



Accounting for dimensional differences in stochastic domain invasion with applications to precancerous cell removal



Thomas E. Woolley^{a,*}, William Hill^{b,c}, Catherine Hogan^c

^a Cardiff School of Mathematics Cardiff University Senghennydd Road, Cardiff CF24 4AG, United Kingdom

^b Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London NW1 1AT, UK

^c European Cancer Stem Cell Research Institute, School of Biosciences, Cardiff University, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK

ARTICLE INFO

Article history:

Received 7 October 2021

Revised 6 January 2022

Accepted 10 January 2022

Available online 30 January 2022

Keywords:

Elimination

Eden model

Juxtacrine

Pancreatic cancer

Agent-based modelling

Stochastic system

Comparing deterministic and stochastic

modelling

Fisher wave

Wave speed

Data fitting

Homeostasis

ABSTRACT

We consider a specific form of domain invasion that is an abstraction of pancreatic tissue eliminating precancerous mutant cells through juxtacrine signalling. The model is explored discretely, continuously, stochastically and deterministically, highlighting unforeseen nonlinear dependencies on the dimension of the solution domain. Specifically, stochastically simulated populations invade with a dimension dependent wave speed that can be over twice as fast as their deterministic analogues. Although the wave speed can be analytically derived in the cases of small domains, the probabilistic state space grows exponentially and, thus, we use numeric simulation and curve fitting to predict limiting dynamics.

© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

All cells in a tissue compete for survival (Alberts et al. (2004)). Normally, this competition is regulated by homeostasis (Brown et al. (2017); Qiu et al. (2019); Picco et al. (2019)), however, as new cells develop and divide from old cells mutations can start to appear due to incorrect replication of DNA (Boland and Ricciardiello (1999)). These mutations can lead to cancerous cells and change the balance of competition and other processes between the mutant cells and the surrounding healthy tissue Hogan et al. (2009); Hill et al. (2021); Kajita et al. (2010); Porazinski et al. (2016); Vishwakarma and Piddini (2020).

Although huge numbers of potential incorrect mutations occur throughout life, cancer is a relatively rare event (Boland and Ricciardiello (1999); Brown et al. (2017); Vishwakarma and Piddini (2020)). This rarity does not appear to stem from the possibility that cells replicate with high fidelity, leading to few mutations occurring, rather epithelial tissues possess the ability to

'sense' defective cells harbouring genetic mutations and, via distinct processes, eliminate mutant cells and prevent disease (Pray (2008)).

Our focus is on modelling and investigating a mechanism of boundary-based cell elimination that is consistent with how pancreatic tissue repairs itself upon being seeded with mutant cells that express oncogenic Kras proteins. Kras proteins control many cellular functions including cell proliferation and cell survival; mutations in KRAS are key driver mutations in pancreatic cancer tumour formation (Hill et al. (2021)).

We recently described cell competition in adult mouse pancreas tissues in vivo (Hill et al. (2021)). Using red fluorescent protein (RFP) reporters we traced cells in tissues over time showing that the amount of RFP in tissues significantly decreases when mutant cells are present in tissues in low numbers (see Fig. 1). In contrast, RFP levels remain constant in wild type tissues over the same time course, indicating that Kras mutant cells are not lost via normal tissue turnover/homeostasis. We conclude that Kras mutant cells compete with normal cells for space and survival in epithelial tissues.

* Corresponding author.

E-mail address: woolleyt1@cardiff.ac.uk (T.E. Woolley).

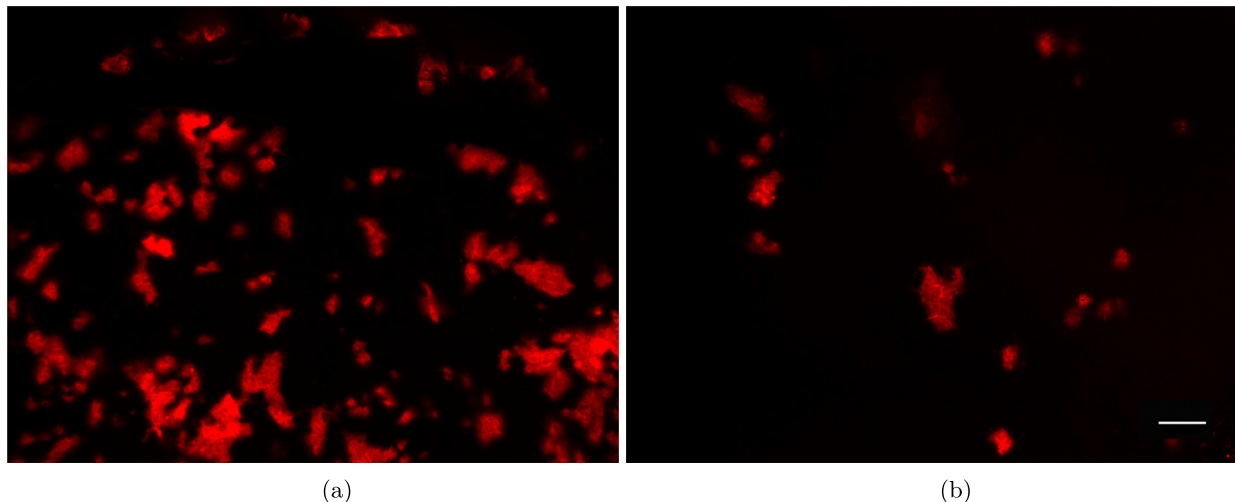


Fig. 1. Red fluorescent protein (RFP) highlighting genomic DNA in mouse pancreas tissue labelling recombined cells in the mouse pancreas. The images are taken at 35 days post-induction (a) represents healthy tissue, whilst (b) represents mutated tissue. There is a clear reduction in the amount of RFP present in the mutated tissue indicating that the healthy cells are eradicating the Kras mutant cells. The scale bar represents $100\mu\text{m}$. See (Hill et al. (2021)) for further experimental details.

The mutant patches were not seen to contain any internal elimination, only ragged boundaries, which suggests there is no long range signalling. Further, research has identified an evolutionarily conserved role of EphA2 receptor tyrosine kinase as an essential regulator of mutant cell elimination (Hill et al. (2021); Porazinski et al. (2016); Hill and Hogan (2019)). In general, EphA2 is activated by membrane-bound ligands, leading to bidirectional cell–cell signalling. Specifically, we observed that cell–cell adhesions at normal–mutant cell–cell boundaries decrease prior to mutant cell loss. Indeed, if we knock-out EphA2 receptor tyrosine kinase in tissues, Kras mutant cells are retained. Although, additional research is required to determine the molecular mechanisms of EphA2 signalling in vivo, we have in vitro data to show that direct cell–cell interactions between normal and mutant cells are required to activate EphA2 signalling and promote mutant cell elimination (Hill et al. (2021); Porazinski et al. (2016); Hill and Hogan (2019)). This data justifies the role of local juxtacrine signalling and that normal cells mobilise to eliminate mutant cells from tissues, occupying space created as mutant cells are eliminated.

In simple epithelial models, normal cells mobilise via cell migration (Porazinski et al. (2016)); however, cell translocation is unlikely in adult pancreas tissues, rather, we show that normal cells, directly adjacent to mutant cells, increase in cell volume, suggesting normal neighbours expand in cell size. A similar phenomenon is described in post mitotic tissues in *Drosophila melanogaster* (Tamori and Deng (2013)).

To model this invasion by elimination we will use an Eden-type process (Eden (1961); Wolf and Kertész (1987); Kertész and Wolf (1988)). Specifically, whenever a wild-type cell and a mutant cell are next to each other the wild-type cell is able to expel the mutant through growing into the space that was previously inhabited by the mutant cell (see Fig. 2). We note that many invasion and elimination models exist with various levels of analytical understanding (Debabrata (2004)). However, most other models include random walk space jumps, as well as interactions (Callaghan et al. (2006); Clifford and Sudbury (1973); Woolley et al. (2014)), whereas we only consider elimination as the means of invasion, because the pancreas cells were not observed to move in such a way.

Our interest in modelling this system is to extract the invasion time scale of healthy cells eliminating mutant cells. Specifically, from modelling this dynamic stochastically we observe that the

averaged system generates a Fisher-like wave of wild type cells expelling mutant cells. Moreover, in one dimension, we are able to derive a continuous differential equation that provides a macroscopic mean-field limit of the averaged stochastic simulations, which is tractable to standard travelling wave analysis. Thus, we are able to derive an algebraic correction to account for the scaling between the deterministic and stochastic invasion rates.

However, extending these ideas to two-dimensions is not as simple as the standard Fisher wave analysis (Belmonte-Beitia et al. (2013)). Specifically, although the stochastic wave front still appears to travel with a constant speed and with an approximately fixed shape the stochastic invasion waves travel faster than the analogous deterministic model and we are unable to derive a similar scaling. Critically, the wave speed can still be extracted through numerical fitting. Hence, we investigate this system of dynamics to understand how the second dimension influences the invasion wave speed in the stochastic simulations. In turn, understanding this relationship will enable us to derive a scaling term that will allow us to correct for the differences between the stochastic and deterministic limits.

It should be noted that this process can be viewed through alternative frameworks, such as Moran processes (Moran (1958); Whigham and Dick (2008)), voter models (Holley and Liggett (1975); Bramson and Griffeath (1980); Klein et al. (2008)), birth-death processes (Jörg et al. (2021); Klein et al. (2007)) and rooted-tree approximations (Hall and Siebert (2021)). Our interest in wave speed would then be interpreted as an extinction time scale and there are many techniques within these fields to extract such time scales. Although these other formalisms may be just as applicable, the authors' skills align with using an Eden formulation (Eden (1961)), which provides an interpretable basis for the cellular agents and offers clarity when comparing the stochastic equations to their deterministic analogues.

This paper is purposefully written to provide a clear introduction to the diverse modelling techniques of stochastic and deterministic modelling, theory and simulation. Each approach allows us to view the problem in a different way and by combining all of the strengths we can provide a more complete understanding of the underlying dynamics and the biological application. Indeed, this work offers a pedagogical demonstration of how different techniques and fields should be considered complementary, rather than conflicting. However, none of the frameworks are new and we

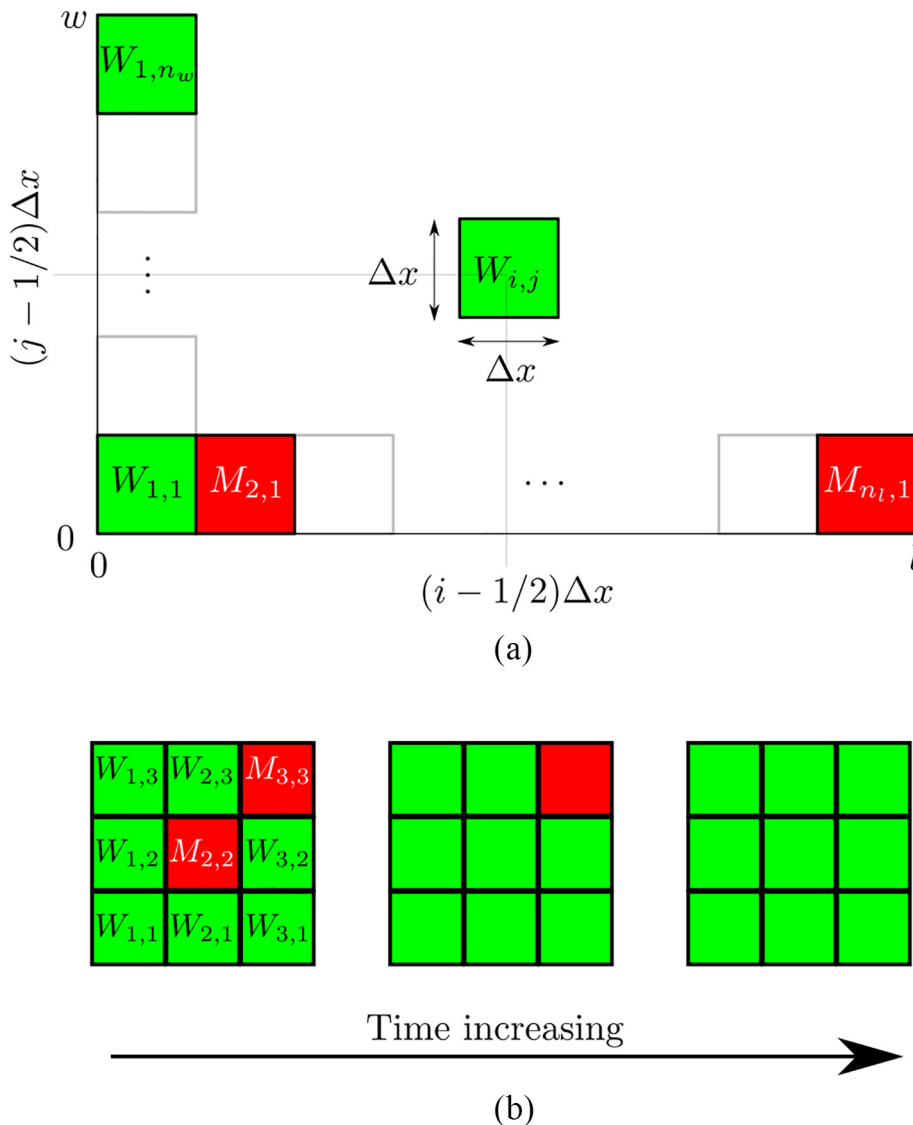


Fig. 2. Schematic diagrams illustrating (a) the domain geometry and (b) the interaction dynamics. The colour and W and M labels indicate where each population is present. Namely, if square (i,j) is green then $W_{ij} = 1$ (and $M_{ij} = 0$), whereas if the square is red then $M_{ij} = 1$ (and $W_{ij} = 0$).

suggest (Erban et al. (2007); van Kampen (2007); Murray (2003a)) as a more complete introduction to the field of linking deterministic and stochastic systems.

Due to the diverse audience that the work may interest we offer a roadmap for the rest of the article, highlighting the sections that new users may want to revise, whilst allowing more advanced users to easily target the information that they require. Firstly, we recommend everyone read Section 2, as we present our theoretical and numerical framework of modelling the invasion dynamic. Namely, we approach this problem using spatially discrete stochastic simulations and analysis to match the microscopic features of the experimental problem of discrete cell elimination. Further, by using various scaling laws, we are able to derive deterministic continuum equations that match the averaged stochastic simulations.

We then suggest that the more novice and advanced readers separate in Section 3. Specifically sections 3.1 and 3.2 are aimed at new users of stochastic and deterministic techniques as they develop our intuition regarding the invasion wave's dependence on dimension by first starting with one dimension, where the

problem is tractable and can be solved explicitly. This is in comparison to when we extend to higher dimensions where we depend more on simulation and curve fitting to understand the limiting dynamics of large stochastic systems. The more advanced reader may want to skip directly to Section 3.2.3, where the results of these sections are collected with those of large two-dimensional grids.

Finally, all readers may be interested in Section 4, where we apply our model to data from Kras mutant expulsion experiments. We demonstrate that not only is the mechanism consistent with the data but we are also able to parameterise the rate at which healthy and mutant cells interact. All codes and data are available at https://github.com/ThomasEWoolley/domain_invasion.

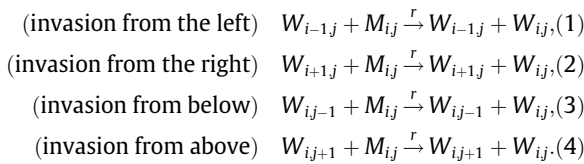
2. Model

We consider a rectangular domain made up of a discretised grid of $n_l \times n_w$ ($n_l, n_w \in \mathbb{N}$) squares of side width Δx . The length and width of the domain are then $l = n_l \Delta x$ and $w = n_w \Delta x$, respectively.

Each grid site is centred at $((i - 1/2)\Delta x, (j - 1/2)\Delta x)$ for $1 \leq i \leq n_l$ and $1 \leq j \leq n_w$ (see Fig. 2a).

These grid sites contain the cellular populations and, thus, each square can contain either a wild-type cell, W_{ij} , or a mutant cell, M_{ij} , but not both. Further, because initially no grid square is empty then no grid square can ever be empty, because this would indicate a hole, or tear, in the tissue. Thus, all grid sites contain exactly one of the populations.

We assume that the populations interact stochastically through a simple competition dynamic. Namely, whenever a wild-type cell orthogonally borders a mutant cell then at a rate r the wild-type cell will expel the mutant cell from its grid site and replace it with a new wild-type cell that is birthed in its place (see Fig. 2b). For $1 < i < n_l$ and $1 < j < n_w$ this interaction can be written as



To close the system fully we must also define what happens to the cells on the boundary, as well as an initial condition.

Multiple types of interactions could be defined on the boundary. The simplest being interactions only occur between squares that are physically next to each other, thus, making the boundary a zero-flux (or homogeneous Neumann) condition. Alternatively, we could consider periodic boundary conditions on one, or more, sets of opposite boundaries.

An illustrative example of the invasion dynamics can be seen in Fig. 2b. Specifically, we see a 3×3 grid with wild-type cells occupying all spaces except the centre and top-right square. If we were to consider the zero-flux boundary conditions as specified above then probabilistically the centre square is more likely to be expelled first as it has four neighbours, whereas the top-right square only has two. However, if opposite boundaries are identified under periodic boundary conditions then both mutant cells would have four wild-type neighbours that could invade them.

Here, to allow for ease of analysis, we apply periodic boundary conditions to the top and bottom boundaries only such that populations with indices $(i, 1)$ interact with populations with indices (i, n_w) in exactly the same ways as defined in Eqs. (3) and (4), *mutatis mutandis*. However we briefly consider zero-flux conditions on all boundaries in Section 3.3.

2.1. Simulation

2.1.1. Propensities and stoichiometric matrices

We intend to approach this problem both deterministically and stochastically. Deterministically, we will see that we can either simulate a large network of coupled ordinary differential equations (ODEs), or a single partial differential equation (PDE). These differential equations will be simulated using either `ode45`, or `pdepe` from Matlab 2019B, as required (Coleman (2013); Stanoyevitch (2005)).

To simulate the system stochastically we use a standard Stochastic Simulation Algorithm as introduced by (Gillespie (1976); Gillespie (1977); Gillespie (2007); Woolley et al. (2011); Woolley et al. (2011); Abdennur (2022)). This uses two components: the stoichiometric matrix, \mathbf{v} , whose i^{th} row informs us as to how the populations are updated once the i^{th} reaction occurs; and the propensity vector, $\boldsymbol{\eta}$, whose i^{th} element is the rate at which the i^{th} reaction occurs.

Firstly, we define a one-dimensional population vector, \mathbf{W} , that flattens the two-dimensional space of populations, namely, the wild-type population is specified by

$$\begin{aligned} \mathbf{W}^T &= (W_{1,1}, W_{1,2}, \dots, W_{1,n_w}, W_{2,1}, W_{2,2}, \dots, W_{2,n_w}, \dots, W_{n_l,n_w}), \\ &= (W_1, W_2, \dots, W_{n_l n_w}), \end{aligned}$$

where we introduce W_i as a one-dimensional indexed population, which is equivalent to $W_{\lceil i/n_w \rceil, i \bmod n_w}$, where $i \bmod n_w \in \{1, 2, \dots, n_w\}$ and $\lceil x \rceil$ is the ceiling function of x , namely, the largest integer that is larger than, or equal to x (see Fig. 3). Equally, W_{ij} maps to $W_{(i-1)n_w+j}$. Finally, the mutant population vector would be defined by $\mathbf{M} = \mathbf{1} - \mathbf{W}$, as specified by the conservation relation.

Next we define the stoichiometric vectors, which represent the change in population that happens whenever one of reactions (1)–(4) occurs. Firstly, because of the periodic boundary conditions on the top and bottom, every grid space has a neighbour above and below them. Hence, there is always the possibility of the grid cell changing from $W_i = 0$ to $W_i = 1$ (depending on its current state and neighbours). Thus, we define the above, \mathbf{v}_{Ai} , and below, \mathbf{v}_{Bi} , invasion stoichiometric vectors to be $\mathbf{v}_{Ai} = \mathbf{1}_i = \mathbf{v}_{Bi}$, $1 \leq i \leq n_l n_w$ where $\mathbf{1}_i$ is a vector of size $n_l n_w \times 1$ which is full of zeros except in the i^{th} position where there is a one. Explicitly, because every grid cell can be invaded from above, or below, the i^{th} above, or below, reaction occurs due to a mutant cell in the i^{th} grid square being displaced by a wild-type cell from above, or below, invading its space, thus, increasing the

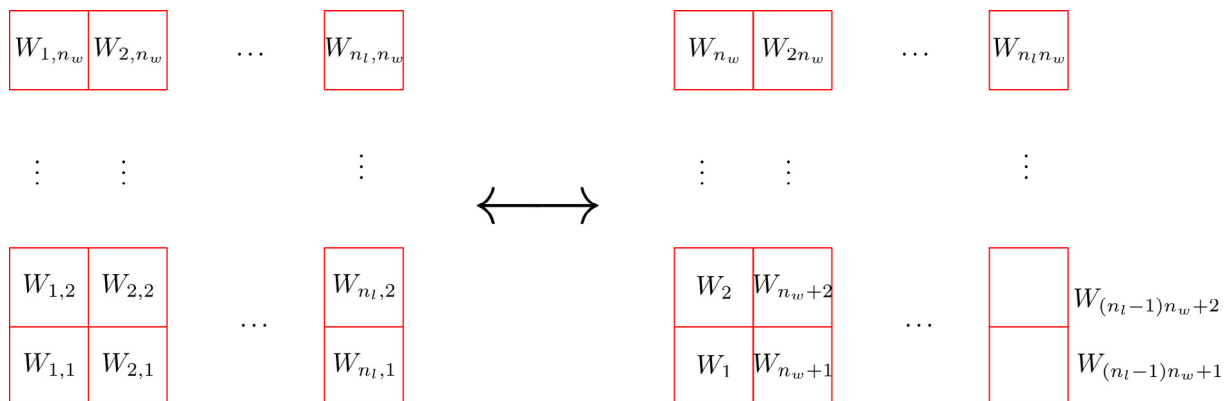


Fig. 3. Mapping between the (right) one- and (left) two-dimensional indexing structure of the grid sites.

wild-type population by one in the i^{th} grid square. The stoichiometric matrices, \mathbf{v}_A and \mathbf{v}_B , of these two interactions are constructed as a matrix whose rows are the \mathbf{v}_{A_i} and \mathbf{v}_{B_i} vectors, respectively. Thus, $\mathbf{v}_A = \mathbf{I} = \mathbf{v}_B$, where \mathbf{I} is the identity matrix of size $n_i n_w \times n_i n_w$.

The stoichiometric matrices for the left, \mathbf{v}_L , and the right, \mathbf{v}_R , invasion actions are similarly defined. However, we have to account for the zero-flux conditions on the left and right boundaries. Namely, there is no invasion from the left in the first column and, equally, there is no invasion from the right in the last column. Thus,

$$\mathbf{v}_L = \mathbf{I}_L = \left. \begin{array}{c} \begin{array}{cc} n_w & n_w(n_i - 1) \\ \begin{array}{cccccc} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{array} \end{array} \right\} n_w(n_i - 1) \quad (5)$$

where \mathbf{I}_L a matrix of size $n_i n_w \times n_w(n_i - 1)$ that can be generated by removing the first n_w rows from \mathbf{I} since the invasion from the left does not occur for the first n_w elements of \mathbf{W} . Similarly,

$$\mathbf{v}_R = \mathbf{I}_R = \left. \begin{array}{c} \begin{array}{cc} n_w(n_i - 1) & n_w \\ \begin{array}{cccccc} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{array} \end{array} \right\} n_w(n_i - 1) \quad (6)$$

where \mathbf{I}_R a matrix of size $n_i n_w \times n_w(n_i - 1)$ that can be generated by removing the last n_w rows from \mathbf{I} since the invasion from the right does not occur for the last n_w elements of \mathbf{W} .

Next, we need to define the propensity vectors for each reaction. The probabilistic reaction rate for each interaction is derived by applying the Law of Mass Action (Guldberg and Waage (1879)) to Eqs. (1)–(4). The above and below invasion propensity vectors, $\boldsymbol{\eta}_A$ and $\boldsymbol{\eta}_B$, respectively, are once again simple to state, the i^{th} elements are $[\boldsymbol{\eta}_A]_i = rW_{i+1 \bmod n_w n_i}(1 - W_i)$ and $[\boldsymbol{\eta}_B]_i = rW_{i-1 \bmod n_w n_i}(1 - W_i)$.

To define the left and right invasion propensity vectors, $\boldsymbol{\eta}_L$ and $\boldsymbol{\eta}_R$, respectively, we construct $[\boldsymbol{\eta}_L]_{i+n_w} = rW_i(1 - W_{i+n_w})$ and $[\boldsymbol{\eta}_R]_i = rW_{i+n_w}(1 - W_i)$ for $1 \leq i \leq n_w(n_i - 1)$. Finally, we collect all of these individual terms together. Namely,

$$\mathbf{v} = \begin{bmatrix} \mathbf{v}_L \\ \mathbf{v}_R \\ \mathbf{v}_B \\ \mathbf{v}_A \end{bmatrix}, \quad \text{and} \quad \boldsymbol{\eta} = \begin{bmatrix} \boldsymbol{\eta}_L \\ \boldsymbol{\eta}_R \\ \boldsymbol{\eta}_B \\ \boldsymbol{\eta}_A \end{bmatrix} \quad (7)$$

are the total stoichiometric matrix and the total propensity vector, respectively.

2.1.2. Stochastic simulation algorithm

Having defined the propensity functions and stoichiometric matrices we can apply a standard Gillespie algorithm to simulate the system (Gillespie (1976); Gillespie (1977); Gillespie (2007); Woolley et al. (2011); Woolley et al. (2011)). Due to the stochasticity of the system any two simulations are likely to be different. However, the Gillespie algorithm is a Monte Carlo procedure that allows us to generate hundreds of simulations, which provide an ensemble of possible outcomes that are statistically accurate in their distribution.

The algorithm is as follows:

1. Initialise a start time, $t = 0$, a final time, t_f , and an initial condition, $\mathbf{W} = \mathbf{W}_0$;

2. Calculate the propensity vector, $\boldsymbol{\eta}$, and its sum $s = \sum_{k=1}^{4n_i n_w - 2n_w} [\boldsymbol{\eta}]_k$;
3. Generate two uniformly random numbers, r_1 and r_2 , from the interval $[0,1]$;
4. Calculate

$$\tau = \frac{1}{s} \ln \left(\frac{1}{r_1} \right)$$

and find the smallest $m \in \mathbb{N}$ such that

$$\frac{1}{s} \sum_{k=1}^m [\boldsymbol{\eta}]_k > r_2;$$

5. Update $t := t + \tau$ and $\mathbf{W} := \mathbf{W} + \mathbf{v}_m$;
6. Repeat lines (2)–(5) until either $s = 0$ or $t \geq t_f$ at which point the simulation is stopped and the recordings of t and \mathbf{W} are returned.

Finally, we note that because we are simulating each reaction explicitly this can result in large final data sets if every time step is recorded. However, we know that there are, on average, going to be $1/r$ reactions per time unit. Since we are going to be fixing $r = 1$ we can normalise the trajectory to lie upon an integer time grid. Namely, although we update the system every τ steps as specified in the algorithm, we only record the state at times $t = 0, 1, \dots, t_f$. This sub-sampling reduces the population output, whilst providing an accurate representation of the solution trajectory over an easily manipulable time grid. Equally, having the simulations on a common time discretisation allows easy extraction of averages. In the simulations presented here the final time point is taken to be large enough to ensure that all mutant cells are eliminated, thus, $s = 0$ since no more reactions can occur.

3. Results

3.1. One dimension, $n_w = 1$

We begin by understanding the simplest case of the system as we fix the domain to be one compartment wide *i.e.* $n_w = 1$. This greatly simplifies the theory and the numerics as we do not have to consider invasions from above, or below. Further, we can analytically approach the system and write down the evolution equation for the probability of a state of the system, $\mathcal{P}(\mathbf{W}, t)$, using the Chemical Master Equation (CME) (Jahnke and Huisinga (2007); Woolley (2011); van Kampen (2007); Schumacher et al. (2013); Picco et al. (2018)). The Chemical Master Equation is a rigorous framework, which allows us to turn reaction schema Eqs. (1)–(4) into an ODE that defines the evolution of the probability distribution function, $\mathcal{P}(\mathbf{W}, t) : \{0, 1\}^{n_i} \times [0, t_f] \rightarrow [0, 1]$.

Normally, the CME is not tractable due to it being a high dimensional and highly non-linear equation. However, even in the case that we cannot derive \mathcal{P} directly we can use the SSA, as defined in Section 2.1, since the ensemble distribution of the SSA trajectories converge to \mathcal{P} . In the current case, where we will be able to derive the solution, we will be able to demonstrate the comparison between simulation and theory.

To aid the readability of the CME we are going to abuse notation slightly and suppress any variable that is not changed, namely to talk about the changes to the i^{th} population of \mathbf{W} we simply discuss W_i . Equally, we explicitly assume that \mathcal{P} is only defined when each of the W_i are equal to 0 or 1. For any value not equal to these $\mathcal{P} = 0$. Further W_i is only defined for $1 \leq i \leq n_i$ and, thus, $W_i = 0$ for any i outside of this range. From these assumptions we can specify the CME,

rate of change of probability of being in a state \mathbf{W} = probability we reach state \mathbf{W} by invading a box from the left - probability we are in state \mathbf{W} and another invasion from the left occurs + probability we reach state \mathbf{W} by invading a box from the right - probability we are in state \mathbf{W} and another invasion from the right occurs

which is interpreted mathematically to be

$$\begin{aligned} \dot{\mathcal{P}}(\mathbf{W}, t) &= \sum_{i=1}^{n_i} rW_{i-1}W_i\mathcal{P}(W_i - 1) - \sum_{i=1}^{n_i} rW_{i-1}(1 - W_i)\mathcal{P}(W_i) \\ &\quad + \sum_{i=1}^{n_i} rW_{i+1}W_i\mathcal{P}(W_i - 1) - \sum_{i=1}^{n_i} rW_{i+1}(1 - W_i)\mathcal{P}(W_i), \\ &= \sum_{i=1}^{n_i} r(W_{i-1} + W_{i+1})(W_i\mathcal{P}(W_i - 1) - (1 - W_i)\mathcal{P}(W_i)). \end{aligned} \quad (8)$$

For clarity we explicitly derive the CME in the case of $n_i = 3$. Thus, we have three populations $\mathbf{W} = (W_1, W_2, W_3)$ and because there are 2 possible states for wild-type population ($W_i = 0$ or 1) there are 8 possible states overall. Thus, in this case, Eq. (8) actually represents the system

$$\dot{\mathcal{P}}(0, 0, 0) = 0, \quad (9)$$

$$\dot{\mathcal{P}}(0, 0, 1) = -r\mathcal{P}(0, 0, 1), \quad (10)$$

$$\dot{\mathcal{P}}(0, 1, 0) = -2r\mathcal{P}(0, 1, 0), \quad (11)$$

$$\dot{\mathcal{P}}(1, 0, 0) = -r\mathcal{P}(1, 0, 0), \quad (12)$$

$$\dot{\mathcal{P}}(0, 1, 1) = r\mathcal{P}(0, 0, 1) - r\mathcal{P}(0, 1, 1) + r\mathcal{P}(0, 1, 0), \quad (13)$$

$$\dot{\mathcal{P}}(1, 1, 0) = r\mathcal{P}(1, 0, 0) - r\mathcal{P}(1, 1, 0) + r\mathcal{P}(0, 1, 0), \quad (14)$$

$$\dot{\mathcal{P}}(1, 0, 1) = -2r\mathcal{P}(1, 0, 1), \quad (15)$$

$$\dot{\mathcal{P}}(1, 1, 1) = r\mathcal{P}(0, 1, 1) + 2r\mathcal{P}(1, 0, 1) + r\mathcal{P}(1, 1, 0). \quad (16)$$

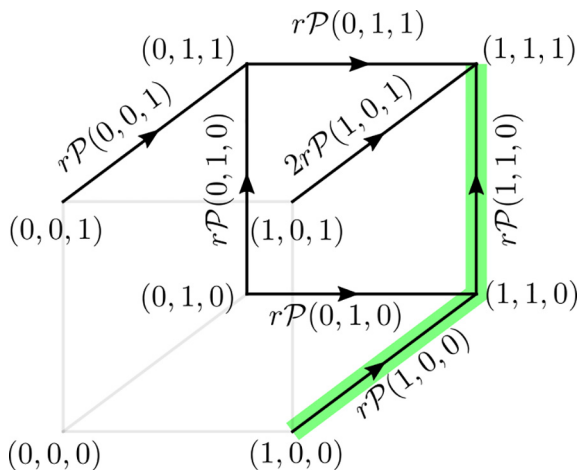


Fig. 4. Illustrating the transitions between states using a cube. States that can transition to one another are linked with black lines and the arrows demonstrate the direction of transition. Further, the lines are labelled with the transition rates. Critically, (0,0,0) cannot be transitioned to, or from, because there are no populations to begin the invasion. Equally, the invasion dynamic only allows invasion to neighbouring cells, thus, (0,0,1) cannot transition to (1,0,1). Any transitions that cannot happen are illustrated as light grey. The only transitions that are possible, due to the chosen initial condition (1,0,0), are highlighted in green.

Because we are only dealing with three compartments and, thus, three dimensions, we can further illustrate this problem by representing these transition on the edges of a cube (see Fig. 4).

Fig. 4 and Eqs. (9)-(16) clearly illustrate the long term dynamics of the underlying description. Namely, since we are considering an invasive form of dynamic the only states that are steady are the wild-type extinct state, (0,0,0), and the wild-type dominated state, (1,1,1). Further, Fig. 4 demonstrates that (0,0,0) is unstable, whilst (1,1,1) is stable, in the sense that any valid non-zero perturbation to at least one of the populations in the state (0,0,0) will cause the system to converge to (1,1,1). Equally, any perturbation to (1,1,1) that eradicates some, but not all, of the wild-type locations will tend back to (1,1,1).

Due to the linearity of Eqs. (9)-(16) we can solve the system explicitly. Specifically, if the initial conditions for the probabilities of being in a state i are p_i where $i \in \{(0,0,0), \dots, (1,1,1)\}$, $0 \leq p_i \leq 1$ and $\sum_i p_i = 1$ then

$$\mathcal{P}(0, 0, 0) = p_{(0,0,0)}, \quad (17)$$

$$\mathcal{P}(0, 0, 1) = p_{(0,0,1)} \exp(-rt), \quad (18)$$

$$\mathcal{P}(0, 1, 0) = p_{(0,1,0)} \exp(-2rt), \quad (19)$$

$$\mathcal{P}(1, 0, 0) = p_{(1,0,0)} \exp(-rt), \quad (20)$$

$$\begin{aligned} \mathcal{P}(0, 1, 1) &= (rtp_{(0,0,1)} + p_{(0,1,0)} + p_{(0,1,1)}) \\ &\quad \exp(-rt) - p_{(0,1,0)} \exp(-2rt), \end{aligned} \quad (21)$$

$$\begin{aligned} \mathcal{P}(1, 1, 0) &= (rtp_{(1,0,0)} + p_{(0,1,0)} + p_{(1,1,0)}) \\ &\quad \exp(-rt) - p_{(0,1,0)} \exp(-2rt), \end{aligned} \quad (22)$$

$$\mathcal{P}(1, 0, 1) = p_{(1,0,1)} \exp(-2rt), \quad (23)$$

$$\begin{aligned} \mathcal{P}(1, 1, 1) &= (p_{(0,1,0)} - p_{(1,0,1)}) \exp(-2rt) \\ &\quad - ((p_{(0,0,1)} + p_{(1,0,0)})rt + p_{(0,0,1)} \\ &\quad + 2p_{(0,1,0)} + p_{(1,0,0)} + p_{(0,1,1)} + p_{(1,1,0)}) \exp(-rt) \\ &\quad + 1 - p_{(0,0,0)}. \end{aligned} \quad (24)$$

We specifically want to consider the case of an infection wave travelling along the domain. Thus, the initial conditions are zero, except for $p_{(1,0,0)} = 1$. In this case the solutions simplify to

$$0 = \mathcal{P}(0, 0, 0) = \mathcal{P}(0, 0, 1) = \mathcal{P}(0, 1, 0) = \mathcal{P}(0, 1, 1) = \mathcal{P}(1, 0, 1), \quad (25)$$

$$\mathcal{P}(1, 0, 0) = \exp(-rt), \quad (26)$$

$$\mathcal{P}(1, 1, 0) = rt \exp(-rt), \quad (27)$$

$$\mathcal{P}(1, 1, 1) = 1 - (rt + 1) \exp(-rt). \quad (28)$$

These solutions are compared with approximations derived from their stochastic averages in Fig. 5. As we can see we get an excellent comparison between the different solution methods.

Although we could solve the general CME form, Eq. (8), with an arbitrary initial condition, extending the analysis is cumbersome and does not provide insight (see Eqs. (9)-(16)). Thus, we focus on solving the one-dimensional system under the specific initial condition of $\mathcal{P}((1, 0, \dots, 0), 0) = 1$ and all other states having zero probability. This further simplifies the possible transitions

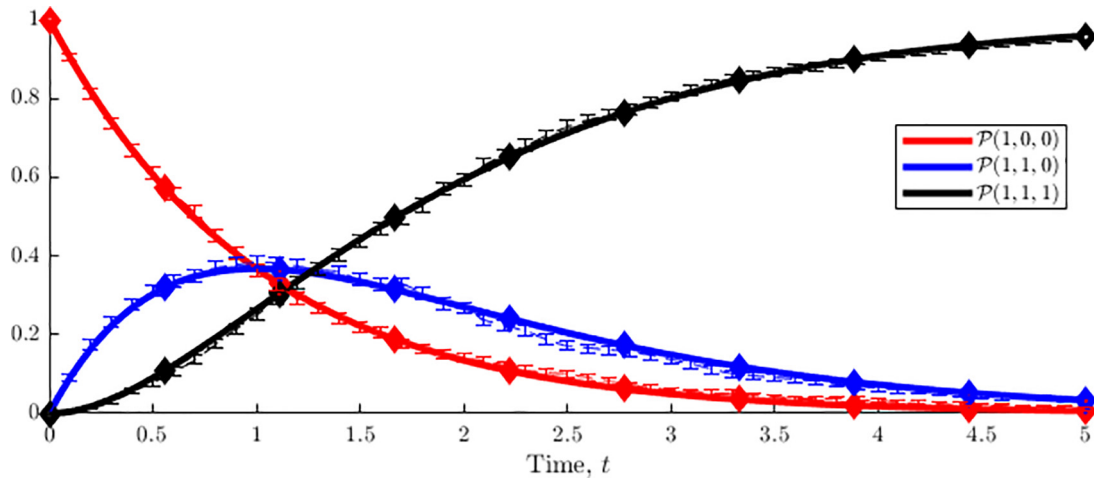


Fig. 5. Calculating \mathcal{P} in three different ways. The colour of the lines determine the state that is being considered, i.e. red is (1,0,0), blue is (1,1,0) and black is (1,1,1). The style of the line determines how \mathcal{P} is being calculated. The solid lines are the numerical solutions of Eqs. (9)-(16), the diamonds are the analytical solutions Eqs. (25)-(28) and the dashed lines with error bars represents the mean and standard errors, respectively, of 1000 stochastic simulations. In all simulations $r = 1$.

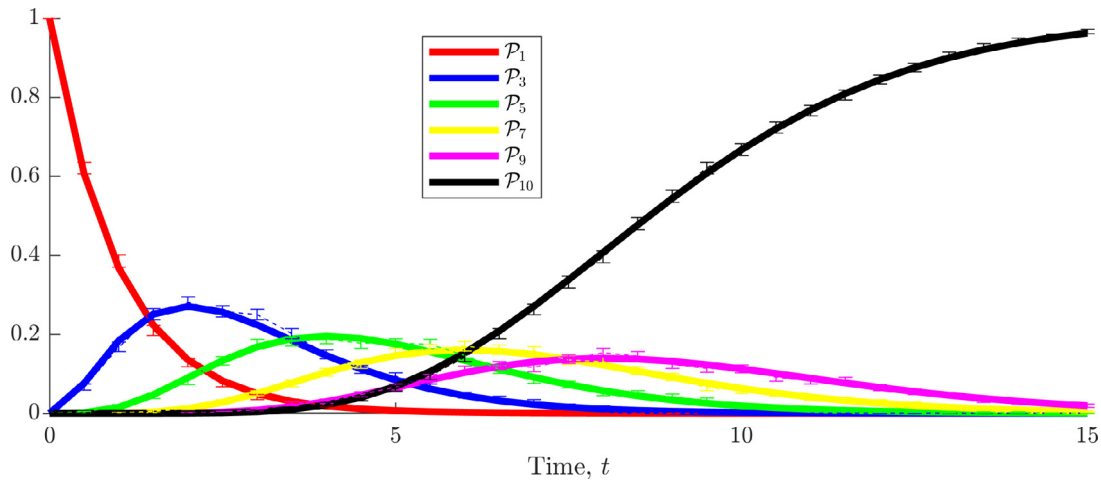


Fig. 6. Calculating \mathcal{P} two different ways. The colour of the lines determine the state that is being considered. The style of the line determines how \mathcal{P} is being calculated. The solid line is the analytical solution of Eqs. (33) and (34) and the dashed lines with error bars represents the mean and standard errors, respectively, of 1000 stochastic simulations. In all simulations $r = 1$ and $n_l = 10$.

that can be observed, namely, the evolution must follow the course of

$$(1, 0, \dots, 0) \rightarrow (1, 1, 0, \dots, 0) \rightarrow \dots \rightarrow (1, 1, \dots, 1_i, 0, \dots, 0) \rightarrow \dots \rightarrow (1, 1, \dots, 1), \tag{29}$$

where $(1, 1, \dots, 1_i, 0, \dots, 0)$ is the state that has ones filling the first i spaces and zeros thereafter. Hence, we only need to consider the evolution of the probabilities of n_l states, rather than 2^{n_l} states.

To simplify the algebra exposition we define $\mathcal{P}_i = \mathcal{P}((1, 1, \dots, 1_i, 0, \dots, 0), t)$. Explicitly, \mathcal{P}_i is the probability that the system is in a state where the wild-type cells occupy all sites up to and including the i^{th} and the mutant cells occupy all sites $i + 1$ to n_l . This simplification allows us to derive the following set of equations for these states,

$$\dot{\mathcal{P}}_1 = -r\mathcal{P}_1, \tag{30}$$

$$\dot{\mathcal{P}}_i = r\mathcal{P}_{i-1} - r\mathcal{P}_i, \quad \text{for } i = 2, \dots, n_l - 1, \tag{31}$$

$$\dot{\mathcal{P}}_{n_l} = r\mathcal{P}_{n_l-1}, \tag{32}$$

and note that any state not in the transition chain (29) has probability zero for all time. Using induction, we find that system Eqs. (30)-(32) is completely solvable,

$$\mathcal{P}_i = \frac{(rt)^{i-1}}{(i-1)!} e^{-rt} \quad i = 1, \dots, n_l - 1, \tag{33}$$

$$\mathcal{P}_{n_w} = 1 - \sum_{i=1}^{n_l-1} \mathcal{P}_i. \tag{34}$$

Solutions (33) and (34) are compared with their stochastically averaged estimates in Fig. 6, where we set $n_l = 10$.

Having demonstrated that the analytical derivation of \mathcal{P} provides a excellent approximation for the averaged solutions of the chemical master equation we consider the invasion characteristics of the population. Specifically, knowing \mathcal{P} allows us to evaluate the average occupancy of compartment i , $\langle W_i \rangle$ for all $i = 1, \dots, n_w$ and $t \geq 0$. Namely,

$$\langle W_i \rangle = \sum_{\mathbf{w}} w_i \mathcal{P}(W_i = w_i), \tag{35}$$

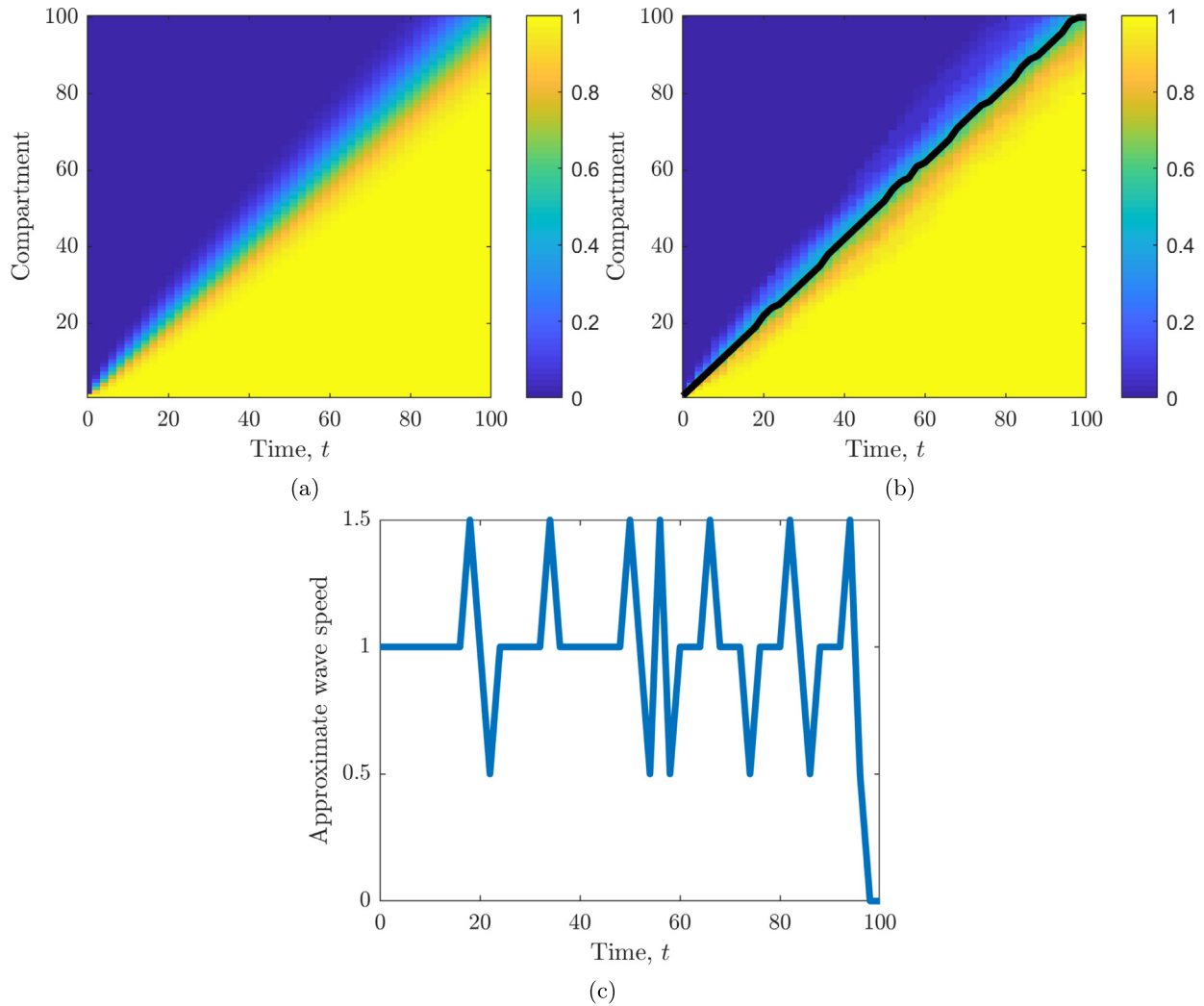


Fig. 7. Simulating the invasion of a one-dimensional space of $n_i = 100$ compartments. (a) is the theoretical solution as derived from Eq. (36) and (b) is the stochastically simulated analogue averaged over 100 simulations. The colour bar illustrates the mean occupancy, $\langle W_i \rangle$, of each compartment, i . The black line, $b(t)$, tracks the compartment at which the occupancy first breaches $\langle W_i \rangle = 0.5$. The discrete derivative of the black line in (b) (see Eq. (39)) is illustrated in (c). Throughout all simulations $r = 1$.

$$\begin{aligned}
 &= \sum_{j=i}^{n_i} \mathcal{P}_j, \\
 &= 1 - \sum_{j=1}^{i-1} \mathcal{P}_j + \sum_{j=i}^{n_i-1} \mathcal{P}_j, \\
 &= 1 - e^{-rt} \sum_{j=1}^{i-1} \frac{(rt)^{j-1}}{(j-1)!}.
 \end{aligned} \tag{36}$$

Critically, $\langle W_i \rangle$ can be interpreted as the probability of finding the i^{th} compartment occupied by a member of the wild-type, W , population. Since the wild-type cells dominate the interaction we would expect $\langle W_i \rangle = 1, \forall i$, eventually, which is confirmed by Eq. (36).

In Fig. 7 we compare theory and simulation once again. Critically, we plot $\langle W_i \rangle$ as a two-dimensional simulation, varying time along the x -axis and i along the y -axis. The colour bar then represents the value of $\langle W_i \rangle$. Specifically, Eq. (36) is illustrated in Fig. 7a, which can be compared with the stochastically derived averaged results Fig. 7b. Not only do we observe the excellent comparison, but we also notice that the W population appears to invade with a constant wave speed. This is evidenced by the approximately

constant gradient of the black line in Fig. 7b, which tracks the point at which the occupancy probability breaches 1/2 (see Fig. 7c).

To derive the wave speed we consider the time at which the i^{th} compartment is most likely to be invaded. This is the time at which \mathcal{P}_i is maximal (see Fig. 6), i.e. when $\dot{\mathcal{P}}_i = 0$, which implies the invasion occurs most likely when

$$\mathcal{P}_i = \mathcal{P}_{i-1}, \tag{37}$$

$$\begin{aligned}
 \Rightarrow \frac{(rt)^{i-1}}{(i-1)!} &= \frac{(rt)^{i-2}}{(i-2)!}, \\
 \Rightarrow t &= \frac{i-1}{r}.
 \end{aligned} \tag{38}$$

Thus, the invasion time scale is proportional to the compartment's location with wave speed given by r . This matches the approximate wave speed data illustrated in Fig. 7c, which has been derived from Fig. 7b. Namely, the black line, $b(t)$, in Fig. 7b tracks the first spatial location at which $\langle W_i \rangle \geq 0.5$, Fig. 7c then illustrates the discrete derivative of the line. Explicitly this is defined to be the average gradient of $b(t)$, i.e.

$$\frac{\Delta x(t)}{\Delta t} = \frac{b(t+1) - b(t-1)}{2}. \tag{39}$$

Fig. 7c shows that whilst the population is far away from the last compartment, $i = n_i$ (i.e. $t < 100$) the gradient of $b(t)$ in Fig. 7b is approximately equal to $r = 1$. Moreover, as the population fills the domain the wave speed drops to zero, which is seen in Fig. 7c at around $t \approx 100$.

Manipulating equations (30)-(32) and equation (35) we are able to see that

$$\langle \dot{W}_1 \rangle = 0, \quad \langle W_1(0) \rangle = 1, \tag{40}$$

$$\langle \dot{W}_i \rangle = -r\Delta x \frac{\langle W_i \rangle - \langle W_{i-1} \rangle}{\Delta x}. \tag{41}$$

Taking the limit of $\Delta x \rightarrow 0$ (i.e. discrete space tending to continuous space) and assuming r scales as to ensure $c = r\Delta x$ remains finite then we are able to derive the following advection equation with Dirichlet boundary conditions,

$$\langle W(x, 0) \rangle = 0, \quad \langle W(0, t) \rangle = 1, \tag{42}$$

$$\frac{\partial \langle W \rangle}{\partial t} + c \frac{\partial \langle W \rangle}{\partial x} = 0. \tag{43}$$

This well-known equation translates solutions at a speed c , which supports our intuition that the discretised case should equally con-

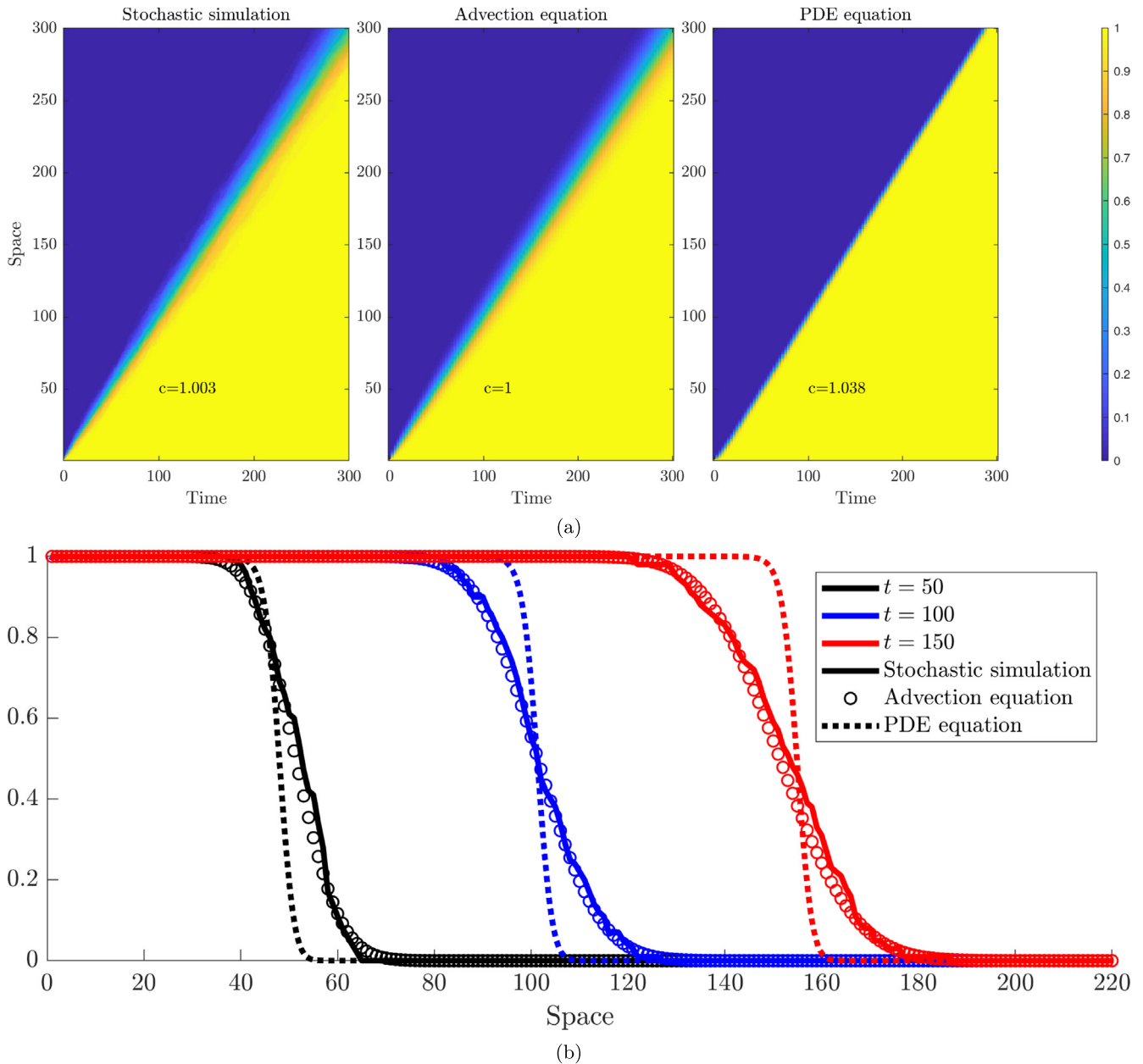


Fig. 8. Comparing travelling waves from multiple descriptions of domain invasion, over multiple scales. (a, left) presents the average dynamics of the stochastic invasion as described in Section 2.1 (averaged over 1000 simulations), (a, middle) presents the simulated solution of the advection system, Eqs. (40) and (41), (a, right) presents the simulated solution of PDE (44). In all case the domain is $[0,300]$ and the space has been discretised into $n_i = 300$ compartments of size $\Delta x = 1$. On each figure the wave speed, c , is noted. The wave speed was derived as in Fig. 7, namely, the point at which each wave is equal to a $1/2$ is tracked over the simulation and a discrete derivative is taken to approximate the gradient. These gradients are then averaged to produce a single value estimate. (b) compares the three descriptions at three specific time points. The colours red, blue and black define the time points $t = 50, 100$, and 150 , respectively. The style represents the simulation, namely, the solid lines are the stochastic simulations, the circles are the advection formulation, Eqs. (40) and (41), and the dotted lines are the solutions to (44). Parameters are $\Delta x = 1, r = 1$ and $\rho = 1/\sqrt{8}$.

tain a travelling wave solution travelling with wave speed, $c = r$, when $\Delta x = 1$.

We now suppose that there were no correlations between the grid sites and so we can use $\langle W_i W_j \rangle = \langle W_j \rangle \langle W_i \rangle$. This is of course an incorrect assumption, but we will find it useful going forward. To highlight that we are now solving an approximate problem we use the variable ω_i instead of $\langle W_i \rangle$ and ρ instead of r . In this case we can model the invasion dynamic from first principles as a solution to,

$$\begin{aligned}\dot{\omega}_i &= \rho(\omega_{i-1} + \omega_{i+1})(1 - \omega_i), \\ &= \rho(\omega_{i-1} - 2\omega_i + \omega_{i+1})(1 - \omega_i) + 2\rho\omega_i(1 - \omega_i), \\ &= \rho\Delta x^2 \frac{(\omega_{i-1} - 2\omega_i + \omega_{i+1})}{\Delta x^2} (1 - \omega_i) + 2\rho\omega_i(1 - \omega_i).\end{aligned}\quad (44)$$

If we abuse our notion of limits further we can see that Eq. (44) is a discretised form of the PDE

$$\frac{\partial \omega}{\partial t} = D(1 - \omega) \frac{\partial^2 \omega}{\partial x^2} + 2R\omega(1 - \omega),$$

$$\omega(0, t) = 1, \quad \frac{\partial \omega}{\partial x}(L, t) = 0, \quad \omega(x, 0) = 0, \quad (46)$$

where L is the length of the domain. To truly make this equivalence we would need $D = \rho\Delta x^2$ and $R = \rho$ to both be non-trivial and finite as $\Delta x \rightarrow 0$, which is impossible. However, although not rigorous in the limit we are able to learn about the discrete system through comparison arguments. Namely, we can check *a posteriori* the intuition gained from understanding Eq. (45) will approximately hold in the case of small, but finite, Δx , where $\rho\Delta x^2$ and ρ both make sense. The question is then how big can Δx be and there still be an approximate comparison between the solution of Eq. (45) and the original discrete invasion problem?

We now notice that Eq. (45) is similar to the form of Fisher's equation, with a non-linear diffusion coefficient. Using standard travelling frame coordinate changes and stability arguments (*i.e.* $\omega \approx \epsilon \exp(\lambda(x - ct))$, with $0 < |\epsilon| \ll 1$) (Murray (2003b)) we can show that if Eq. (45) presents a travelling wave then its wave speed satisfies the inequality $c \geq \sqrt{8DR} = \sqrt{8}\rho\Delta x$.

Comparing this to the original formulation, we know from Eq. (38) that the stochastic wave travels with speed $c = r\Delta x$. Thus, if our tower of assumptions is to hold then $\rho \leq r/\sqrt{8}$. Further, at least in the case where the wave front is sharp, we would expect that the inequality would be close to equality. In the case of near equality we observe the this means the stochastic rate of reaction, r has to be scaled by a factor $1/\sqrt{8}$ to be equivalent to its deterministic counter-part ρ , *i.e.* the deterministic reaction rate is slower than its stochastic analogue, which is observed in Fig. 8.

Fig. 8 compares our multiple solution approaches. Specifically, we compare the average of 1000 stochastic simulations; a solution to the discretised advection formulation, Eqs. (40) and (41); and a solution to the discretised PDE formulation, Eq. (44). Immediately, we observe that Fig. 8a shows that all solutions compare favourably over long times, in that the wave front moves at a constant rate and each simulation provides a wave speed of $c \approx 1$.

However Fig. 8b provides a different way of observing the wave fronts, which highlights the differences between the waves. Namely, we present isolated profiles from each solution method at three different time points. Here we see that although the point $\langle W_i \rangle = 1/2 = \omega_i$ is approximately consistent between the simulations (highlighting the comparable wave speeds across the simulation methods) the shape of the wave fronts differ between the solution methods. Specifically, the stochastic and advection solutions tend to spread out, whilst the PDE solution retains a sharp transition profile. These similarities and differences are to be

expected because the advection equation as derived in Eqs. (40) and (41), is an exact formulation of the mean of the stochastic simulation. Hence we expect these two simulation approaches to match well. However, the discretised PDE approach of (44) ignores the covariances between neighbouring grid points. Explicitly, we are ignoring second order and higher moments, assuming that each population is independent of its neighbours. Thus, although it can provide an accurate prediction for the mid point of the wave, it cannot provide a good approximation for the spread of the wave because these ignored higher order moments would account for understanding of terms such as variance, skewness, etc. Hence, we conclude that although the discretised PDE provides good wave speed data, it also provides poor wave shape data.

Finally, we comment on the wave speeds and the accompanying parameterisation. As mentioned above, the advection equation is an exact transport equation with wave speed $c = 1$, which is reproduced in the simulation (middle of Fig. 8a). This approximately matches the derived wave speed from the stochastic solution $c = 1.003$. However, as predicted by the theory above, even though we have accounted for the factor of $\rho = r/\sqrt{8}$ in the simulations, the wave speed calculated from the discretised PDE simulation is slightly faster than expected, $c = 1.038$. Even though this is only 3.5% faster than the stochastic simulation over the long time simulation of Fig. 8 we are able to see that the discretised PDE wave starts to pull ahead (see the red curves in Fig. 8b in particular). Thus, even though we have used a number of approximations to be able to compare the deterministic PDE and stochastic simulation, we conclude that these approximations hold well enough for standard travelling wave speed analysis to provide excellent insight into discrete stochastic simulations.

3.2. Two dimensions

The second dimension greatly increases the complexity of the system. Theoretically, the problem is fairly simple: we are able to write down a closed system of ODEs modelling the CME, which completely defines the occupation probabilities and therefore we can derive the speed of invasion. Further, all of the ODEs are linear, thus, solving them is trivial. However, as we will show, the number of equations that are required to be solved grow exponentially, even after symmetries have been removed. For example considering a relatively small 4×10 grid would require us to remove the symmetries from $2^{4 \times 9} - 1 \approx 10^{12}$ equations and subsequently solve a similar order of coupled ODEs.

One positive feature about the ODE system's structure is that we do not have to solve all equations simultaneously, the solution can be generated sequentially. Namely, as we saw in Section 3.1, the ODE for $\mathcal{P}(1, 1, 1)$ linearly depended only on the probabilities of $\mathcal{P}(1, 1, 0)$ and $\mathcal{P}(1, 1, 0)$. Similarly, the ODE for $\mathcal{P}(1, 1, 0)$ linearly depended on $\mathcal{P}(1, 1, 0)$ and $\mathcal{P}(1, 0, 0)$. Finally, the ODE for $\mathcal{P}(1, 0, 0)$ only depends on $\mathcal{P}(1, 0, 0)$. Hence, by reversing these observations, deriving the closed form solution of $\mathcal{P}(1, 0, 0)$ allows us to immediately derive the closed form solution of $\mathcal{P}(1, 1, 0)$. Equally, in simple two dimensional cases the integrations can be solved in such an inductive manner (see Fig. 9). Critically, because of the sequential nature of the solution dependence, not even parallel processing can help speed up the computation.

Moreover, even if we were able to solve algebraically solve 10^{12} ODEs, the sheer number of solutions would be ungainly and not provide any real insight. Thus, we gain what insights we can from small two-dimensional grids, growing the second dimension in a consecutive fashion until we must rely on stochastic simulation to provide the rest of the picture.

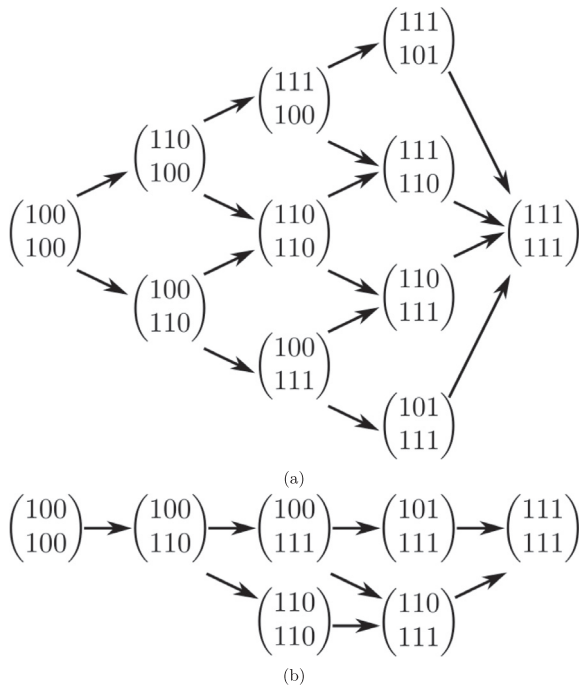


Fig. 9. Schematic diagrams presenting how each state feeds into the calculation of the probability of the next state. Namely, each matrix represents a possible state that a simulation on a 2×3 grid could achieve. To calculate the probability of being in specific state we must calculate the probability of all states that are possible precursors (denoted by the arrows). The full set of possible states are shown in (a) whilst symmetries have been removed in (b).

3.2.1. Grid size $2 \times n_t$

We begin by considering a domain that is two grid squares wide and n_t long. The possible states of this grid will be represented by a binary $2 \times n_t$ matrix where the ones represent spaces occupied by the wild type cells and zeros represent mutant spaces. To cement our intuition let us first consider a small grid of six squares, when $n_t = 3$.

Fig. 9a illustrates all possible states that our system can inhabit. Firstly, we note that, because of the initial condition and invasive dynamic we are considering (e.g. remember that the top and bottom of the domain are periodic), any possible state must have a contiguous set of ones in the matrix, e.g.

$$\mathcal{P} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} = 0.$$

Further, we note that due to the symmetry of the initial condition the probabilities will also contain a number of symmetries. Namely, the solutions are independent of matrix reflections. For example, in the 2×3 case we are considering in Fig. 9

$$\mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}.$$

Alongside the reflection symmetries that can be removed, we can also remove cyclic permutations of the matrix rows because the boundary conditions of the domain are periodic. In the case domains of the size $2 \times n_t$ the reflection symmetries are the same as the cyclic permutations, however, in grids of length $n_w \geq 3$ these two symmetries will be different. For example

$$\mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}.$$

By using the contiguous and symmetry arguments stated above we can reduce the number of ODEs that we need to solve. Such a reduction can be seen in Fig. 9, where although there are $2^{2 \times 3} = 64$ possible 2×3 binary matrices there are only 11 matrices that satisfy the initial condition and have a contiguous group of ones as seen in Fig. 9a. We then use the symmetry arguments to further reduce the number of states we need to consider to only the 7 states illustrated in Fig. 9b.

Table 1 illustrates the minimum (and maximum) number of equations that we would need to solve as the domain grows longer. Critically, as we see from Table 1, and its plotting in Fig. 10, even though we are able to dramatically reduce the number of equations we need to consider, the number still grows exponentially (note the logarithmic scale on the y-axis).

Investigating, the patterns presented in Table 1 we find that they are well-known to be related to self-avoiding random walks on grids (Sloane and Plouffe (1995); Courant and Robbins (1996)). Further, we can derive the following recursion relations

$$C(n_t) = 2C(n_t - 1) + C(n_t - 2) + 2, \quad C(1) = 1, C(2) = 4, \quad (47)$$

$$U(n_t) = 2U(n_t - 1) + U(n_t - 2) + 3 - n_t, \quad U(1) = 1, U(2) = 3, \quad (48)$$

which can be solved explicitly as

$$C(n_t) = \frac{2 - \sqrt{2}}{4} (1 - \sqrt{2})^{n_t} + \frac{2 + \sqrt{2}}{4} (1 + \sqrt{2})^{n_t} - 1, \quad (49)$$

$$U(n_t) = \frac{2 - \sqrt{2}}{8} (1 - \sqrt{2})^{n_t} + \frac{2 + \sqrt{2}}{8} (1 + \sqrt{2})^{n_t} + \frac{n_t - 1}{2}, \quad (50)$$

which concretely defines the exponential growth. Although knowing how many equations there are going to help us to know that we have enumerated all possible states, it does not help us derive the accompanying ODEs, nor solve them. Thus, we rely on numerical algorithms to derive the required equations.

Specifically, the ODE system for the 2×3 grid is

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = -2\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad (51)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} - 4\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad (52)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} - 5\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad (53)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} = 6\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} - 2\mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \quad (54)$$

Table 1

Table of total number of states that have to be considered in a $2 \times n_t$ solution domain. In a domain of length n_t we have to check $2^{2 \times (n_t - 1)}$ possible states (first row of table). Note that the index is $n_t - 1$ and not n_t because the first column of the domain is specified by the initial condition. From these states we only select those states that have a contiguous region of ones (third row) and from this we remove any states that have a reflection and cyclic symmetries. Thus, the fourth row specifies exactly how many ODEs would be required to be solved to provide a complete analytical solution.

Domain length, n_t	1	2	3	4	5	6	7	8	9	10
Total number of states, $2^{2 \times (n_t - 1)}$	1	4	16	64	256	1024	4096	16384	65536	262144
Number of states with contiguous 1s, $C(n_t)$	1	4	11	28	69	168	407	984	2377	5740
Number of unique contiguous states up to reflection and cyclic symmetries, $U(n_t)$	1	3	7	16	37	87	207	496	1193	2875

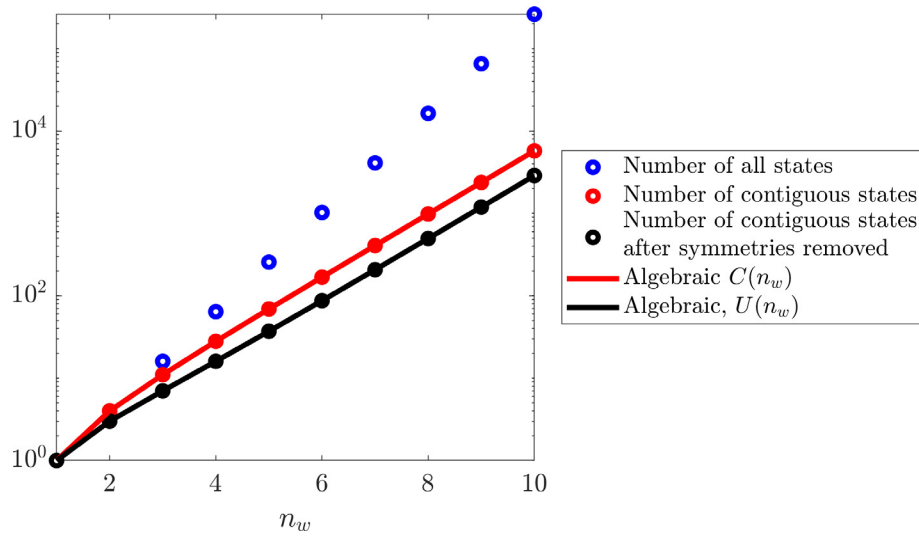


Fig. 10. Plot of the data (hollow circles) from Table 1 and the accompanying algebraic solutions of Eqs. (49) and (50) (continuous lines).

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 2\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} - 4\mathcal{P} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix}, \quad (55)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} + 3\mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} - 3\mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad (56)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 8\mathcal{P} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} + 6\mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}, \quad (57)$$

where the initial condition for each ODE is zero, except Eq. (51), which has initial probability one and we have set $r = 1$, or alternatively, scaled the time ($t \mapsto rt$) to remove this extra complicating factor. The solutions of Eqs. (51)–(57) are

$$\mathcal{P}_1 = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \exp(-2t), \quad (58)$$

$$\mathcal{P}_2 = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix} = \frac{\exp(-2t) - \exp(-4t)}{2}, \quad (59)$$

$$\mathcal{P}_3 = \mathcal{P} \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} = \frac{\exp(-2t)}{6} - \frac{\exp(-4t)}{2} + \frac{\exp(-5t)}{3}, \quad (60)$$

$$\mathcal{P}_4 = \mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} = \left(3t - \frac{3}{2}\right) \exp(-2t) + \frac{3\exp(-4t)}{2}, \quad (61)$$

$$\mathcal{P}_5 = \mathcal{P} \begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 1 \end{pmatrix} = \frac{\exp(-2t)}{6} + \left(\frac{1}{2} - t\right) \exp(-4t) - \frac{2\exp(-5t)}{3}, \quad (62)$$

$$\mathcal{P}_6 = \mathcal{P} \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} = (3t - 4) \exp(-2t) + \frac{9\exp(-3t)}{2} - \frac{\exp(-5t)}{2}, \quad (63)$$

$$\mathcal{P}_7 = \mathcal{P} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} = 1 + \left(\frac{41}{6} - 9t\right) \exp(-2t) - 9\exp(-3t) + \left(2t - \frac{1}{2}\right) \exp(-4t) + \frac{5\exp(-5t)}{3}. \quad (64)$$

By considering Eqs. (58)–(64), which are illustrated in Fig. 11, we start to see the influence of the second dimension. Specifically, because the domain is two grid points wide the transition out of the initial condition occurs faster. Namely, when we were considering just a one-dimensional domain the probability of inhabiting the initial condition was $\exp(-t)$ (see Eq. (26) with $r = 1$), whereas the probability of inhabiting the two-dimensional initial condition is $\exp(-2t)$ (see Eq. (58)). This increase in rate of leaving the initial condition is to be expected because there are now two initial cells that can begin the invasion, rather than one.

Up to reflective symmetry the first invasion action causes the system to transition from

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ to } \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0^* & 0 \end{pmatrix}. \quad (65)$$

However, the asterisked entry is now surrounded by three invasion sites (due to periodic boundary conditions), thus, the probability of transitioning to the

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$

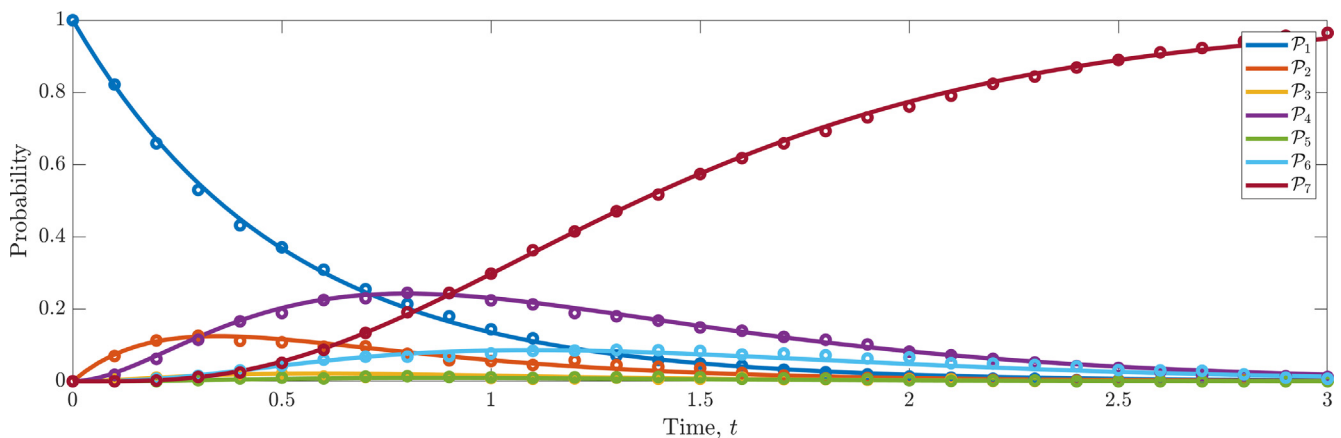


Fig. 11. Calculating $\mathcal{P}_i, i = 1, \dots, 7$ in two ways. The colour indicates which probability is being calculated (see legend) and the style represents the method of calculation. Namely, solid lines are the solution to Eqs. (58)–(64), whilst the circles are the stochastic solutions averaged over 1000 simulations.

state is three times greater than transitioning to the

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

state. Thus, although the transition from

$$\begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \text{ to } \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix} \tag{66}$$

requires two transitions, rather than the one transition that it takes to go from

$$(1 \ 0 \ 0) \text{ to } (1 \ 1 \ 0), \tag{67}$$

these two-dimensional transitions actually occur faster than the one-dimensional transitions because there are more possible routes to create the two-dimensional transition (66) than the one-dimensional transition (67). Thus, the two dimensional wave invades the domain faster than the one-dimensional wave.

This idea is illustrated in Fig. 12 where we compare the one- and two-dimensional occupancy probabilities of

$$\mathcal{P}(1,1,0) \text{ and } \mathcal{P}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \tag{68}$$

respectively. Firstly, we see that

$$\mathcal{P}(1,1,0) \geq \mathcal{P}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}. \tag{69}$$

This makes sense because the one-dimensional system must transition through the (1,1,0), thus the integral over all time (i.e. $t \in (0, \infty)$) must sum to one (which can be checked by integrating Eq. (27)). Comparatively, the integral of Eq. (61) is less than one because the system does not have to transition through this state; there are other state options available. Specifically, the asterisked zero in the matrix of transition (65) will eventually convert to a one, but it may not be the second transition to happen.

For our purposes of understanding the travelling wave speed the relative sizes of probability are irrelevant. The more important metric is the time at which these probabilities obtain their maxima. As we saw in Section 3.1 the one-dimensional wave invades the domain at a constant wave speed of $c = r$ ($c = 1$ in the non-dimensional case that we are considering). This can be seen in Fig. 12 by noting that the maxima of $\mathcal{P}(1, 1, 0)$ occurs at $t = 1$, (blue dashed line). This is in contrast to the maximum of

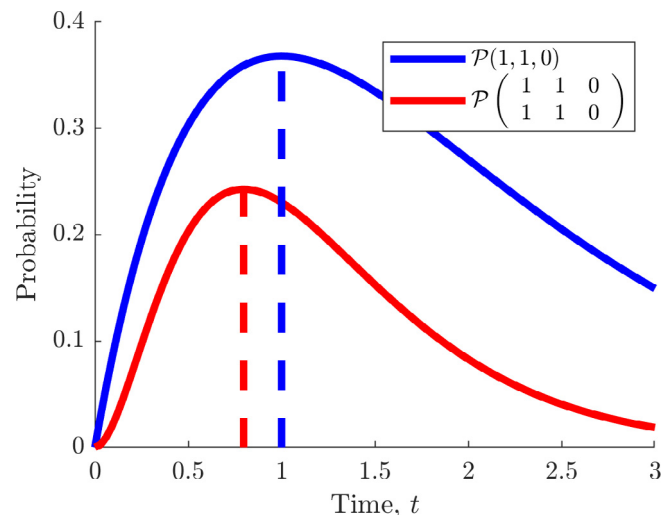


Fig. 12. Comparing the transient probabilities of the first invasion step. Plots of Eqs. (27) and (61).

$$\mathcal{P}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix},$$

which occurs at $t = t_c \approx 0.80$ (red dashed line in Fig. 12), where t_c is the solution of

$$\begin{aligned} 0 &= \dot{\mathcal{P}}\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \end{pmatrix}, \\ &= 6(1 - t_c) \exp(-2t_c) - 6 \exp(-4t_c), \\ \Rightarrow 1 &= (1 - t_c) \exp(2t_c). \end{aligned} \tag{70}$$

Notably, Eq. (70) is a transcendental equation, which does not have a closed form algebraic solution (Corless et al. (1997)). However, it should be noted that the properties of such equations have been investigated through the use of ‘‘Lambert functions’’ (Mezö and Keady (2016); Fukushima (2013)), which can also be linked to the solution of delay differential equations Asl and Ulsoy (2000). By taking inspiration from the 1D case it may be possible to recast this problem, more generally, as a form of discrete spatial delay. Namely, to find the wave speed our question is generically: how long does it take the system to evolve from the state

$$\begin{pmatrix} 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 & 0 & 0 & \dots & 0 \end{pmatrix} \tag{71}$$

to the state

$$\begin{pmatrix} 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \end{pmatrix}? \tag{72}$$

However, for a very long domain, state (71) would be practically the same as state (72). Alternatively, state (71) could be thought of as a spatially delayed form of state (72). Whether either of these two ideas lead to a fruitful reinterpretation of this problem has yet to be determined and will be developed later in future work.

When working with the stochastic simulations we normalised the stochastic trajectories such that the output times are the integers, thus, when calculating the approximate wave speed from the data we always knew the time step over which we were measuring a change in space. However, in this problem, we do not know the time that an invasion step will take place, but we do know that spatial length scale that the step occurs. Explicitly, we now estimate the wave speed by $c(i) = \Delta x / \Delta t(i)$ where $\Delta x = 1$ is the width of the discretised compartment and $\Delta t(i) = t_i - t_{i-1}$ where t_i is the time at which

$$\mathcal{P}\begin{pmatrix} 1 & 2 & \dots & i-1 & i & i+1 & \dots & n_i \\ 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & \dots & 1 & 1 & 0 & \dots & 0 \end{pmatrix} = 0.$$

Namely, the time at which ‘‘the probability of inhabiting the state where all columns up to and including the i^{th} column are ones and all the columns after this are zero’’ is maximal. Additionally, we define $t_1 = 0$.

As mentioned above, although the ability to derive the analytical formulation of this time is possible, the growth in the required number of formulas makes the process intractable as the domain grows (see Fig. 10). Equally, as we see from this simple case, the equations that have to be solved are not algebraic. Thus, after algebraically solving all ODEs we must numerically solve for each t_i .

Specifically, in the case of $n_i = 10$ we set $t_1 = 0$ and evaluate t_2, \dots, t_9 through numerically finding the roots of eight transcendental equations. Note that we do not evaluate t_{10} , because the final ‘full’ domain state probability does not have a local maximum as it monotonically tends to one. The equations for t_i rapidly grow

in size, for example, the equation for t_9 contains the nine even exponential powers from $\exp(-2t)$ to $\exp(-18t)$ the coefficients of which are then polynomials of up to degree seven. Presenting such equations algebraically offers little in terms of insight into the results, thus, we suppress the full set of equations and supply just the first three to demonstrate the growth in complexity, namely, t_2, t_3 and t_4 satisfy

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix} = 0 = 6(1 - t_2)e^{-2t_2} - 6e^{-4t_2}, \quad (73)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & 0 & \dots & 0 \end{pmatrix} = 0 = \left(3t_3 + \frac{13}{16}\right)e^{-6t_3} + \left(9t_3 + \frac{65}{4}\right)e^{-4t_3} + \left(-9t_3^2 + \frac{95t_3}{4} - \frac{273}{16}\right)e^{-2t_3}, \quad (74)$$

$$\dot{\mathcal{P}} \begin{pmatrix} 1 & 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 1 & 1 & 0 & 0 & \dots & 0 \end{pmatrix} = 0 = \left(\frac{7t_4}{3} + \frac{259}{144}\right)e^{-8t_4} + \left(\frac{9t_4^2}{2} + \frac{21t_4}{4} + \frac{9}{16}\right)e^{-6t_4} + \left(-\frac{27t_4^2}{4} - \frac{87t_4}{2} - \frac{1047}{16}\right)e^{-4t_4} + \left(-9t_4^3 + \frac{177t_4^2}{4} - \frac{983t_4}{12} + \frac{9083}{144}\right)e^{-2t_4}, \quad (75)$$

respectively. We observe that each term in the equations is a product of a polynomial and an exponential in t . Since the exponentials decay faster than the polynomials grow then, by inspection, we note that we obtain a good approximation for t_i in all cases by just considering the coefficients of the $\exp(-2t)$ and $\exp(-4t)$ terms (see Fig. 13). Although this does offer a way to simplify the presentation of the equations, we are already going to be evaluating them numerically thus we evaluate the full solution, rather than the approximation. Equally, in using the approximation we have to be careful to take the correct root, because, as seen in Fig. 13 the approximation has two positive roots, where the original solution only has one positive root (as well a root at zero, which is easy to eliminate).

From evaluating the t_i as defined we can extract the wave speed throughout the domain, which we compare against simulation data. Namely, we simulate 1000 cases of the domain invasion algorithm as specified in Section 2.1. The average of these simulations can be seen in Fig. 14a, where we also compare the results of the $n_w = 2$ simulations with the one-dimensional, $n_w = 1$, case. This information is simplified in Fig. 14c, where we have averaged over the domain length to produce a single value for the population profile over the domain.

We immediately observe from Figs. 14a and 14b that, in the $n_w = 2$ case, the invading population has been able to move further into the domain than compared to the $n_w = 1$ case. Namely, as predicted, the invasion wave moves faster than we would have

expected had we simply considered the invasion deterministically, where we would not observe such a dimensional influence.

From the averaged simulations (see Fig. 14b) we are able to extract the wave speed from tracking the position at which the averaged population is approximately 0.5. Specifically, from Fig. 14c we are able to observe that, similar to the 1D wave invasion, the 2D wave invasion occurs at an approximately constant rate (the lines are approximately straight). Equally, the gradients of the lines tell us the wave speed of invasion. Specifically, the $n_w = 2$ wave invades with a speed of approximately $c \approx 1.6$, which is greater than the $n_w = 1$ case of $c = 1$.

Finally, in Fig. 14d, we see that the simulated wave speed compares favourably with the value extracted from the algebraic solutions. Namely, after a period of transient behaviour the wave speed as approximated by $1/(t_i - t_{i-1})$ tends to 1.6 (to 1 decimal place). Since we are using the algebraic solutions we are able to easily generate more accuracy for this value than we would be able to through stochastic simulation alone. Namely, to three decimal places we can predict that $c \approx 1.585$.

Overall, this section has demonstrated that by passing into the second dimension a stochastic invasion wave moves faster than its one dimensional analogue. A result that would not be apparent in a deterministic simulation. Moreover, as seen in Section 3.1, although applying a deterministic formulation is possible to calculate a wave speed, the parameter values have to be corrected carefully. Equally, although the wave speed may be extractable, the transition in the wave shape would be too sharp. In the next section we extend the width even further to see whether if the increase in wave speed continues.

3.2.2. $n_w = 3, 4$ and 5

As seen in the last section, the number of ODEs that are required to solve the CME entirely grows exponentially, meaning that we can only solve the problem analytically for problems of moderate widths. By making the domain wider the problem size will grow quicker and, thus, our algebraic insights will have to be gained from domains of smaller lengths. The number of contiguous, $C(n_w, n_l)$, and unique, $U(n_w, n_l)$, ODEs that are required for various values of (n_w, n_l) are specified in Table 2. The empty spaces illustrate regions where deriving the equations took longer than simulating 1000 of the accompanying stochastic simulations.

Critically, being able to generate fewer data points means that it is harder to derive the recurrence relations that define C and U . For example, from Table 2 we are only able to conjecture that $C(3, n_l)$ satisfies

$$C(3, n_l) = 5C(3, n_l - 1) + 3C(3, n_l - 2) - C(3, n_l - 3) + 2, \\ C(3, 1) = 1, C(3, 2) = 8, C(3, 3) = 45. \quad (76)$$

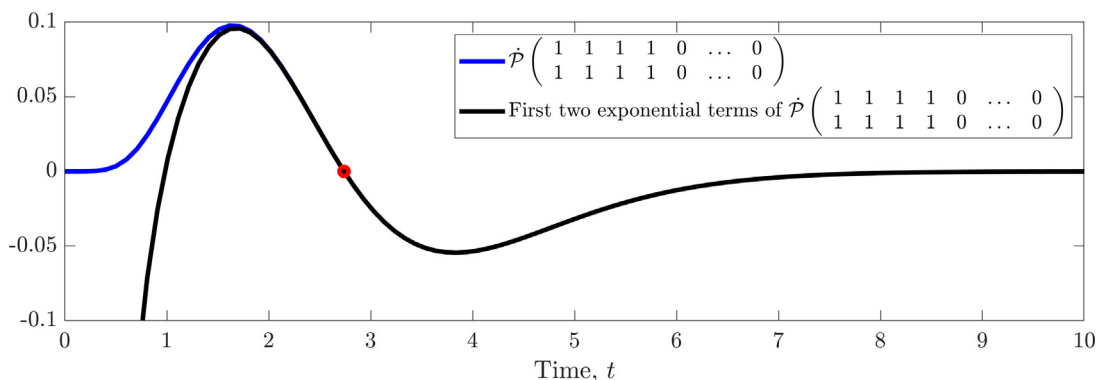


Fig. 13. Plot of Eq. (75) (blue line) and its approximation to only the $\exp(-2t)$ and $\exp(-4t)$ terms (black line). The red circle at $t_4 \approx 2.74$ demonstrates the close approximation of the root of the two curves.

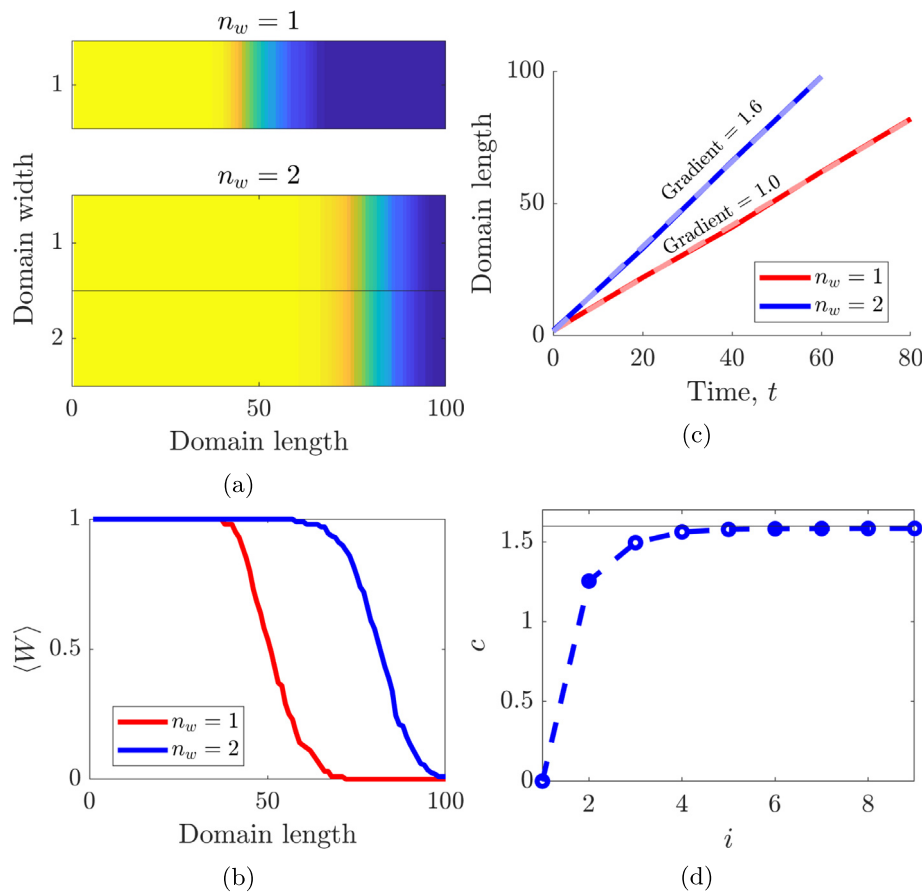


Fig. 14. Comparing invasion dynamics and deriving wave speed when $n_w = 1$ and 2 and $n_l = 100$. All simulated data has been averaged over 1000 runs and the simulations were run until all mutant cells were eliminated. (a) Illustration of the invasion wave at a single time point, $t = 50$, in the $n_w = 1$ (top) and $n_w = 2$ (bottom) cases. The colour scale is the same as in Fig. 7. (b) Illustration of the same information as (a) but we have averaged the data over the domain length so we can collapse the second spatial dimension and compare the curves directly. (c) Tracked position of the point at which $\langle W \rangle = 0.5$ over all time points. The solid lines are the simulated data and the lighter dashed lines are least square fitted lines. The gradient of these lines are noted on the figure next to their, respective, lines. (d) Convergence of the wave speed to $c \approx 1.6$ as the wave speed is evaluated further into the domain. The black line marks the asymptote is fixed at $c = 1.585$, as derived from the analytical solution. The data only exists at the integer values of i , the dashed line has been added to aid visualisation.

Table 2

Table of total number of states that have a contiguous region of ones, $C(n_w, n_l)$, and the number of states after reflections and cyclic symmetries have been removed, $U(n_w, n_l)$. The domain length is stated along the top and we consider $n_w = 3, 4$ and 5. The gaps in the lower right of the table illustrates the region where the number of states becomes too large to calculate within a reasonable time frame.

Domain length, n_l	1	2	3	4	5	6	7	8
$C(3, n_l)$	1	8	45	250	1379	7602	41899	230924
$U(3, n_l)$	1	4	14	57	266	1353	7189	38982
$C(4, n_l)$	1	16	175	1976	22441	255072		
$U(4, n_l)$	1	6	37	310	3092	33262		
$C(5, n_l)$	1	32	653	14484	328785			
$U(5, n_l)$	1	8	87	1553	33400			

We note that Eq. (76) is solvable, but the direct formula depends on the roots of

$$Z^3 - 3Z^2 - 5Z + 1 = 0,$$

which, although real, require the general cubic formula to be used and, thus, written in terms of complex conjugate pairs. Equally, using the intuition gained from this section and Eqs. (47) and (48) we conjecture that the recursion relations for domains of width $n_w = 4$ and 5 will be fourth and fifth order, respectively. Thus, not only would we need more data to fit these recurrence relations

the accompanying auxiliary polynomial would no longer be solvable in terms of radicals for $n_w \geq 5$ and, so, we would have to use numerical approximations anyway.

Fig. 15 illustrates similar information to that of Fig. 14. Namely, we compare simulated and algebraically derived information and inspect our ability to derive the wave speed in domains of increasing width, whilst fixing the length to be 100. Firstly, from Fig. 15a we see the two-dimensional invasion wave averaged over 1000 stochastic simulations. Specifically, as noted in Section 3.2.1, as the domain width increases (top to bottom) we can see that the wave travels further in the same time of $t = 25$ units. This compar-

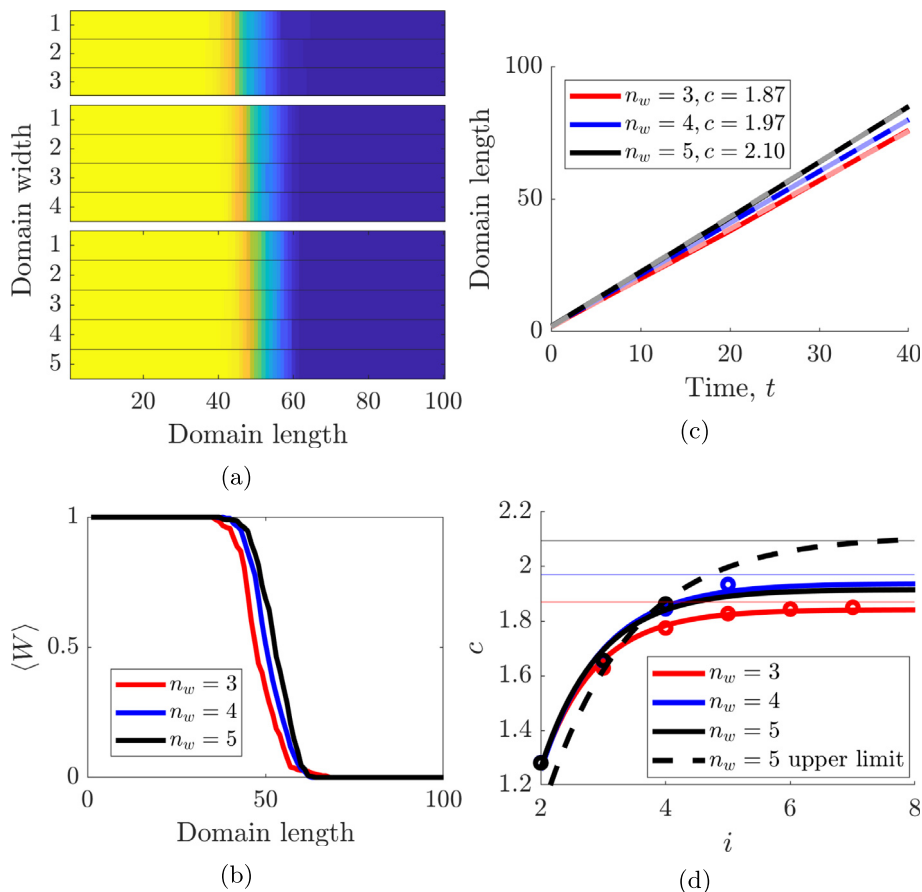


Fig. 15. Comparing invasion dynamics and deriving wave speed when $n_w = 3, 4$ and 5 and $n_l = 100$. All simulated data has been averaged over 1000 runs and the simulations were run until all mutant cells were eliminated. (a) Illustration of the invasion wave at a single time point $t = 25$ in the $n_w = 3$ (top), $n_w = 4$ (middle) and $n_w = 5$ (bottom) cases. The colour scale is the same as in Fig. 7. (b) Illustration of the same information as (a) but we have averaged the data over the domain width so we can collapse the second spatial dimension and compare the curves directly. (c) Tracked position of the point at which $\langle W \rangle = 0.5$ over all time points. The solid lines are the simulated data and the lighter dashed lines are least square fitted lines. The gradient and, thus, wave speed, c , is denoted for each line in the legend. (d) Convergence of the algebraically derived wave speeds. The dots indicate the algebraically derived data, whilst the thick solid lines are least squared fits of the equation $c(1 - \exp(b_1(1 - i)))$ to these data points, which has then been extrapolated. The thin horizontal lines indicate the limiting wave speeds as calculated from the gradients of the lines in (c) in the order $n_l = 3, 4$ and 5 , bottom to top, respectively. The dashed black line illustrates the $c(1 - \exp(b_1(1 - i)))$ but c takes the upper limit from its 95% confidence interval, rather than the mean.

ison is easier to see in Fig. 15c, where we have, once again, averaged over the width to observe that the wave speed increases as n_w increases. By tracking the point at which $\langle W \rangle = 0.5$ over time (see Fig. 15c) we can extract the wave speed for each domain length and this data is presented in Fig. 15c.

The dots in Fig. 15c are the data derived from the 30,000–40,000 ODEs that we can solve in each of these cases. Specifically, we once again derive an approximate wave speed using the same procedure illustrated in Section 3.2.1. Namely, we calculate the difference in times at which the probability of being in consecutive vertically homogeneous state is maximal.

Critically, note that the data represented by the red, blue and black dots extend to different lengths due to the decreased domain size over which we can calculate them (see Table 2). Although we are able to algebraically derive enough points in the $n_w = 3$ to see that the wave speed curve converges, this is not true in the $n_w = 4$ and 5 case. Thus, we fit an exponentially decaying curve,

$$f(i) = c(1 - \exp(-b_1(i - 1))), \tag{77}$$

to these cases to try and leverage the information we have as a way of predicting the asymptotic wave speed, which we can then compare with the wave speed as derived from the stochastic simulations (see Fig. 15c). We note that $f(1) = 0$ and the parameters c and b_1 are interpretable in terms of the problem's features. Namely, c is the asymptotic wave speed after the transients have been

ignored (i.e. the gradient of the curves in Fig. 15c), whilst $1/b_1$ provides a decay length scale, which is an estimate of how long the domain would have to be before the asymptote would be reached.

We note that the red ($n_w = 3$) and blue ($n_w = 4$) lines fit rather well and provide good estimations for the asymptotic wave speed. However, we note that with only 4 data points the black fitted curve greatly underestimates the $n_w = 5$ asymptote, being below even the blue line. However, along with any fitting we can provide a confidence interval for the parameters that are being used. Specifically, in the $n_w = 5$ case the expected value of $c = 1.916$, but this has a 95% confidence interval of (1.720, 2.112). If we instead use the upper estimate of $c = 2.112$ we generate the black dashed line, which provides a much better estimate for the asymptotic wave speed.

Thus, as the domain grows, exponentially more ODEs would be required to estimate the wave speed to the same precision. However, even in the case that we cannot generate enough data algebraically, curve fitting still offers a way to estimate the asymptotic wave speed with fewer data points such that the parameter values are within the 95% calculated confidence interval.

In any case, our result here is that the speed of a stochastically invading wave does increase as the size of the width increase. Critically, for a two-dimensional domain of width 5 the wave speed is over 2, which is over twice that of the

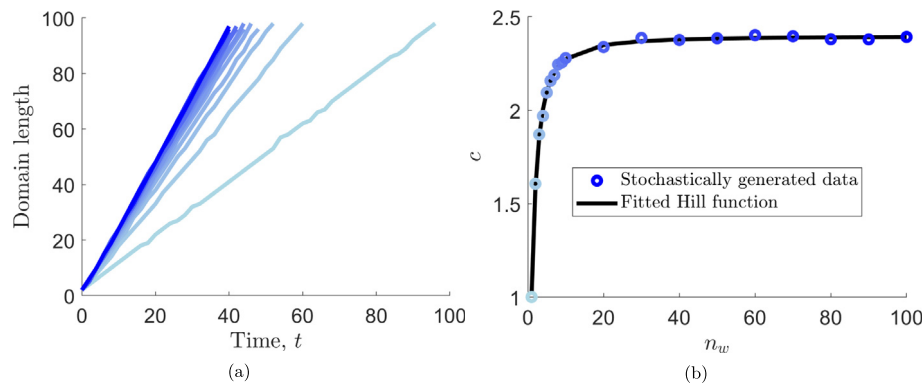


Fig. 16. Comparing invasion dynamics and deriving wave speed when $n_w \in \{1, 2, \dots, 10, 20, \dots, 100\}$. All simulated data has been averaged over 1000 runs and in all cases $n_l = 100$. (a) Tracked position of the point at which $\langle W \rangle = 0.5$. The colour gradient transitions from light blue, $n_w = 1$, to dark blue, $n_w = 100$, with larger n_w correlating with darker colours. (b) Convergence of the algebraically derived wave speeds. The dots indicate the gradient data derived from (a). The colour of the line in (a). The thick solid line is a least squared fit of the Hill function, $c_\infty n_w^n / (b_2 + n_w^n)$.

expected value derived from a one-dimensional case. We may have expected the one- and two-dimensional wave speeds to have been the same since they are in the deterministic case and, on average, the two-dimensional stochastic system is homogeneous in the vertical direction.

We have now reached the limits of what we can algebraically achieve. In the next section we depend on the stochastic simulations to push our intuition to its limits.

3.2.3. Large two-dimensional grids

Figs. 14c and 15c immediately suggest a question regarding the relationship between domain size and wave speed. Namely, although the wave speed increases with increasing domain size, does the wave speed converge to a limit, or does it increase without bound? Answering such a question would not be possible with the algebraic methods presented due to the exponential growth in the ODEs required to be solved. Thus, we turn to simulation and provide evidence, rather than a rigorous argument that the speed does converge. However, although not rigorous, our results are still valid and accurate because of the close alignment of theory and simulation that was demonstrated in the earlier sections.

We generate 100 domain invasion simulations for each domain width $n_w \in \{1, 2, \dots, 10, 20, \dots, 100\}$. As before we average over

Table 3
Simulated wave speed extracted from the data in Fig. 16a.

n_w	c as calculated from the gradient of Fig. 16a to three decimal places with 95% confidence intervals
1	1.001 ± 0.005
2	1.607 ± 0.010
3	1.871 ± 0.011
4	1.970 ± 0.011
5	2.095 ± 0.013
6	2.168 ± 0.018
7	2.188 ± 0.013
8	2.245 ± 0.014
9	2.256 ± 0.013
10	2.279 ± 0.017
20	2.338 ± 0.016
30	2.387 ± 0.018
40	2.376 ± 0.020
50	2.385 ± 0.018
60	2.402 ± 0.021
70	2.396 ± 0.021
80	2.380 ± 0.021
90	2.379 ± 0.019
100	2.392 ± 0.020

the simulations and the domain width and track the front of the wave at the point where $\langle W \rangle = 0.5$. This data is observed in Fig. 16a, where the colour of the lines forms a smooth gradient from light blue, $n_w = 1$, to dark blue, $n_w = 100$. For each stochastic data line in Fig. 16a we fit a least squares straight line through the data and extract the gradient, stated explicitly in Table 3 and plotted in Fig. 16b. The colour of the points in Fig. 16b matches the colour of the data in Fig. 16a, from which the point was derived.

The wave speed data, as plotted in Fig. 16b, illustrates the evidence that the wave speed appears to converge to approximately 2.39. To confirm this we fitted a Hill function of the form

$$H(n_w) = \frac{c_\infty n_w^n}{b_2 + n_w^n}. \tag{78}$$

If $n_1 > n_2$ then we would expect the a wave speed to continue to grow as the domain length increases. The fitting suggests that, to three decimal places, $n_1 = 1.416$ and $n_2 = 1.414$, which suggests that even if the wave speed does grow it grows extremely slowly with domain length. Specifically, to an excellent approximation $n_1 \approx n_2$ and, so, upon refitting Hill function (78) assuming $n_1 = n_2 = n$ we find that $c_\infty = 2.40, b_2 = 1.36$ and $n = 1.40$, thus $\lim_{n_w \rightarrow \infty} c \approx c_\infty = 2.40$.

Using the same ideas as in Section 3.1 we can compare this discrete two-dimensional wave to the continuous and deterministic analogue, where covariances are ignored. Namely, for $1 \leq i \leq n_l$ and $1 \leq j \leq n_w$ the approximate discrete equation for the mean

$$\langle \dot{W}_{ij} \rangle = r \langle (W_{i-1,j} + W_{i,j+1} + W_{i,j-1} + W_{i+1,j})(1 - W_i) \rangle,$$

can be manipulated into a form that compares with a discretisation of

$$\frac{\partial \omega}{\partial t} = D(1 - \omega) \nabla^2 \omega + 4R\omega(1 - \omega), \tag{79}$$

$$\begin{aligned} \omega(0, y, t) = 1, \quad \frac{\partial \omega}{\partial x}(L_x, y, t) = 0, \quad \omega(x, 0, t) \\ = \omega(x, L_y, t), \quad \omega(x, y, 0) = 0, \end{aligned} \tag{80}$$

on the square domain $[0, L_x] \times [0, L_y]$, where R is the deterministic transition rate and $D = R\Delta x^2$ is, once again, interpreted as approximation for small but finite Δx .

Critically, it should be noted that although we have been developing our intuition through comparing our knowledge with spatially extended form of Fisher's equation, here we see that our invasion dynamic is fundamentally different in higher dimensions. Namely, in Eq. (79) we have to account for a coefficient of 4, rather than the coefficient of 2 that appears in the one-dimensional case

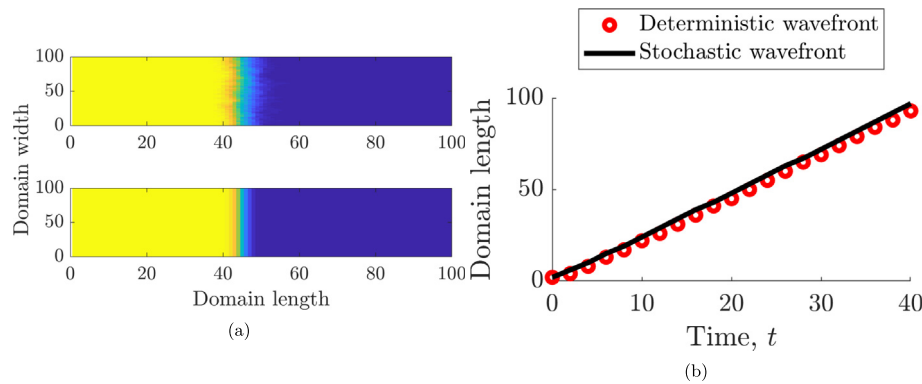


Fig. 17. Comparing deterministic and stochastic invasion dynamics. (a) A simulated wave profile on a domain of size $[0, 100] \times [0, 100]$ at time $t = 20$. (top) Average of 100 stochastic simulations with $r = 1$ and $\Delta x = 1$. (bottom) Solution to discretised PDE Eq. (79) with $R = 0.545$. (b) Tracked position of the point at which $\langle W \rangle = 0.5 = \omega$.

of Eq. (45). This is in contrast with Fisher’s equation, where the kinetic term is the same in all dimensions. Further, in three dimensions, this coefficient would be 6 for a cubic lattice, suggesting that the dependence of wave speed on dimension is exaggerated further in higher dimensions. Namely, we would expect the invasion wave to be even faster in three dimensions.

Assuming that standard Fisher wave analysis works as equally well as in Section 3.1 we derive that if there is a travelling wave then its speed c must satisfy the inequality $c \geq 4R\Delta x$. If we are to match the limiting stochastic wave speed, where $\Delta x = 1$ and $c = 2.4r$, then we would then need to satisfy $R \leq 3r/5$.

In one dimension we demonstrated that the simulations suggest that the derived inequality is approximately an equality, *i.e.* $\rho \approx r/\sqrt{8}$, within the tolerances of noise. However, in the two-dimensional case simulations suggest that the inequality is strict. Specifically, to provide a deterministic invasion wave that matches the stochastic wave, with $r = 1$ (see Fig. 17a) we ran a parameter optimisation on R and found that $R = 0.543$ (to 3 decimal places), which produces a deterministic simulation wave speed of $c = 2.391$ and we observe that $R = 0.543 < 3r/5 = 0.6$, thus, satisfying the derived wave speed inequality.

We finish this section by highlighting the results that we have presented. Firstly, although the infection wave speed does increase with a growth in the width, there does appear to be a limiting wave speed of approximately 2.4. Further, the stochastic and deterministic formulations can once again be compared approximately, such that intuition from analytic theory from the deterministic theory approximately holds for the stochastic simulations. Critically, this highlights that a deterministic modelling approach to modelling cancer elimination requires subtle scaling to match the individual-level cellular activity.

3.3. Influence of boundary conditions

Another unexpected feature that arose through comparing our invasion dynamic to that of Fisher’s equations was the influence of periodic versus no flux Neumann boundary conditions. Specifically, in the Fisher wave case a zero-flux boundary condition is a sub-symmetry of periodic boundary conditions. Thus, if we start with an initial condition that is homogeneous in the y -direction then the solution is the same under both sets of boundary conditions. However, this is not true in our domain invasion case.

The periodic boundary conditions, assumed up until now, have allowed the invasion to occur through the top and bottom borders. Zero-flux boundary conditions stop this from occurring, meaning that the boundary cells can only be invaded from one side, *i.e.* only from below on the top boundary, or only from above on the bottom boundary. This reduces the probability of invasion when compared to the middle cells that can be invaded from all directions.

On converting to zero-flux boundary conditions (see Fig. 18) we find that, in both the stochastic and deterministic simulations, the wave front is no longer homogeneous across the y -direction. The edges of the wave lag behind the front. This is most notable in the top figures of Fig. 18 as the width of these grids are $n_w = 10$ and, thus, much smaller than the figures on the bottom, which are simulated on grids of width $n_w = 100$. Moreover, the stochastic nature of the wave front in Fig. 18a appears to amplify the heterogeneity of the wave front along the y -axis as we can observe that the black line in the bottom figure of Fig. 18a, which represents the front of the stochastic simulation is more curved than its deterministic analogue (bottom image of Fig. 18b), where the heterogeneity is restricted to within a boundary layer of the edges.

Similar to the periodic boundary condition simulations in Section 3.2 the wave speed of the front in the stochastic simulations

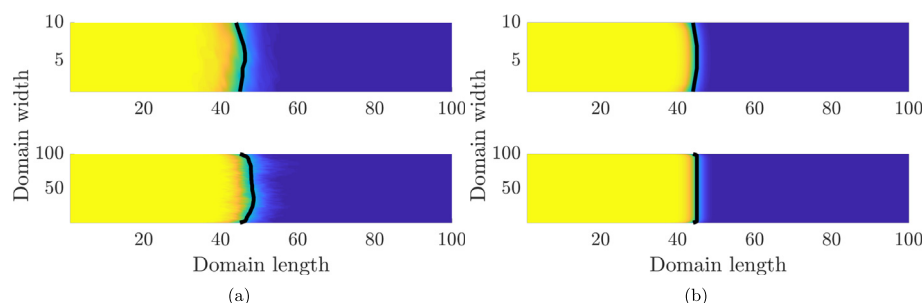


Fig. 18. Comparing (a) stochastic and (b) deterministic simulations under zero-flux boundary conditions. The top simulations are on a grid of width $n_w = 10$, whilst the bottom simulations are on a grid of width $n_w = 100$. The black lines mark the positions at which the (stochastic, or deterministic) density is 0.5.

is faster on domains of larger width than on thinner domains. Explicitly, the simulations in Fig. 18 are all run to 20 time units and, in the stochastic simulations of Fig. 18a the wave fronts in the simulations on the bottom are all slightly further ahead of the wave fronts in the top simulations, because the bottom domain is wider than the top domain. However, in the deterministic simulations of Fig. 18b, the wave fronts are approximately at the same location.

We do not attempt to make analytic progress in understanding this situation as it is fraught with difficulties. For example, even without considering the combinatorial explosion of ODEs, defining what a wave front is in this case would require explicit care, which is outside the scope of the current article. However, we present this as a point of interest and the source of future work.

4. Application to Kras mutant elimination

Now that we better understand the influence of dimension on the speed of invasion, we apply this model to our Kras mutant data. Experimental data was generated by tracing the fate of fluorescently labelled KrasG12D mutant cells in murine pancreas tissues over time (Hill et al. (2021)). We use a genetically engineered mouse model of pancreatic cancer ($Pdx-1Cre^{ER}LSL-Kras^{G12D/+}; Rosa26^{LSL-RFP}$; (Hingorani et al. (2003))). By administering a single low dose of tamoxifen, we induced KrasG12D (and RFP) expression in low numbers of cells in an otherwise healthy

epithelium, allowing us to study normal-mutant cell-cell interactions in vivo. By measuring the amount of RFP fluorescence in tissues at 7 days post tamoxifen induction, we found that a single low dose induced transgene expression in approximately 25% of the tissue, with RFP-labelled cell clusters varying across a range of sizes (Hill et al. (2021)). Based on previous research (Hogan et al. (2009); Porazinski et al. (2016)), we hypothesised that KrasG12D cells compete with normal cells for space and survival in tissues and elimination of mutant cells would occur over protracted time points (Morton et al. (2010)). We chose 35 days post tamoxifen induction as an end point and compared the amount of RFP (fluorescence and genomic DNA) in tissues harvested at this time point, to 7-day data. This analysis revealed a significant decrease in the amount of RFP in KrasG12D-expressing tissues at 35-days compared to 7-days post induction, indicating that KrasG12D cells are eliminated from tissues over this time period (see Fig. 1).

Specifically, we initially assume that the mutant cells form approximately square patches of different side length. The reason we are using initial square patches is because, as we will see later, the density scales as the inverse of the patch size, suggesting that initially the patches are highly regular shapes. It should be noted that other regular shapes were chosen (see (Hill et al. (2021)) for the circular case), but these did not significantly change the results.

Fig. 19a illustrates a single stochastic simulation with a square initial condition of side length 90. Due to the healthy cells dominating the interaction we see the mutant patch size decreases over

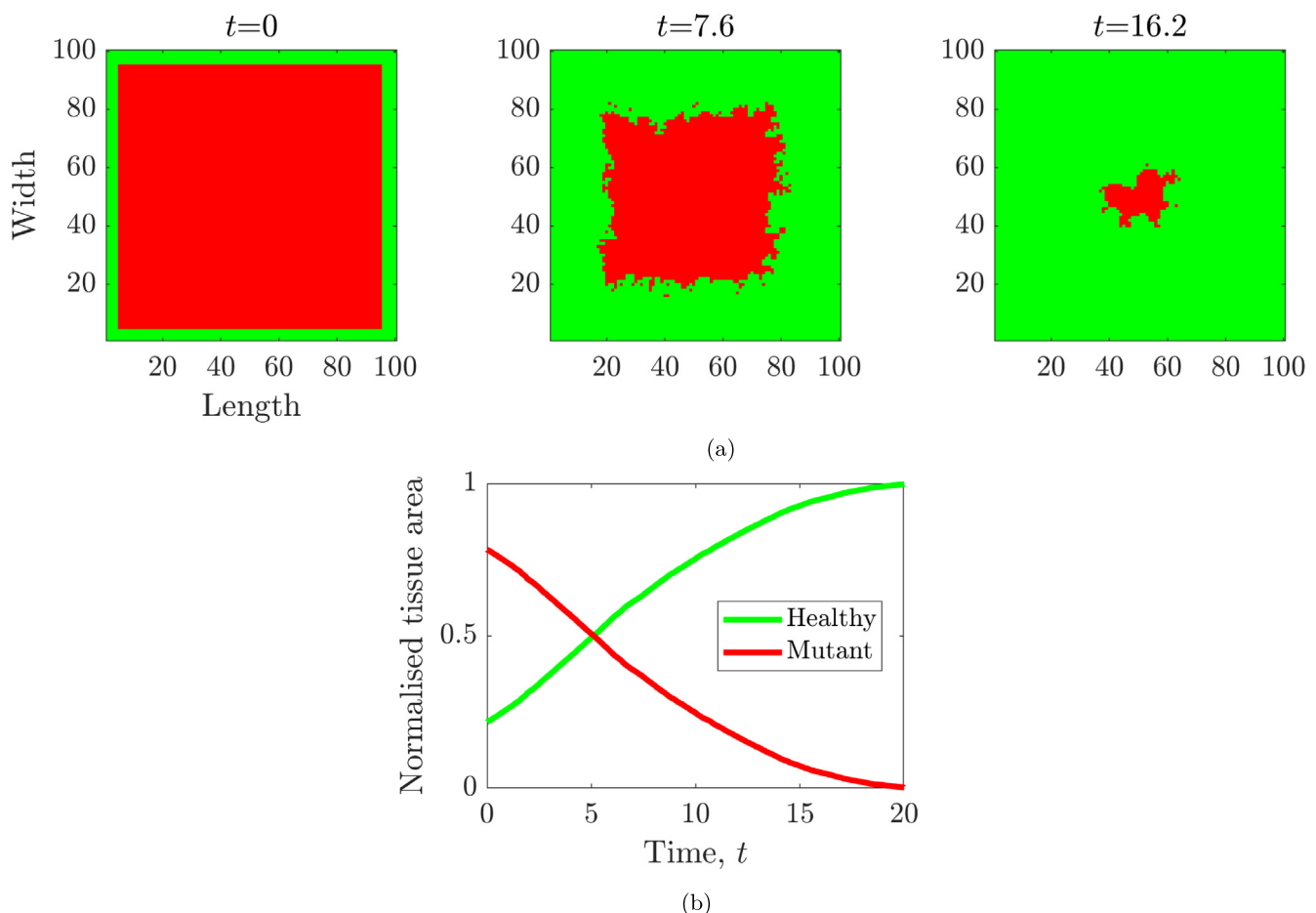


Fig. 19. Example simulation of Kras mutant cells (red) being eliminated by healthy cells (green). The domain is a 100×100 grid we start with a square initial condition of side length 90. The stochastic invasion dynamics are then simulated causing the square to degrade over time (in arbitrary units), as shown left to right in (a). In (b) we illustrate the evolving normalised population area fraction. Namely, the green and red grid spaces are tallied and normalised by the domain size of 100^2 . Since the mutant area is square the red line starts at $90^2/100^2 = 0.81$.

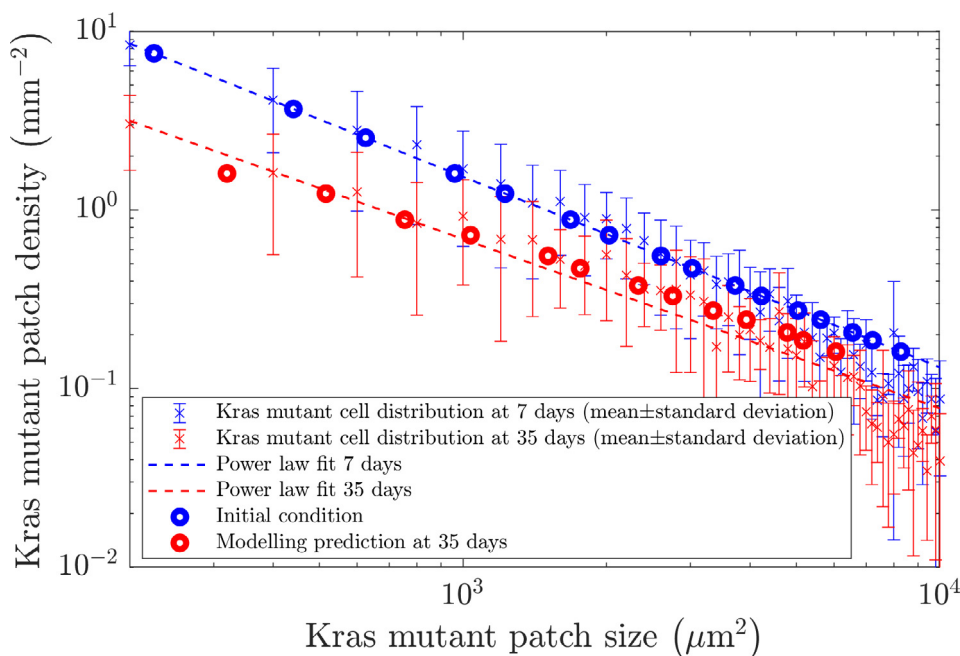


Fig. 20. Modelling results and comparative data for the density and size of Kras mutant cell patches within healthy pancreatic tissue. Note the logarithmic scales. The star arpoints and error bars represent the 7- and 35-day data, as taken from (Hill et al. (2021)). The blue and red dashed lines are equations of the form (81), which were fitted to the 7- and 35-day data sets, respectively. The blue circles are sampled from the blue dashed line and used as an initial condition for the invasion dynamics. The parameter r was then optimised so that the model output data (red circles) were the closest match to the red star points.

time, left to right. However, we also observe the influence of stochasticity. Namely, the square symmetry of the initial condition is not maintained during the expulsion of the cancerous cells. Moreover, after approximately 20 time units, the mutant cells are completely wiped out (see Fig. 19b).

The experimental set-up focused on peppering healthy pancreatic tissue with patches of cells that over expressed the protein KrasG12D. Such cells are thought to be precancerous and, hence, the body tries to eliminate them from the tissue. Amongst other statistics that were taken, the size and density of the Kras mutant patches were taken at 7 and 35 days (blue and red stars with error bars in Fig. 20, respectively).

A power law of the form

$$\text{patch density} = a(\text{patch size})^b \quad (81)$$

was fitted to these data. For the 7 day data $a = 2436$, $b = -1.07$ and $R^2 = 0.993$, where R^2 is a goodness of fit statistic known as the coefficient of determination. The R^2 value measures the proportion of the variance in the dependent variable (patch density) that is predicted by the independent variable (patch size). Specifically, $R^2 = 1$ indicates a perfect fit between data and regression curve, whilst the value reduces as the fit becomes worse (Glantz and Slinker (2012)). For the 35 day data $a = 480$, $b = -0.948$ and $R^2 = 0.972$. These power laws are indicated as blue and red dashed lines on Fig. 20.

The power law line for 7 days was used as an initial condition. Specifically, we defined a grid square to be $1\mu\text{m}^2$ and sampled along this initial distribution to choose initial patch sizes, with the correct patch density (blue circles in Fig. 20). Namely, we used 18 square patches of side length 15, 21, 25, 31, ..., 81, 85, 95. Each mutant patch was assumed to be surrounded by healthy cells (as in Fig. 19) and our stochastic invasion dynamics were applied to the system of healthy and mutant cells. Namely, any mutant cell that was orthogonally adjacent to a healthy cell was eliminated at a rate proportional to the number of its healthy neighbours, with coeffi-

cient r . As the stochastic simulations ran the size of the patches reduced, causing the blue point to shift left. Finally, the 35-day data can be used to provide a least squares best fit value of $r = 0.115\mu\text{m}^{-1}\text{day}^{-1}$. Consequently, from Fig. 20 we see that our suggested boundary elimination mechanism is consistent with the Kras mutant data.

5. Conclusions

Motivated by the pancreas' ability to eliminate precancerous cells we have developed and investigated a discrete form of domain invasion that is consistent with experimental data. Critically, this boundary elimination mechanism is parameterised by only one variable r , which scales time. Thus, r can be simply scaled out of the mathematical analysis and parameterised exactly via data over two time points, as seen in Section 4. Fitting this time scale allows us to estimate the rate at which healthy cells can clear cancerous patches. Specifically, in the case that mutant cells on the perimeter have approximately the same number of healthy cell neighbours we can say that approximately 10% of the mutant cells that are on the perimeter will be eliminated per day.

Our study is consistent with the idea that KrasG12D cells are outcompeted from healthy pancreas tissue and competition is tumour preventative (Hill et al. (2021)). However, mutations in the oncogene KRas (KRasG12D): drive the most common form of pancreatic ductal adenocarcinoma (PDAC); are detected in $> 90\%$ of human tumours (Kleeff et al. (2016)); and are required at all stages of PDAC (Collins et al. (2012)). This work indicates that KrasG12D cells must first overrule competition signals (via undefined mechanisms) to initiate and drive pancreatic cancer. Cell competition is a complex process, regulated by multiple mechanisms that are context dependent (Morata (2021); Bowling et al. (2019)). Future studies will be required to determine factors that impact on the rate at which pancreas tissues clear mutant cells from tissues. It is plausible that exogenous factors (e.g. inflamma-

tion, aging) would negatively impact on mutant cell clearance, increasing the number of oncogenic niches in tissues and disease risk. A better understanding of the biology underlying how mutant cells expand in pancreas tissues will underpin the development of new and improved early detection cancer strategies.

Beyond the biological interpretation the mathematical analysis has demonstrated a number of issues that arise from comparing stochastic and deterministic modelling approaches (Woolley et al. (2012); Woolley et al. (2011)). Specifically, a stochastically invading wave can move more than two times faster than its deterministic analogue if the two models were parameterised with the same interaction rate values. This insight stems from the two results from the paper. Firstly, unless parameters are altered with increasing width, the deterministic wave speed would not change for wider domains. Secondly, in the stochastic setting we have calculated that as the two-dimensional domain width grows the wave speed tends to 2.4 times the wave speed of the one-dimensional case. Thus, if a deterministic simulation was parameterised using one-dimensional information, but applied to the two-dimensional case then it would predict a wave that was much slower than its stochastic analogue.

Further we have also seen that assumptions of dimension can play a critical role in model definition. Namely, we have seen that the speed of invasion is directly linked to a domain's dimension, thus, we must be careful in our abstraction of a domain because one-dimensional, pseudo-one-dimensional and two-dimensional domains can act very differently in the stochastic setting, when compared to their deterministic analogues.

These mathematical insights are particularly pertinent in the case of periodic boundary conditions, which are often used in conjunction with small domains in order to make the small domain mimic the action of being embedded in a larger domain, without simulating the larger field (Woolley (2017); Woolley et al. (2017); Maini and Woolley (2019); Woolley et al. (2017)). We have illustrated that care should be taken to fully justify biological system abstraction to ensure that assumptions on dimension and boundary conditions are not severely influencing the results.

Future directions for the mathematical theory are clear in terms of extending this analysis to different boundary conditions, as suggested in Section 3.3, or extending the solution space to the third (or even higher) dimension. Critically, our work suggests that invasion waves in higher dimensions may even be faster than currently expected. However, higher dimensions would greatly slow down the stochastic simulations meaning that we would only be able to investigate small cubic grid sizes. Moreover, even in the two-dimensional case, there are still questions to be answered, e.g. how does the wave speed depend on initial condition density and order? Namely, a set of mutant cells diluted amongst wild-type cells will be eliminated much more quickly than if those cells are in a one-dimensional line.

As highlighted in the introduction there are potential alternative frameworks to view and extend this work (Moran (1958); Holley and Liggett (1975); Klein et al. (2007)). It is possible that these different frameworks would allow us an easier route to extract the dependency of the wave speed on the initial density and order (Chopp et al. (2003); Parsons et al. (2008); Antal and Scheuring (2006)). Future work could focus on comparing results of these different techniques with the approach presented here. Our hope is that old techniques may be repurposed to answer questions in different fields.

Finally, we note that, although mechanistic, the model is still phenomenological, in that it provides a simple model of juxtacrine signalling. Therefore we suggest that future experimental work be focused on understanding the signalling pathways between healthy and Kras mutant cells, because, at the moment, we can only specify what is happening in the healthy case (*i.e.* mutant cells

are expelled). As yet, we do not understand what causes this signalling to fail, which, in turn, allows the Kras mutant cells to colonise the pancreas, leading to cancer.

Codes and data

All codes and data are available at https://github.com/ThomasEWoolley/domain_invasion.

CRedit authorship contribution statement

Thomas E. Woolley: Conceptualization, Methodology, Software, Formal analysis, Visualization. **William Hill:** Validation, Data curation. **Catherine Hogan:** Conceptualization, Resources, Validation, Data curation.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

W.H. was supported by MRC DTG PhD studentship and Wellcome Trust ISSF consolidator fund. Research in C.H. lab is supported by CRUK early detection project grant to C.H. (A27838), Amser Justin Time funding (registered charity 1124951). This project started with a Pancreatic Cancer UK Research Innovation Fund (A16868) to C.H. and was supported by European Cancer Stem Cell Research Institute (Cardiff University) fellowship to C.H. We thank all members of C.H. lab (past and present) and co-authors for their contributions to the original *in vivo* study.

References

- Alberts, B., Bray, D., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, F., 2004. *Essential cell biology*. Garland Sci.
- Brown, S., Pineda, C.M., Xin, T., Boucher, J., Suozzi, K.C., Park, S., Matte-Martone, C., Gonzalez, D.G., Rytlewski, J., Beronja, S., Greco, V., 2017. Correction of aberrant growth preserves tissue homeostasis. *Nature* 548 (7667), 334–337.
- Qiu, Y., Chen, W., Nie, Q., 2019. Stochastic dynamics of cell lineage in tissue homeostasis. *Discrete and continuous dynamical systems. Series B* 24 (8), 3971.
- Picco, N., Hippenmeyer, S., Rodarte, J., Streicher, C., Molnár, Z., Maini, P.K., Woolley, T.E., 2019. A mathematical insight into cell labelling experiments for clonal analysis. *J. Anat.* 235 (3), 687–696.
- Boland, C.R., Ricciardiello, L., 1999. How many mutations does it take to make a tumor? *Proc. Nat. Acad. Sci.* 96 (26), 14675–14677.
- Hogan, C., Dupré-Crochet, S., Norman, M., Kajita, M., Zimmermann, C., Pelling, A.E., Piddini, E., Baena-López, L.A., Vincent, J.-P., Itoh, Y., Hosoya, H., Pichaud, F., Fujita, Y., 2009. Characterization of the interface between normal and transformed epithelial cells 11, 460–467.
- Hill, W., Zaragkoulias, A., Salvador-Barbero, B., Parfitt, G.J., Alatsianos, M., Padilha, A., Porzinski, S., Woolley, T.E., Morton, J.P., Sansom, O.J., Hogan, C., 2021. EphA2-dependent outcompetition of KRASG12D mutant cells by wild-type neighbors in the adult pancreas. *Curr. Biol.*
- M. Kajita, C. Hogan, A.R. Harris, S. Dupre-Crochet, N. Itasaki, K. Kawakami, G. Charras, M. Tada, and Y. Fujita. Interaction with surrounding normal epithelial cells influences signalling pathways and behaviour of Src-transformed cells. 123: 171–180, 2010.**.
- Porzinski, S., De Navascues, J., Yako, Y., Hill, W., Jones, M.R., Maddison, R., Fujita, Y., Hogan, C., 2016. EphA2 drives the segregation of Ras-transformed epithelial cells from normal neighbors. *Curr. Biol.* 26 (23), 3220–3229.
- Vishwakarma, M., Piddini, E., 2020. Outcompeting cancer 20, 187–198.
- Pray, L., 2008. DNA replication and causes of mutation. *Nature Educ.* 1 (1), 214.
- Hill, W., Hogan, C., 2019. Normal epithelial cells trigger EphA2-dependent RasV12 cell repulsion at the single cell level. *Small GTPases* 10 (4), 305–310.
- Tamori, Y., Deng, W.-M., 2013. Tissue repair through cell competition and compensatory cellular hypertrophy in postmitotic epithelia. *Developmental Cell* 25 (4), 350–363.
- Eden, M., 1961. A two-dimensional growth process. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, vol. 4, pages 223–239, 1961.

- Wolf, D.E., Kertész, J., 1987. Noise reduction in Eden models. I. *J. Phys. A: Math. Gen.* 20 (4), L257.
- Kertész, J., Wolf, D.E., 1988. Noise reduction in Eden models: II. Surface structure and intrinsic width. *J. Phys. A: Math. Gen.* 21 (3), 747.
- Debabrata, P., 2004. Effects of fluctuations on propagating fronts. *Phys. Rep.* 393 (2), 87–174.
- Callaghan, T., Khain, E., Sander, L.M., Ziff, R.M., 2006. A stochastic model for wound healing. *J. Stat. Phys.* 122 (5), 909–924.
- Clifford, P., Sudbury, A., 1973. A model for spatial conflict. *Biometrika* 60 (3), 581–588.
- Woolley, T.E., Gaffney, E.A., Waters, S.L., Oliver, J.M., Baker, R.E., Goriely, A., 2014. Three mechanical models for blebbing and multi-blebbing. *IMA J. Appl. Math.* 79 (4), 636–660.
- Belmonte-Beitia, J., Woolley, T.E., Scott, J.G., Maini, P.K., Gaffney, E.A., 2013. Modelling biological invasions: Individual to population scales at interfaces. *J. Theor. Biol.* 334, 1–12.
- Moran, P.A.P., 1958. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* 54, 60–71.
- Whigham, P.A., Dick, G., 2008. Evolutionary dynamics for the spatial Moran process. *Genet. Program. Evolvable Mach.* 9 (2), 157–170.
- Holley, R.A., Liggett, T.M., 1975. Ergodic theorems for weakly interacting infinite systems and the voter model. *Ann. Probab.*, 643–663
- Bramson, M., Griffeath, D., 1980. Asymptotics for interacting particle systems on Z^d . *Z. Wahrscheinlichkeit.* 53 (2), 183–196.
- Klein, A.M., Doupé, D.P., Jones, P.H., Simons, B.D., 2008. Mechanism of murine epidermal maintenance: Cell division and the voter model. *Phys. Rev. E* 77 (3) 031907.
- Jörg, D.J., Kitadate, Y., Yoshida, S., Simons, B.D., 2021. Stem cell populations as self-renewing many-particle systems. *Annu. Rev. Condensed Matter Phys.* 12, 135–153.
- Klein, A.M., Doupé, D.P., Jones, P.H., Simons, B.D., 2007. Kinetics of cell division in epidermal maintenance. *Phys. Rev. E* 76 (2) 021910.
- Hall, C.L., Siebert, B.A., 2021. Exact solutions and bounds for network SIR and SEIR models using a rooted-tree approximation. *arXiv, math.DS* (2108.07128).
- R. Erban, S.J. Chapman, and P.K. Maini. A practical guide to stochastic simulations of reaction-diffusion processes. *arXiv:0704.1908v2 [q-bio.SC]*, 2007.**
- van Kampen, N.G., 2007. *Stochastic Processes in Physics and Chemistry*. North Holland.
- Murray, J.D., 2003. *Mathematical Biology I: An Introduction*, vol. 1. Springer-Verlag, third ed.
- Coleman, M.P., 2013. *An Introduction to Partial Differential Equations with MATLAB*. Chapman and Hall/CRC.
- Stanoyevitch, A., 2005. *Introduction to Numerical Ordinary and Partial Differential Equations using MATLAB*. Wiley-Interscience.
- Gillespie, D.T., 1976. A general method for numerically simulating stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22, 403–434.
- Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81 (25), 2340–2361.
- Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.* 58 (1), 35–55.
- Woolley, T.E., Baker, R.E., Gaffney, E.A., Maini, P.K., 2011. Power spectra methods for a stochastic description of diffusion on deterministically growing domains. *Phys. Rev. E* 84 (2) 021915.
- Woolley, T.E., Baker, R.E., Gaffney, E.A., Maini, P.K., 2011. Influence of stochastic domain growth on pattern nucleation for diffusive systems with internal noise. *Phys. Rev. E* 84 (4) 041905.
- Abdennur, N., 2022. Gillespie Stochastic Simulation Algorithm. GitHub repository. [url:https://github.com/invictus/Gillespie](https://github.com/invictus/Gillespie).
- Guldberg, C.M., Waage, P., 1879. Concerning chemical affinity. *J. Prakt. Chem.* 19 (2), 69–114.
- Jahnke, T., Huisinga, W., 2007. Solving the chemical master equation for monomolecular reaction systems analytically. *J. Math. Biol.* 54 (1), 1–26.
- Woolley, T.E., 2011. *Spatiotemporal Behaviour of Stochastic and Continuum Models for Biological Signalling on Stationary and Growing Domains*. PhD thesis, University of Oxford.
- Schumacher, L.J., Woolley, T.E., Baker, R.E., 2013. Noise-induced temporal dynamics in Turing systems. *Phys. Rev. E* 87 (4) 042719.
- Picco, N., Garcia-Moreno, F., Maini, P.K., Woolley, T.E., Molnár, Z., 2018. Mathematical modeling of cortical neurogenesis reveals that the founder population does not necessarily scale with neurogenic output. *Cereb. Cortex*.
- Murray, J.D., 2003b. *Mathematical Biology II: Spatial Models and Biomedical Applications*, volume 2. Springer-Verlag, third ed.
- Sloane, N.J.A., Plouffe, S., 1995. *The Encyclopedia of Integer Sequences*. Academic Press.
- Courant, R., Robbins, H., 1996. *What Is Mathematics? An Elementary Approach to Ideas and Methods*. Oxford University Press.
- Corless, R.M., Jeffrey, D.J., Knuth, D.E., 1997. A sequence of series for the Lambert W function. In: *In Proceedings of the 1997 international symposium on Symbolic and algebraic computation*, pp. 197–204.
- Mező, I., Keady, G., 2016. Some physical applications of generalized Lambert functions. *Eur. J. Phys.* 37 (6) 065802.
- Fukushima, T., 2013. Precise and fast computation of Lambert W-functions without transcendental function evaluations. *J. Comput. Appl. Math.*, 244: 77–89.
- Asl, F.M., Ulsoy, A.G., 2000. Analytical solution of a system of homogeneous delay differential equations via the Lambert function. In *Proceedings of the 2000 American Control Conference*. ACC (IEEE Cat. No. 00CH36334), volume 4, pp. 2496–2500.
- Hingorani, S.R., Petricoin, E.F., Maitra, A., Rajapakse, V., King, C., Jacobetz, M.A., Ross, S., Conrads, T.P., Veenstra, T.D., Hitt, B.A., Kawaguchi, Y., Johann, D., Liotta, L.A., Crawford, H.C., Putt, M.E., Jacks, T., Wright, C.V.E., Hruban, R.H., Lowy, A.M., Tuveson, D.A., 2003. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* 4 (6), 437–450.
- Morton, J.P., Timpson, P., Karim, S.A., Ridgway, R.A., Athineos, D., Doyle, B., Jamieson, N.B., Oien, K.A., Lowy, A.M., Brunton, V.G., Frame, M.C., Evans, T.R.J., Sansom, O.J., 2010. Mutant p53 drives metastasis and overcomes growth arrest/senescence in pancreatic cancer. *Proc. Nat. Acad. Sci.* 107 (1), 246–251.
- Glantz, S.A., Slinker, B.K., 2012. *Primer of Applied Regression & Analysis of Variance*. McGraw-Hill Education.
- Kleeff, J., Korc, M., Apte, M., La Vecchia, C., Johnson, C.D., Biankin, A.V., Neale, R.E., Tempero, M., Tuveson, D.A., Hruban, R.H., Neoptolemos, J.P., 2016. Pancreatic cancer. *Nat. Rev. Dis. Primers* 2 (1), 1–22.
- Collins, M.A., Bednar, F., Zhang, Y., Brisset, J.-C., Galbán, S., Galbán, C.J., Rakshit, S., Flannagan, K.S., Adsay, N.V., di Magliano, M.P., 2012. Oncogenic Kras is required for both the initiation and maintenance of pancreatic cancer in mice. *J. Clin. Invest.* 122 (2), 639–653.
- Morata, G., 2021. Cell competition: A historical perspective. *Dev. Biol.*
- Bowling, S., Lawlor, K., Rodríguez, T.A., 2019. Cell competition: the winners and losers of fitness selection. *Development* 146 (13), dev167486.
- Woolley, T.E., Baker, R.E., Gaffney, E.A., Maini, P.K., Seirin-Lee, S., 2012. Effects of intrinsic stochasticity on delayed reaction-diffusion patterning systems. *Phys. Rev. E* 85 (5) 051914.
- Woolley, T.E., Baker, R.E., Gaffney, E.A., Maini, P.K., 2011. Stochastic reaction and diffusion on growing domains: Understanding the breakdown of robust pattern formation. *Phys. Rev. E* 84 (4) 046216.
- Woolley, T.E., 2017. Pattern production through a chiral chasing mechanism. *Phys. Rev. E* 96 (3) 032401.
- Woolley, T.E., Baker, R.E., Maini, P.K., 2017. *The Turing Guide*, chapter 35: Turing's theory of morphogenesis. Oxford Univ Press.
- Maini, P.K., Woolley, T.E., 2019. *The Turing Model for Biological Pattern Formation*. Springer, pp. 189–204.
- Woolley, T.E., Baker, R.E., Maini, P.K., 2017. *The Incomputable*, chapter Turing's theory of morphogenesis: where we started, where we are and where we want to go. Springer, pp. 219–235.
- Chopp, D.L., Kirisits, M.J., Moran, B., Parsek, M.R., 2003. The dependence of quorum sensing on the depth of a growing biofilm. *Bull. Math. Biol.* 65 (6), 1053–1079.
- Parsons, T.L., Quince, C., Plotkin, J.B., 2008. Absorption and fixation times for neutral and quasi-neutral populations with density dependence. *Theor. Popul. Biol.* 74 (4), 302–310.
- Antal, T., Scheuring, I., 2006. Fixation of strategies for an evolutionary game in finite populations. *Bull. Math. Biol.* 68 (8), 1923–1944.