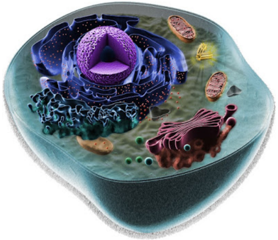
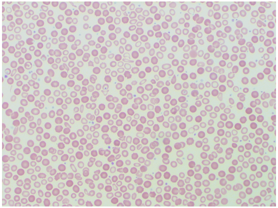
Stochastic modeling for the single cell and the cell population: considerations for data-driven methodologies





Stefan Engblom

Div of Scientific Computing, Dept of Information Technology, Uppsala University

Systems Biology Seminar, Stuttgart, Germany, November 7th, 2019

Outline

Intro: data for inspiration & the modeling challenge

1. Computational modeling...

2. ...numerical analysis

3. Worked examples

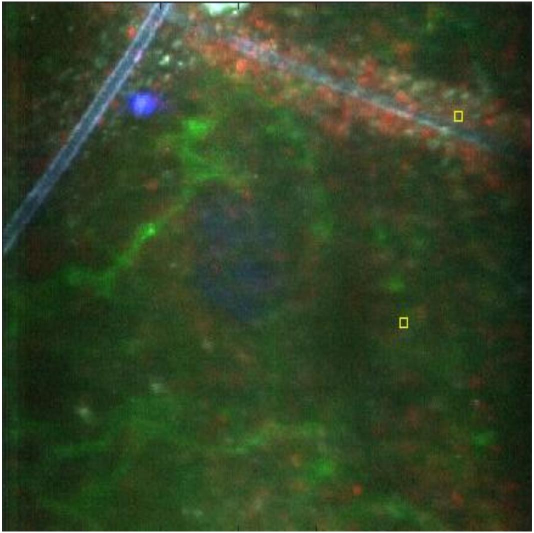
Summary

Joint work with and/or input from:

- **Mia Phillipson**, **Gustaf Christoffersson**, **Femke Heindryckx** @ Medical Cell Biology, Uppsala university
- **Ruth Baker**, **Dan Wilson** @ Math Institute, University of Oxford
- **Augustin Chevallier** @ ENS Cachan/INRIA Sophia Antipolis
  Jonas R. Umaras @ Scientific computing, Uppsala university

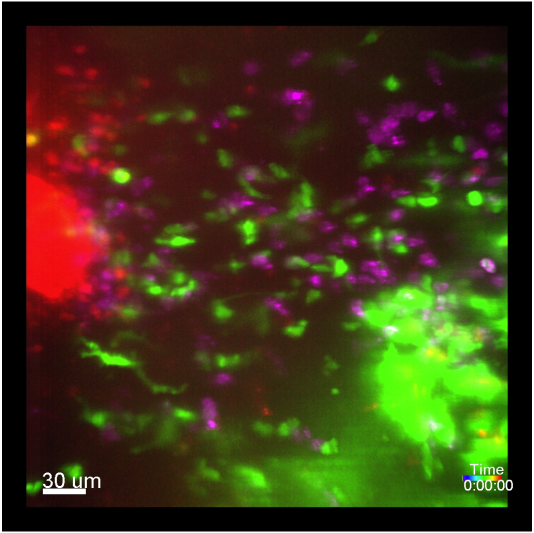
Wound healing around transplant

Recruitment and coordination of white blood-cells



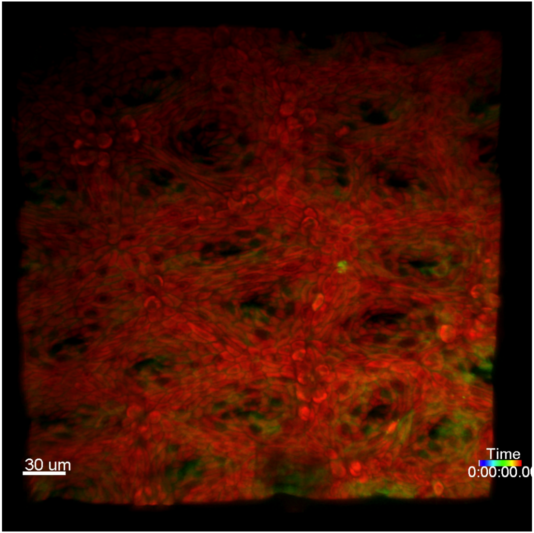
Migrating cells

Sensing gradients (lactic acid)



Colon crypts

Stem cells coordination in a noisy environment



Quorum sensing

Synthetic circuit *in vivo* from Danino, *et al.*, Nature 463, 2010

The modeling challenge

"How to think"

Aim: **realistic** and **useful** computational models of populations of living cells.

The modeling challenge

"How to think"

Aim: **realistic** and **useful** computational models of populations of living cells.

"Realistic" flexible and understandable (= analyzable) numerical models, that in perspective can incorporate all relevant processes

The modeling challenge

"How to think"

Aim: **realistic** and **useful** computational models of populations of living cells.

"Realistic" flexible and understandable (= analyzable) numerical models, that in perspective can incorporate all relevant processes

"Useful" (1) explanatory (incl. emergent behavior), (2) test hypotheses, (3) predictive value, (4) help to build an argument in cases where many factors are unknown

The modeling challenge
"How to think"

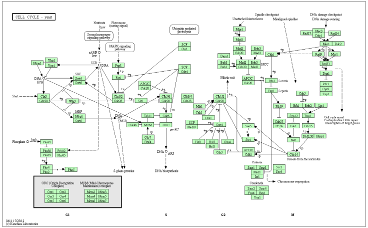
Aim: **realistic** and **useful** computational models of populations of living cells.

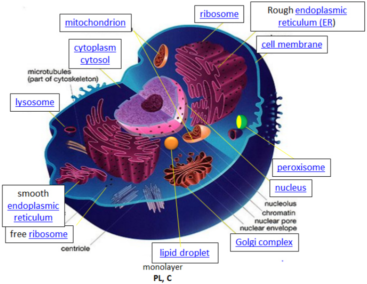
- "Realistic" flexible and understandable (= analyzable) numerical models, that in perspective can incorporate all relevant processes
- "Useful" (1) explanatory (incl. emergent behavior), (2) test hypotheses, (3) predictive value, (4) help to build an argument in cases where many factors are unknown

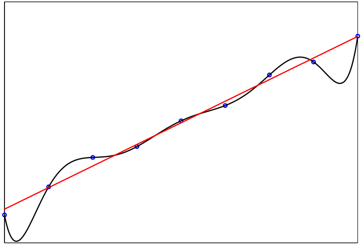
(1) is about modeling consistency & power, (2)+(3)+(4) mainly about being able to incorporate data *and* about simulation performance

Risk of over-modeling

"...help to build an argument in cases where many factors are unknown..."







Caution:

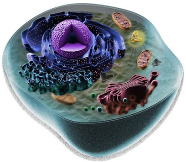
- *really* detailed, or,
- *imaginary* accuracy, or,
- just a plain overfit?

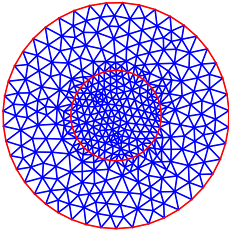
Rest of the talk

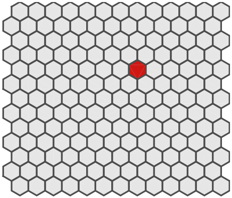
1. Computational modeling: aim for a single scalable framework
2. Analysis in that framework: propagation of uncertainties & errors
3. Illustrations

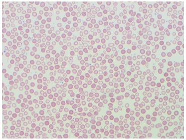
Computational modeling

inner-outer idea





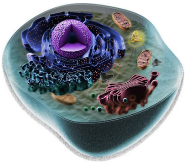
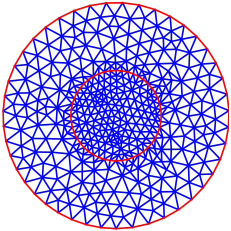
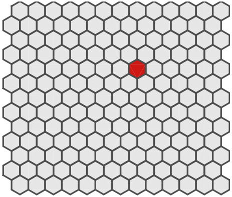
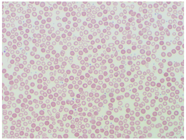




Immediate idea: one type of model describing an individual cell ("inner scale"), coupled together with a population level model ("outer scale").

Computational modeling

inner-outer idea

Immediate idea: one type of model describing an individual cell ("inner scale"), coupled together with a population level model ("outer scale").

Challenge: the aim is a single (analyzable) framework. So: {inner workings of singel cells, sensatory input/output, extracellular space, population mechanics, ...} — also *fast*!

A single framework

Properties

Real-world property	Model implication
"noisy"	stochastic
species discreteness	discrete state
spatial inhomogeneous	grid-based

A single framework

Properties

Real-world property	Model implication
"noisy"	stochastic
species discreteness	discrete state
spatial inhomogeneous	grid-based

The RDME

-A spatial continuous-time Markov chain stand out as a promising alternative. This is the Reaction-Diffusion Master Equation, (a kind of "discretized SPDE").

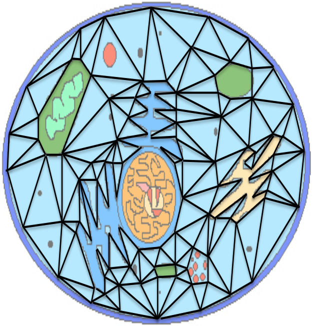
The idea 1

inner scale: RDME

Inside a cell, reactions and diffusion of various molecules take place.

The rates for these events determines *what* happens and *when* in a stochastic, event-driven simulation.

repeat
  pick a random number
  sample what happens and when
  execute this event
until done



www.urdme.org

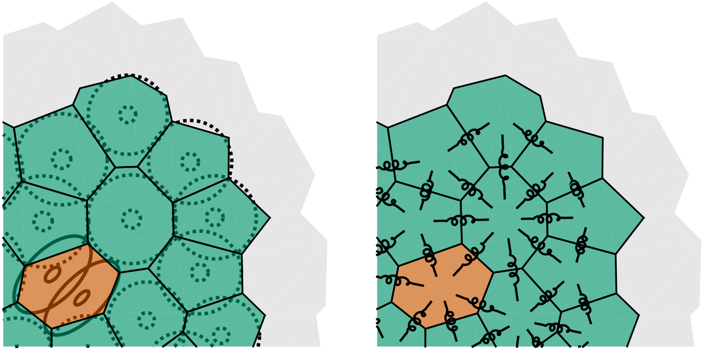
A single framework?

-Cells are also discrete noisy objects, occuping space. Is there a "cell-population RDME"?

-Differences: cells move due to (1) mechanics/pushing, (2) active movements/crawling, and (3) experience adhesion.

The idea 2

outer scale



Cellular pressure, propagated by a connecting spring model. The “flow” of cells is driven by a gradient in this pressure (Darcy’s law).

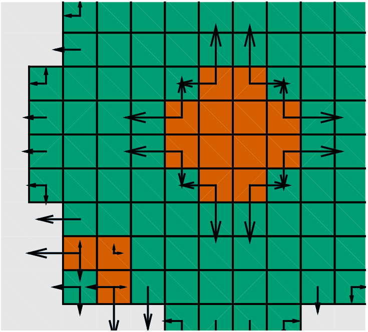
The idea 2

outer scale: DLCM

From three basic assumptions:

1. thermal movements are ignored
2. rapid equilibrium of pressure
3. movements only into less crowded voxels

one derives a (discrete) Laplacian with certain BCs and source terms. Hence rates... hence events in continuous time.



"Discrete Laplacian Cell Mechanics" (DLCM).
"Darcy's Law Cell Mechanics"...

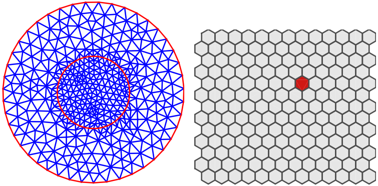
Coupling of scales

Observation #1: since both the inner scale and the outer scale are formed in continuous time, there is *one and only one* way of correctly coupling them together.

Coupling of scales

Observation #1: since both the inner scale and the outer scale are formed in continuous time, there is *one and only one* way of correctly coupling them together.

Observation #2: the two types of models can be expected to take place at different temporal scales. *Approximation:* evolve the inner scales one step in time (e.g., in parallel), then connect at the outer scale.



-*In fact*, one can think of all sorts of computational tricks like this. Often: accept a small(?) error for computational efficiency.

Analysis message

Terms & conditions

Want to use these models when any combination of

- stochasticity
- nonlinearity
- species discreteness
- spatial inhomogeneities

makes a difference. The model itself is therefore likely going to be sensitive to perturbations in any of the above.

$\Longrightarrow$ A computational framework should allow for error estimates of useful approximations.

A priori

Long story, but short

Notation: $\mathbb{X}_{ij}$ =#molecules of species i in voxel j (RDME, but a similar notation for the DLCM works too), $\|\mathbb{X}\|^2 \equiv \sum_{i,j} \mathbb{X}_{ij}^2$.

A priori

Long story, but short

Notation: $\mathbb{X}_{ij}$ =#molecules of species i in voxel j (RDME, but a similar notation for the DLCM works too), $\|\mathbb{X}\|^2 \equiv \sum_{i,j} \mathbb{X}_{ij}^2$.

$\Longrightarrow$ *a priori*: with suitable initial data and under certain assumptions on the model formulation and the rates, one can show that the problem is strongly well-posed, i.e., $\mathbb{X}$ exists and behaves well.

- $\mathbb{E}[\sup_{s \in [0,t]} \|\mathbb{X}(s)\|^p]$ bounded, any $p \geq 1$
- if $\mathbb{X}(0) = \mathbb{Y}(0)$ a.s., and if $\mathbb{Y}(t)$ is obtained by δ-perturbing the rate intensities ($r \to (1 \pm \delta)r$), then

$$\lim_{\delta \to 0} \mathbb{E}[\|\mathbb{X}(t) - \mathbb{Y}(t)\|^2] = 0.$$

Analysis: Multiscale variable splitting

Set-up: ϵ, h

Consider a separation of scales:

- species are either abundant $\sim \epsilon^{-1}$, or appear in low copy numbers ~ 1
- rate constants are either fast ~ 1, or slow ϵ

$\Longrightarrow$ rescaled variable $\bar{\mathbb{X}}(t) \sim 1$.

Analysis: Multiscale variable splitting

Set-up: ϵ, h

Consider a separation of scales:

- species are either abundant $\sim \epsilon^{-1}$, or appear in low copy numbers ~ 1
- rate constants are either fast ~ 1, or slow ϵ

$\Longrightarrow$ rescaled variable $\bar{\mathbb{X}}(t) \sim 1$.

Multiscale splitting methods:

"Hybrid", $\bar{\mathbb{Y}}(t)$ all stochastic processes driving an abundant species are replaced with mean drift terms, a "deterministic-stochastic hybrid"

"Numerical", $\bar{\mathbb{Y}}^{(h)}(t)$ discrete step h; low copy number variables are first simulated in $[t, t+h)$ letting abundant species be frozen at time t, next abundant species are integrated in $[t, t+h)$

Analysis of errors

Results

For certain explicit exponents (u, v)...

Multiscale error

Under certain assumptions,

- $\mathbb{E}[\|\bar{\mathbb{Y}}(t) - \bar{\mathbb{X}}(t)\|^2] = O(\epsilon^{1+v} + \epsilon^{1/2+v/2+u})$

Analysis of errors

Results

For certain explicit exponents (u, v)...

Multiscale error

Under certain assumptions,

- $\mathbb{E}[\|\bar{\mathbb{Y}}(t) - \bar{\mathbb{X}}(t)\|^2] = O(\epsilon^{1+v} + \epsilon^{1/2+v/2+u})$

Time-discretization error

Under the same assumptions, then if the processes are bounded,

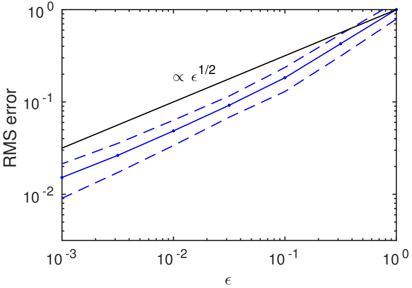
- $\mathbb{E}[\|\bar{\mathbb{Y}}^{(h)}(t) - \bar{\mathbb{Y}}(t)\|^2] = O\left(h(\epsilon^{2u} + \epsilon^{u+v})\right) + O\left(h^2\epsilon^{2v}\right)$

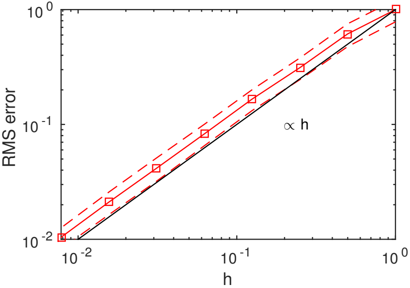
Example: catalytic process

"Stress test" of theory

$(A, C) \sim \epsilon^{-1}$, $(B, D) \sim 1$, diffusion$_{A,C} \sim \epsilon$, diffusion$_{B,D} \sim 1$.

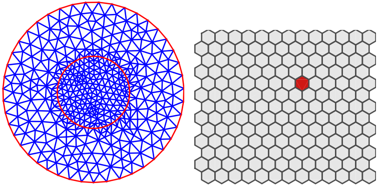
$$\begin{array}{lcl}
A + B \xrightarrow{kAB} C + B & \qquad & A \underset{b_a}{\overset{\epsilon d_a A}{\rightleftharpoons}} \emptyset \\
C + D \xrightarrow{kCD} A + D & \qquad & B \underset{b_b}{\overset{d_b AB}{\rightleftharpoons}} \emptyset \\
B + B \underset{k_d D}{\overset{k_b B(B-1)}{\rightleftharpoons}} D & &
\end{array}$$





Proposed modeling framework

RDME & DLCM



Outer scale DLCM, pressure-driven (passive) cellular movements

Inner scale ODEs, SDEs, or the RDME for the highest resolution

-*Clearly doable:* analyze an inner/outer RDME/DLCM split-step method following the outlined RDME theory.

Cellular communication: Notch Delta

Classical model from Coller *et al.* J. theor. Biol. 183, 1996

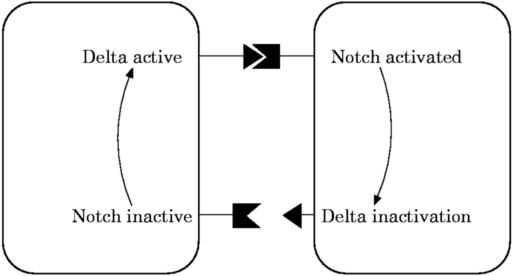
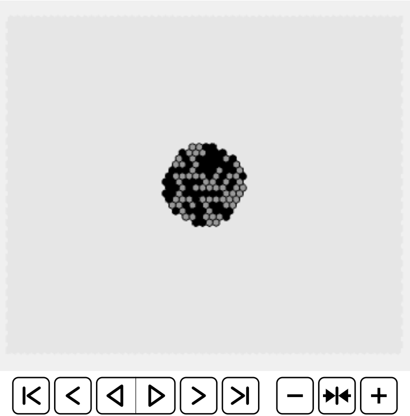


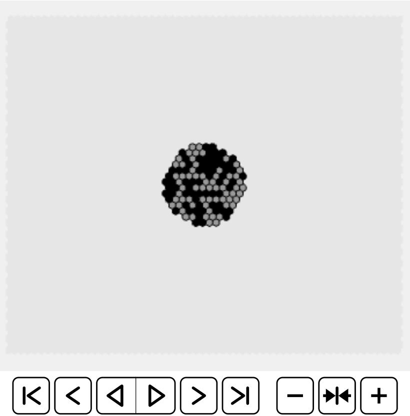
FIG. 1. Diagrammatic representation of the effective feedback loop between Notch and Delta in neighbouring cells. Details of the Notch signalling pathway are omitted for clarity. *Key*: Delta; Notch.

-One cell develops high Notch, the other low Notch (black/white patterning).

Cellular communication: Notch Delta

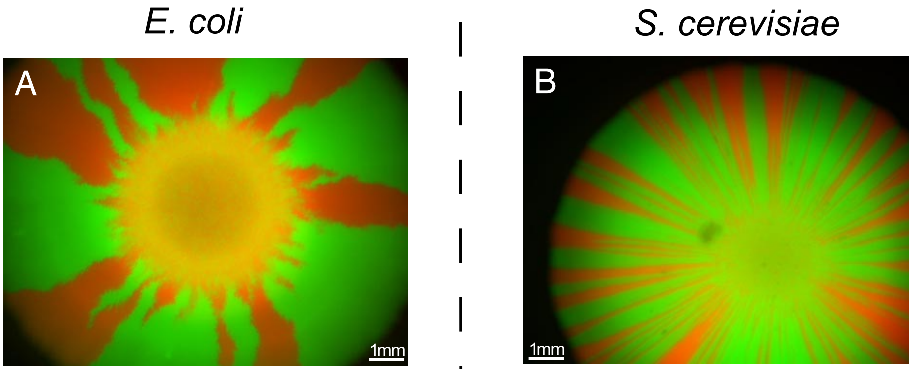
Inner scale: ODE, outer scale: spatial stochastic





Pattern formation 1: colonization

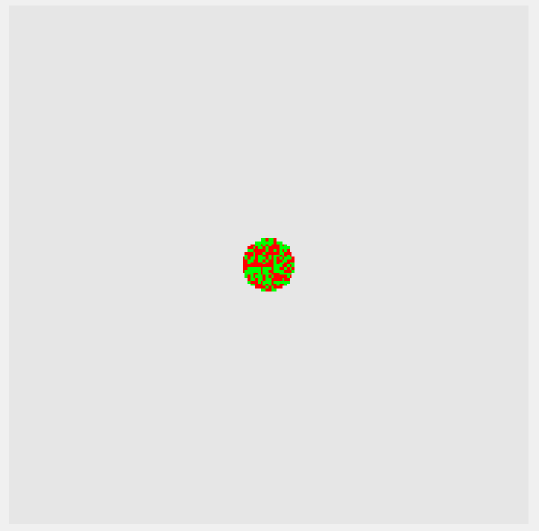
In vitro results from Hallatschek, *et al.*, PNAS 104, 2007



-Through colonization the red/green gene wins.

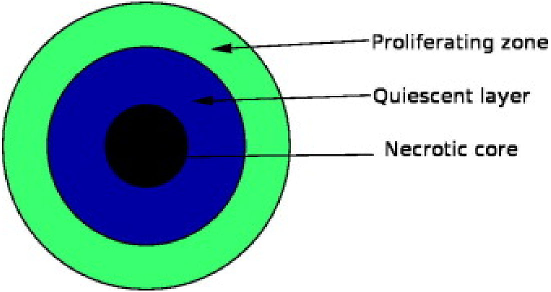
In silico colonization

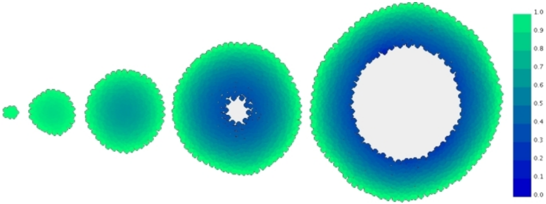
Inner scale: non-spatial stochastic, outer scale: spatial stochastic



Non-trivial dynamics in tumour

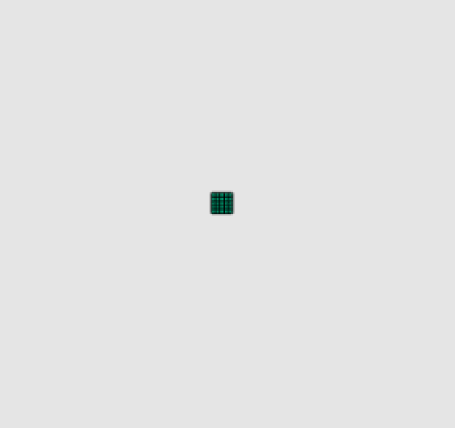
Mambili-Mamboundou *et al.*, Math. Bio. 249, 2014, & *Chaste*





Non-trivial dynamics in tumour

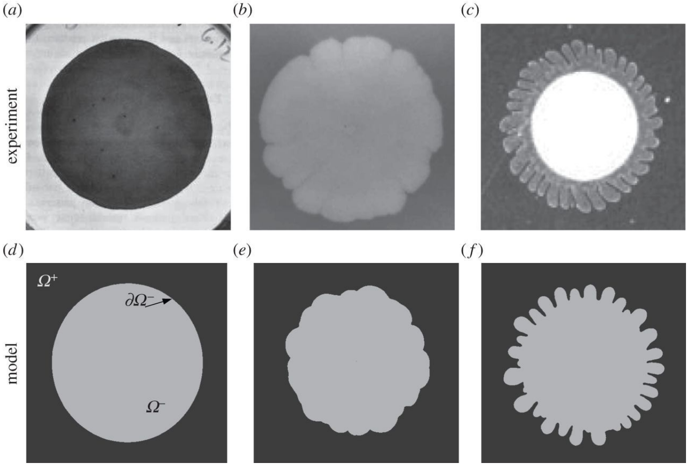
Inner scale: non-spatial stochastic, outer scale: spatial stochastic



-Finding (emergent behavior): increasing the surface means increasing oxygen intake $\Longrightarrow$ steady-state is unstable.

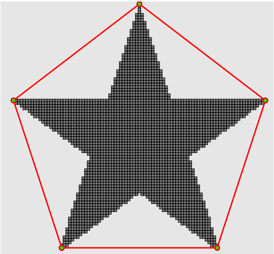
Sidenote: instability

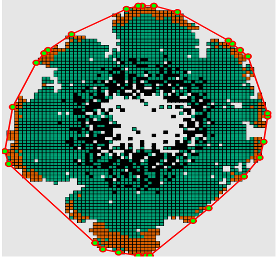
In vitro and *in silico* results from Giverso, *et al.*, J. R. Soc. Interface 12, 2015

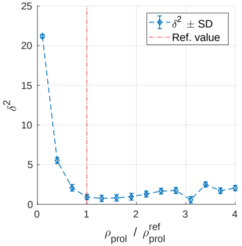


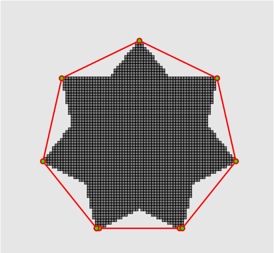
Ongoing work...

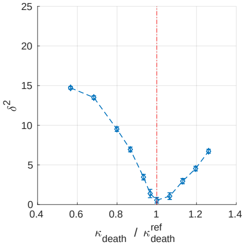
ABC parameter inversion of tumour model





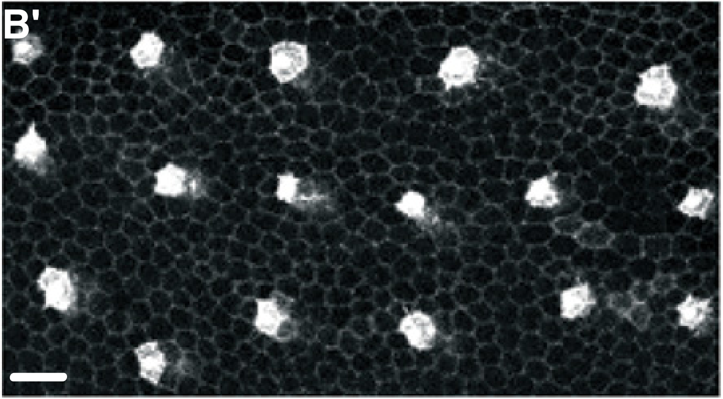






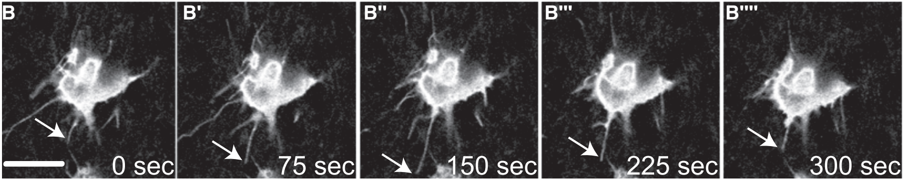
Pattern formation 2: Notch Delta & protrusions

In vivo results from Cohen, *et al.*, Cell 19, 2010



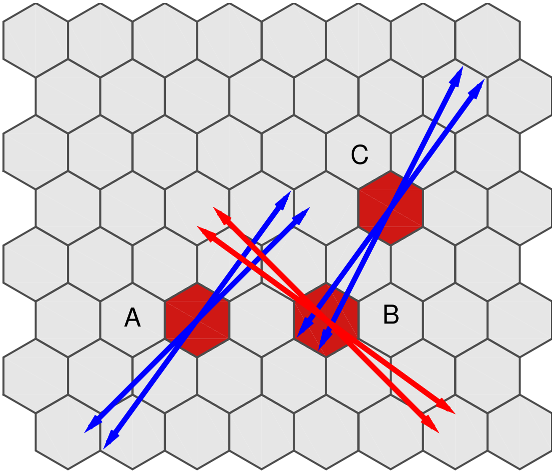
Protrusions

In vivo results from Cohen, *et al.*, Cell 19, 2010



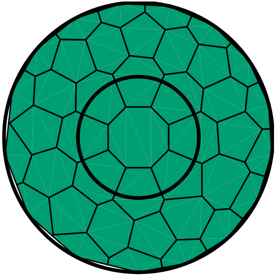
Protrusion interactions model

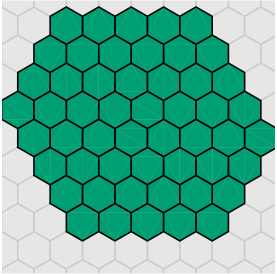
In silico model from Hadjivasiliou, *et al.*, J. R. Soc. Interface 13, 2016

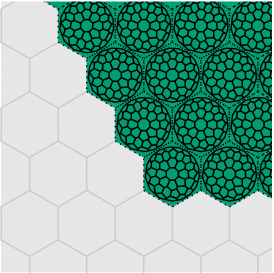


Direct (neighbor↔neighbor), via protrusions (A↔B), and non-symmetric (B↔C).

Spatial discretization







Left: single cell discretization, *middle:* cell population layer, *right:* grids combined.

Notch Delta: differential weighting of signals

Inner scale: spatial stochastic, outer scale: spatial stochastic

Summary

- Microscopy data, mostly for inspiration...
- "How to think": realistic & useful models, through flexible/understandable/generalizable

Summary

- Microscopy data, mostly for inspiration...
- "How to think": realistic & useful models, through flexible/understandable/generalizable
- 1. Modeling: inner/outer scale, RDME/DLCM one suitable such combination, consistency through time-continuous coupling, **event-based computational framework** *(fast!)*

Summary

- Microscopy data, mostly for inspiration...
- "How to think": realistic & useful models, through flexible/understandable/generalizable
- 1. Modeling: inner/outer scale, RDME/DLCM one suitable such combination, consistency through time-continuous coupling, **event-based computational framework** *(fast!)*
- 2. Analysis: the RDME framework, stability, analysis of basic numerical methods, *doable*: bring this to the RDME/DLCM combination.

Summary

- Microscopy data, mostly for inspiration...
- "How to think": realistic & useful models, through flexible/understandable/generalizable
- 1. Modeling: inner/outer scale, RDME/DLCM one suitable such combination, consistency through time-continuous coupling, **event-based computational framework** *(fast!)*
- 2. Analysis: the RDME framework, stability, analysis of basic numerical methods, *doable*: bring this to the RDME/DLCM combination.
- 3. Examples: flexible coupling cell-to-cell/cell-to-environment (solutions in URDME @ GitHub, `www.urdme.org`)

Thanks

Programs, Papers, and Preprints are available from my web-page.
Thank you for the attention!