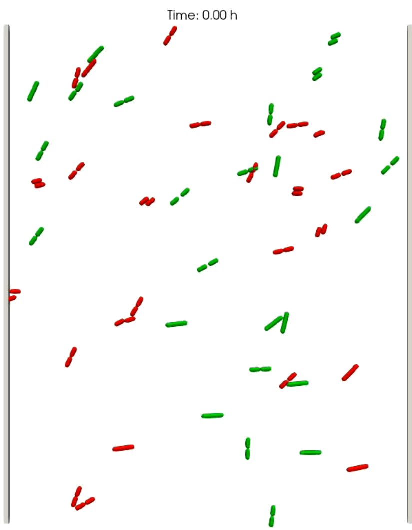
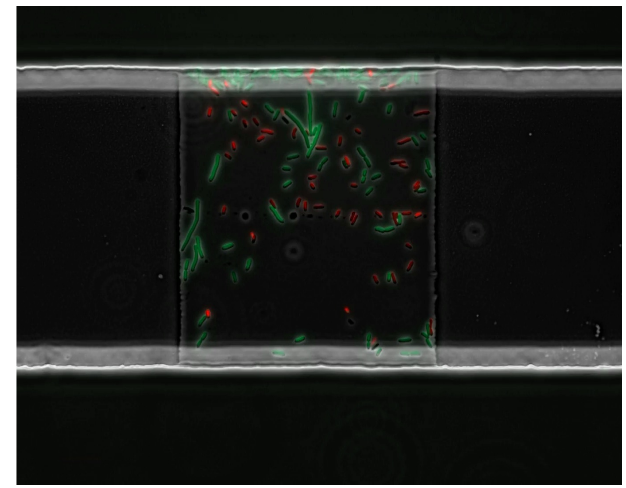
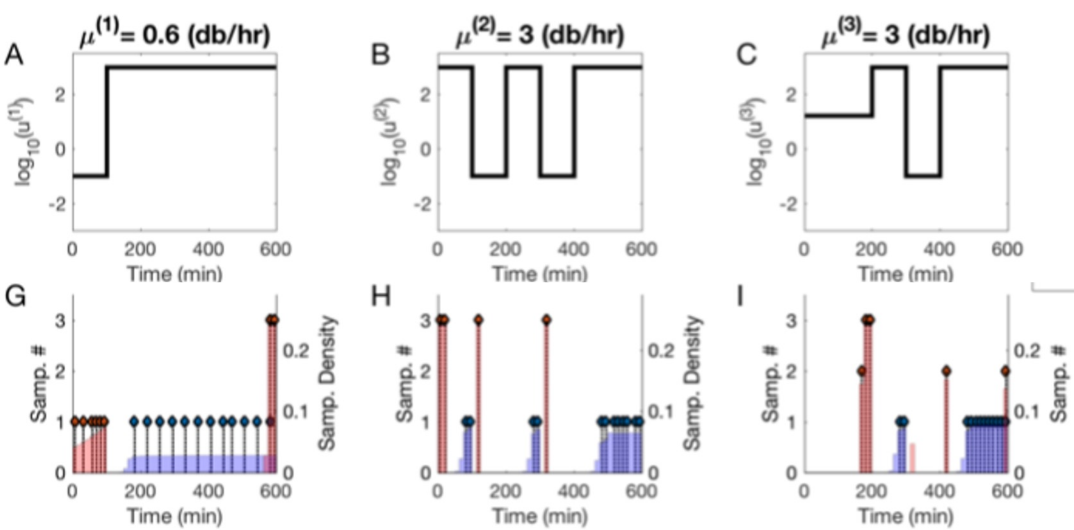
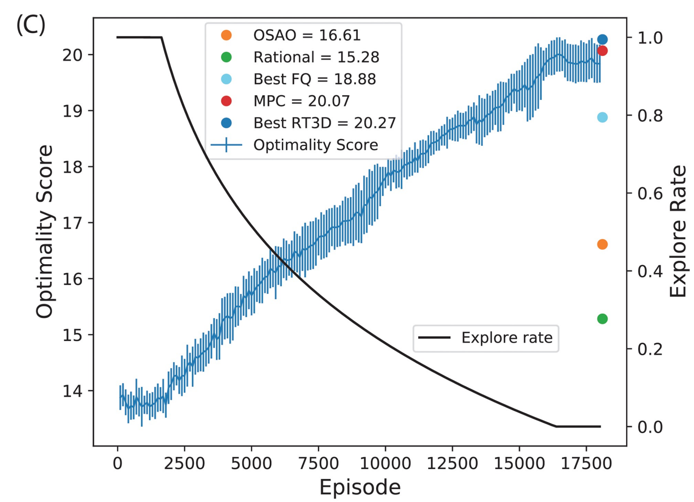
Challenges and opportunities for model calibration



Brian Ingalls (he/him)
Department of
Applied Mathematics
University of Waterloo
Waterloo, Canada
@bpingalls







Outline

1) Calibration strategies for agent-based population models of mixed bacterial populations

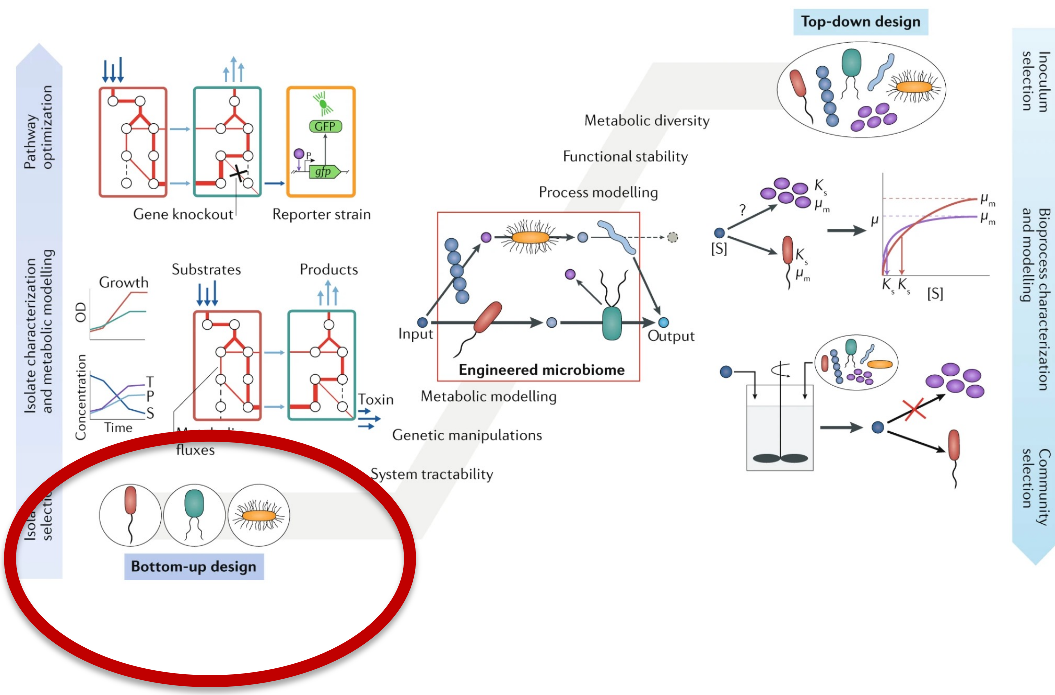
2) Optimal experimental design tools for systems and synthetic biology

Outline

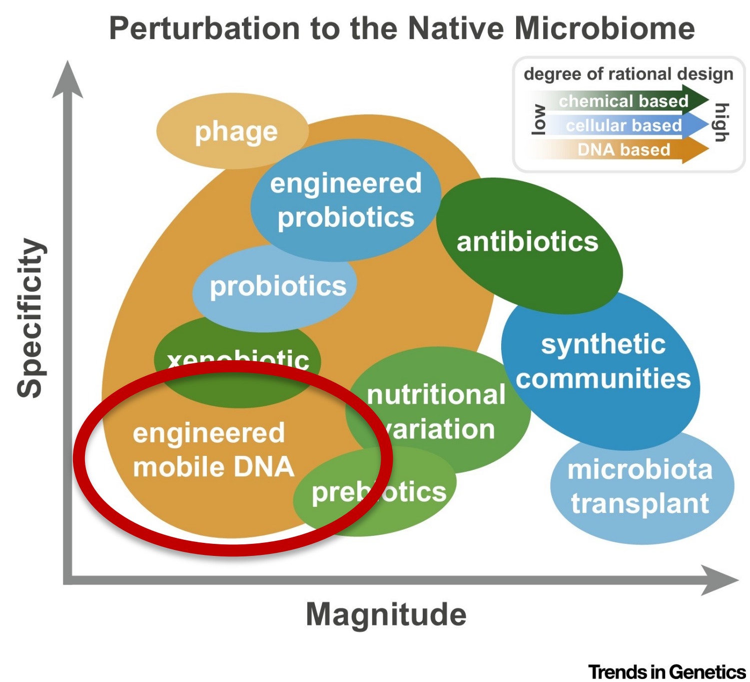
1) **Calibration strategies for agent-based population models of mixed bacterial populations**

2) Optimal experimental design tools for systems and synthetic biology

Goal: model-based design for manipulation of mixed microbial communities

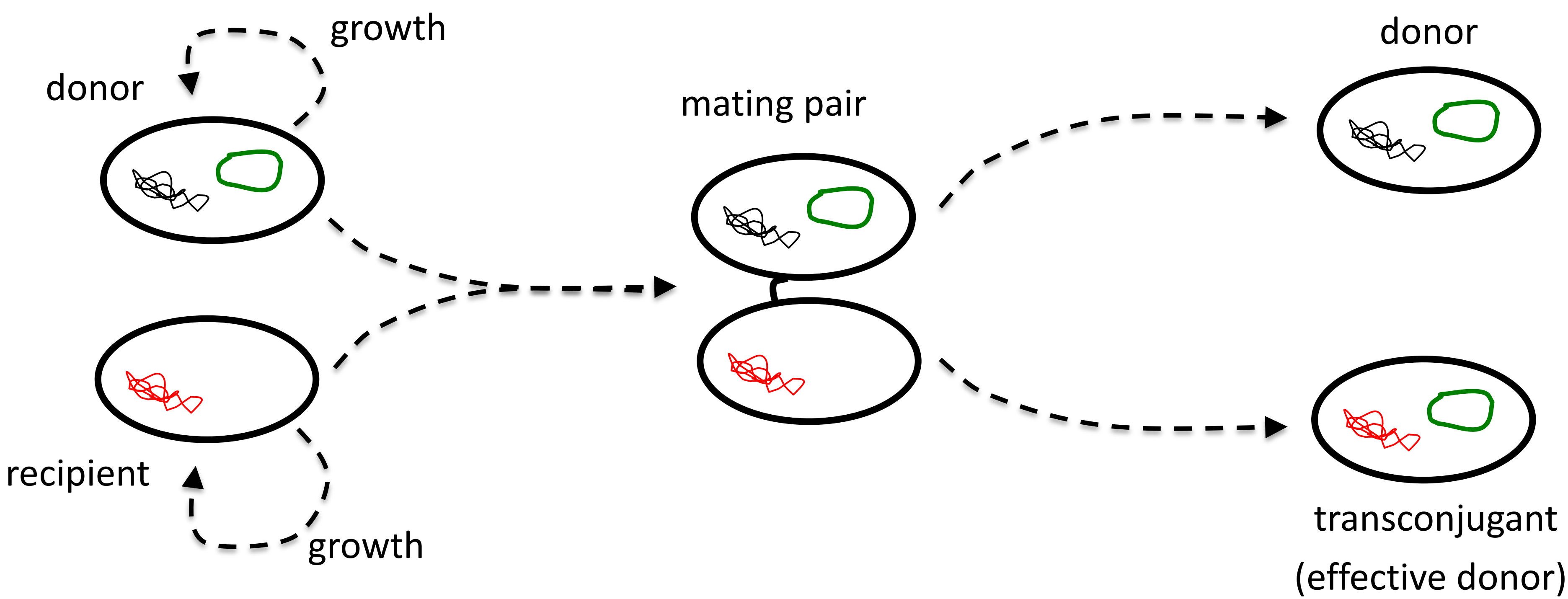


Lawson, et al, *Nature Reviews Microbiology*, 2019



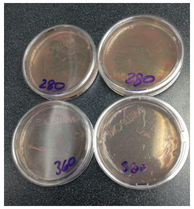
Sheth, et al. *Trends in Genetics*, 2016

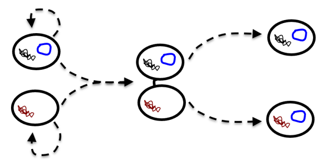
Modelling of plasmid delivery by conjugation

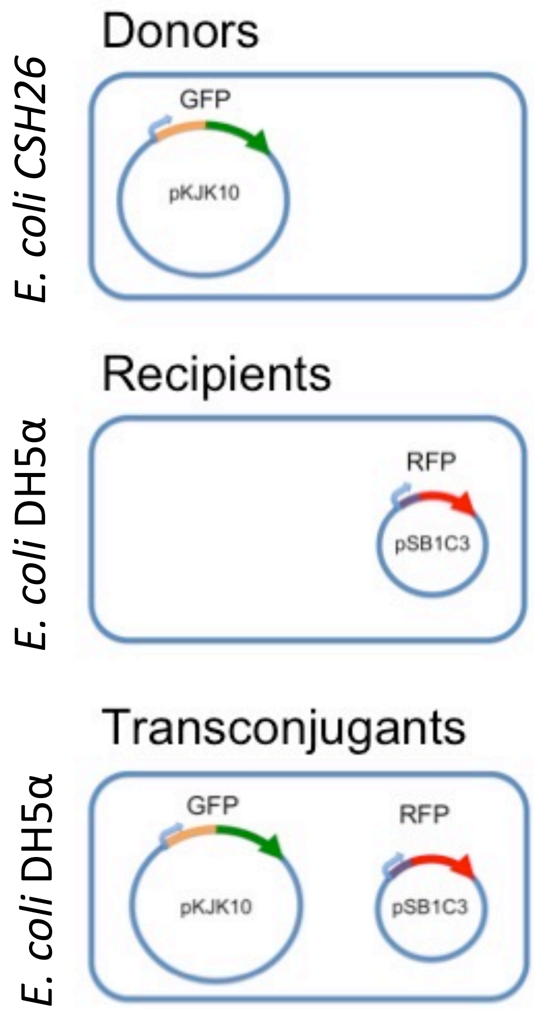


Malwade, Nguyen, Sadat-Mousavi, Ingalls (2017) *Frontiers in Microbiology*

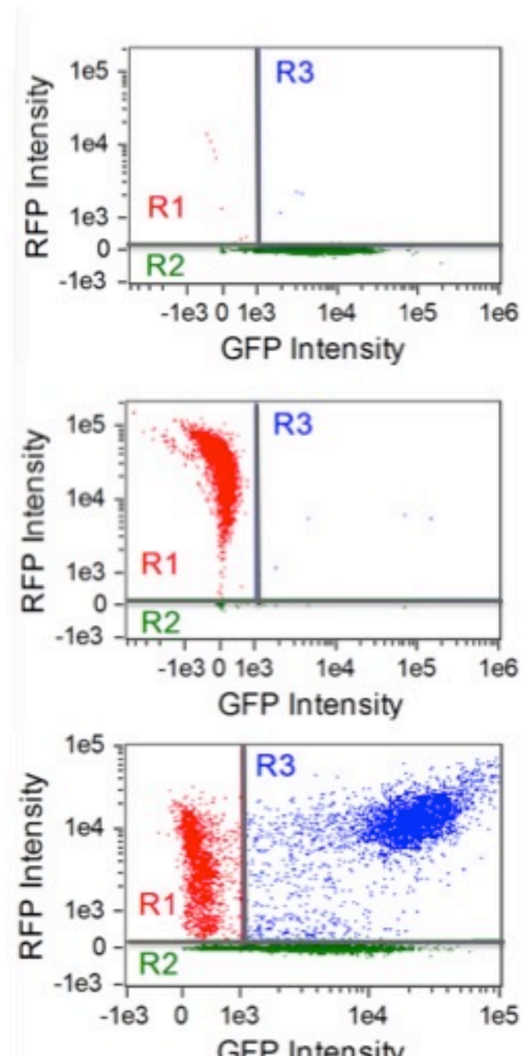
Approach 1: Filter Mating Experiments



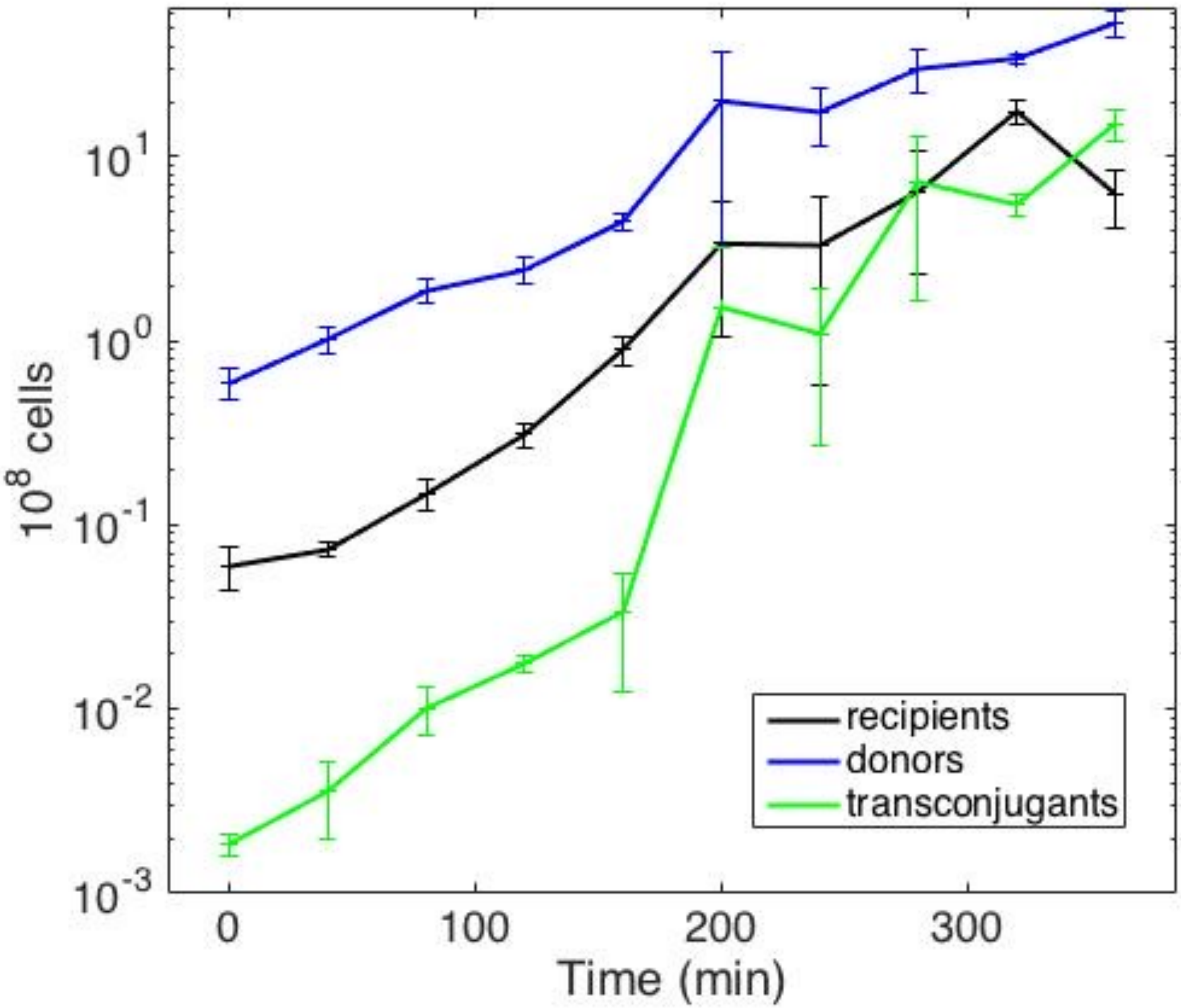




Populations binned by fluorescence signature



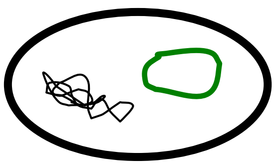
Time point observations



Differential Equation Model (Levin et al., 1979)

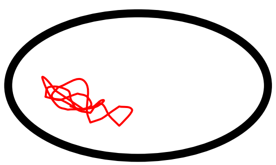
$$\text{donor population: } D$$
$$\text{recipient population: } R$$
$$\text{transconjugant population: } T$$

Balance equations:



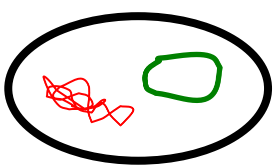
Donors

$$\frac{d}{dt}D(t) = \underbrace{\alpha D(t)}_{\text{growth}}$$



Recipients

$$\frac{d}{dt}R(t) = \underbrace{\alpha R(t)}_{\text{growth}} - \underbrace{\gamma \overbrace{(D(t) + T(t))}^{\text{effective donors}} R(t)}_{\text{conjugation}}$$

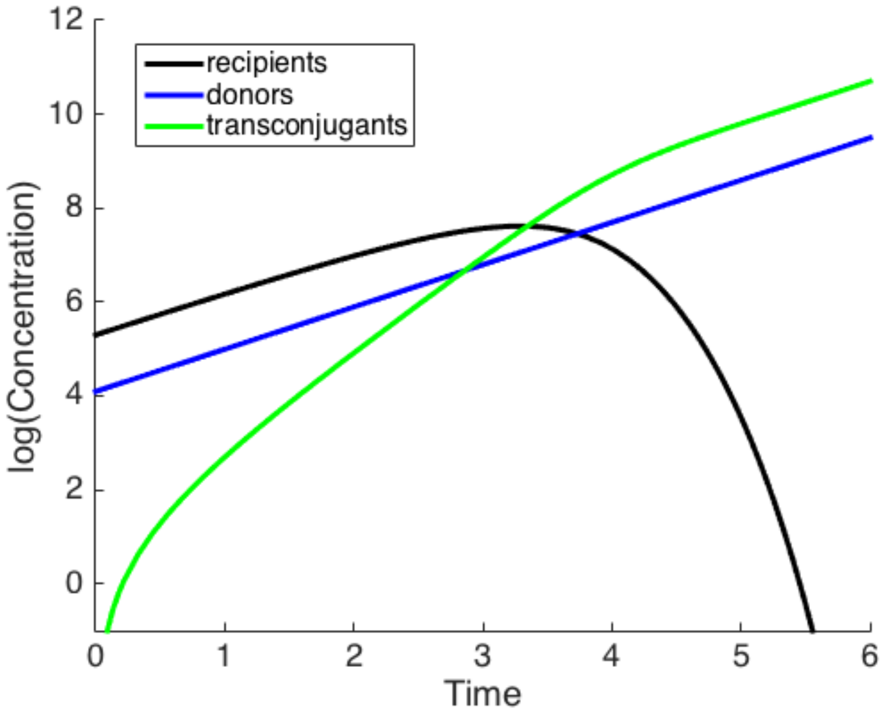


Transconjugants

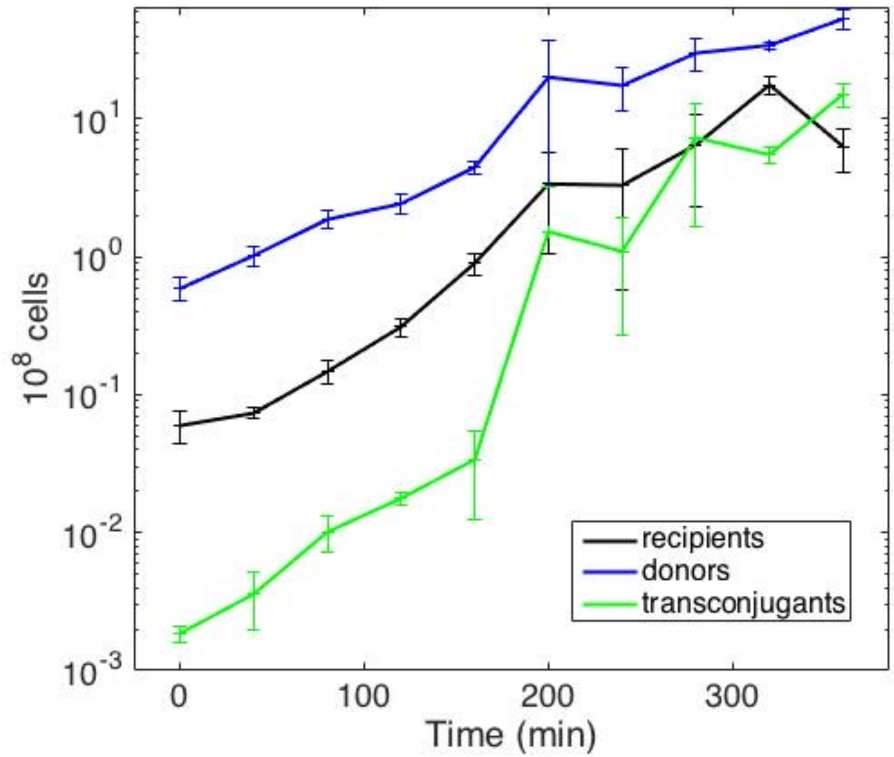
$$\frac{d}{dt}T(t) = \underbrace{\alpha T(t)}_{\text{growth}} + \underbrace{\gamma (D(t) + T(t)) R(t)}_{\text{conjugation}}$$

Analogous to susceptible-infectious (SI) epidemiological models

Levin model



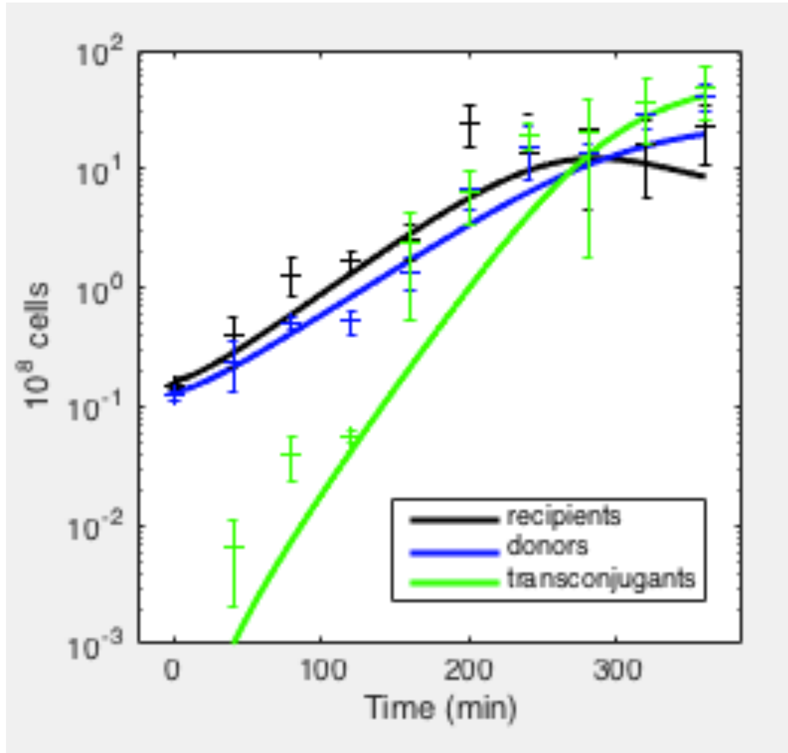
Filter mating data



Adjustments required:

- Distinct kinetics for donors, recipients, and transconjugants
- Lag in initial growth (lag phase) [Baranyi and Roberts, 1994]
- Nutrient limitation (stationary phase) [Simonsen et al., 1990]

Parameter Fitting

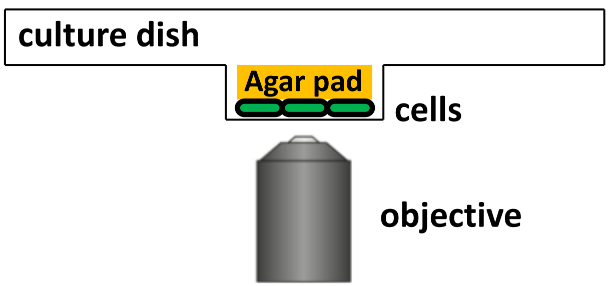


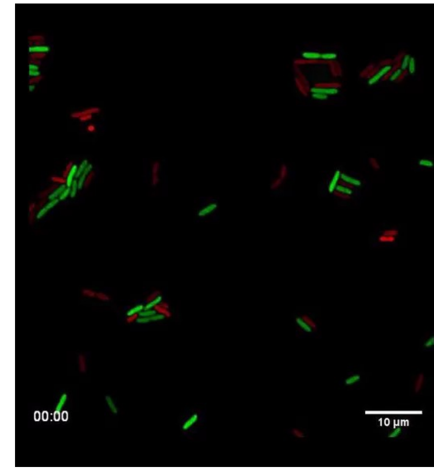
Quality of fit: weighted sum of squared errors

$$\text{Error}(\mathbf{p}) = \sum \left(\frac{\text{observation} - \text{prediction}}{\text{standard deviation}} \right)^2$$

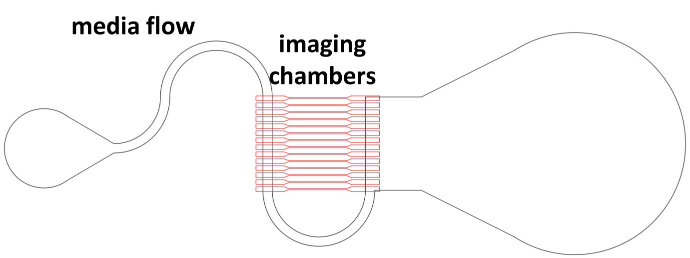
Approach 2: Collection of spatiotemporal data

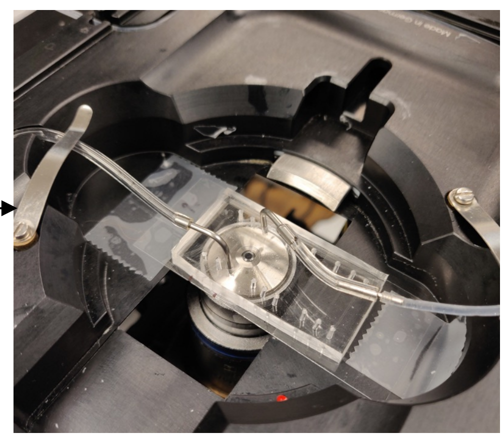
Agar pad

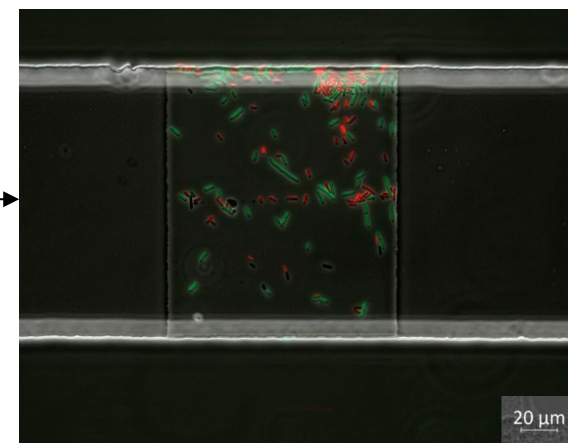




Microfluidic chip







Waterloo microfluidics lab

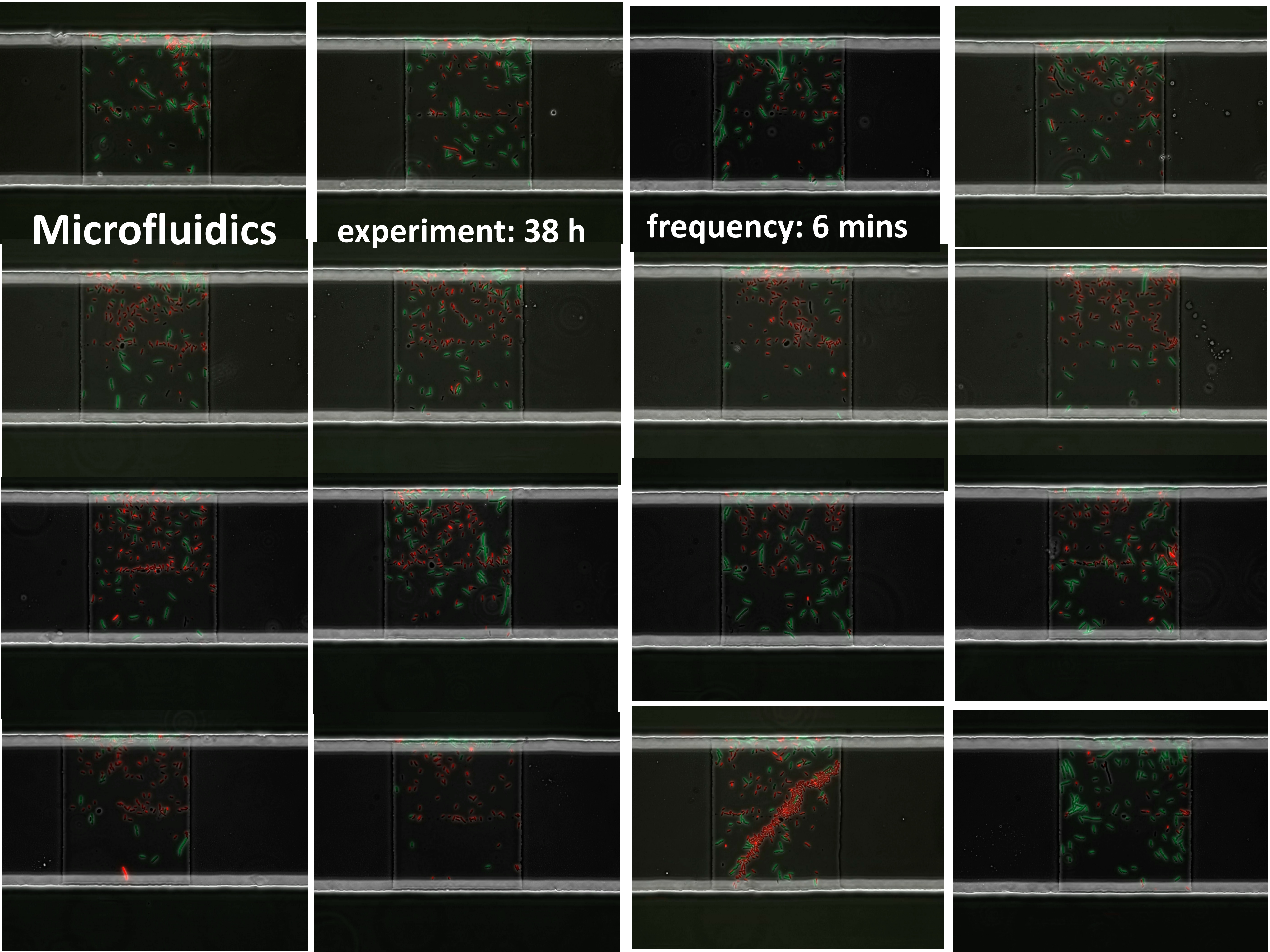
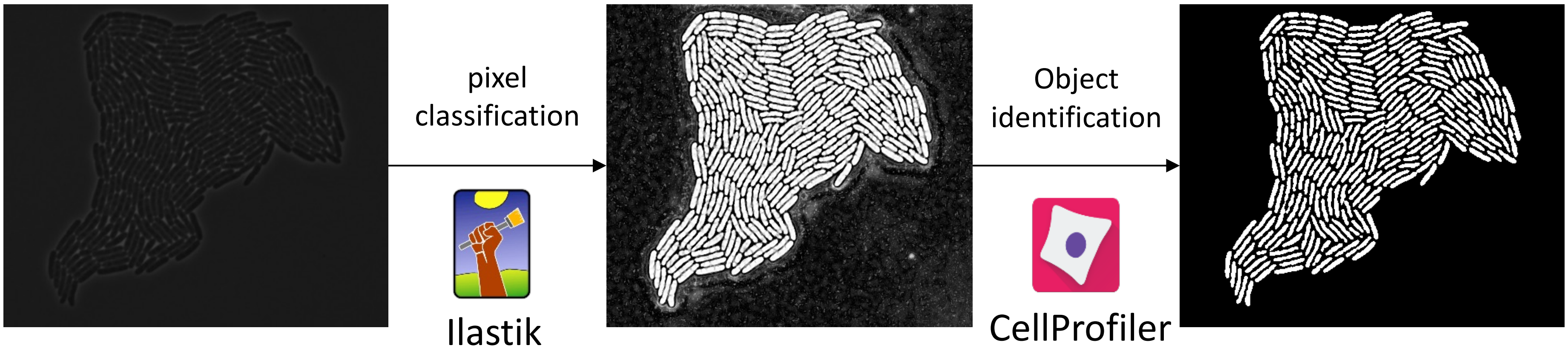
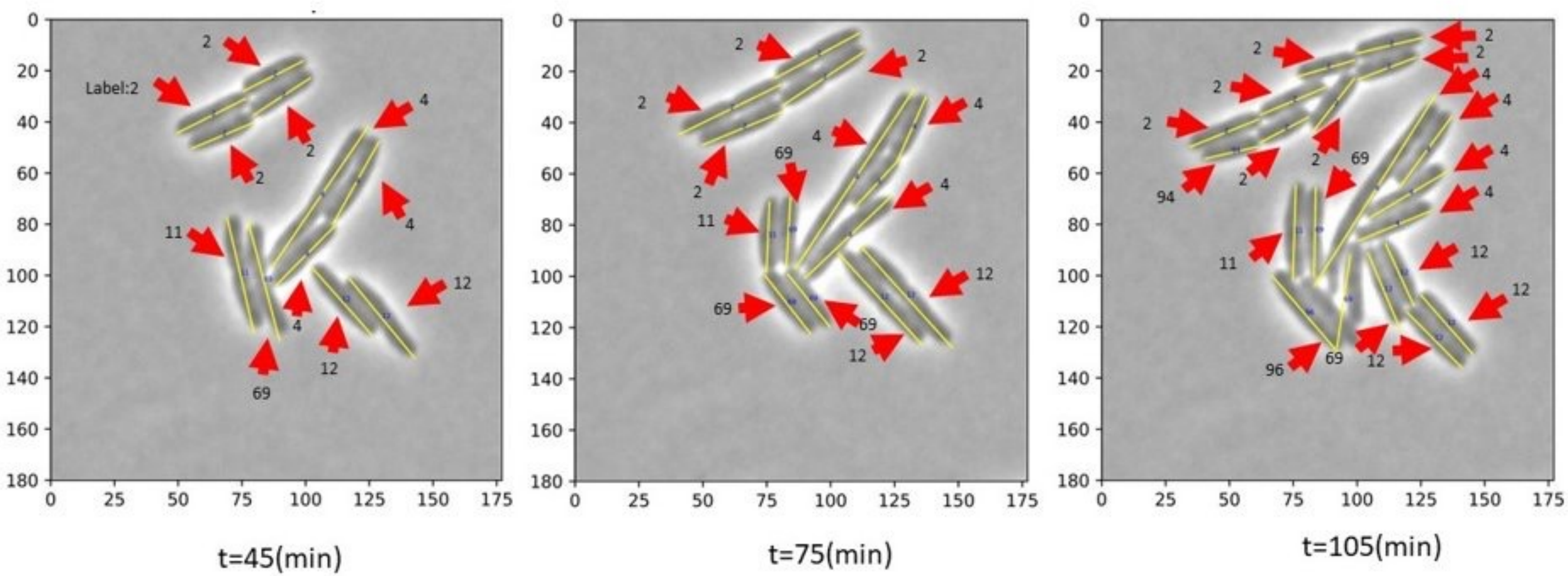

Microfluidics
experiment: 38 h
frequency: 6 mins

Image processing

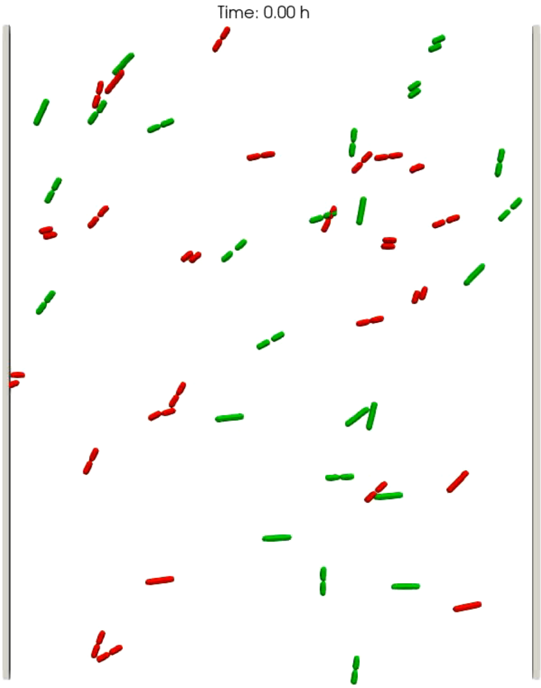
Segmentation

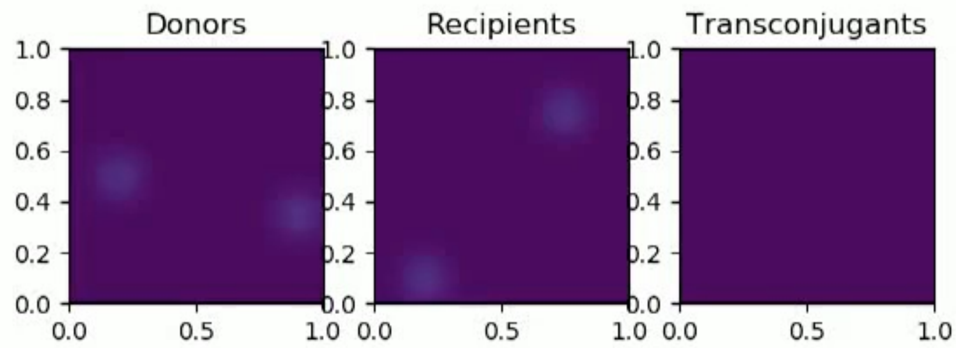




Frame-to-frame: track cells and identify division events

Modelling approaches





Single-cell: **Individual/Agent-based model**

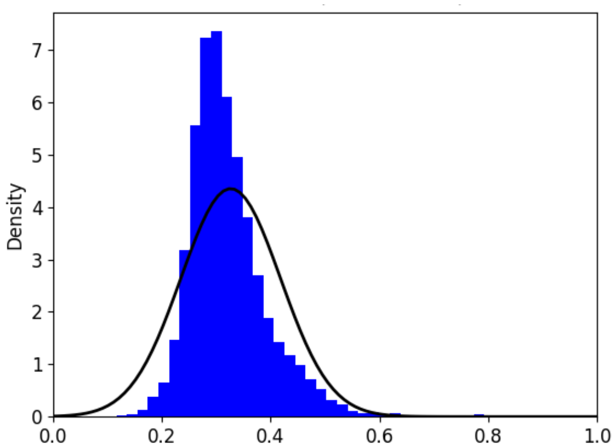
Coarse-grained (density measure): **partial differential equation**

https://github.com/cellmodeller/CellModeller

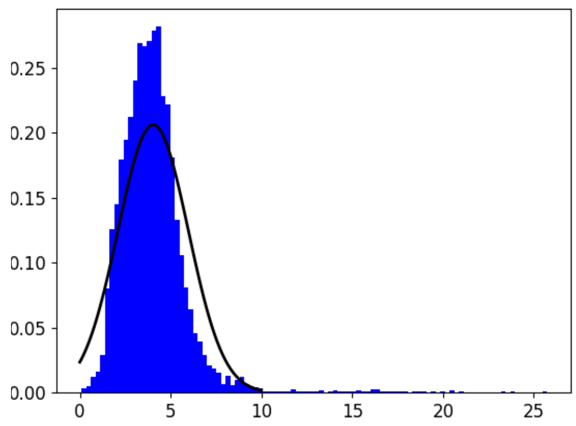
Tim Rudge

Directly observable parameters

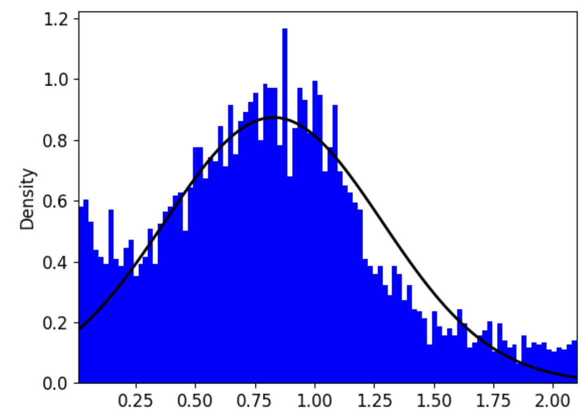
Individual cellular measurements



Radius (um)

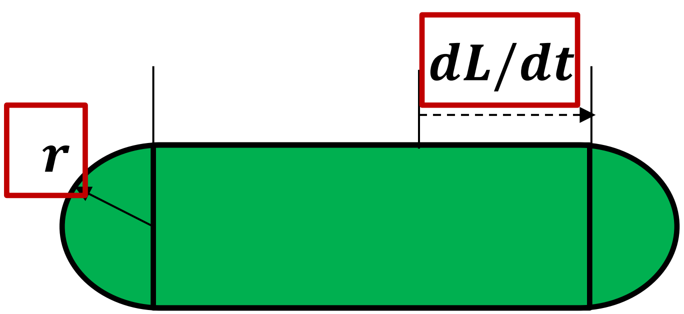


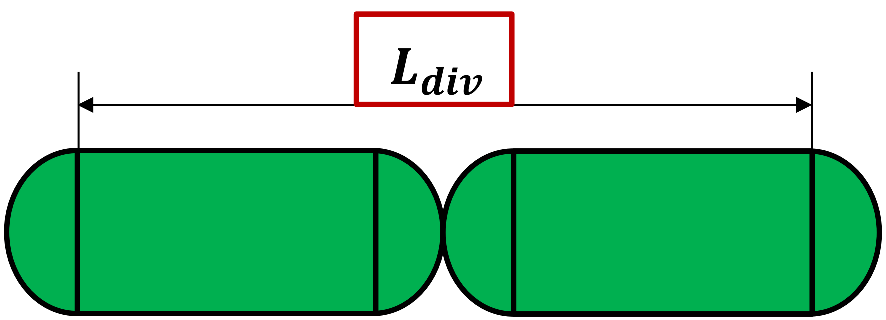
Length at division (um)



Elongation rate (um/h)

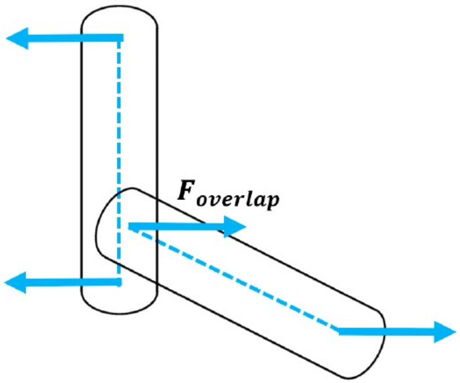
These can be incorporated directly into the ABM formulation





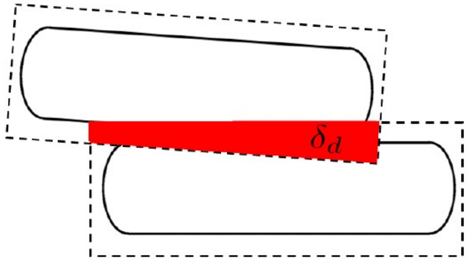
Parameters to be inferred

Biophysical:



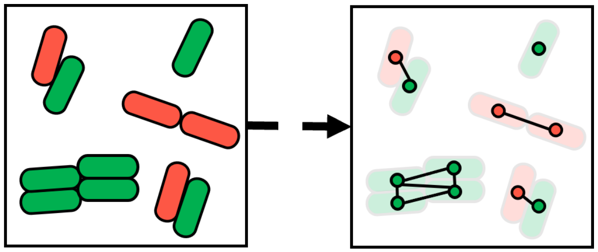
cell stiffness
(exclusion force)

growth pressure



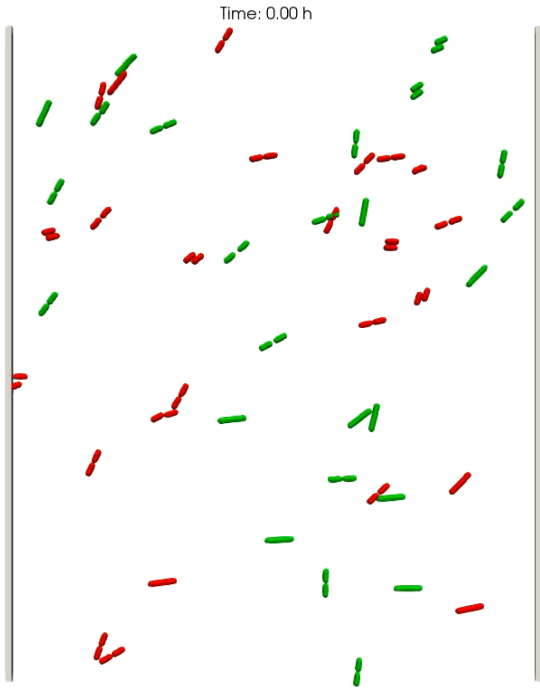
adhesion strength

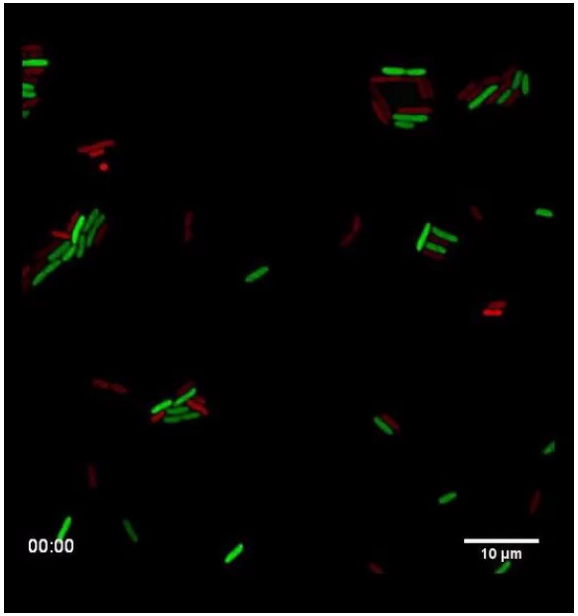
Process-specific:



conjugation process
(degree of contact, delay, zygotic induction)

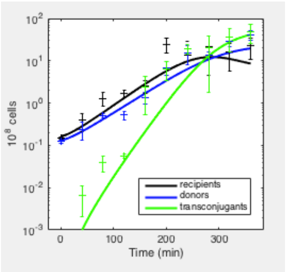
Agent-based model calibration challenges





Stochasticity
Lack of ‘obvious’ goodness-of-fit measure

Strategy (from ecology):
“Pattern-oriented Modelling”

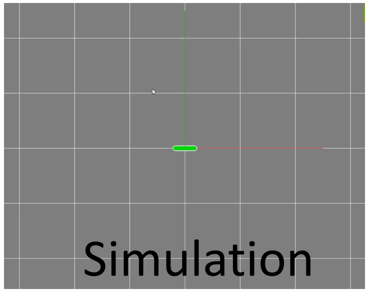


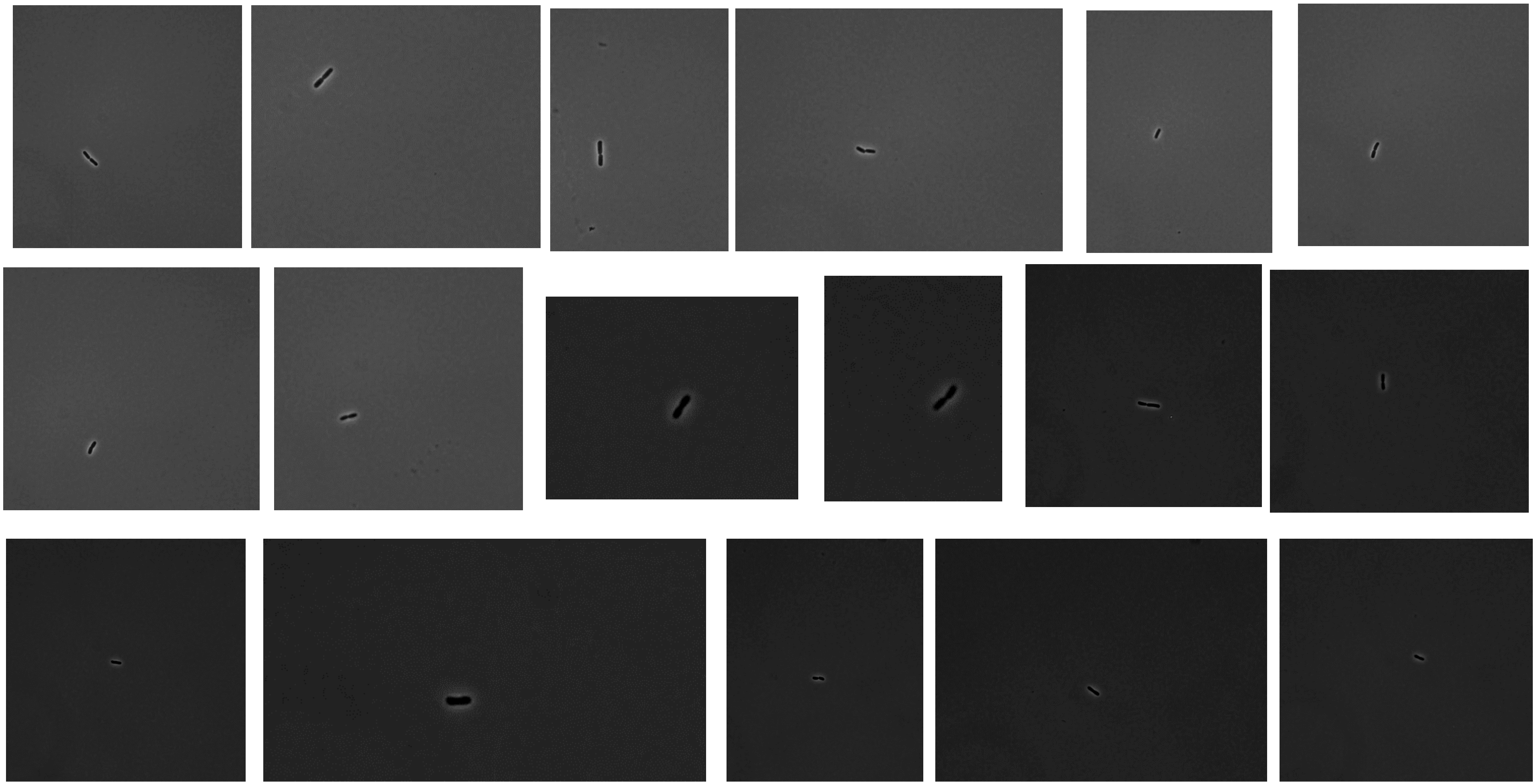
$$\mathrm{Error}(\mathbf{p}) = \sum \left(\frac{\text{observation} - \text{prediction}}{\text{standard deviation}} \right)^2$$

non-spatial case

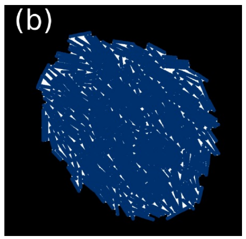
Observation of system behavior

Biophysical parameters: growth of isolated microcolonies

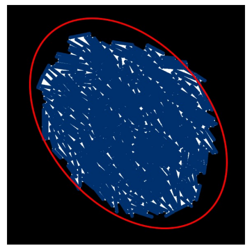




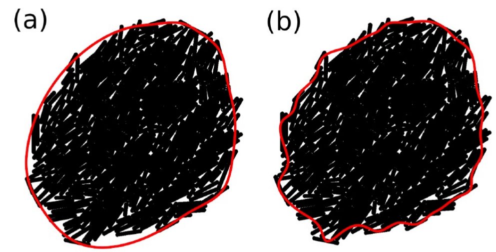
Biophysical Features



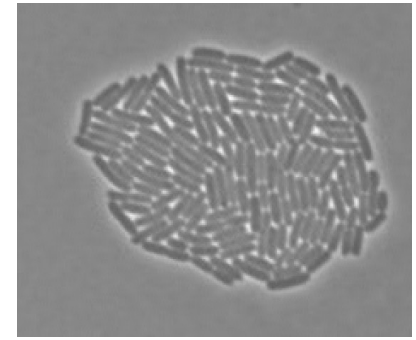
colony density

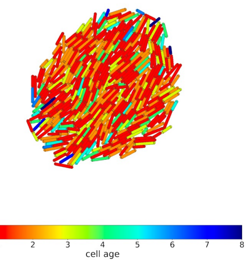


colony aspect ratio

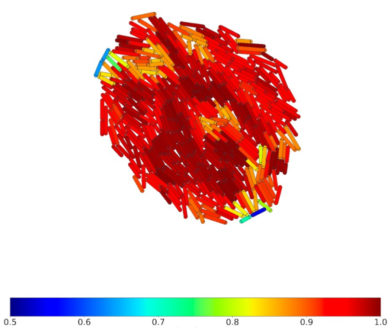


boundary jaggedness

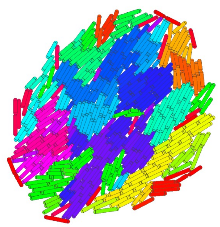




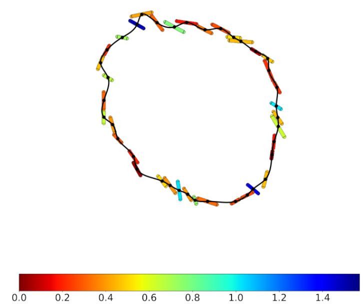
distribution of cell (pole) ages



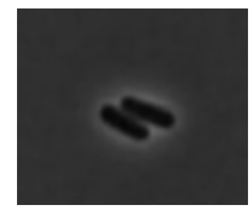
anisotropy of cell orientation

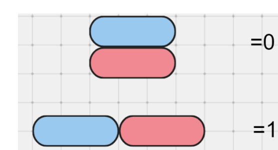


distribution of oriented ‘patches’



orientation of boundary cells



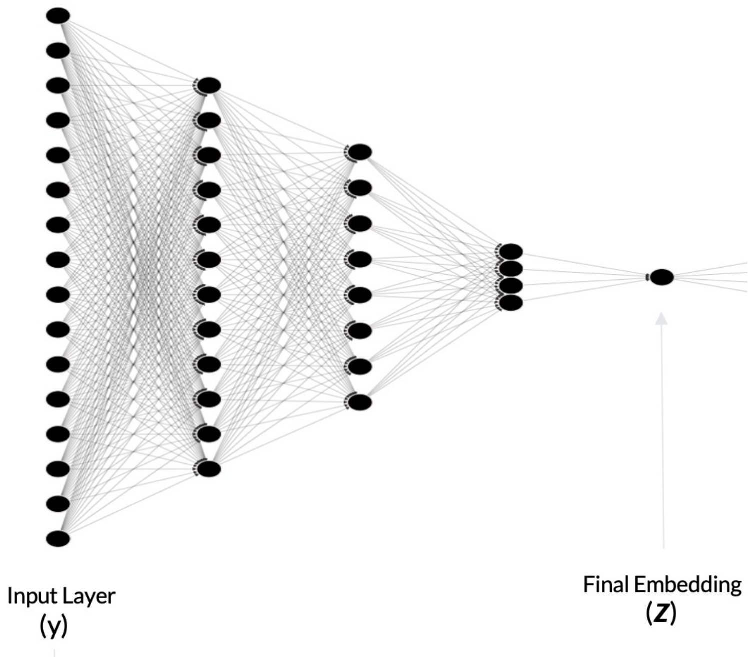


initial dyad orientation

Yip et al. (2022) Calibrating spatiotemporal models of microbial communities to microscopy data: a review. PLOS Computational Biology

Deep representation learning to identify features

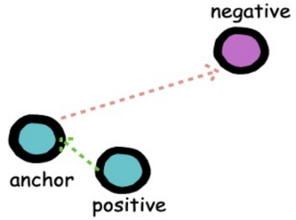
Inspired by: Cess and Finley "Calibrating agent-based models to tumor images using representation learning." *PLOS Comp. Biol.* 2023

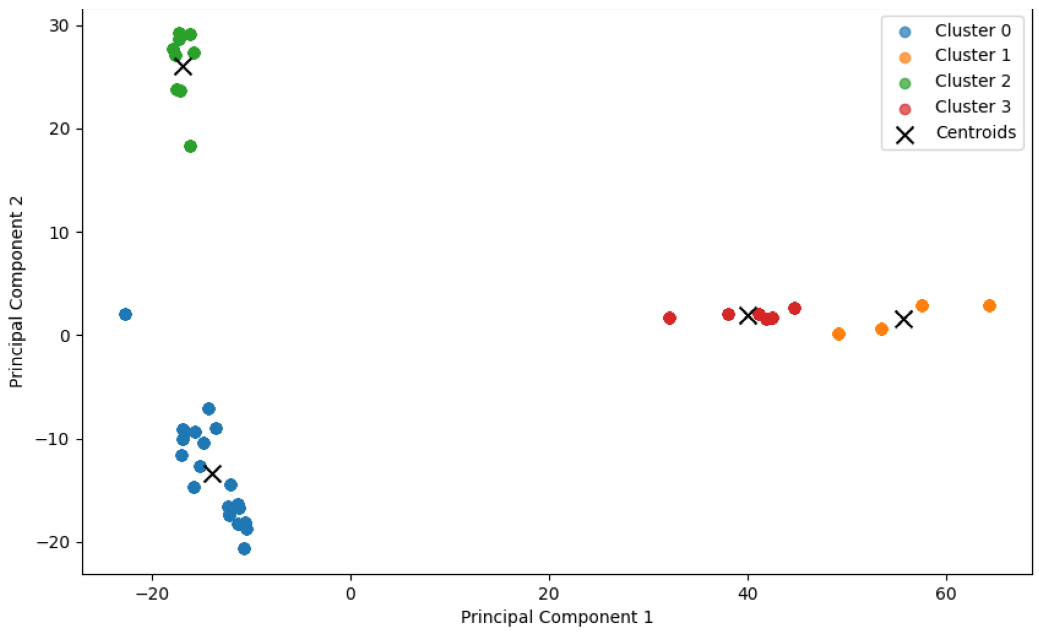


Successful preliminary results with low-dimensional embeddings (simulated data)

Layers: LSTM or convolutional

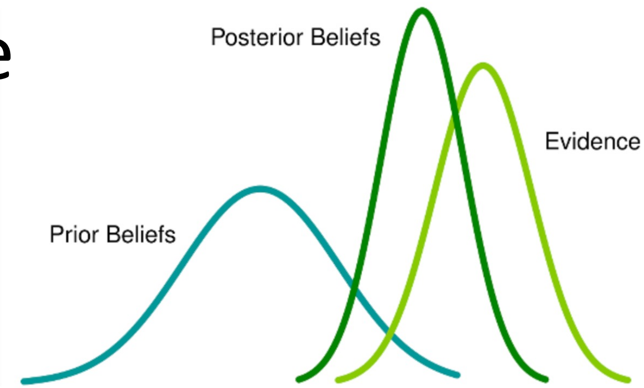
Triplet loss (supervised learning)



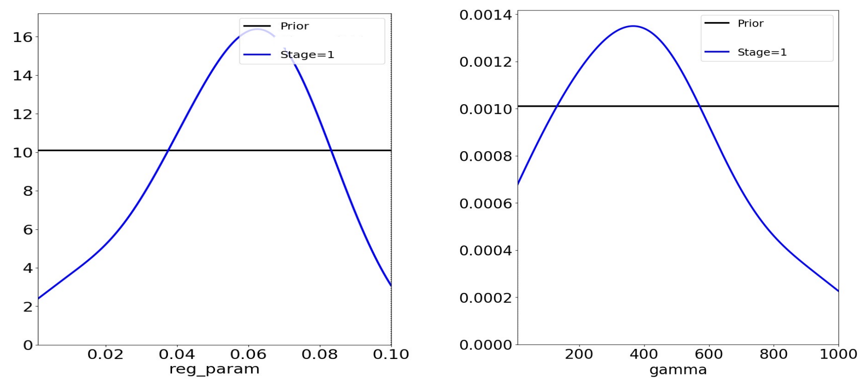


Calibration procedure

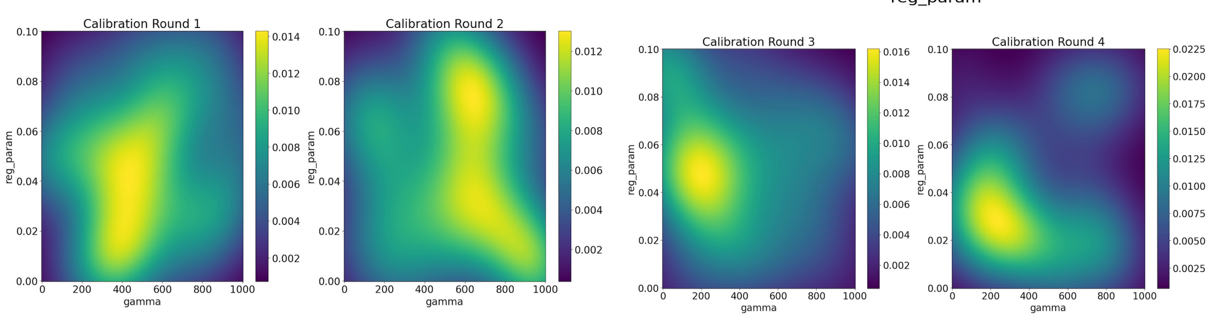
Approach: Bayesian inference



Method: Approximate Bayesian Computation (ABC)



Results:

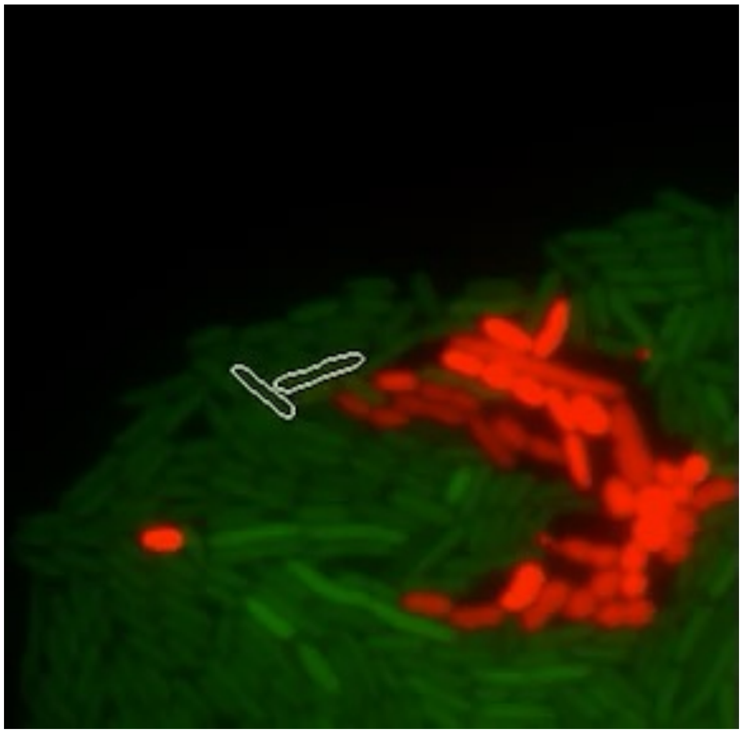


Also exploring deep regression models

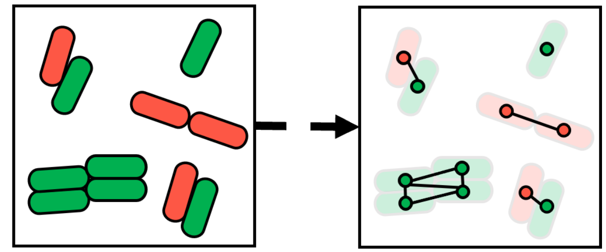
Next step: calibration of conjugation parameters

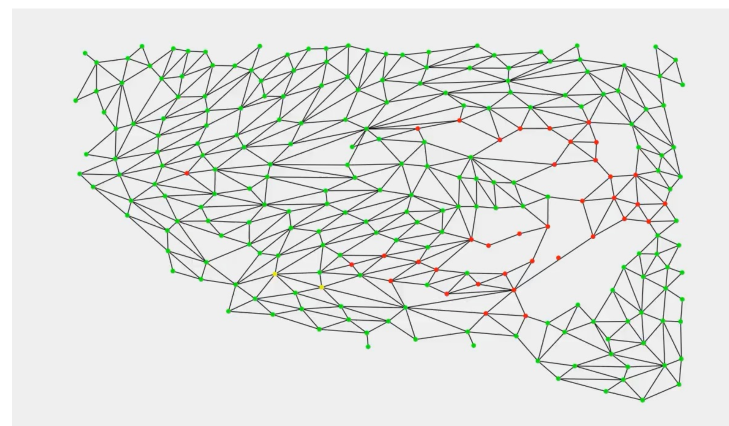
(incubation period*, degree of contact, zygotic induction)

Delayed conjugation events



Contact network





Identification of conjugation events: integer programming approach inspired by epidemiological contact tracing analysis

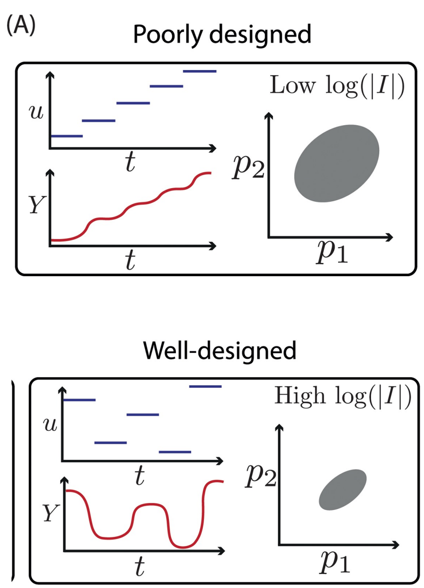
Outline

1) Calibration strategies for agent-based population models of mixed bacterial populations

2) **Optimal experimental design tools for systems and synthetic biology**

Model Based Optimal Experimental Design

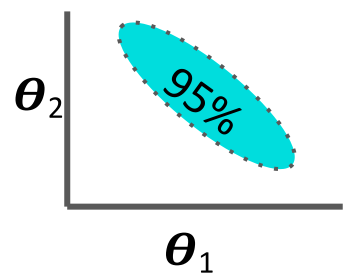
Identification of maximally informative experiments



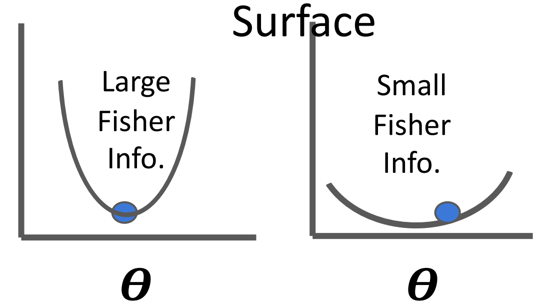
Fisher Information-based approach

D-optimality

Minimize the Confidence Region Volume



Find an Experiment that Yields a ‘Pointy’ Optimization Surface



$$\mathrm{cov}(\theta) \sim I^{-1}$$

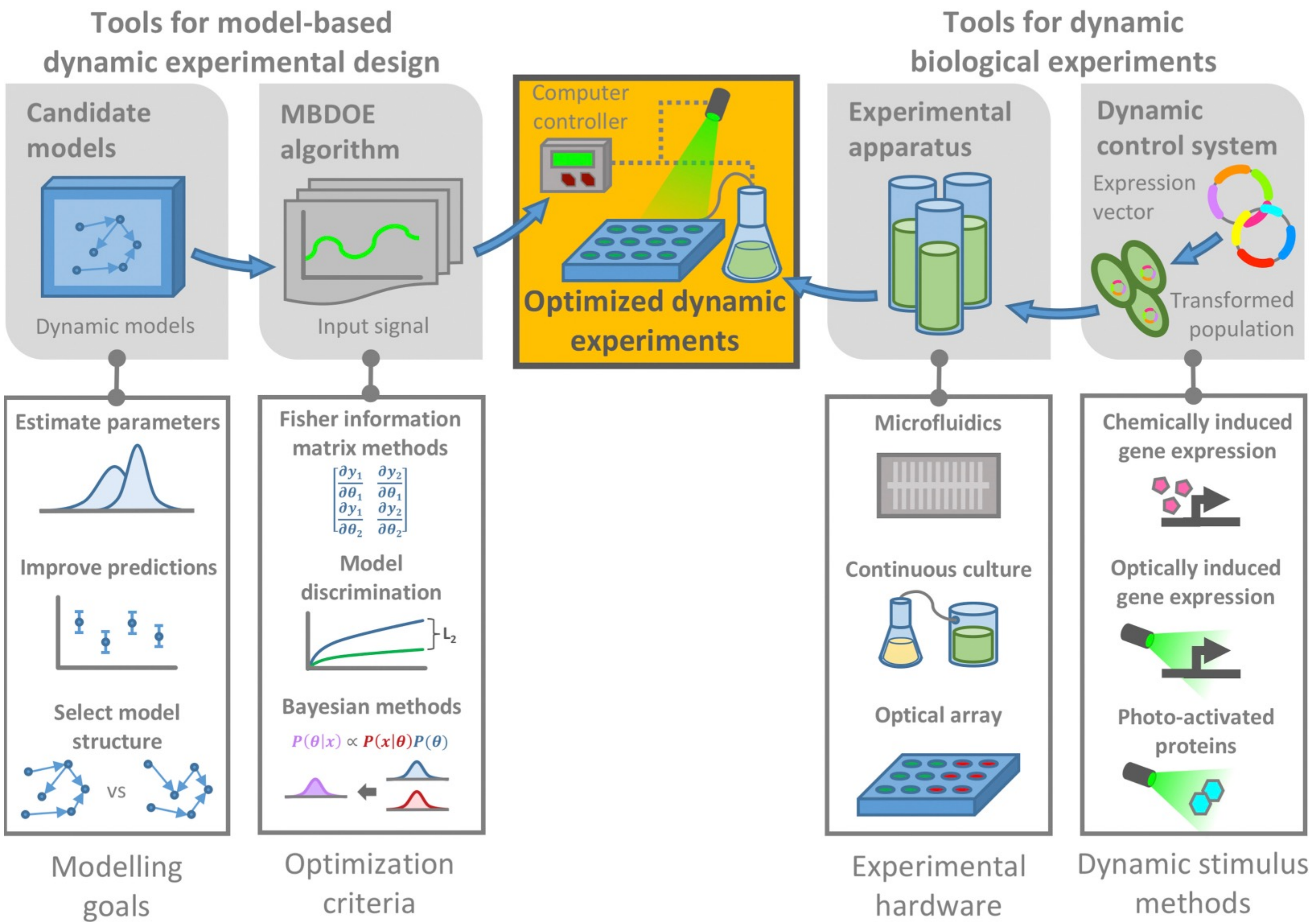
$$\det(\mathrm{cov}(\theta))$$

Minimize **D**eterminant of the Parameter Covariance Matrix

$$\det(I)$$

Maximize **D**eterminant of Fisher Information Matrix*

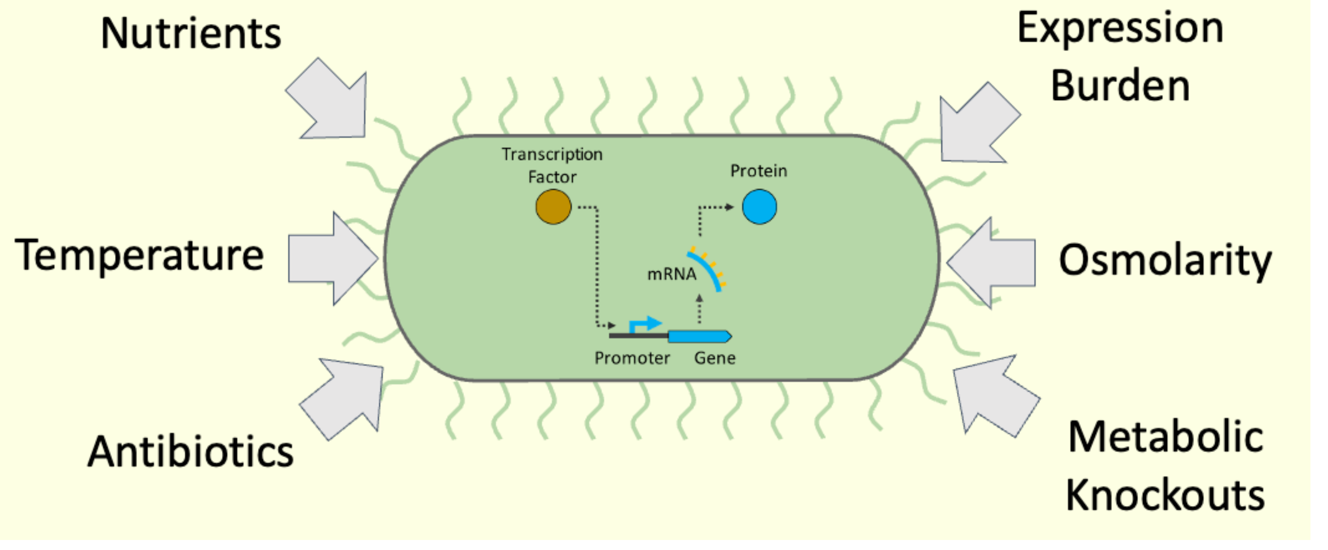
*Sensitivities of outputs to parameter values scaled by confidence in measurements



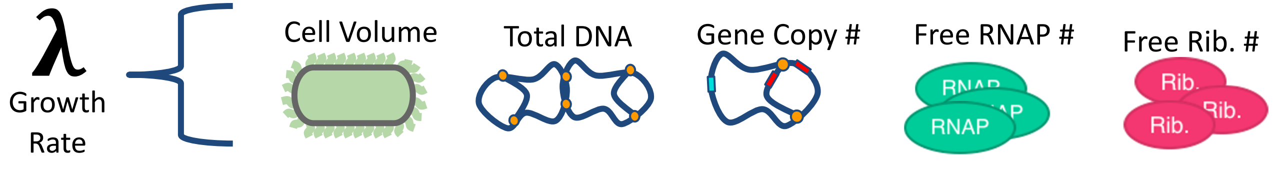
Braniff and Ingalls, New opportunities for optimal design of dynamic experiments in systems and synthetic biology *Curr. Op. Sys. Biol.* 9 (2018)

Optimal design: unravelling the effects of physiology on gene expression dynamics

Modelling goal: assess the effect of environmental factors on the dynamics of gene regulatory networks

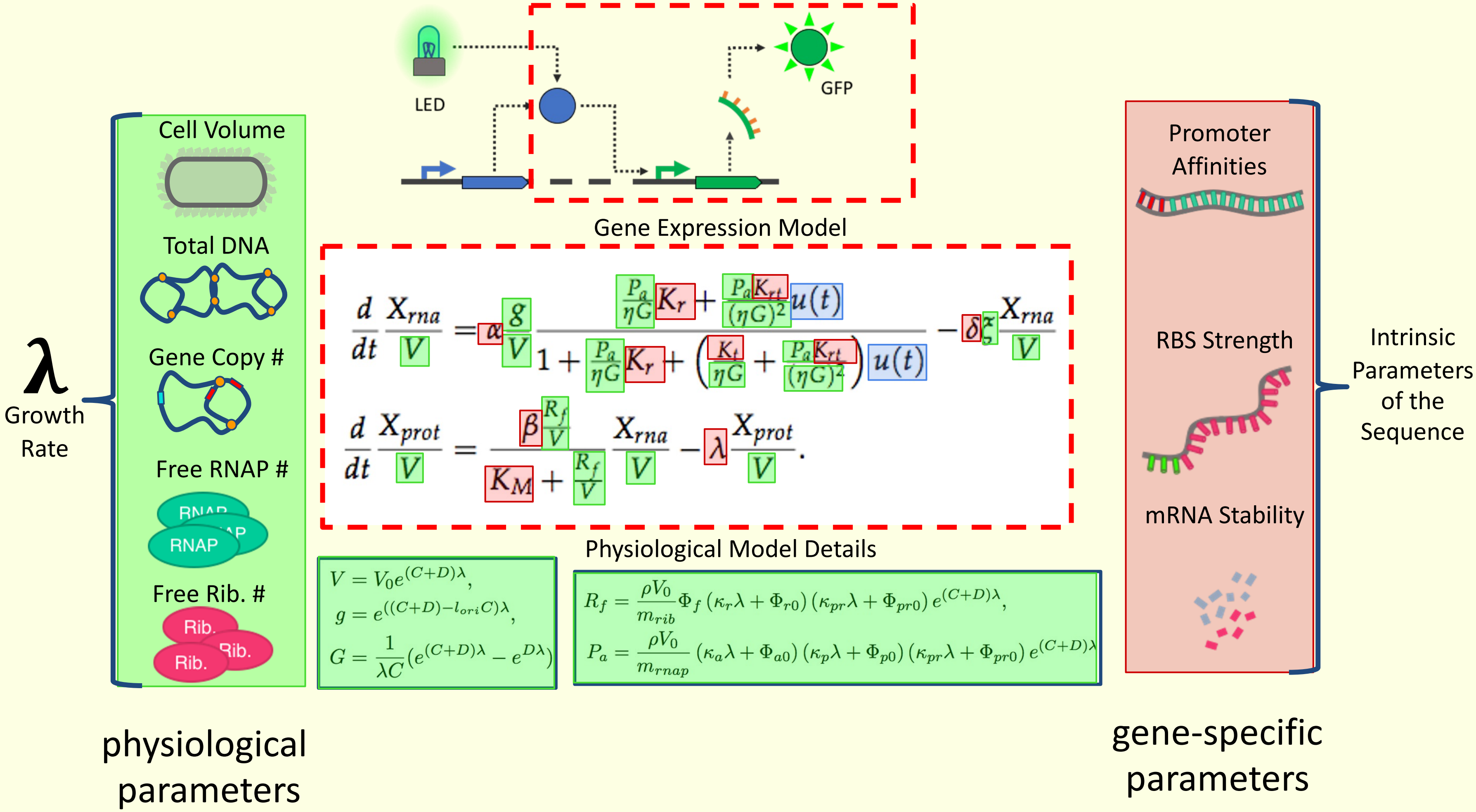


Klumpp and Hwa, (2014): Growth rate is a **sufficient statistic** for *E. coli* host physiology

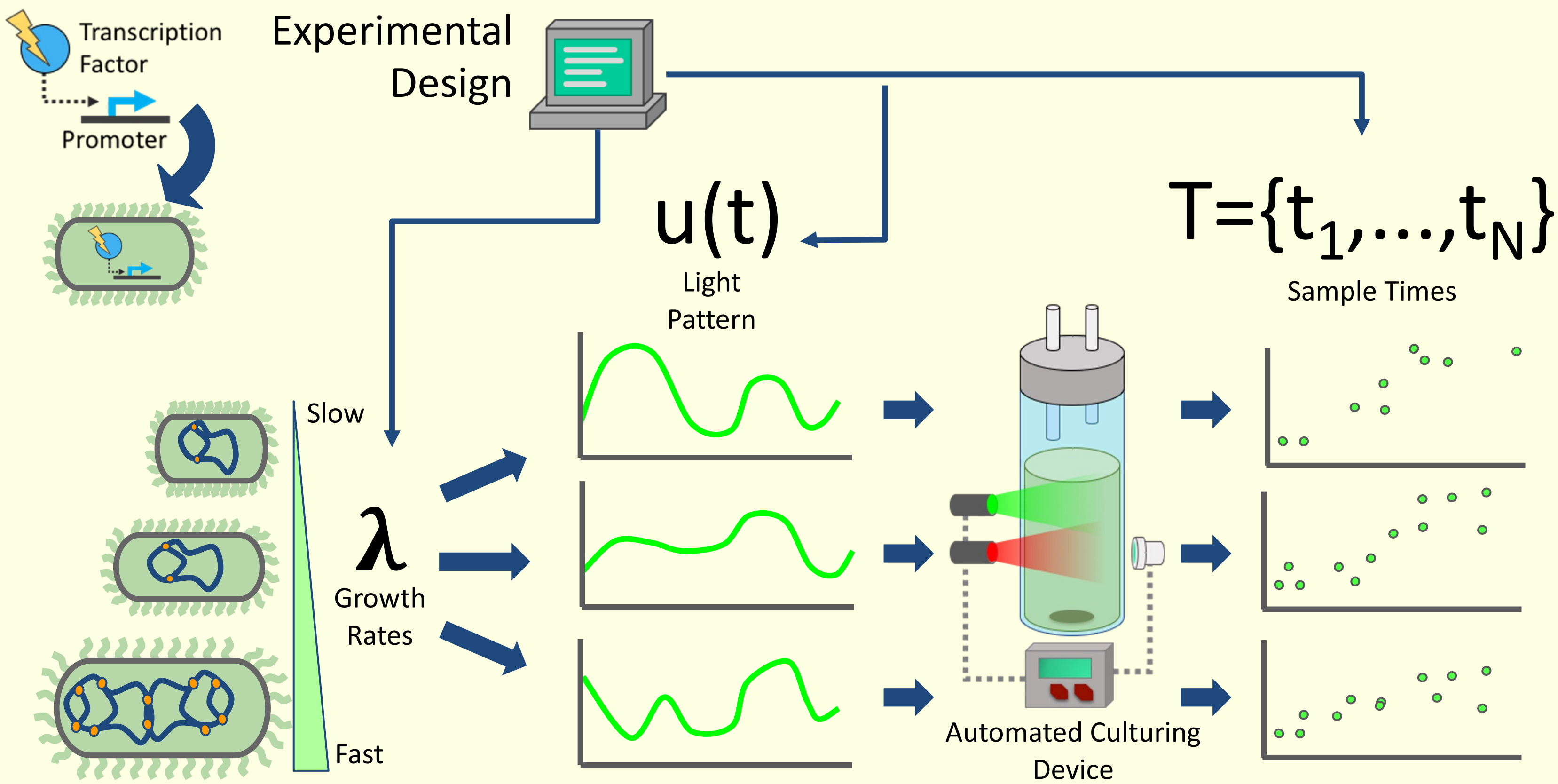


Braniff, Scott, and Ingalls. Component characterization in a growth-dependent physiological context: optimal experimental design. *Processes* (2019)

Physiologically-aware gene expression model

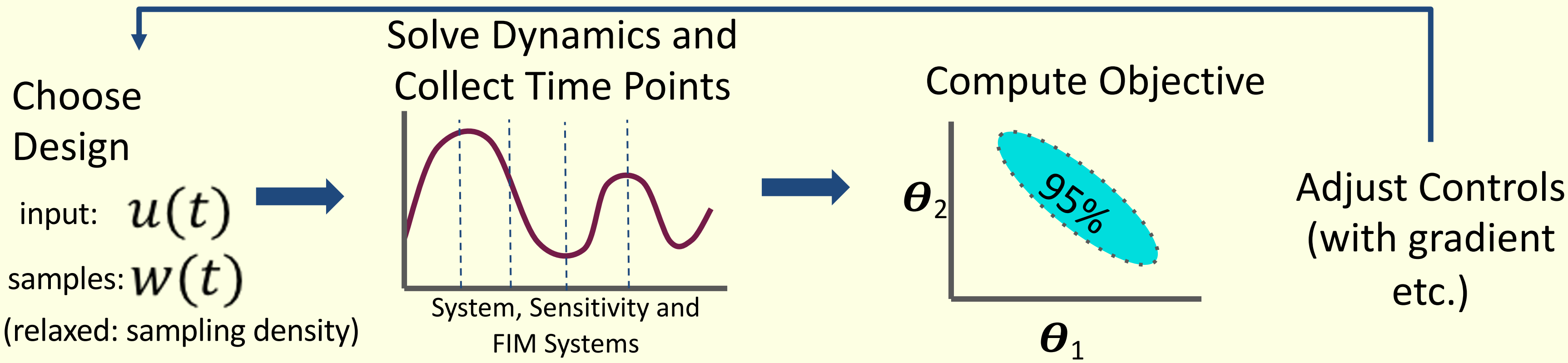


Experimental Design

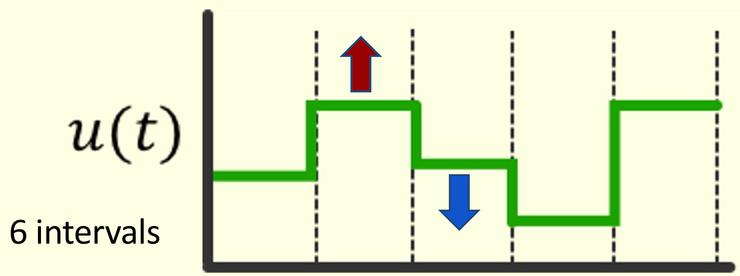


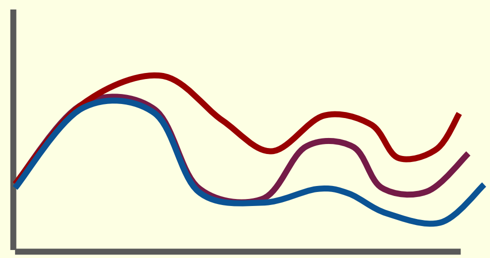
Optimization

Naive Approach



Computational challenges





Each control's effect is:

1) Distributed
2) Non-linear

Problem is poorly structured, gradient isn't useful

Solution: recast as optimal control problem

Experimental Design as Optimal Control

System Dynamics

$$\dot{\boldsymbol{y}} = \boldsymbol{F}(\boldsymbol{y}, \boldsymbol{\theta}, u(t), \lambda)$$

$$\frac{d}{dt}\frac{X_{rna}}{V} = \alpha \frac{g}{V} \frac{\frac{P_a}{\eta G}K_r + \frac{P_a K_{rt}}{(\eta G)^2}u}{1 + \frac{P_a}{\eta G}K_r + \left(\frac{K_t}{\eta G} + \frac{P_a K_{rt}}{(\eta G)^2}\right)u} - \delta \frac{\xi}{V}\frac{X_{rna}}{V}$$

$$\frac{d}{dt}\frac{X_{prot}}{V} = \beta \frac{R_f}{V}\frac{X_{rna}}{V} - \lambda \frac{X_{prot}}{V}$$

$$\dot{\boldsymbol{S}} = \frac{\partial \boldsymbol{F}(\boldsymbol{y}, \boldsymbol{\theta}, u(t), \lambda)}{\partial \boldsymbol{\theta}} + \frac{\partial \boldsymbol{F}(\boldsymbol{y}, \boldsymbol{\theta}, u(t), \lambda)}{\partial \boldsymbol{y}}\boldsymbol{S}$$

$$\boldsymbol{S} = [\frac{\partial \boldsymbol{y}}{\partial \theta_1}, \dots, \frac{\partial \boldsymbol{y}}{\partial \theta_N}]$$

$$\dot{\mathcal{J}} = w(t)\boldsymbol{S}^{\boldsymbol{T}}\boldsymbol{S}$$

Telen et al., *Computers and Chemical Engineering*, 2014

Objective Function

$$\det(\mathcal{J}(t_f))$$

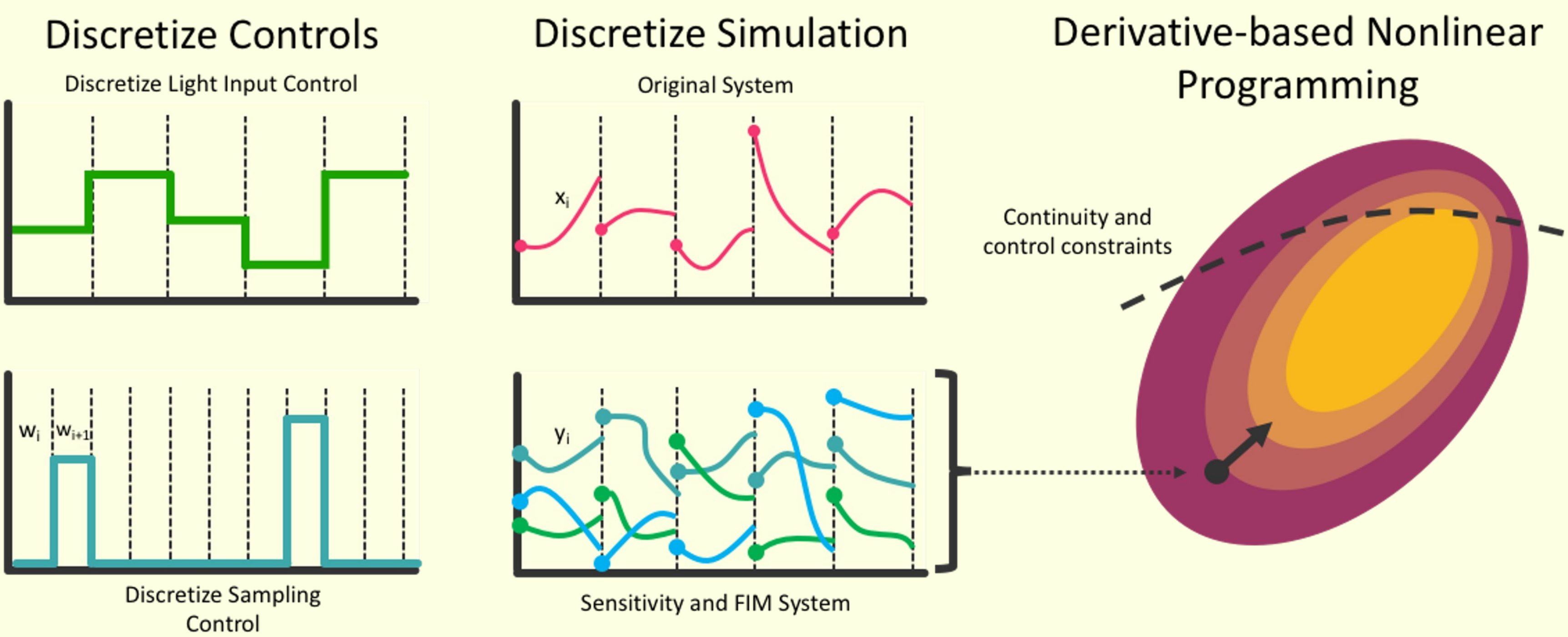
Controls

$u(t)$ Induction Control

$\boldsymbol{\lambda}$ Growth Rate

$w(t)$ Sampling (Continuous Relaxation)

Solution method: Multiple Shooting



CasADi symbolics for sensitivities of constraints (including initial conditions) and simulation steps (4th order Runge-Kutta)

Janka et al., *Multiple Shooting and Time Domain Decomposition Methods*, 2015

Optimal Experiment

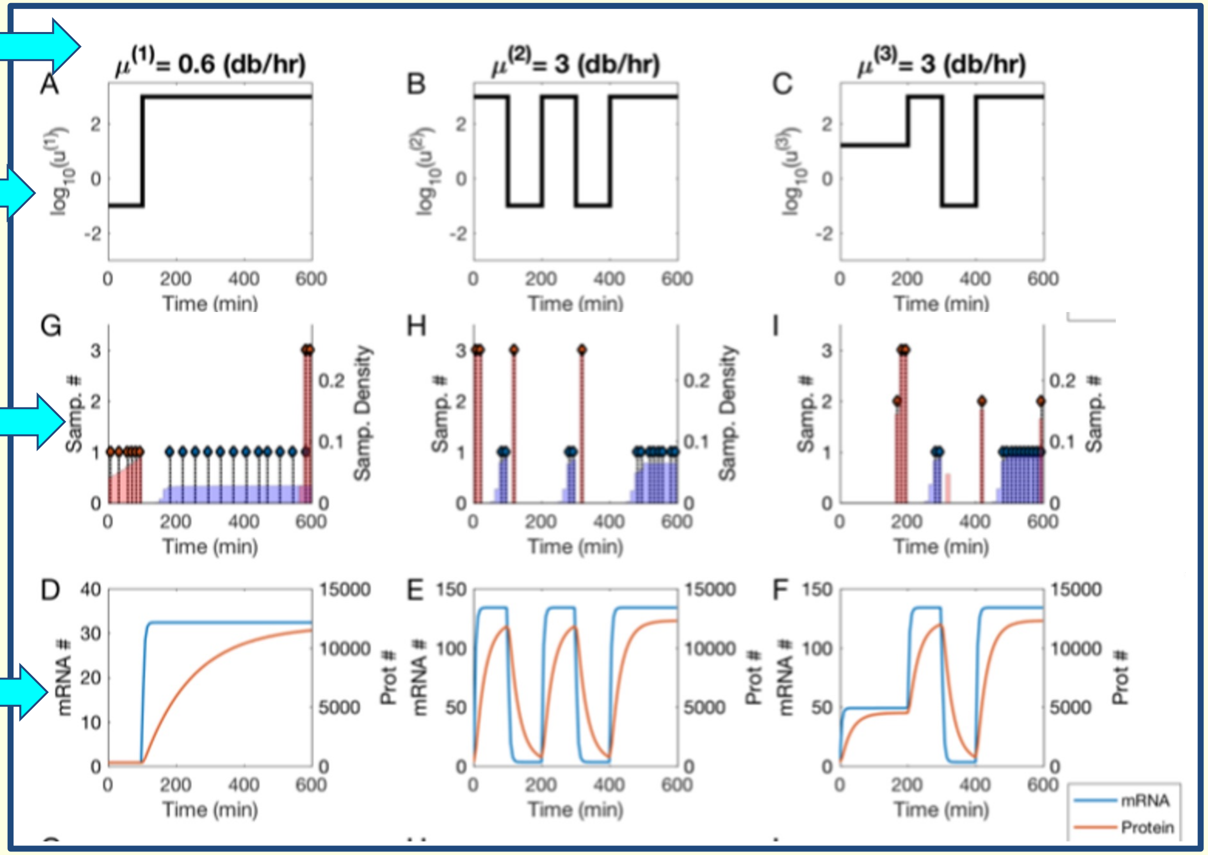
λ Growth Rates

u(t) Light Pattern

$T=\{t_1,\ldots,t_N\}$

Sample Times

System Response

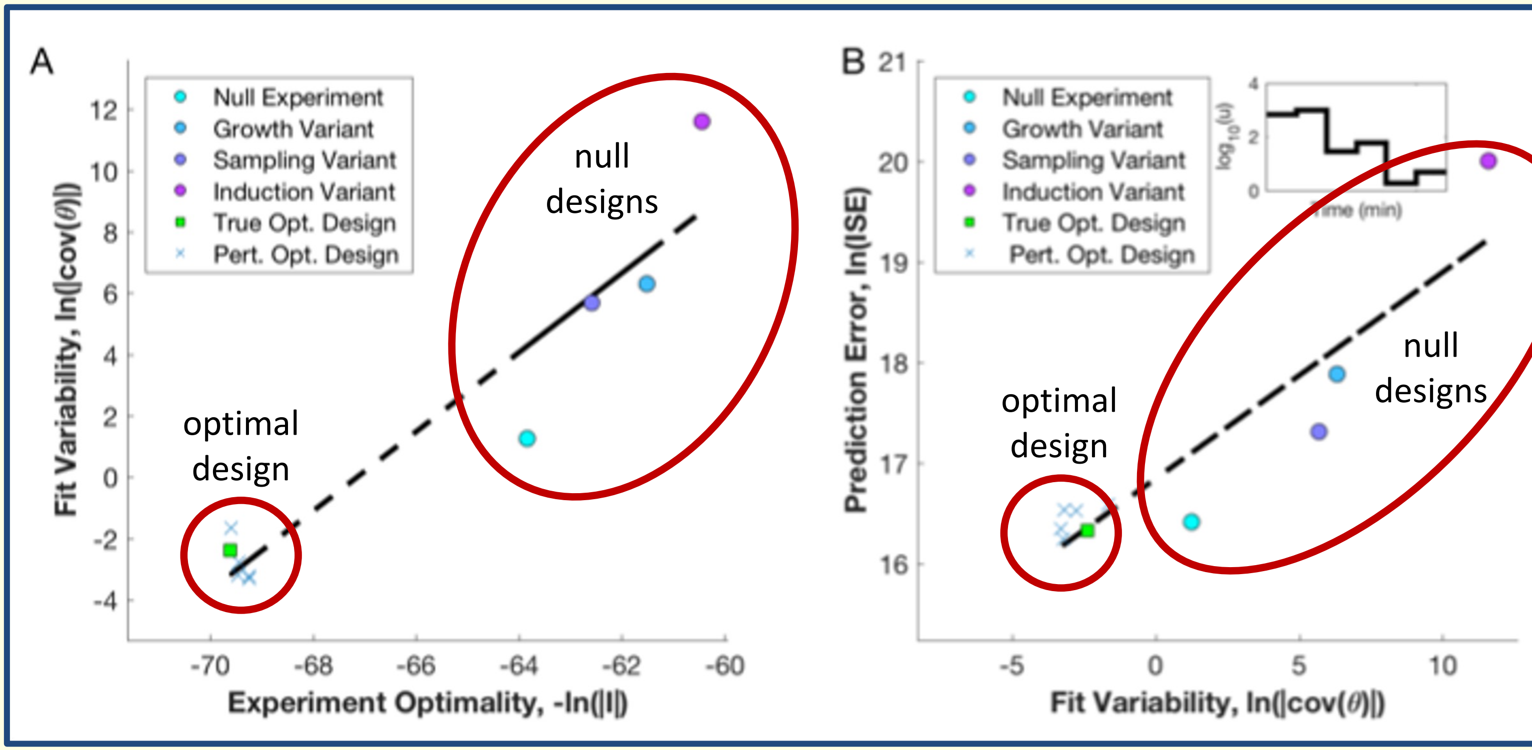


Validation

Parameter estimation

Prediction Accuracy
(out of sample experiment)

Parameter Variance

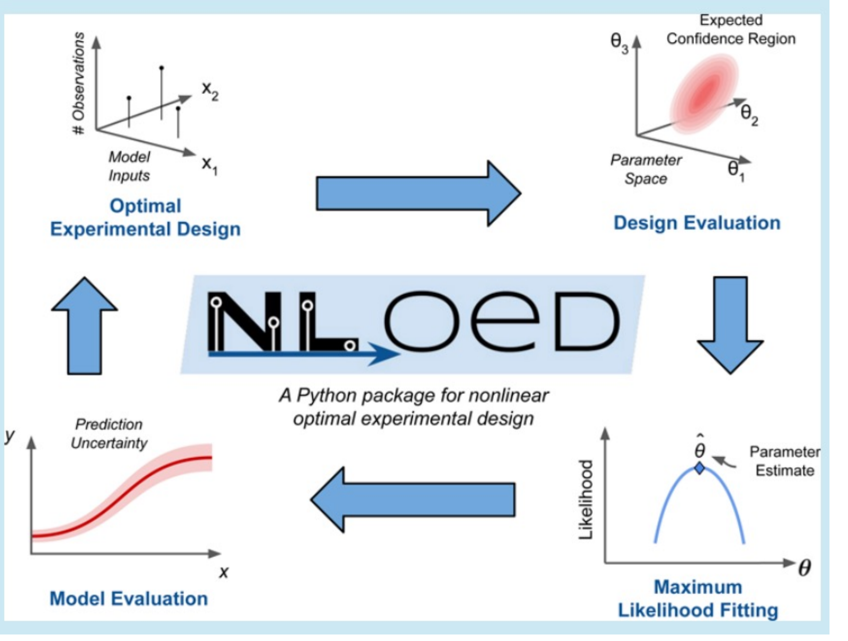


D-optimality score

Parameter Variance

Software package

Python package: one-stop-shop for FIM-based (local) Model-Based Optimal Experimental Design

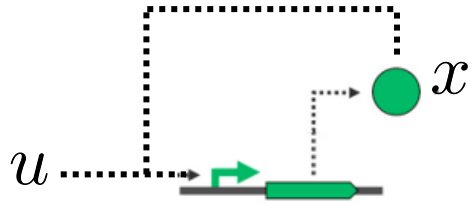


- -Sequential design workflow
- -Nonlinear models
- -Non-Gaussian distributions, Poisson (e.g. plate counts), log-normal (e.g. gene expression), Bernoulli or binomial (e.g. viability assays)
- -Symbolic model construction
- -Sensitivities: automatic differentiation with CasADi
- -Nonlinear programming: IPOPT
- -D-optimal design over sampling and input profiles
- -Integer sample counts relaxed to real-valued weights, then rounded
- -Auxiliary methods: Maximum likelihood model fitting, sensitivity analysis, model simulation, data sampling, and design evaluation

Braniff, et al. "NLoed: A Python package for nonlinear optimal experimental design in systems biology." *ACS SynBio* 2022

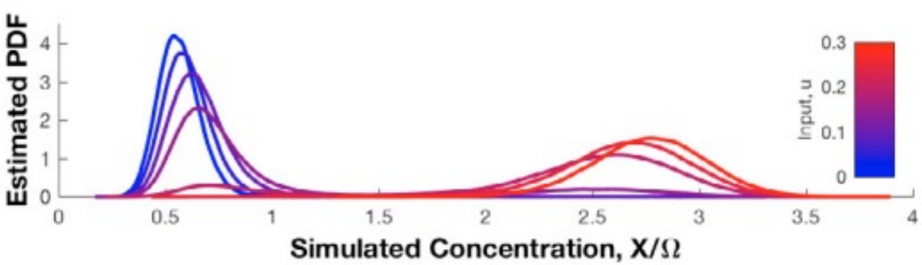
OED for multimodal gene expression system

Bistable autoactivating gene expression with activating external input

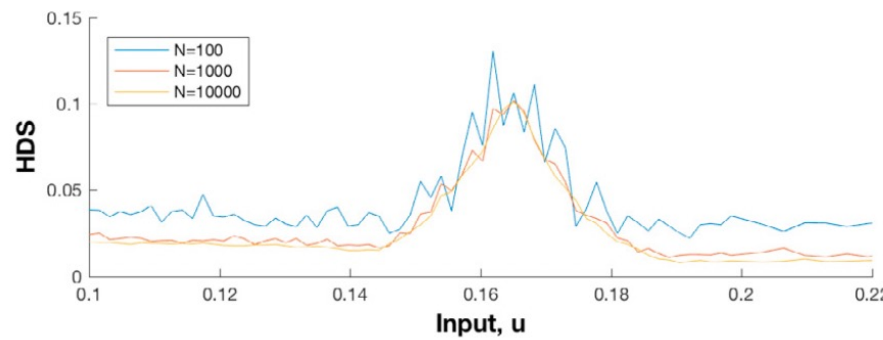


$$\frac{d}{dt}x(t) = \alpha_0 + \alpha\frac{(u+x(t))^n}{K^n + (u+x(t))^n} - x(t)$$

Observations bimodally distributed



Identified from steady state data via Hartigan dip statistic



Approximate log-likelihood: Gaussian mixture mode

$$\ell(\theta|D,U) = \sum_i \log\{\rho(u_i)\cdot\varphi_T(y_i|u_i,\theta) + [1-\rho(u_i)]\cdot\varphi_B(y_i|u_i,\theta)\}$$

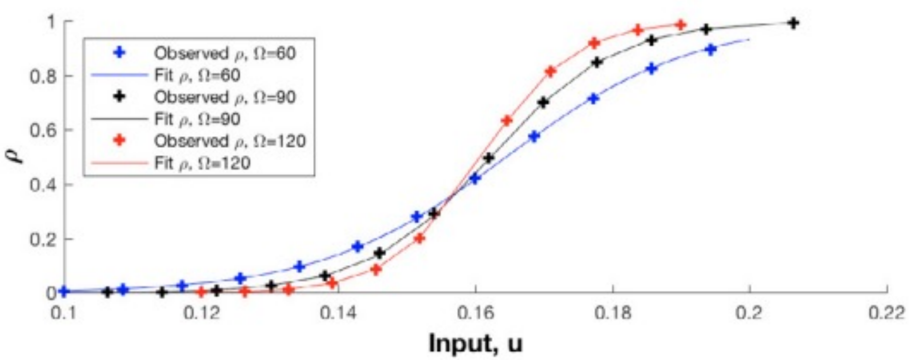
Braniff, Richards and Ingalls "Optimal experimental design for a bistable gene regulatory network." *IFAC-PapersOnLine* (2019)

OED for multimodal gene expression system

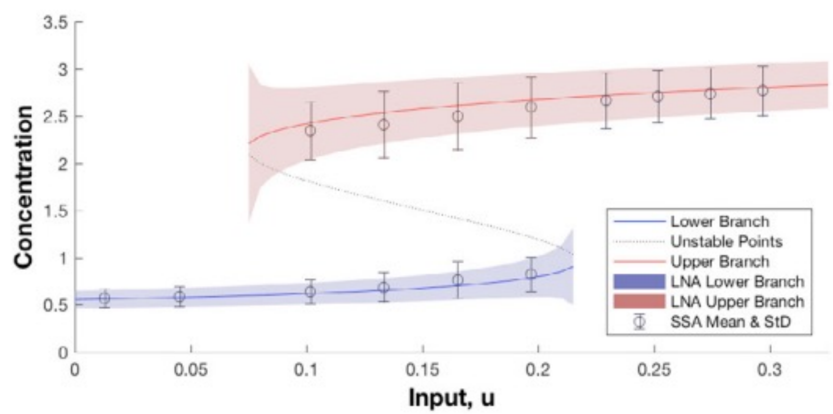
Approximate log-likelihood: Gaussian mixture mode

$$\ell(\theta|D,U) = \sum_i \log\{\rho(u_i)\cdot\varphi_T(y_i|u_i,\theta) + [1-\rho(u_i)]\cdot\varphi_B(y_i|u_i,\theta)\}$$

Logistic approximation of probability of each mode (based on Kramers-Moyal approximation of escape times)



Linear Noise Approximation to estimate normal distribution around each mode:



Validated by SSA

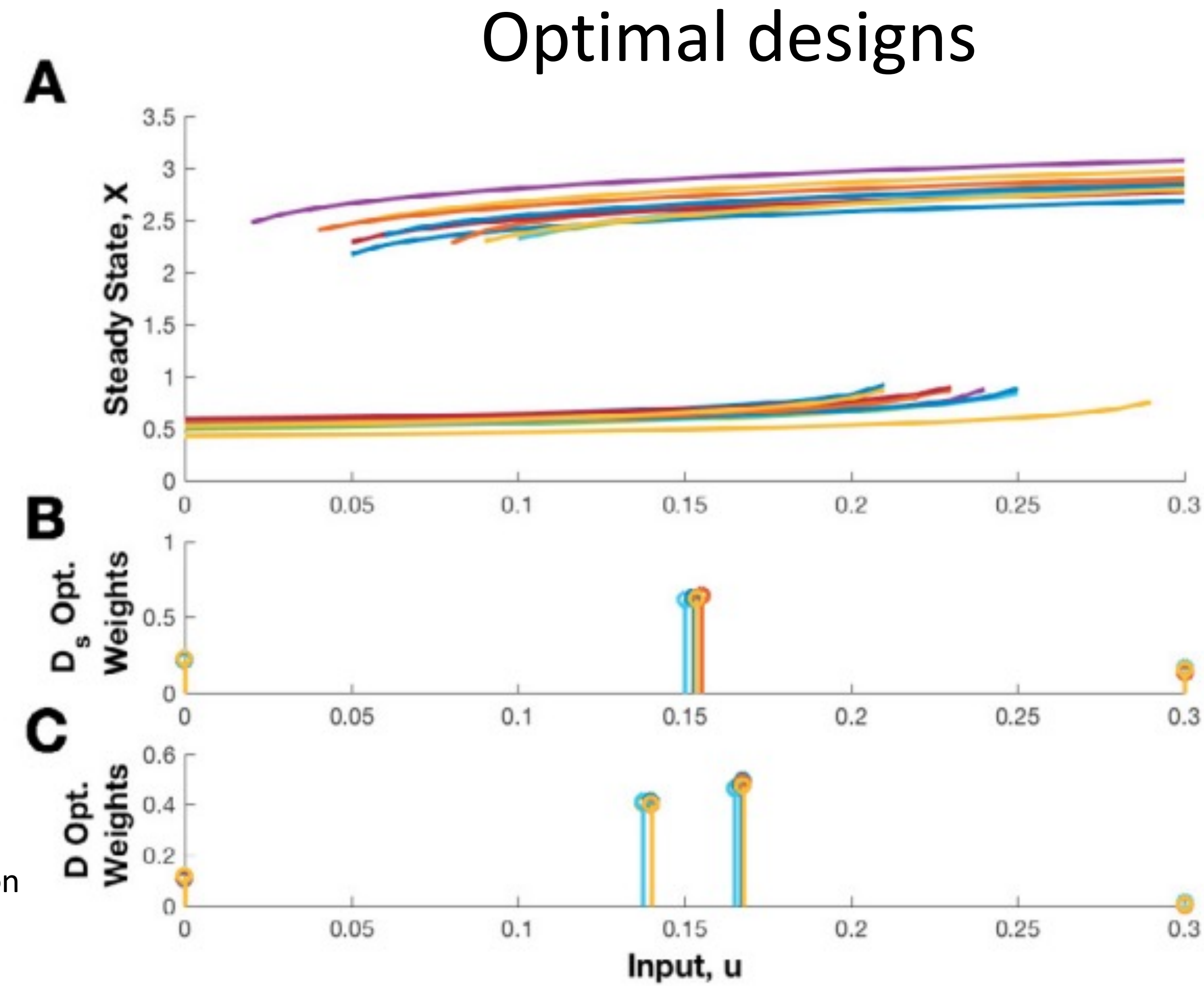
Braniff, Richards and Ingalls "Optimal experimental design for a bistable gene regulatory network." *IFAC-PapersOnLine* (2019)

OED for multimodal gene expression system

Stable branches
(10 sampled
parameterizations)

Optimal for inferring
kinetic parameters

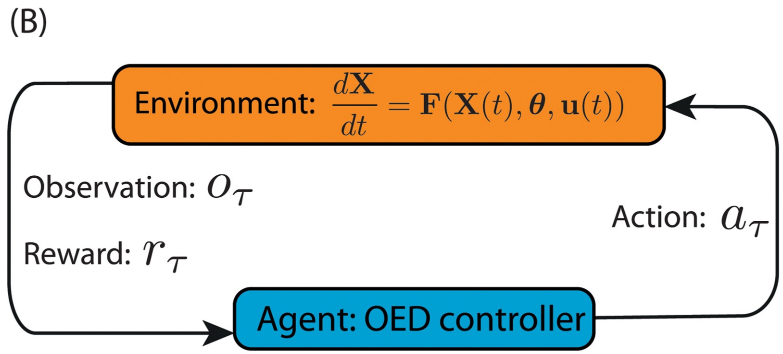
Optimal for inferring
kinetic parameters and
location of bistable region



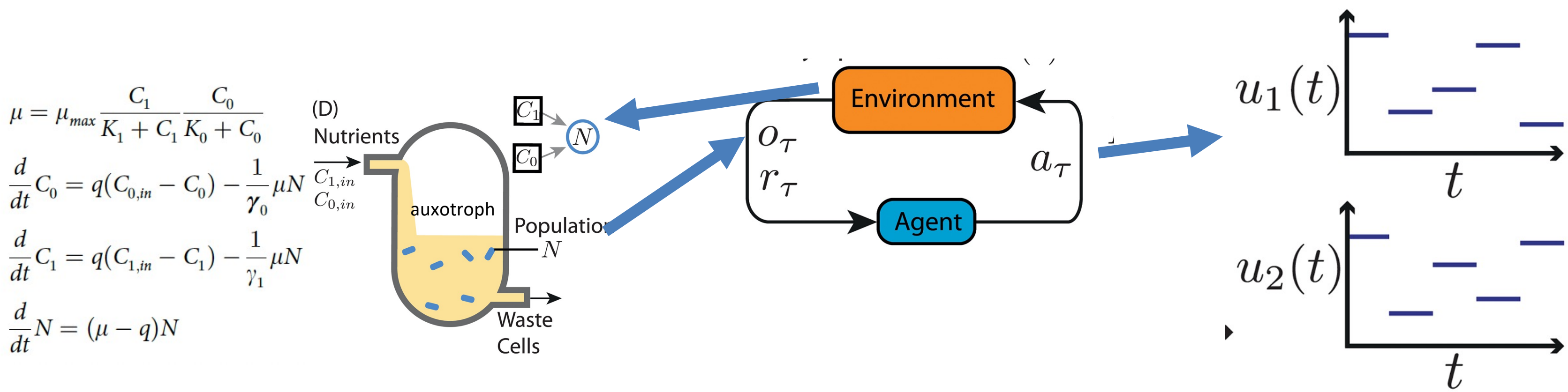
Braniff, Richards and Ingalls "Optimal experimental design for a bistable gene regulatory network." *IFAC-PapersOnLine* (2019)

Deep reinforcement learning for OED

Reinforcement learning:
Agent receives **observation** and **reward;**
Implements **action** on **environment**



environment (system)
observation (time step, state measurement, estimate of FIM)
reward (increment in FIM)
action (experimental input)



Advantage: learn model parameters while optimizing designs

Limitation: data hungry

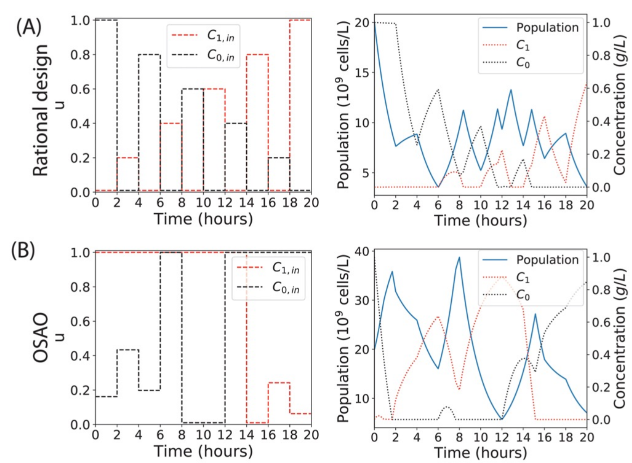
Treloar, Braniff, Ingalls, and Barnes. "Deep reinforcement learning for optimal experimental design in biology." *PLOS Computational Biology* (2022)

Fitted-Q learning over discrete action space: value function as deep neural network

Baseline performance: access to true parameter values

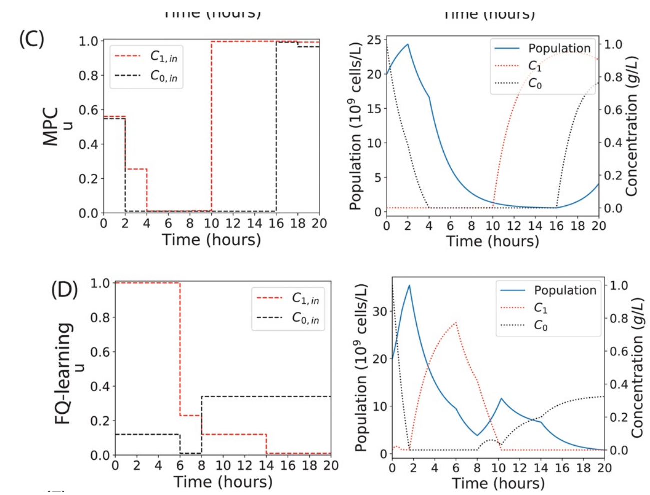
Rational (human) design

Greedy optimization



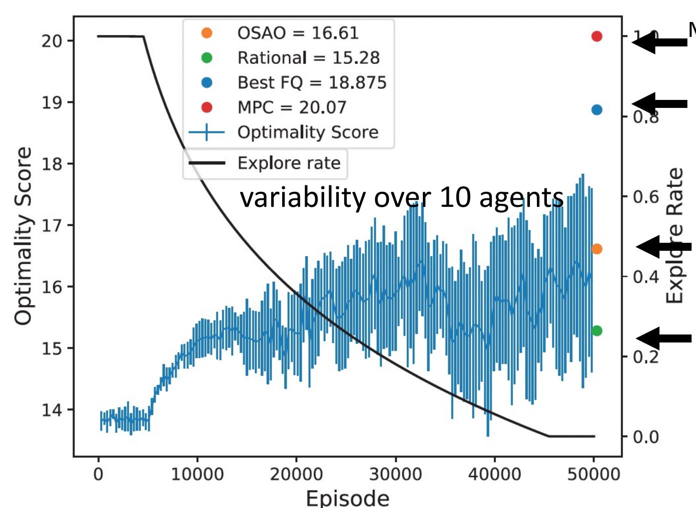
Model predictive control

Reinforcement learning



Performance:

Training over 50000 simulated experiments



Model predictive control

Reinforcement learning

Greedy optimization

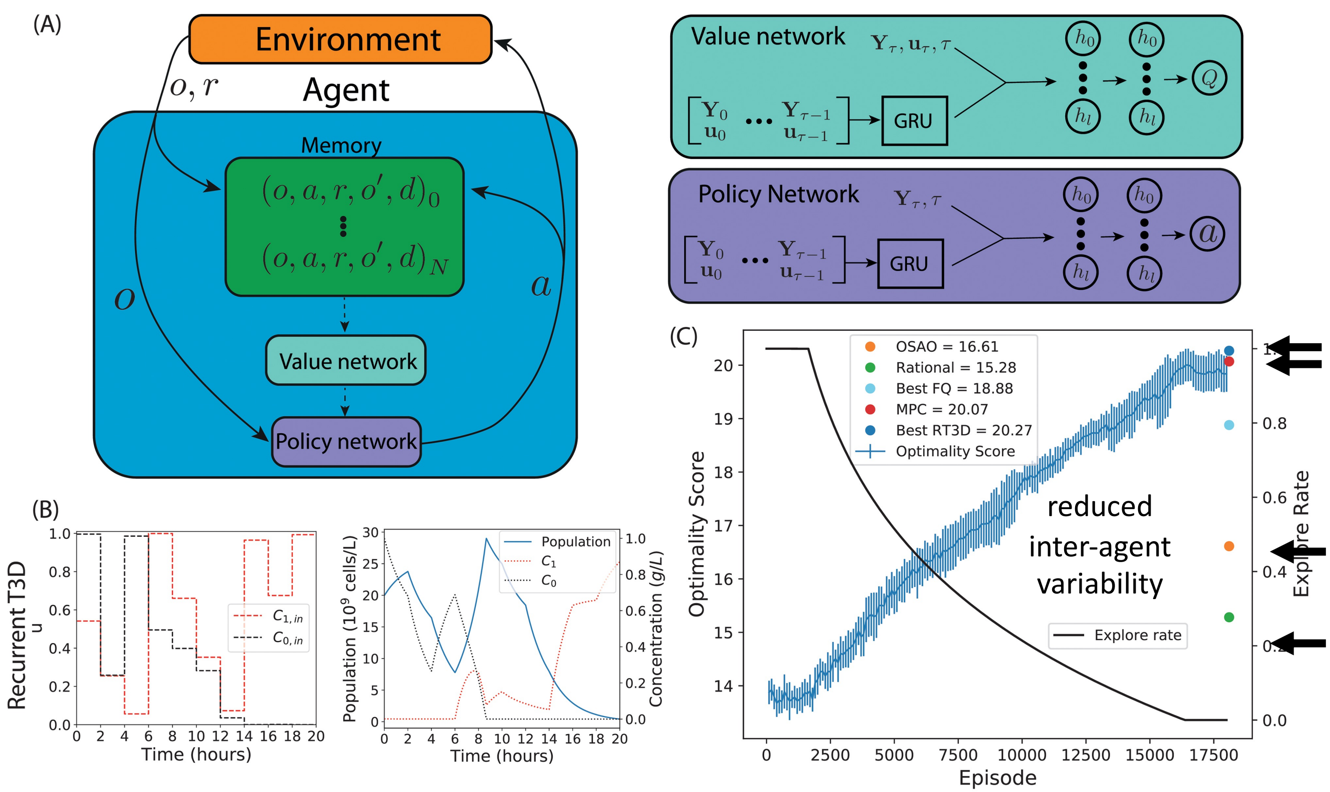
Rational design

Similar results for parameter covariance and estimation accuracy

RL: promising results

Algorithm refinement: Recurrent Twin Delayed Deep Deterministic Policy Gradient (RT3D)

-Observation includes past history (allows learning of unknown parameter values)
-Continuous action space (requires additional recurrent neural network for feedback policy)



Baseline Performance

Training over 17500 simulated experiments

access to true parameter values

RT3D Reinforcement learning
Model predictive control

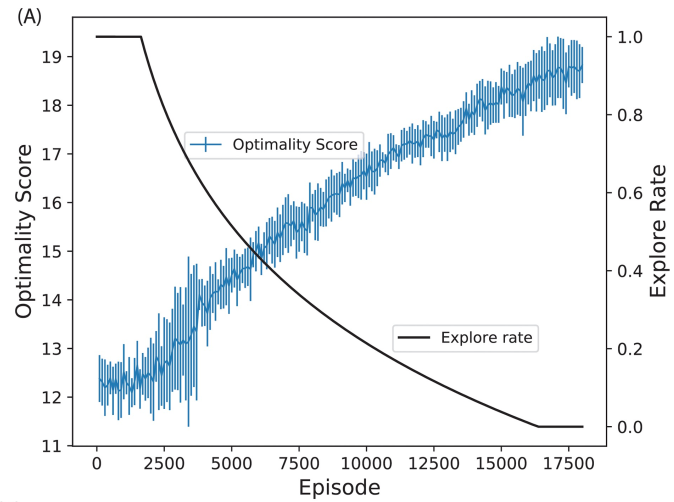
Greedy optimization

Human design

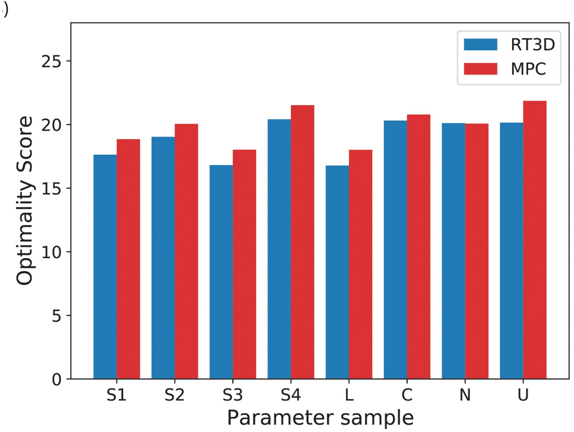
RL: equivalent to MPC

Agent performance over parameter distribution

10 agents. Each training simulation sampled from a uniform distribution



Performance comparison with **MPC acting with knowledge** of true sampled parameter values

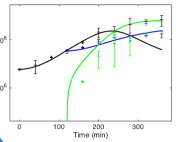


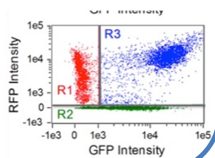
equivalent performance despite lack of a priori knowledge of parameter values

RL: improved robustness to parameter uncertainty in comparison to MPC

Acknowledgments

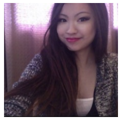
Nonspatial characterization





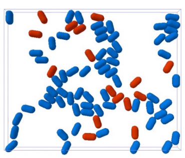
Akshay Malwade

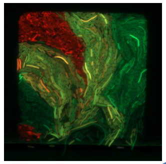
Angel Nguyen



Peivand Sadat-Mousavi

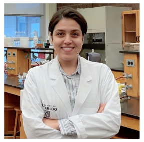
Single-cell analysis





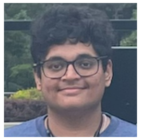
Aaron Yip

Atiyeh Ahmadi



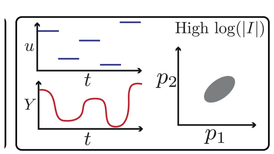
Matt Courtney

Dhruva Rajwade



Nat Kendal-Freedman

Experimental Design



Nathan Braniff

Addison Richards

Neythen Treloar

WATERLOO iGEM

NSERC CRSNG

INNOVATION.CA
CANADA FOUNDATION FOR INNOVATION | FONDATION CANADIENNE POUR L'INNOVATION

Mitacs Globalink

Introductory modelling textbook:

PDF freely available at www.math.uwaterloo.ca/~bingalls/MMSB/