u_x u_{xxx}, (u_{xx})^2 + u_x u_{xx}, 2u_{xx} u_{xxx}\} \end{aligned}$$

Note that here, since the solution data  $u$  is reconstructed as a KDE,  $u$  and its derivatives satisfy the vanishing criterion by inheriting it from the Gaussian kernel  $K(x)$ .

### 5.3.3 Results

For four different levels of IFN production – `ifn_prod=1,2,5`, and `10` – kernel density estimation was used to reconstruct the solution surface  $u$  from  $N = 10^5$  independent simulation of the CA model for 150 time steps. The PDE-Find algorithm with STRidge regression was used to recover a sparse model.

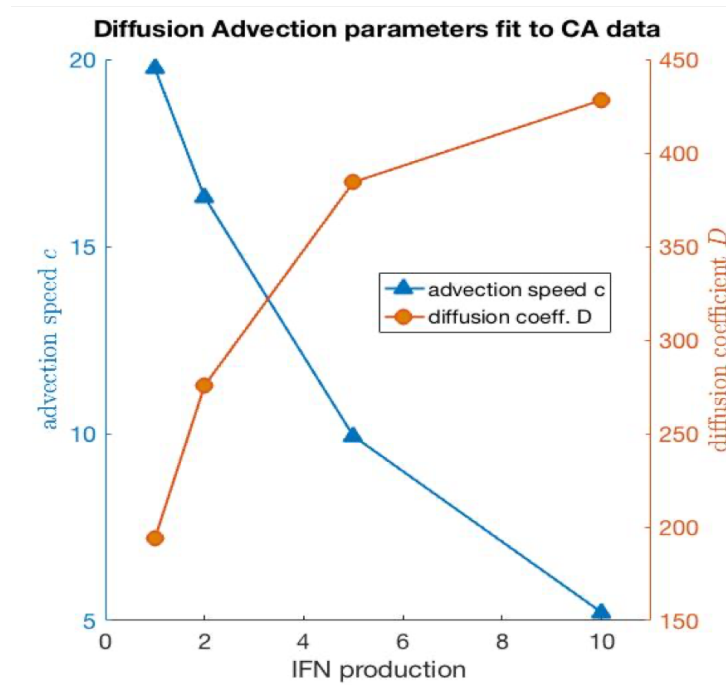
We find that when using the entire time course of 150 steps, no consistent model is learned across

the different levels of IFN response. This suggests that either our library is insufficiently inclusive or that the parametric response of the right-hand side  $F$  to changes of `ifn_prod` is non-linear. However, we find that when only a portion of the solution surface corresponding to 40 time steps is used, PDE-Find with STRidge regression consistently recovers a diffusion advection equation  $u_t = -c u_x + D u_{xx}$  across all levels of IFN production. This result is conducive to our stated goal of interpreting the impact of IFN signaling on the uncertainty of the spread of infection. In this case, since the form of the distribution is nearly Gaussian, we can interpret the recovered coefficients  $c$  and  $D$  as the rate of change of the mean and variance of the distribution, respectively. In Figure 5.4 we show the recovered coefficients  $c$  and  $D$  across all four levels of IFN production. We see that the advection rate  $c$  decreases with `ifn_prod`, since more frequent IFN signaling decreases the number of successful infection events. We also see that the diffusion coefficient  $D$  of the distribution increases with `ifn_prod`, suggesting that the variance of the distribution increases more rapidly as IFN signaling increases. This can be interpreted as IFN signaling having the effect of making the spread of infection more stochastic.

However, we find that the fit of the diffusion-advection equation breaks down with time and does not describe the time evolution of the probability distribution for large numbers of time steps. If we recall from Chapter 4, the number of infected cells on the CA lattice increases as a power law in time. However, this is inconsistent with a constant advection rate  $c$  of the distribution, which would indicate that the number of infected cells grows linearly with time. Rather, we should interpret the  $c$  we recover to represent the instantaneous growth rate of the mean of the distribution.

## 5.4 A novel equation learning framework with trainable candidate terms

As we saw in the previous section, depending on the nature of the dataset, it may be a combinatorially intractable problem to create a sufficiently broad library of candidate terms to capture the dynamics of the data. Specifically, we may find that the true right-hand side contains non-linear terms with non-integer exponents. In this section we will develop a novel framework for using a linear regression equation learning framework coupled with trainable candidate terms. Rather than creating an overly inclusive library of candidate terms, we will presuppose a non-linear parametrization for the right hand side terms and allow the parameters to be trained on the data in addition to recovering a



**Figure 5.4 Coefficients of recovered diffusion-advection equation vary with IFN production.** Coefficients  $c$  and  $D$  of diffusion-advection equation  $u_t = -c u_x + D u_{xx}$  recovered for varying levels of IFN production. The advection rate  $c$  can be interpreted as the average rate at which the number of infections increases. The diffusion coefficient  $D$  can be interpreted as the rate at which the variance of the distribution increases. The trend suggests that increasing IFN production has the expected consequence of decreasing the average rate at which the number of infections increase, but furthermore increases the rate at which the variance of the distribution increases, suggesting that the stochasticity of the infection process increases with IFN production.

sparse set of coefficients via linear regression. We will limit ourselves to the ordinary differential equation (ODE) setting, where from longitudinal time series data  $\{(t_j, y_j, \dot{y}_j)\}$  we will learn an ODE of the form:

$$\dot{y} = \sum_{i=1}^p \xi_i y^{\gamma_i}.$$

Now, rather than pre-supposing a “sufficiently inclusive” library of candidate right-hand side terms  $y^{\gamma_i}$  such as  $\{\phi\} = \{y, \sqrt{y}, \frac{1}{y}, y^{3/2}, y^2, \dots\}$ , we propose a candidate library  $\{y^{\gamma_1}, y^{\gamma_2}, y^{\gamma_3}, \dots\}$  where the vector of exponent parameters  $\gamma = [\gamma_1 \ \gamma_2 \ \gamma_3 \ \dots]^T$  will be learned from the data in tandem with the coefficient vector  $\xi$ . The details of our algorithm are discussed below.

Given a time series data  $\{(t_j, y_j, \dot{y}_j)\}$ , we will construct the linear system

$$\dot{\mathbf{y}} = \Phi(\gamma)\xi \quad \text{where} \quad \dot{\mathbf{y}} = \begin{bmatrix} \vdots \\ \dot{y}_j \\ \vdots \end{bmatrix}, \quad \Phi(\gamma) = \begin{bmatrix} \vdots & \vdots & \vdots & \vdots \\ y_j^{\gamma_1} & y_j^{\gamma_2} & y_j^{\gamma_3} & \dots \\ \vdots & \vdots & \vdots & \vdots \end{bmatrix}.$$

We will alternate between performing gradient descent to train the exponent parameters  $\gamma$  and solving the linear regression problem  $\dot{y} = \Phi(\gamma)\xi$  to update the coefficients  $\xi$ . We must be careful to update the coefficients  $\xi$  frequently enough so that the optimization landscape in  $\gamma$  does not change too drastically between subsequent applications of gradient descent. In consideration of the sum of squared error loss function

$$\mathcal{L}(\gamma) = \|\Phi(\gamma)\xi - \dot{\mathbf{y}}\|_2^2$$

we can analytically compute the gradient with respect to  $\gamma$  to be

$$\begin{aligned} \nabla_{\gamma} \mathcal{L} &= 2(\nabla_{\gamma}(\Phi\xi))^T (\Phi\xi - \dot{\mathbf{y}}) \\ &= 2\xi^T (\Phi \odot \log(\mathbf{y}))^T (\Phi\xi - \dot{\mathbf{y}}) \end{aligned}$$

where  $\odot$  is used here to represent entry-wise vector multiplication from the right applied to all columns. As gradient descent is performed, the columns of  $\Phi$  may become close to parallel if two  $\gamma_i$

approach the same value. To combat this, we will “merge” columns of  $\Phi$  when two candidate terms become too close. This will act as a natural of achieving sparsity in the learned equation by merging redundant terms.

---

**Algorithm 5.1:** Trainable candidate sparse equation learning algorithm.

---

Load data:  $\mathbf{y}, \dot{\mathbf{y}}$ . Initialize  $\gamma \in \mathbb{R}^{c \times 1}$ .

Build library  $\Phi = [y^{\gamma_1} | \dots | y^{\gamma_c}]$ .

1. Solve for least squares coefficients:  $\xi = (\Phi^T \Phi)^{-1} \Phi^T \dot{\mathbf{y}}$
  2. Update  $\gamma$  with Gradient Descent for num\_GD\_its iterations.
  3. if  $|\gamma_i - \gamma_j| < \text{merge\_thresh}$  for some  $i \neq j$ , merge columns.
  4. Remove columns with  $|\xi_i| < \text{rm\_thresh}$ .
  5. Update library.  $\Phi = [y^{\gamma_1} | \dots | y^{\gamma_c}]$
- 

The linear regression can either be made to enforce sparsity via a method such as lasso or can be taken as the less expensive least squares regression via pseudo-inverse. Since the recovered coefficients will change as  $\gamma$  is trained, we will opt to recover non-sparse  $\xi$  via pseudo-inverse and allow the library to shrink by sequentially thresholding small terms to 0 and by implementing the aforementioned merging. The complete algorithm is shown in Table 5.1.

#### 5.4.1 The Logistic Equation

We will use the example of the Logistic Equation to show the performance of our algorithm on a well understood case. We generate data  $\{t_j, y_j, \dot{y}_j\}$  for the logistic equation  $\dot{y} = r y (1 - \frac{y}{K})$  using the well-known sigmoidal analytic solution

$$y(t) = \frac{K y_0}{P y_0 + (K - y_0) e^{-r t}}$$

evaluated at all time sample points  $t_j$  and subsequently constructing the derivative data as  $\dot{y}_j = r y_j (1 - \frac{y_j}{K})$ . We can see that the logistic equation has the desired form  $\dot{y} = \sum_{i=1}^p \xi_i y^{\gamma_i}$ , where  $\gamma = [1, 2]^T$  and  $\xi = [r, -r/K]^T$ .

**Table 5.1 Table showing recovered equation for logistic data**

Parameter Set	True Equation	Recovered Equation
$r = 3, K = 100, y_0 = 1$	$\dot{y} = 3y - 0.03y^2$	$\dot{y} = 3.08y^{0.983} - 0.0236y^{2.04}$
$r = 5, K = 100, y_0 = 1$	$\dot{y} = 5y - 0.05y^2$	$\dot{y} = 5.01y^{0.998} - 0.0485y^{2.01}$
$r = 2, K = 10, y_0 = 1$	$\dot{y} = 2y - 0.2y^2$	$\dot{y} = 1.99y^{0.998} - 0.196y^{2.01}$
$r = 5, K = 10, y_0 = 0.1$	$\dot{y} = 5y - 0.5y^2$	$\dot{y} = 4.99y^{0.999} - 0.497y^{2.00}$
$r = 5, K = 10, y_0 = 2$	$\dot{y} = 5y - 0.5y^2$	$\dot{y} = 4.84y^{0.947} - 0.295y^{2.21} + 0.0036y^{3.10}$

We will use our equation learning framework to recover the logistic equation. We initialize the algorithm with five candidate terms and initialize the exponents as  $\gamma = [0.5, 1.4, 2.5, 3, 3.4]^T$ . The results are shown in Table 5.1. We see that the algorithm does a good job of removing the three unnecessary candidate terms and recovering the logistic equation for all but one case. In Figure 5.5 we show simulations of the recovered equation compared to the ground truth and show the training of the candidate terms for three of the examples shown in Table 5.1. You can see on the left column of plots that the learned equation captures the data very well. In the right column, you see the training of the candidate terms for the corresponding data from the left column. Shown in blue are the exponents  $\gamma_i$  and in red are the coefficients  $\xi_i$ , with the true values shown in dotted lines of the same color. We see that when a candidate term is removed from the library either by thresholding or merging (seen as a line that terminates) the error rises abruptly but then decreases as the candidate terms approach that of the true equation. In Figure 5.5(c), you see that the data reaches equilibrium very quickly, meaning that the data is not as information-dense as the other examples shown, perhaps explaining why an extra unnecessary term is recovered

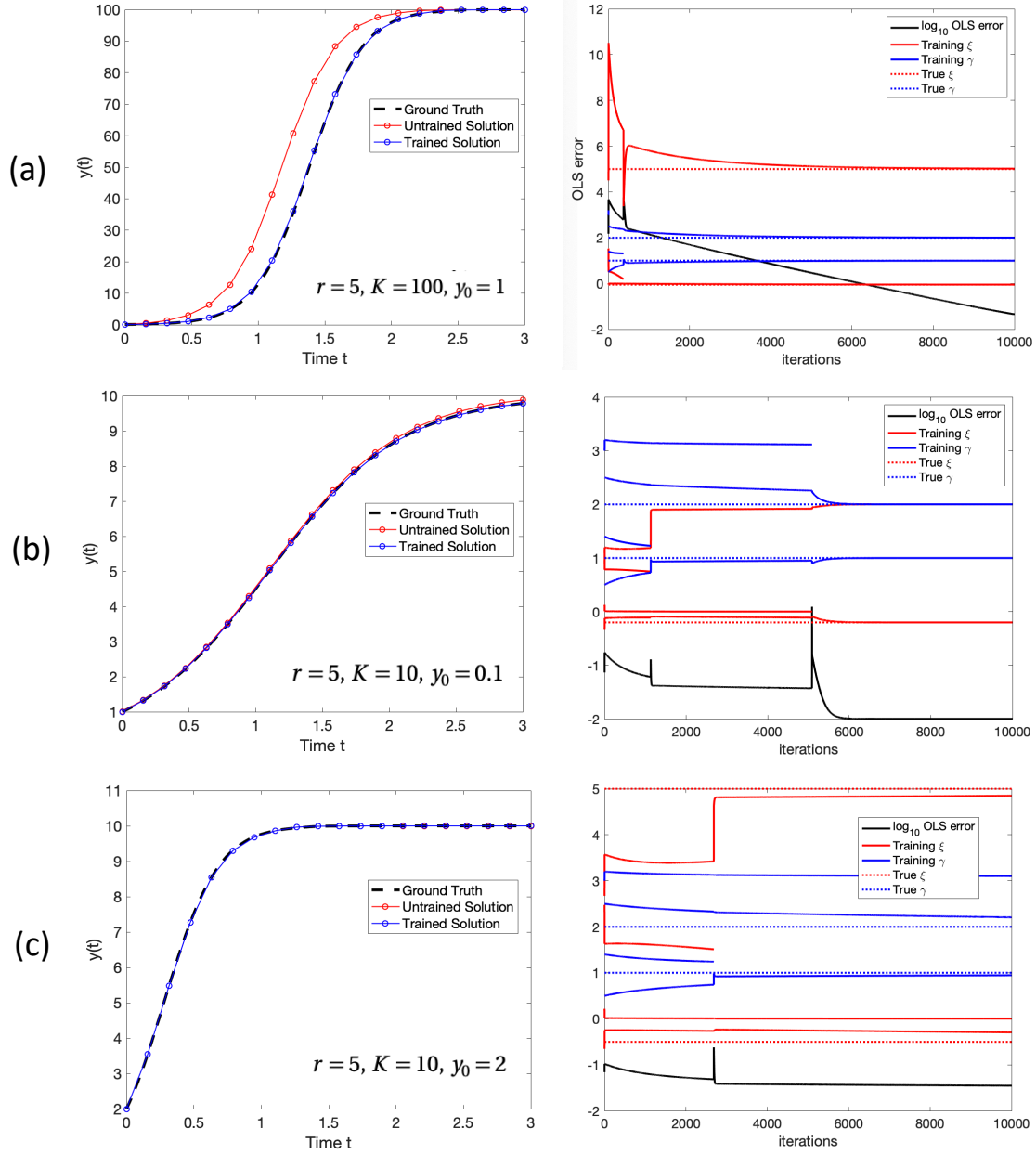
#### 5.4.2 Moment Equations

We will now use our novel equation learning framework to learn moment equations for the probability distribution  $u(x, t)$  studied in section 5.3. That is, for a given statistical moment  $\mu$  that can be computed for the distribution  $u(x, t_j)$  at some time  $t_j$ , we will construct a time series data set  $\mu_j = \mu(t_j)$  and learn from it an governing equation of the form

$$\dot{\mu} = \sum_{i=1}^p \xi_i \mu^{\gamma_i}$$

We build the data set of moment data for the mean (first raw moment) for the distribution collected





**Figure 5.5 Representative simulations of learned equations from logistic data and candidate term training.** The left column shows the ground truth data (black dash), solution of equation fit using untrained initial candidate terms (red circle), and solution of final learned equation (blue circle) for three representative logistic datasets. The right column shows the training of the candidate terms corresponding to the data on the left. Shown in blue are the exponents  $\gamma_i$  and in red are the coefficients  $\xi_i$ , with the true values shown in dotted lines of the same color. Terminated lines represent candidate terms that were removed from the library via merging or thresholding.

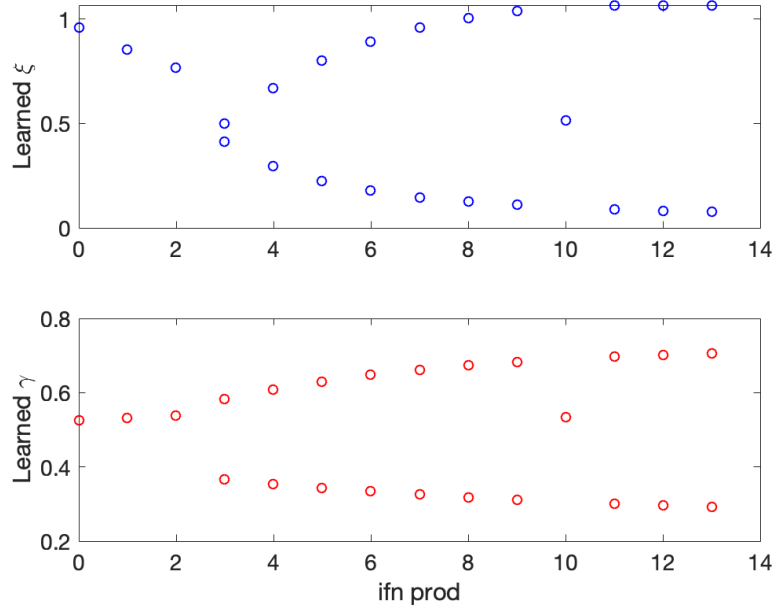
from  $10^5$  independent CA simulations with `ifn_prod` = 0, 1, 2, 3, ..., 13 with cell death. We will study the trends in the learned equations to understand the impact of IFN production on the spread of infection.

We applied our framework detailed in Table 5.1 to these moments data and report the results in Table 5.2 and Figure 5.6. The results show that only one term is recovered at low levels of IFN production, while two terms are recovered at higher levels of IFN production. This result suggests that a simple power law describes the data well at low levels of IFN production, but may break down at higher levels. We observe that a single term is recovered at `ifn_prod`=10. This is due likely to a sensitivity to hyperparameters such as thresholds and learning rate and does not indicate any special properties of this value of `ifn_prod`. Recall that the spread of infection is radial and symmetric at low levels of IFN (see Figure 4.4). This suggests that the number of infections should grow as a power law in time  $\mu(t) \propto t^\beta$  with  $\beta \approx 2$ . However, as the spread is interrupted by higher levels of IFN production and the infected cell lifespan becomes comparable to the expected waiting time to create a daughter infection, extinction becomes a possibility and that power law may fail to hold. Our equation learning results seem to support the observation that the power law resulting from radial 2D spread is valid at low IFN levels, but begins to break down as IFN causes the spread of infection to become more erratic. We furthermore note that the form and parameters of the recovered equation appear to stagnate as `ifn_prod` increases past 10. One can think of each infected cell as the center of a neighborhood of influence on the lattice the size of which depends on the diffusivity of IFN and virions. IFN acts to prevent infection by saturating this neighborhood with protected cells. One can imagine that as `ifn_prod` grows relative to the size of the neighborhood, the impact of further increases in IFN production has diminishing returns once the current level of production is already sufficient to saturate the neighborhood. This local saturation explains why the form of the learned equations is insensitive to further increases in IFN production after `ifn_prod`=11.

We attempted a similar study on the variance (second central moment) of the distributions, but found gradient descent to be causing blow-ups of the right hand side terms. We suspect the data might need to be scaled or normalized to allow for such a study, as the variance of the distribution grows very rapidly as the total number of infections easily approaches  $10^5$  within 150 time steps. Upon devising an appropriate normalization protocol or making the gradient descent step more robust to the scale of the data, we would be interested in performing this study on the variance data.

**Table 5.2 Learned moment equations from CA distribution data.** Table shows the recovered equation from CA data across 14 levels of IFN production. Trend shows that one term is recovered at low levels of IFN production, while two terms are recovered at higher levels of IFN production. This suggests that a simple power law describes the data well at low levels of IFN production, but may break down at higher levels.

ifn_prod	Recovered Equation	# Terms Recovered
0	$\dot{\mu} = 0.9577\mu^{0.52588}$	1
1	$\dot{\mu} = 0.85292\mu^{0.53176}$	1
2	$\dot{\mu} = 0.76762\mu^{0.53742}$	1
3	$\dot{\mu} = 0.49811\mu^{0.36594} + 0.41158\mu^{0.58346}$	2
4	$\dot{\mu} = 0.66911\mu^{0.35423} + 0.29769\mu^{0.60822}$	2
5	$\dot{\mu} = 0.79957\mu^{0.34368} + 0.22422\mu^{0.63}$	2
6	$\dot{\mu} = 0.89269\mu^{0.33428} + 0.17758\mu^{0.64765}$	2
7	$\dot{\mu} = 0.95902\mu^{0.3257} + 0.1465\mu^{0.66197}$	2
8	$\dot{\mu} = 1.0056\mu^{0.3179} + 0.12487\mu^{0.67366}$	2
9	$\dot{\mu} = 1.0368\mu^{0.31103} + 0.10938\mu^{0.68305}$	2
10	$\dot{\mu} = 0.51406\mu^{0.53369}$	1
11	$\dot{\mu} = 1.0645\mu^{0.30006} + 0.089287\mu^{0.6965}$	2
12	$\dot{\mu} = 1.0665\mu^{0.29585} + 0.08252\mu^{0.70118}$	2
13	$\dot{\mu} = 1.0636\mu^{0.29231} + 0.077089\mu^{0.70488}$	2



**Figure 5.6 Parameters of learned moment equations from CA distribution data.**

## 5.5 Contributions and Future Work

In this chapter we have explored the use of data-driven equation learning frameworks to draw conclusions from the cellular automata model developed in Chapter 4. We have adapted the PDE-Find framework developed in [46] to learn Fokker-Plank-type equations governing the time evolution of the probability distribution of the total number of infected cells found on the CA grid. Our novel contributions here are the use of kernel density estimation as a means of reconstructing a smooth solution surface in tandem with derivative surfaces and the implementation of a library of candidate terms that are designed to enforce that the learned equation be mass conserving.

Using PDE-Find, we were able to recover a diffusion-advection equation that describes the data for all levels of IFN production. From these equations we were able to observe that, not surprisingly, IFN has the effect of slowing the growth of the total number of infections. Furthermore, we concluded that IFN production has the effect of increasing the stochasticity of the spread of the infection. That is to say, one can be less certain of the number of infections that will occur downstream under high IFN production.

We also developed a novel equation learning framework that employs trainable candidate terms to avoid the combinatorially large problem of defining a sufficiently inclusive library of candidate terms. This allowed for a smaller number of non-linearly parametrized right-hand side terms to be trained to fit the data and to be removed if redundant. Our framework was limited to the ODE setting, but could be easily extended to a spatio-temporal setting. We demonstrated proof-of-concept by using this approach to recover the logistic equation across differing parameter regimes from an initial library of 5 nonlinearly parametrized terms that was trained and pruned until the original equation was recovered.

We lastly employed our novel equation learning framework to learn equations describing the time-evolution of the mean of the distribution of the number of infected cells. We found that the number of terms in the recovered moment equation increased with IFN production. This suggested that the power law that seems to hold for low levels of IFN production breaks down as IFN production increases.

Moving forward, we hope to expand our novel equation learning framework to the PDE setting to allow for more general equations to be learned from spatial-temporal data while taking advantage of

the computational efficiency of the sparse linear regression framework. Specifically, we are interested in learning equation that govern the phenotypic drift observed in the CA model with viral evolution shown in Figure 4.8, with the goal of studying the parametric response of such governing equations to varying levels of IFN production.

Furthermore, we are interested in expanding the Trainable Candidate EQL framework to handle libraries of a more diverse array of non-linear functions that often appear in biological models, such as Michaelis-Menten and generalized Hill functions.

## CHAPTER

# 6

## DISCUSSION

In this thesis, we have developed and analyzed a series of models of the in-vivo viral infection that seek to characterize the spatial strategies employed by the host interferon response. The IFN response is a critical element of the epithelial transmission barrier, and is a contributing factor to virus viability in the host [6, 47]. Our models have shown how autocrine and paracrine IFN signaling can act to stop an infection before it becomes systemic. Specifically, we have shown that in a spatially structured infection such as epithelial infections at the site of initial viral entry, the IFN response can halt an infection by rapidly inducing an anti-viral state in susceptible cells close to infected cells, thus inhibiting the ability of the infection to spread. This is likely one important mechanism by which IFN signaling is effective in quashing early epithelial infections. Furthermore, as we will argue below, the local targeting of cells near to areas of infection is reminiscent of the "ring vaccination" and "contact tracing" strategies in epidemiological control. In that way, despite this work focusing on in-host innate immune response, a quantitative understanding of the IFN signaling process may provide new insights for developing effective control strategies at the epidemiological scale.

Our work characterizes the different roles of autocrine and paracrine signaling pathways and

quantifies their impacts on the initiation and spread of viral infection in different host cell environments. When infection occurs in well-mixed environment where the host cells do not exist in a static spatial structure (e.g., blood), our ODE model developed in Chapter 2 predicts the role of autocrine signaling to be much more important than that of paracrine signaling in stymieing viral growth during early infection. However, when there exists a strong host-cell spatial structure where virion and IFN are active only in a neighborhood where they are produced (e.g. epithelial infections), our PDE and CA models – developed in Chapters 3 and 4, respectively – the impact of paracrine signaling in shaping the progression of spreading infection becomes remarkably strong due to its ability to act locally. Our work therefore suggests that the two signaling schemes may act complementary to one another, though one may be more or less impactful than the other based on the context of the infection.

The conclusions of our work about the importance of spatial structure in the efficacy of paracrine IFN signaling are consistent with many experimental findings. *In vivo* experimentation has shown that IFN response can arrest the spread of VSV infection in a monolayer culture of human lung epithelial cells by rapidly inducing anti-viral states in proximal cells, even after an initial delay in the production of IFN [48]. Another example of experimentally observed spatial structure in IFN response comes from chronic liver HCV infection. Chronic HCV infection is characterized by small clusters of infected cells that are continuously seeded into the liver tissue from the blood, leading to a spatially inhomogeneous distribution of infection. It is shown in [34] that single-cell analysis of fresh-frozen biopsies of HCV infected liver show that presence of ISG RNA in uninfected hepatocytes is highly correlated with close proximity to HCV-positive hepatocytes, whereas ISG RNA negative uninfected hepatocytes were significantly less likely to have infected neighbors. This reflects the tendency of IFN to act locally to combat infection by segregating infected hepatocytes from healthy tissue with a cluster of anti-viral cells [49]. It has been shown that diffusion-consumption mechanisms such as the IFN signaling modeled here exhibit spatial clustering in the body generally, resulting in the creational of highly localized cytokine niches [50]. This suggests that the study of reaction diffusion systems could have far reaching implications for the study of spatial arrangement at the tissue level more broadly.

The highly localized response to viral infection at the intracellular level can be compared to epidemiological control strategies at the population level. Especially in non-motile cell populations,

we can see the local production and diffusion of IFN as a control strategy targeted at cells that are most likely to become infected or to already be infected. This can be roughly compared to strategies such as quarantine or contact tracing, where the control strategy is only imposed on the individuals deemed to be at highest risk [51]. Just as over-production of inflammatory cytokines in the body can be deleterious to the host, overzealous epidemic control strategies on the population scale can have deleterious economic impact, which is why on both scales it is an important feature of these control strategies that they be highly localized to the most at-risk individuals. Furthermore, the spatial structure in which an insulating layer of protected cells isolated infected regions from healthy tissue can be seen as a parallel to ring vaccination or ring culling [25, 26, 24], where spatially proximal susceptible populations are removed to create a firewall against continued spatial spread of disease. We also can consider how IFN signaling parallels the behavioral modulation of individuals in response to disease, such as increased hand washing as individuals become increasingly aware of the presence of disease in their community. In [52] it is proposed that the spread of high-quality public health information can dampen the ability of an outbreak to spread by reducing the average susceptibility of the population. In that work, preventative information is given to individuals with infected contacts who in turn educate others by word of mouth, creating a firewall of cautious contacts who are unlikely to become infected, paralleling the paracrine signaling of IFN to uninfected cells near to infection. However, a natural extension of this work would include the self-awareness of infected individuals, where, in response to being diagnosed, infected individuals become less infectious as they try to isolate themselves to spare their contacts from infections. This self-regulation could be seen as a parallel to autocrine signaling, wherein an individual's infectivity drops in response to their own infection status. Altogether, we posit that the spread of disease on these macro and micro scales exhibits some meaningful self-similarity, and, though we can not assume them to be wholly analogous, the continued study of each could continue to yield insights about the other.

Biological mechanistic models such as the ODE, PDE and CA models developed in this thesis can be valuable tools for understanding the qualitative features of disease spread. However the development such mathematical models becomes impractical as the number of coupled phenomena rises or as the underlying mechanisms become unclear. In Chapter 5 we use a series of established and novel data-driven equation learning frameworks adapted from those developed in [45] and [46] to learn governing equations from aggregated data from the CA model developed in Chapter 4.



These frameworks use front-end reconstruction of derivative data and creation of a so-called library of candidate terms so that linear regression may be used to learn equations from the data. This strategy allows one to avoid the cost of repeatedly numerically solving a differential equation to fit the data and, most importantly, allows the functional form of the equation to be determined by the data. This approach allowed us to draw conclusions about the effect of IFN response on the spread of infection in the CA model, specifically how IFN response seems to increase stochasticity in the spread of infection. However, we found that the parametric response of the learned equations to IFN response was non-linear, meaning that in general no finite library of candidate terms would be sufficiently inclusive. This led us to develop a novel framework where the candidate terms themselves are learned, thus avoiding the issue of pre-defining a sufficiently inclusive library. Indeed, many canonical models in mathematical biology are based on generalizations or idealizations that do not hold up under all experimental circumstances. For example, fickian diffusion may not describe well the movements of individuals in highly crowded environments such as in cell proliferation in wound healing. Under such circumstances, non-linear diffusion terms with non-integer exponent parameters may be needed to describe experimental data [53]. In cases such as these where the learned equation does not vary only in the value of the coefficients in response to experimental conditions, uncountably many candidate terms are needed to create a “sufficiently inclusive” library. We believe that the novel learnable candidate framework proposed in this thesis is a first step towards a more general equation learning framework that will allow governing equations to be learned from experimental data.

## BIBLIOGRAPHY

- [1] Alan Mathison Turing. "The chemical basis of morphogenesis". In: *Bulletin of mathematical biology* 52.1-2 (1990), pp. 153–197.
- [2] Amanda J Lee and Ali A Ashkar. "The dual nature of type I and type II interferons". In: *Frontiers in immunology* 9 (2018).
- [3] Juan Ayllon and Adolfo Garcia-Sastre. "The NS1 protein: a multitasking virulence factor". In: *Influenza Pathogenesis and Control-Volume II*. Springer, 2014, pp. 73–107.
- [4] Gijs A Versteeg and Adolfo Garcia-Sastre. "Viral tricks to grid-lock the type I interferon system". In: *Current opinion in microbiology* 13.4 (2010), pp. 508–516.
- [5] Dia C Beachboard and Stacy M Horner. "Innate immune evasion strategies of DNA and RNA viruses". In: *Current opinion in microbiology* 32 (2016), pp. 113–119.
- [6] Shilpa S Iyer et al. "Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness". In: *Proceedings of the National Academy of Sciences* 114.4 (2017), E590–E599.
- [7] Volker Fensterl, Saurabh Chattopadhyay, and Ganes C Sen. "No love lost between viruses and interferons". In: *Annual review of virology* 2 (2015), pp. 549–572.
- [8] Bin Jia et al. "Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2". In: *PLoS pathogens* 5.5 (2009), e1000429.
- [9] Joseph Ashour et al. "Mouse STAT2 restricts early dengue virus replication". In: *Cell host & microbe* 8.5 (2010), pp. 410–421.
- [10] Ningguo Feng et al. "Roles of VP4 and NSP1 in determining the distinctive replication capacities of simian rotavirus RRV and bovine rotavirus UK in the mouse biliary tract". In: *Journal of virology* 85.6 (2011), pp. 2686–2694.
- [11] Masaharu Iwasaki and Yusuke Yanagi. "Expression of the Sendai (murine parainfluenza) virus C protein alleviates restriction of measles virus growth in mouse cells". In: *Proceedings of the National Academy of Sciences* 108.37 (2011), pp. 15384–15389.
- [12] Haripriya Sridharan, Chen Zhao, and Robert M Krug. "Species Specificity of the NS1 Protein of Influenza B Virus NS1 Binds only human and non-human primate ubiquitin-like isg15 proteins". In: *Journal of Biological Chemistry* 285.11 (2010), pp. 7852–7856.
- [13] Lionel B Ivashkiv and Laura T Donlin. "Regulation of type I interferon responses". In: *Nature reviews Immunology* 14.1 (2014), p. 36.
- [14] Sandra N Lester and Kui Li. "Toll-like receptors in antiviral innate immunity". In: *Journal of molecular biology* 426.6 (2014), pp. 1246–1264.
- [15] Stefanie Luecke and Søren R Paludan. "Molecular requirements for sensing of intracellular microbial nucleic acids by the innate immune system". In: *Cytokine* 98 (2017), pp. 4–14.
- [16] Christopher T Jones et al. "Real-time imaging of hepatitis C virus infection using a fluorescent cell-based reporter system". In: *Nature biotechnology* 28.2 (2010), p. 167.
- [17] Vincent Madelain et al. "Ebola viral dynamics in nonhuman primates provides insights into virus immuno-pathogenesis and antiviral strategies". In: *Nature communications* 9.1 (2018), p. 4013.

- [18] Prasith Baccam et al. “Kinetics of influenza A virus infection in humans”. In: *Journal of virology* 80.15 (2006), pp. 7590–7599.
- [19] Roberto A Saenz et al. “Dynamics of influenza virus infection and pathology”. In: *Journal of virology* 84.8 (2010), pp. 3974–3983.
- [20] Kasia A Pawelek et al. “Modeling within-host dynamics of influenza virus infection including immune responses”. In: *PLoS computational biology* 8.6 (2012), e1002588.
- [21] Molly Gallagher et al. “Causes and Consequences of Spatial Within-Host Viral Spread”. In: *Viruses* 10.11 (2018), p. 627.
- [22] Ashley Sobel Leonard et al. “Transmission bottleneck size estimation from pathogen deep-sequencing data, with an application to human influenza A virus”. In: *Journal of virology* 91.14 (2017), e00171–17.
- [23] Alexey E Masharsky et al. “A substantial transmission bottleneck among newly and recently HIV-1-infected injection drug users in St Petersburg, Russia”. In: *The Journal of infectious diseases* 201.11 (2010), pp. 1697–1702.
- [24] Edgar Diaz et al. “Ring vaccination as a control strategy for foot-and-mouth disease”. In: *2005 MTBI Summer Program* (2005).
- [25] Michael J Tildesley et al. “Impact of spatial clustering on disease transmission and optimal control”. In: *Proceedings of the National Academy of Sciences* 107.3 (2010), pp. 1041–1046.
- [26] Michael J Tildesley et al. “The role of pre-emptive culling in the control of foot-and-mouth disease”. In: *Proceedings of the Royal Society B: Biological Sciences* 276.1671 (2009), pp. 3239–3248.
- [27] Tom J Howat et al. “Modelling dynamics of the type I interferon response to in vitro viral infection”. In: *Journal of The Royal Society Interface* 3.10 (2006), pp. 699–709.
- [28] Odo Diekmann, JAP Heesterbeek, and Michael G Roberts. “The construction of next-generation matrices for compartmental epidemic models”. In: *Journal of the Royal Society Interface* 7.47 (2009), pp. 873–885.
- [29] Martin A Nowak et al. “Viral dynamics of primary viremia and antiretroviral therapy in simian immunodeficiency virus infection.” In: *Journal of virology* 71.10 (1997), pp. 7518–7525.
- [30] Ulfert Rand et al. “Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response”. In: *Molecular systems biology* 8.1 (2012).
- [31] Bianca Schmid et al. “Live cell analysis and mathematical modeling identify determinants of attenuation of dengue virus 2’-O-Methylation mutant”. In: *PLoS pathogens* 11.12 (2015), e1005345.
- [32] Soheil Rastgou Talemi and Thomas Höfer. “Antiviral interferon response at single-cell resolution”. In: *Immunological reviews* 285.1 (2018), pp. 72–80.
- [33] Mikhail N Matrosovich et al. “Human and avian influenza viruses target different cell types in cultures of human airway epithelium”. In: *Proceedings of the National Academy of Sciences* 101.13 (2004), pp. 4620–4624.
- [34] Stefan Wieland et al. “Simultaneous detection of hepatitis C virus and interferon stimulated gene expression in infected human liver”. In: *Hepatology* 59.6 (2014), pp. 2121–2130.
- [35] Charles A Janeway et al. *Immunobiology: the immune system in health and disease*. Vol. 7. Current Biology London, 1996.

- [36] Grégoire Altan-Bonnet and Ratnadeep Mukherjee. “Cytokine-mediated communication: a quantitative appraisal of immune complexity”. In: *Nature reviews Immunology* 19.4 (2019), pp. 205–217.
- [37] Philip Rosen and Dayton S Misfeldt. “Cell density determines epithelial migration in culture”. In: *Proceedings of the National Academy of Sciences* 77.8 (1980), pp. 4760–4763.
- [38] Ronald Aylmer Fisher. “The wave of advance of advantageous genes”. In: *Annals of eugenics* 7.4 (1937), pp. 355–369.
- [39] Tomás Alarcón, Helen M Byrne, and Philip K Maini. “A cellular automaton model for tumour growth in inhomogeneous environment”. In: *Journal of theoretical biology* 225.2 (2003), pp. 257–274.
- [40] Catherine Beauchemin, John Samuel, and Jack Tuszynski. “A simple cellular automaton model for influenza A viral infections”. In: *Journal of theoretical biology* 232.2 (2005), pp. 223–234.
- [41] Catherine Beauchemin. “Probing the effects of the well-mixed assumption on viral infection dynamics”. In: *Journal of theoretical biology* 242.2 (2006), pp. 464–477.
- [42] Ioannis G Kevrekidis et al. “Equation-free, coarse-grained multiscale computation: Enabling microscopic simulators to perform system-level analysis”. In: *Communications in Mathematical Sciences* 1.4 (2003), pp. 715–762.
- [43] R Gonzalez-Garcia, R Rico-Martinez, and IG Kevrekidis. “Identification of distributed parameter systems: A neural net based approach”. In: *Computers & chemical engineering* 22 (1998), S965–S968.
- [44] Michael Schmidt and Hod Lipson. “Distilling free-form natural laws from experimental data”. In: *science* 324.5923 (2009), pp. 81–85.
- [45] Steven L Brunton, Joshua L Proctor, and J Nathan Kutz. “Discovering governing equations from data by sparse identification of nonlinear dynamical systems”. In: *Proceedings of the national academy of sciences* 113.15 (2016), pp. 3932–3937.
- [46] Samuel H Rudy et al. “Data-driven discovery of partial differential equations”. In: *Science Advances* 3.4 (2017), e1602614.
- [47] Brian R Wasik et al. “Generalized selection to overcome innate immunity selects for host breadth in an RNA virus”. In: *Evolution* 70.2 (2016), pp. 270–281.
- [48] Emily A Voigt, Adam Swick, and John Yin. “Rapid induction and persistence of paracrine-induced cellular antiviral states arrest viral infection spread in A549 cells”. In: *Virology* 496 (2016), pp. 59–66.
- [49] Frederik Graw et al. “Inferring viral dynamics in chronically HCV infected patients from the spatial distribution of infected hepatocytes”. In: *PLoS computational biology* 10.11 (2014), e1003934.
- [50] Alon Oyler-Yaniv et al. “A tunable diffusion-consumption mechanism of cytokine propagation enables plasticity in cell-to-cell communication in the immune system”. In: *Immunity* 46.4 (2017), pp. 609–620.
- [51] Ken TD Eames and Matt J Keeling. “Contact tracing and disease control”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 270.1533 (2003), pp. 2565–2571.
- [52] Sebastian Funk et al. “The spread of awareness and its impact on epidemic outbreaks”. In: *Proceedings of the National Academy of Sciences* 106.16 (2009), pp. 6872–6877.

- [53] Jonathan A Sherratt and James Dickson Murray. “Models of epidermal wound healing”. In: *Proceedings of the Royal Society of London. Series B: Biological Sciences* 241.1300 (1990), pp. 29–36.

## ACRONYMS

**CA** Cellular Automata

**HCV** Hepatitis C Virus

**IAV** Influenza A Virus

**IFN** Interferon

**ISG** Interferon Stimulated Gene

**NGM** Next Generation Matrix

**ODE** Ordinary Differential Equation

**PDE** Partial Differential Equation

**VSV** Vesicular Stomatitis Virus

## GLOSSARY

**autocrine signal** a signal that is produced and received by the same cell

**epithelium** the thin tissue forming the outer layer of a body's surface and lining the alimentary canal and other hollow structures, typically consisting of a cell monolayer with lattice structure

**paracrine signal** a signal that is produced by one cell and received by another

**species tropism** phenotypic shift in response to a selection pressure

**virion** the complete, infective form of a virus outside a host cell, with a core of RNA or DNA and a capsid

**quasi-equilibrium** An assumption on an ODE compartment model under which “fast” compartments are taken to always be at equilibrium value given the current values of “slow” compartments