Harnessing trajectory ensembles for rates, reaction coordinates, and mechanism

Jeremy Copperman

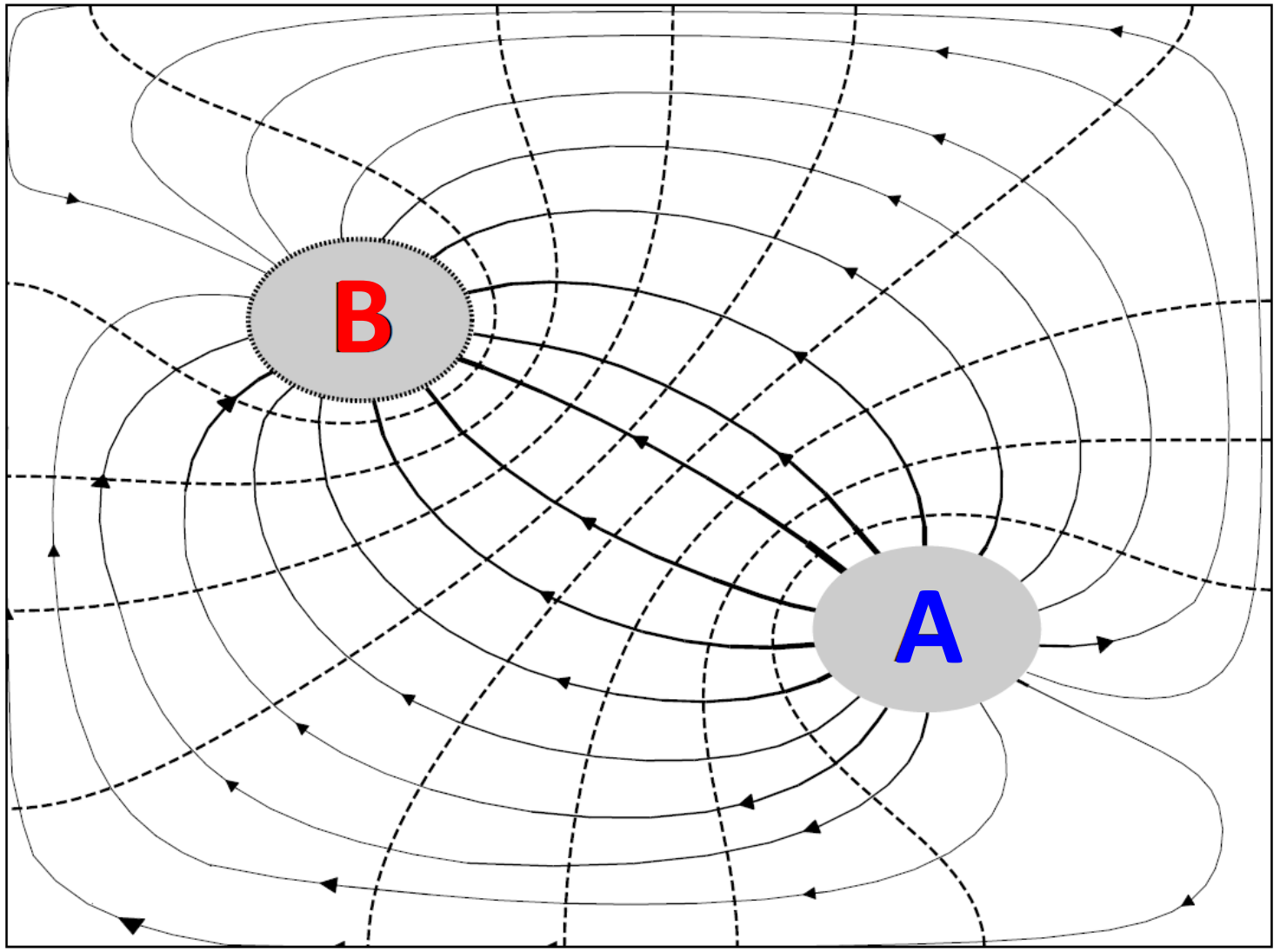
Open Eye Cup, March 10, 2022

School of Medicine
Biomedical Engineering

KNIGHT CANCER Institute

DAMON RUNYON CANCER RESEARCH FOUNDATION

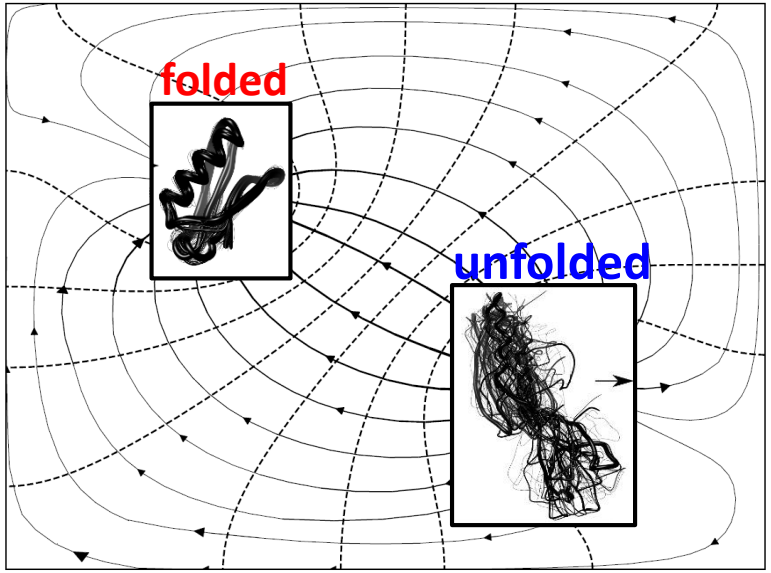
the
A-to-B
dipole



A trajectory-based framework to determine the mechanism, dynamics, and control of complex systems.

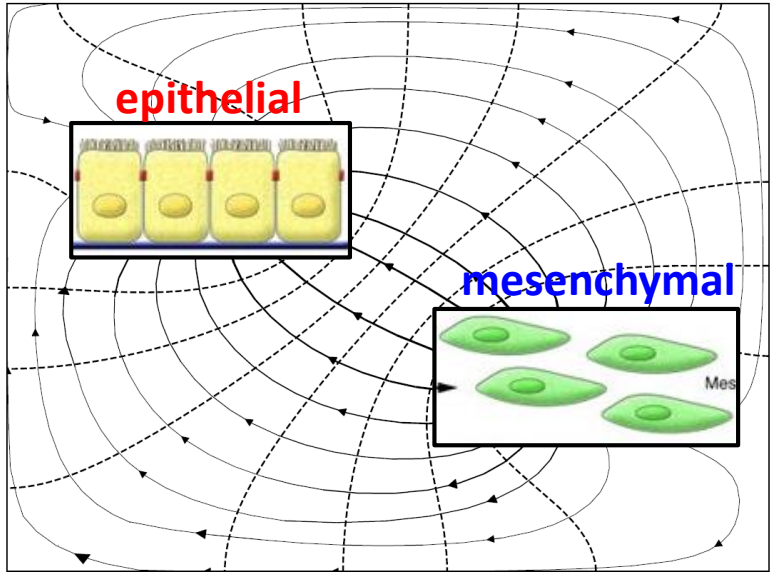
1. Strategies for rate estimation using weighted ensemble: history-augmented Markov State Models (haMSMs) and optimal binning

   with John Russo, David Aristoff, Gideon Simpson, and Daniel Zuckerman



2. Do cells have transition states which can be leveraged to control cell-state transitions?

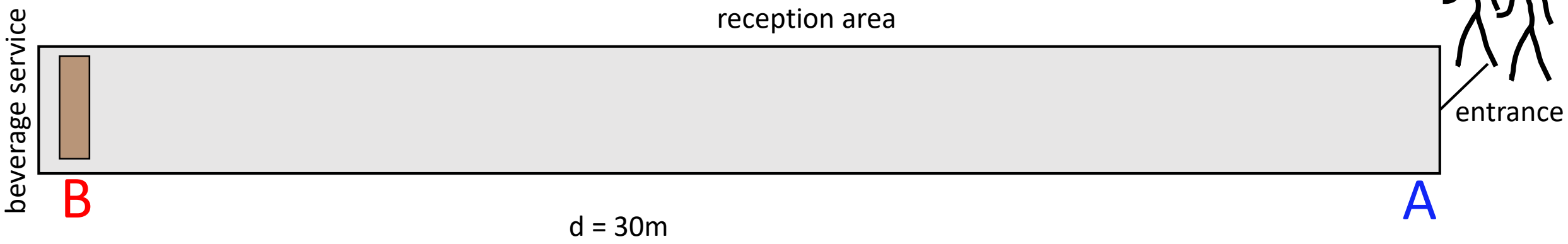
   with Young Hwan Chang, Laura Heiser, and Daniel Zuckerman



Relevant applications of one-way A-to-B ensembles

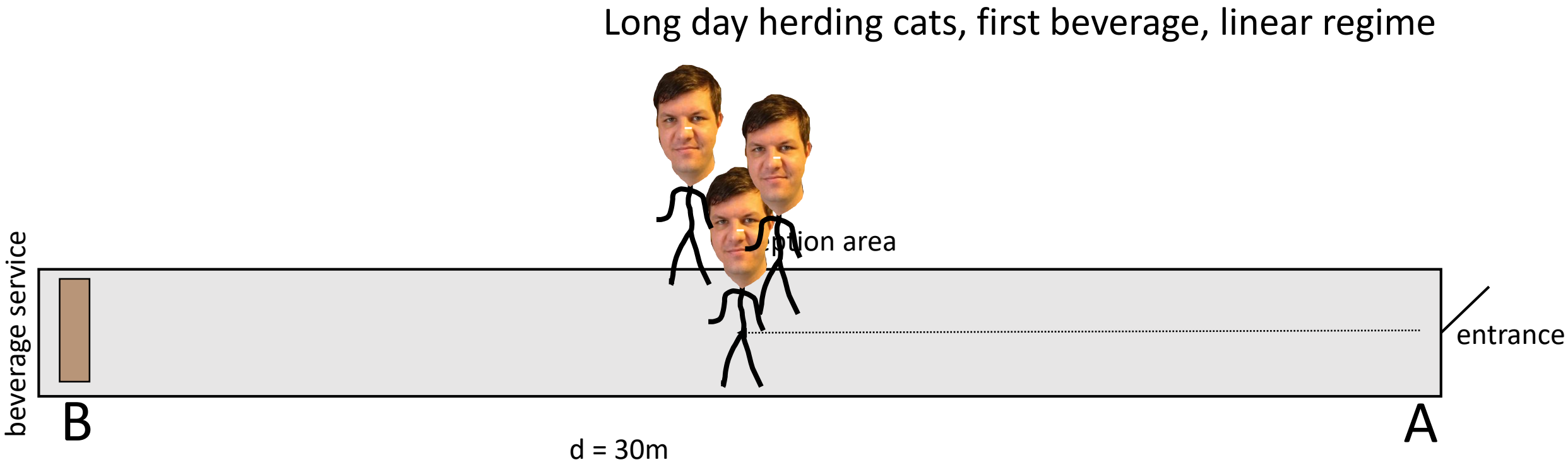
Step 1: Launch non-interacting Dr. LeBard ensemble from A

Step 2: Wait for arrival at B



- Q: How can we determine the mechanism, kinetics, and control of Dr. LeBard's post-conference behavior?
- A: Observe one-way (A-to-B) trajectories traversing the reception area to the beverage service
- Thank you to our session organizer, Dr. David LeBard!

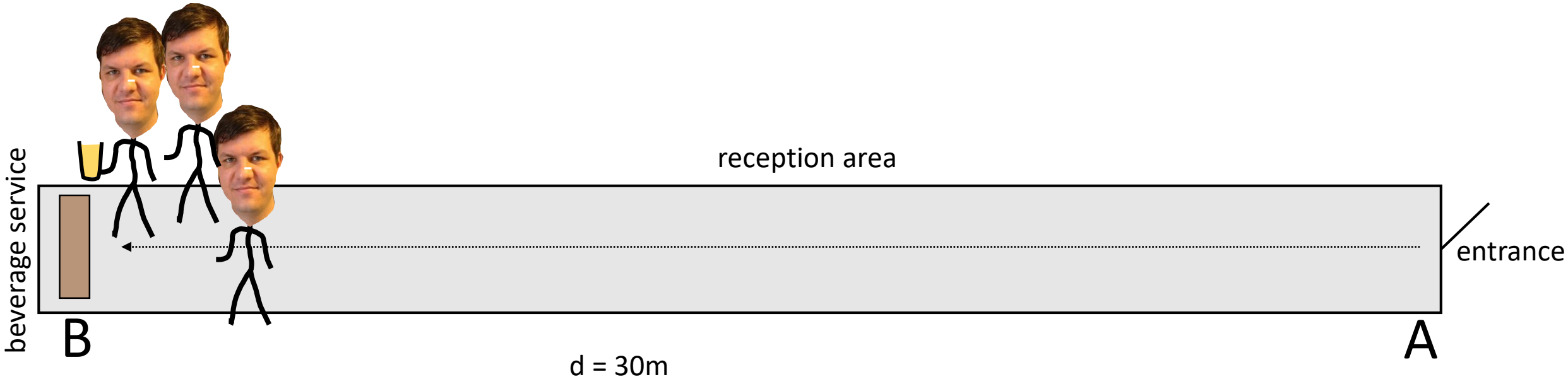
Direct trajectory collection: Start at A, wait for B



- mean first-passage time T = distance / velocity
- d=10m, v=1 m/s, T = 30 seconds
- easy to observe successful trajectories, low variance

Direct trajectory collection: Start at A, wait for B

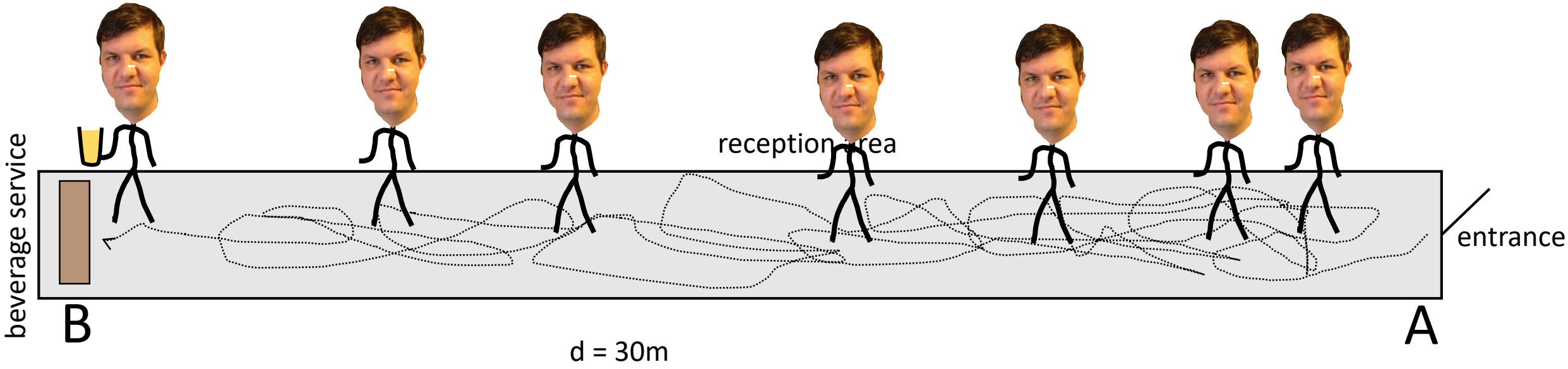
Long day herding cats, first beverage, linear regime



- mean first-passage time T = distance / velocity
- d=10m, v=1 m/s, T = 30 seconds
- easy to observe successful trajectories, low variance

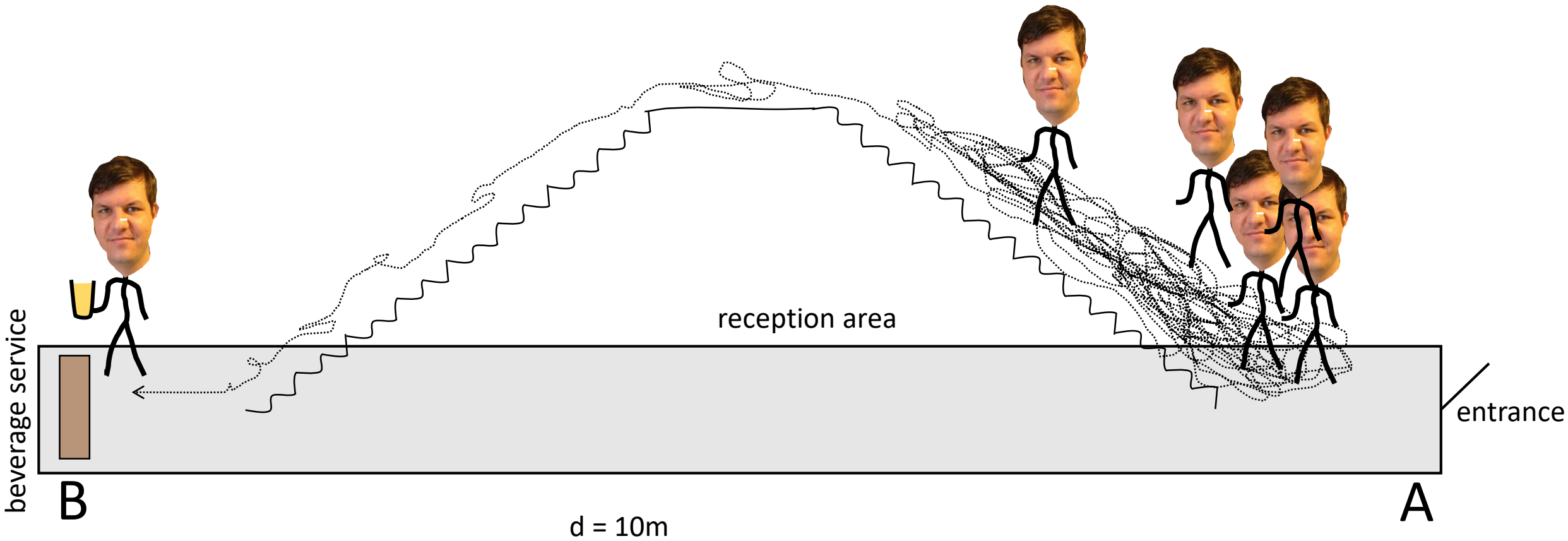
Direct trajectory collection: start at A, wait for B

Long day herding cats, infinite beverage, diffusive regime



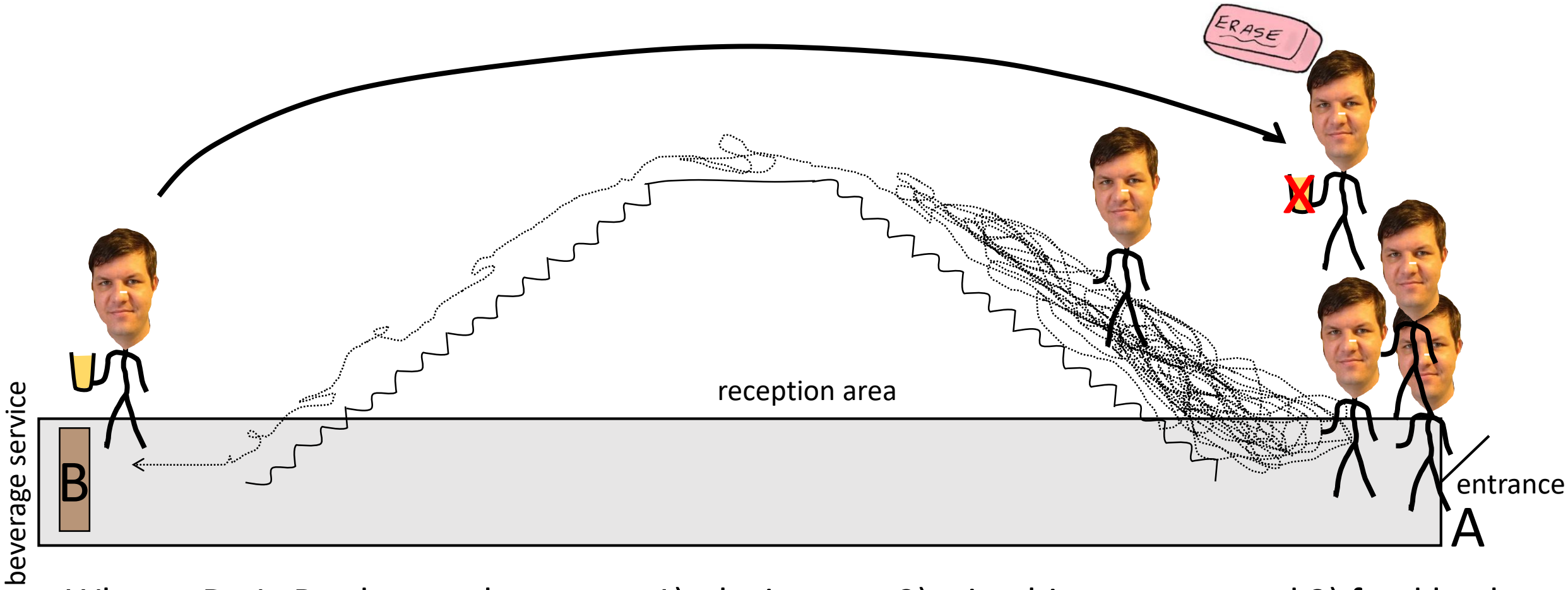
- Diffusion rate D = 0.5 m^2/s
- d=10m, T = $d^2/2D$ ~ 900 seconds
- Not too hard to observe successful trajectories, multiple trajectories needed

Direct trajectory collection: start at A, wait for B



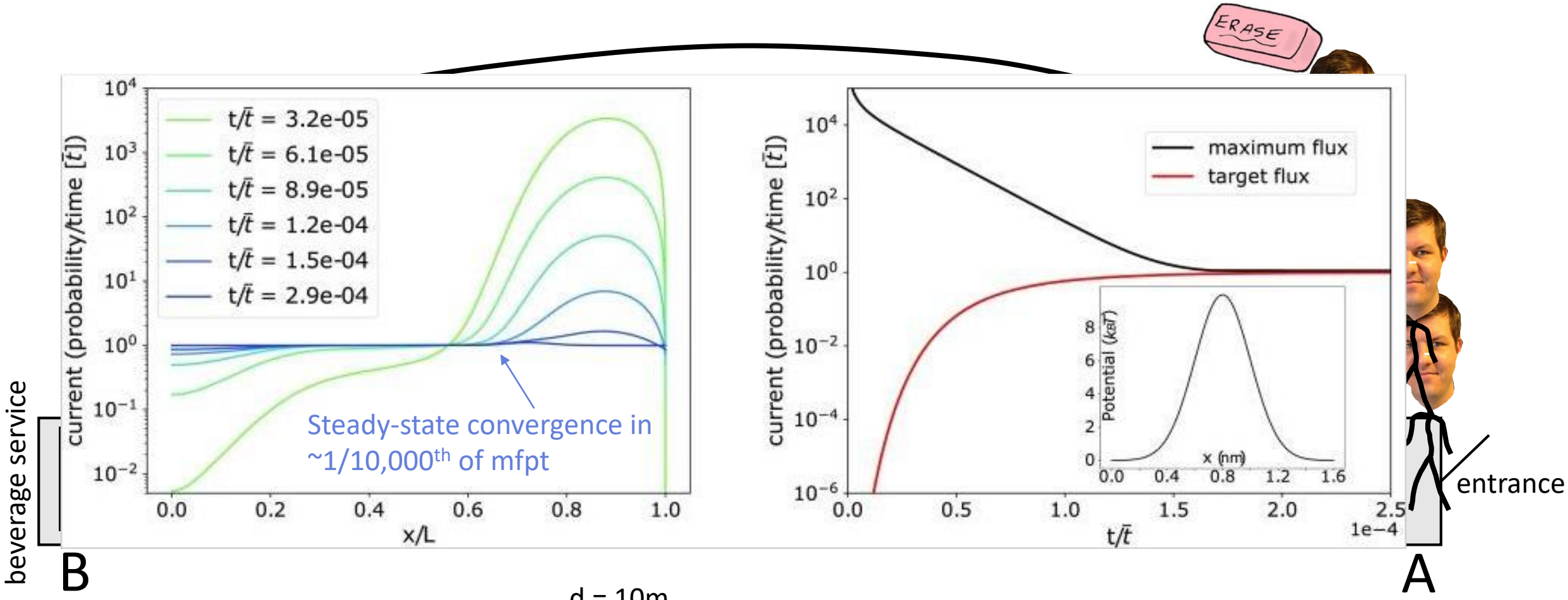
- Barrier height h, T $\propto \exp(h)$
- way too long to observe enough successful trajectories

Sampling the one-way ensemble using feedback



- When a Dr. LeBard gets a beverage, 1)take it away, 2) wipe his memory, and 3) feed back to A
- steady-state (SS) Dr. Lebard flux (rate constant) at B is 1/mfpt (Hill relation)
- SS density proportional to the sum of all one-way trajectories $\rho^{ss}_{\text{recycling}}(x) \propto \int \rho_{\text{absorbing}}(x,t)dt$

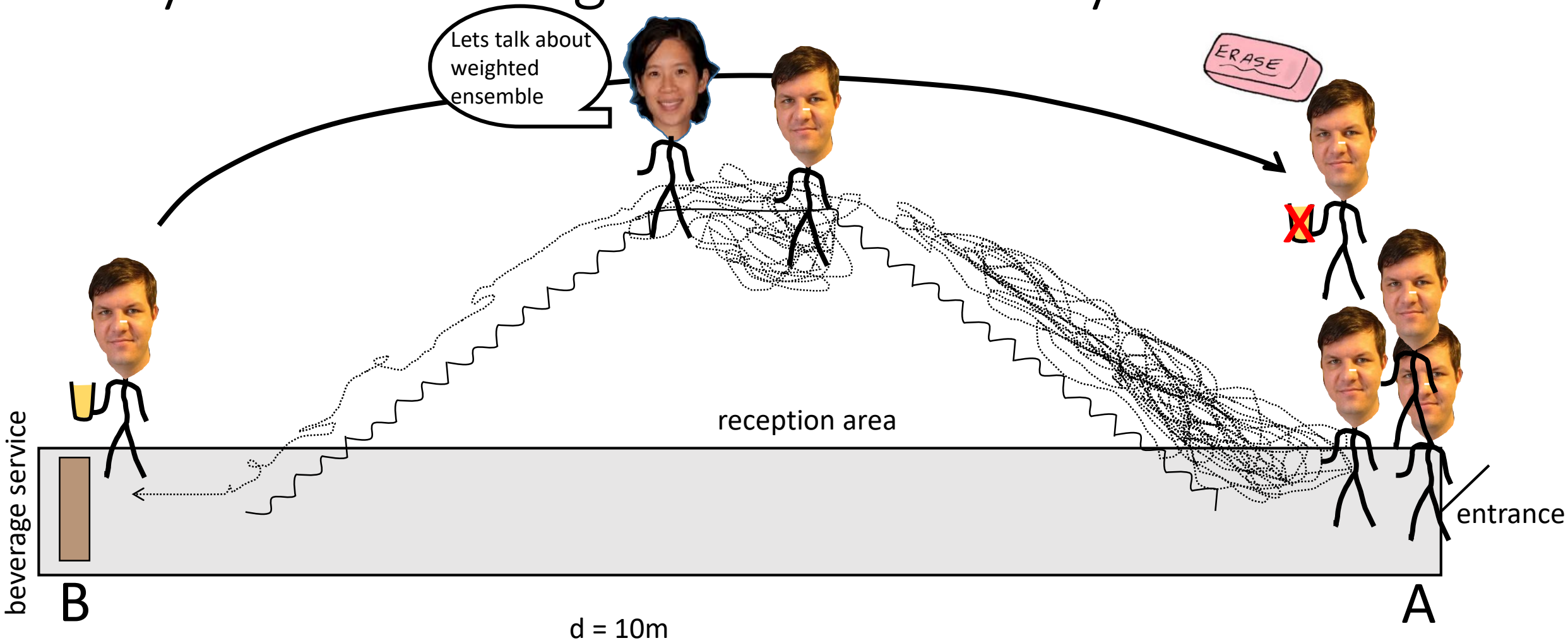
Sampling the one-way ensemble using feedback



- Steady-state convergence can be *arbitrarily* faster than the mean first-passage time $\bar{t}$

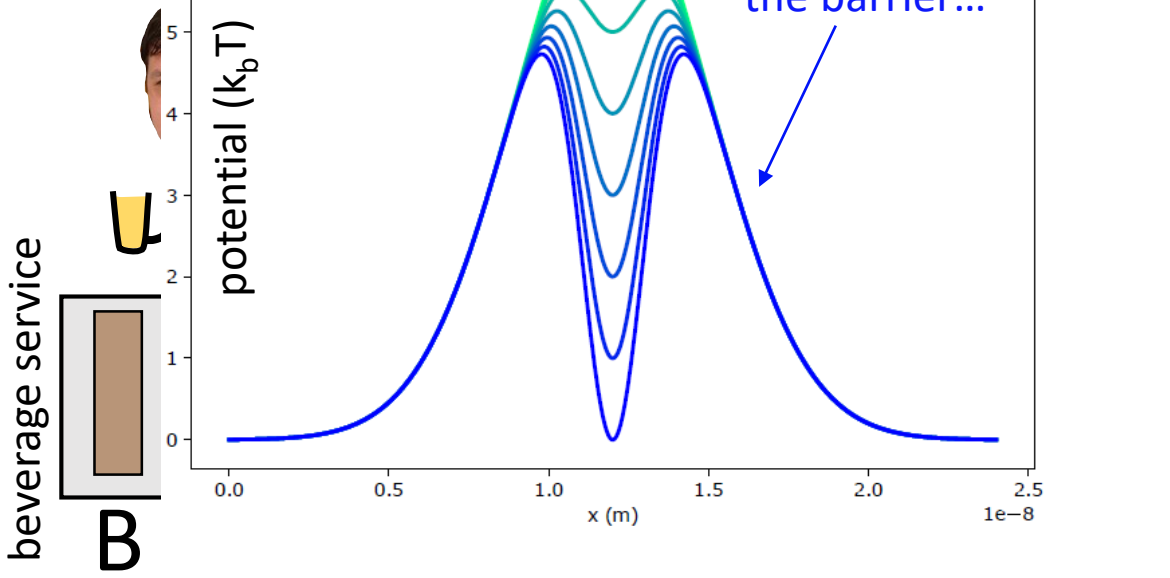
Copperman, J., Aristoff, D., Makarov, D. E., Simpson, G., & Zuckerman, D. M. JCP (2019).

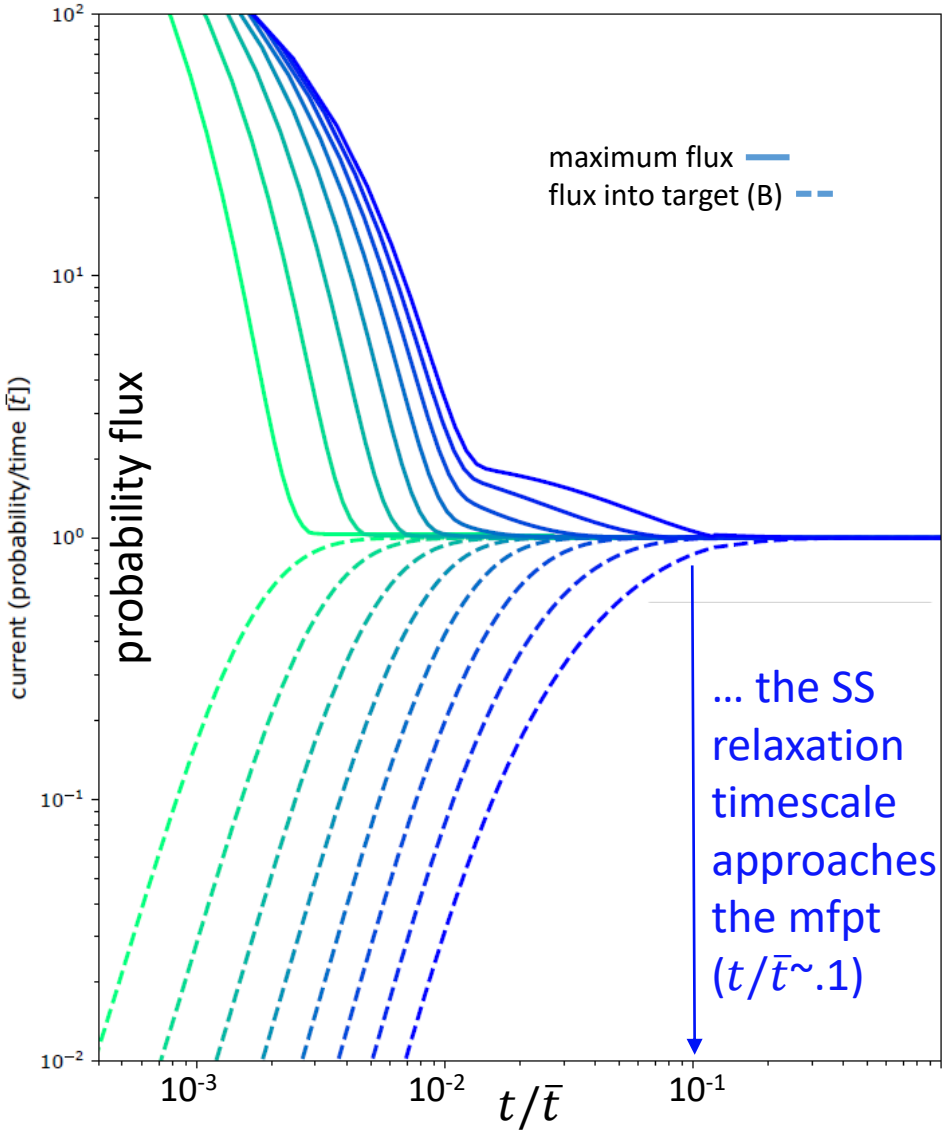
Steady-state convergence is not always fast



- metastable states along path slow SS convergence
- Thank you to session co-organizer Dr. Lillian Chong

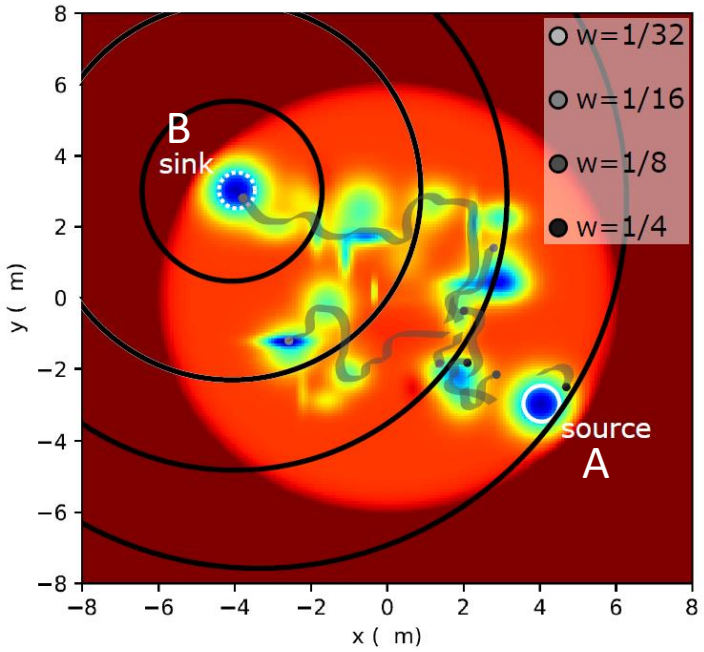
Steady-state convergence is not always fast



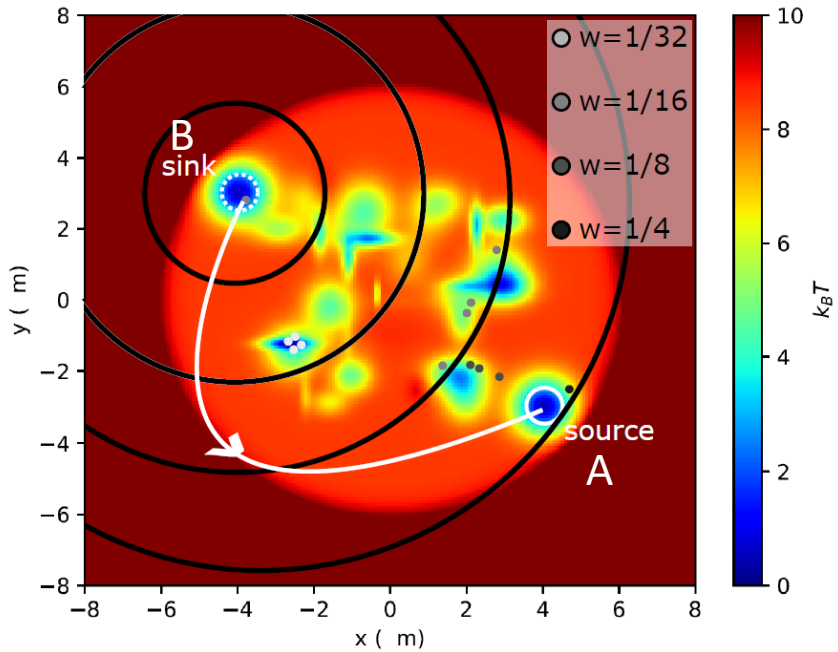


- metastable states along path make SS relaxation time same scale as mfpt

Weighted ensemble + feedback

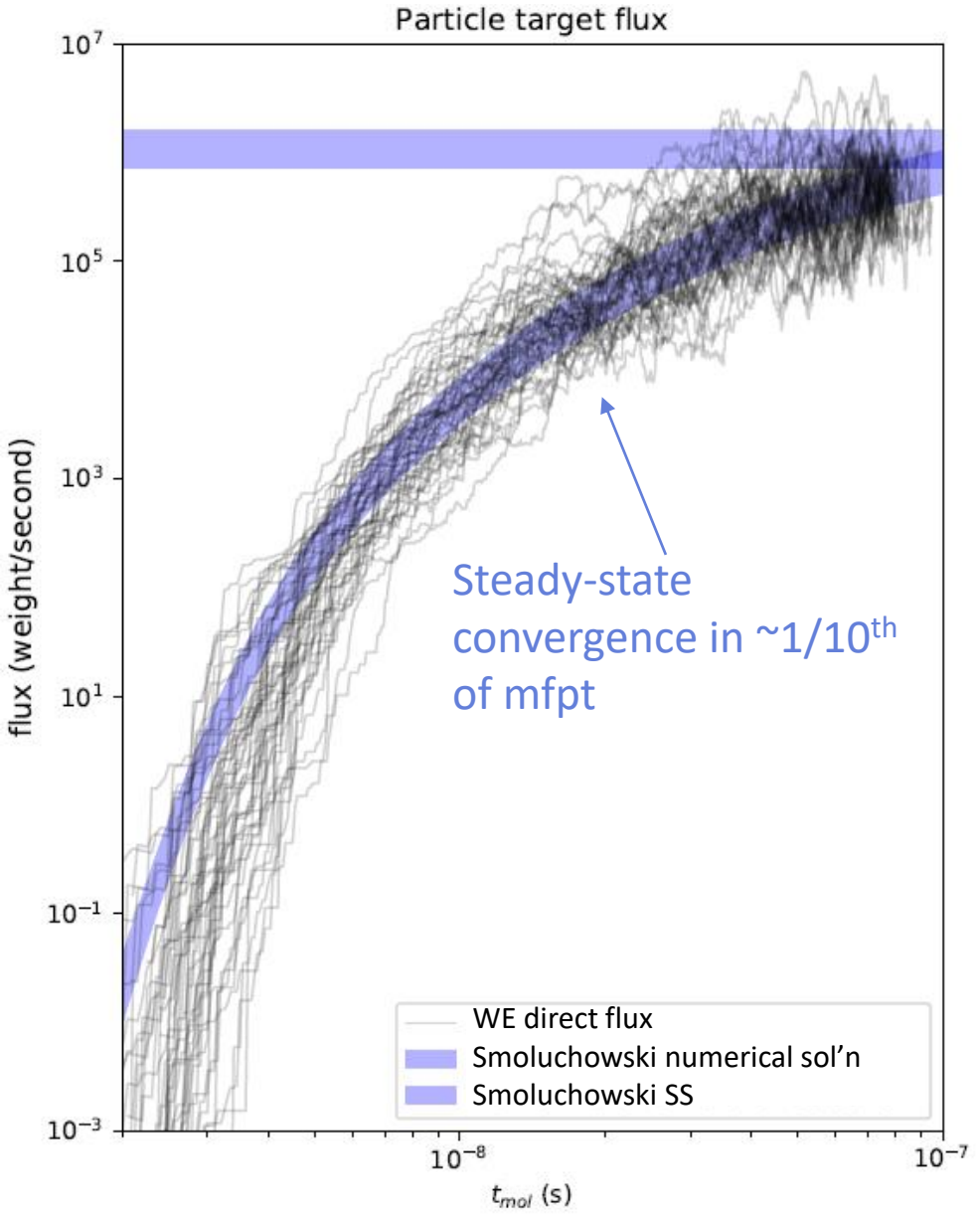


weighted ensemble trajectory weights sum to 1 and are split and merged to sample across bins without bias



One-way ensemble enforced--trajectories reaching the sink (B) feed back to the source (A)

Bhatt, D., Zhang, B. W., & Zuckerman, D. M. JCP (2010).

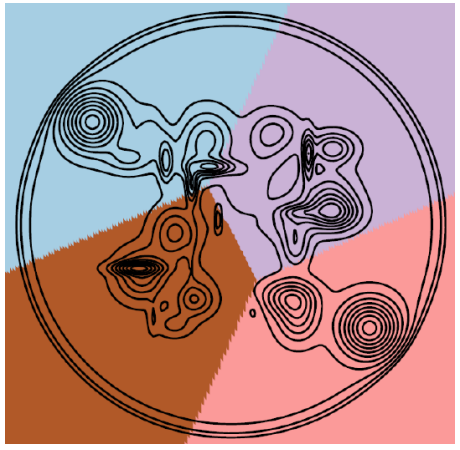


- Steady-state convergence may be as computationally expensive as brute force

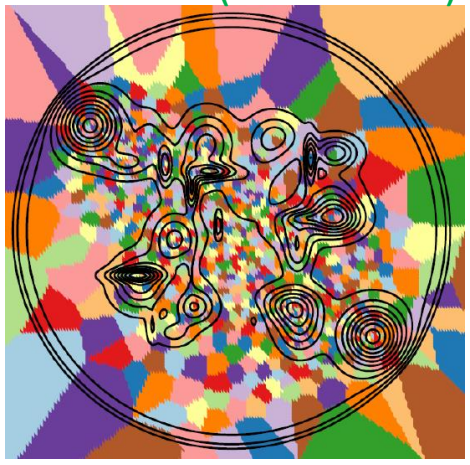
haMSM accelerated rate estimation

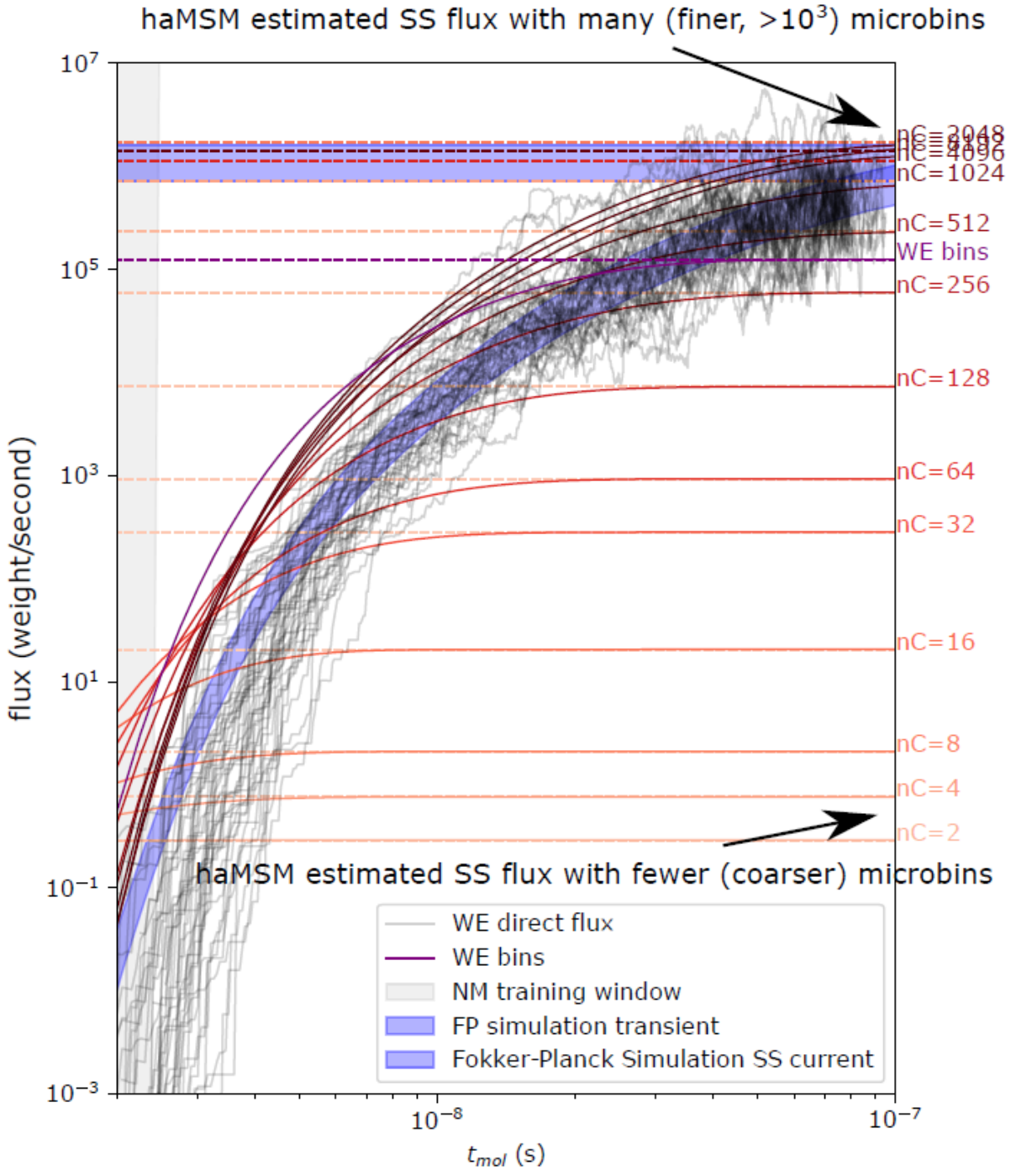
- history-augmented MSM is just a transition matrix built from A-to-B trajectories
- In the steady-state limit yields the unbiased A-to-B mfpt regardless of bin definitions
- Suarez, Lettieri, Zwier, Stringer, Subramanian, Chong, and Zuckerman, JCTC (2014).
- Only requires intrabin local SS convergence

no acceleration in bins with slow internal convergence (4 bins)

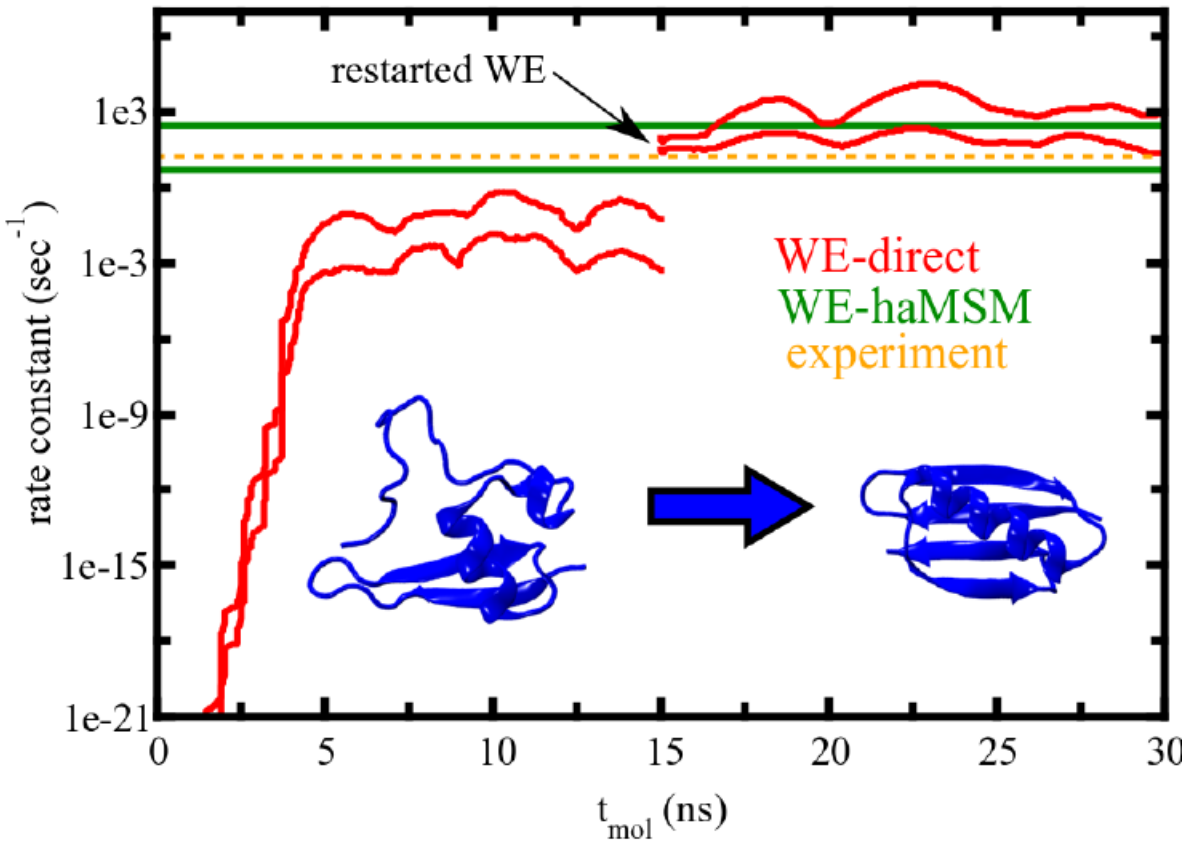


40x acceleration of rate estimation (~1000 bins)





haMSM accelerated rate estimation

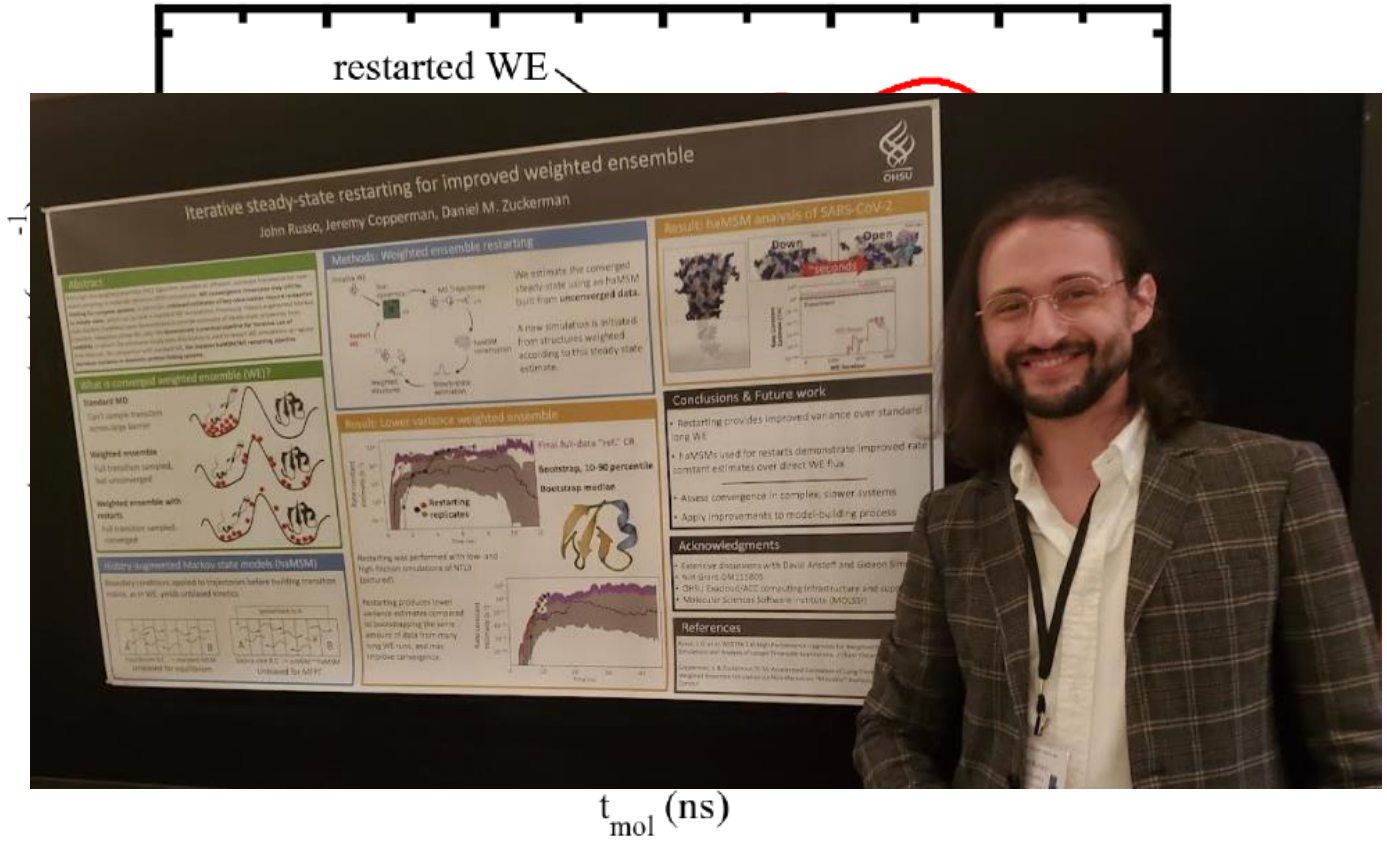


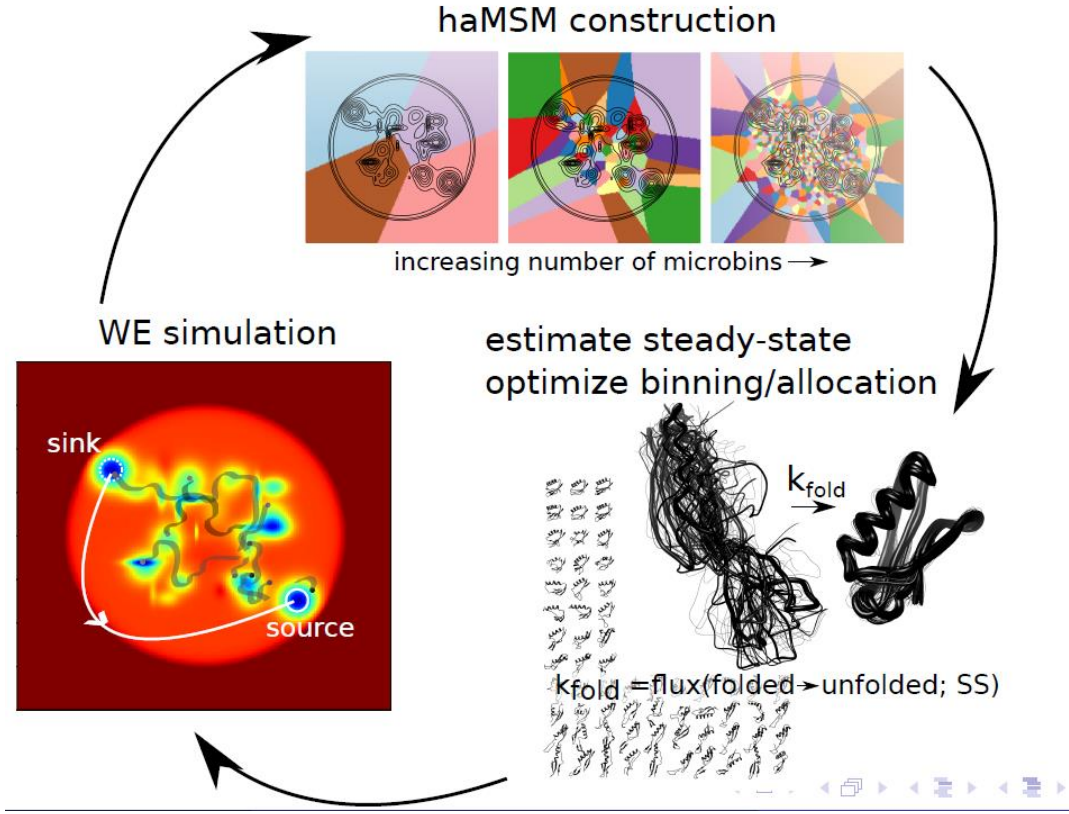
- Efficient estimation of millisecond-scale protein folding rates

Adhikari, Mostofian, Copperman, Subramanian, Petersen, and Zuckerman. JACS (2019).

Copperman and Zuckerman. JCTC (2020).

haMSM accelerated rate estimation

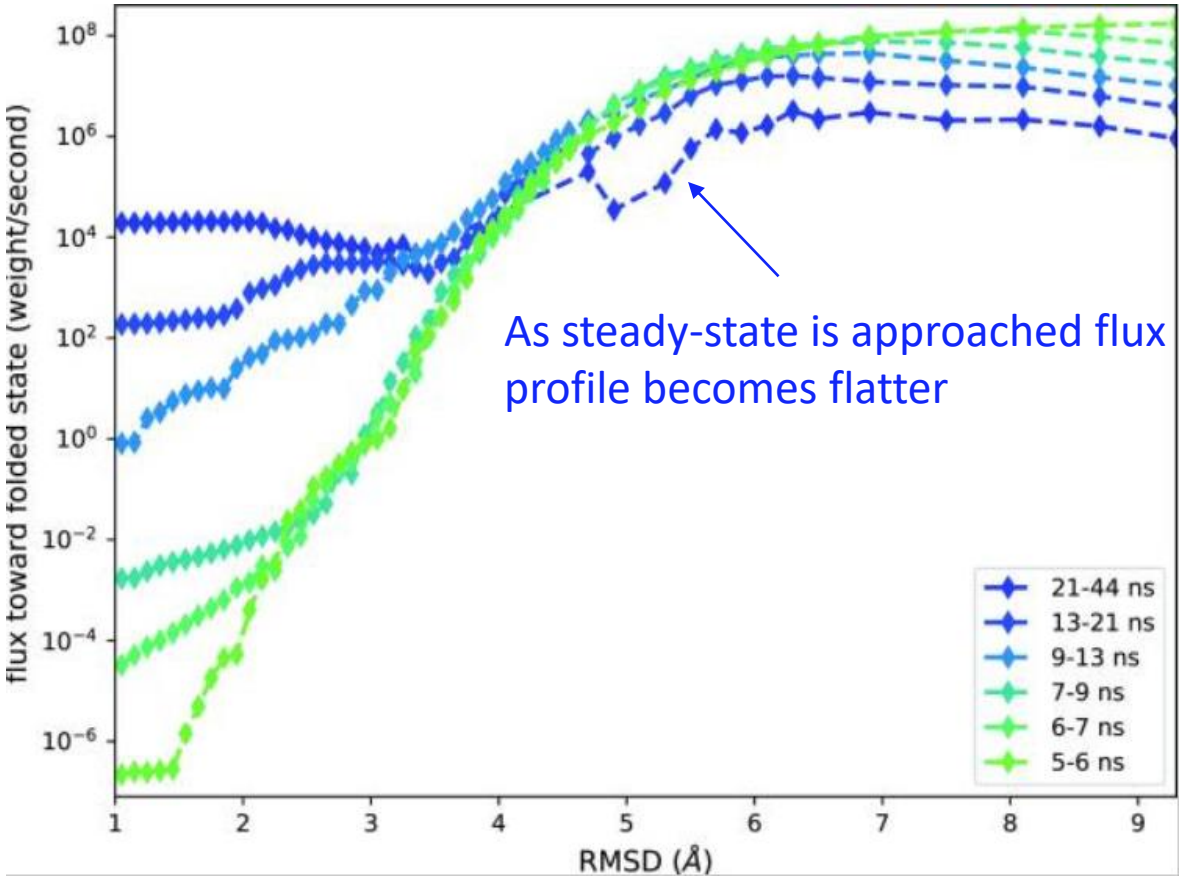


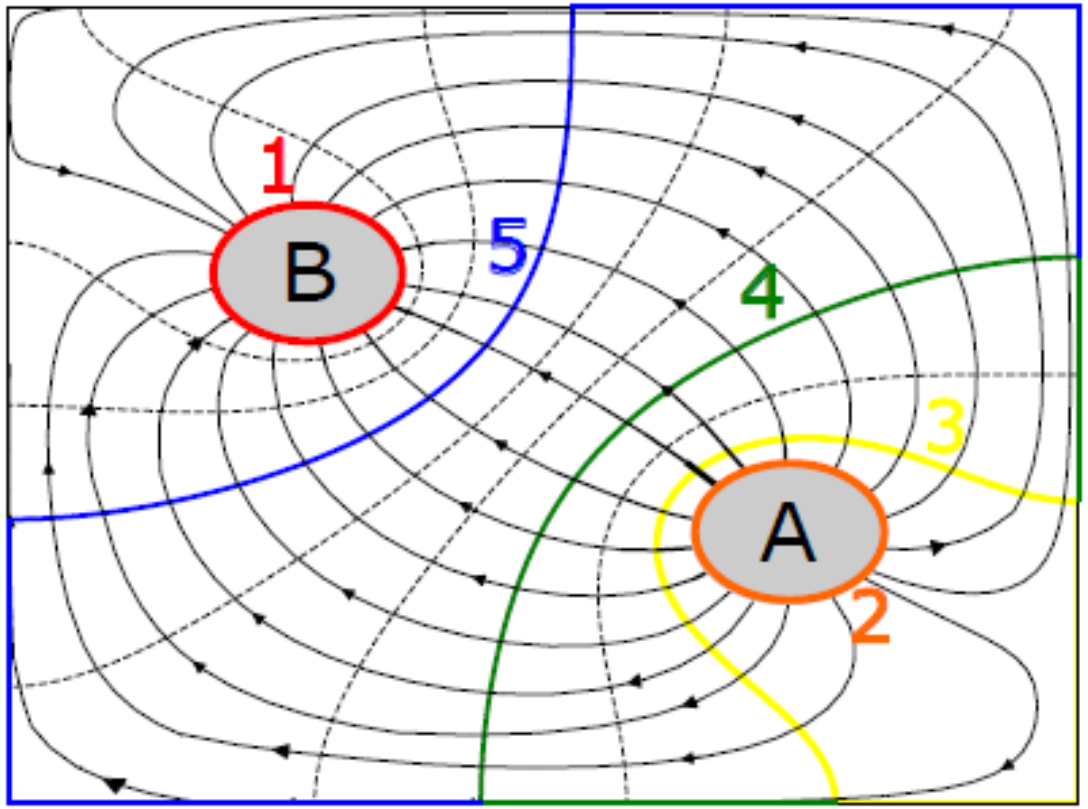


Improved workflow and tools, integration into WESTPA 2.0, and iterative restarting capability! Talk to John Russo

When is a trajectory ensemble converged?

Gauss's law for the A-to-B dipole: at steady-state the flux through any surface separating initial state A and final state B is constant



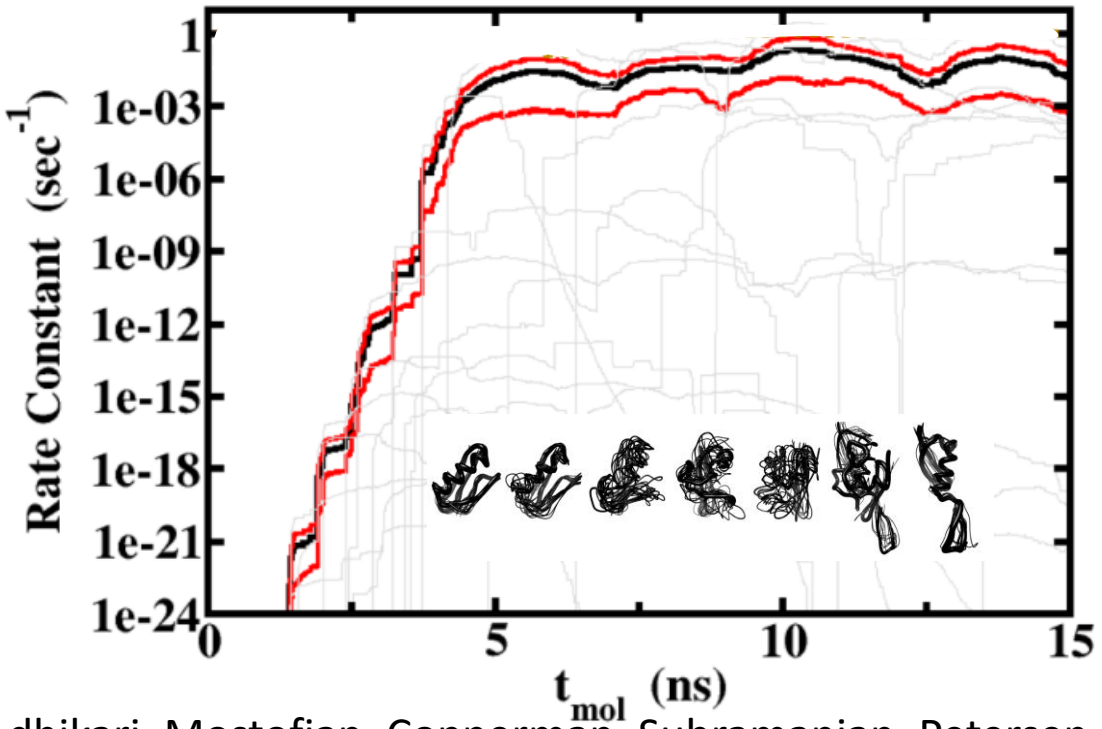


completely flat flux profile is an absolute measure of SS convergence… but may be overly restrictive

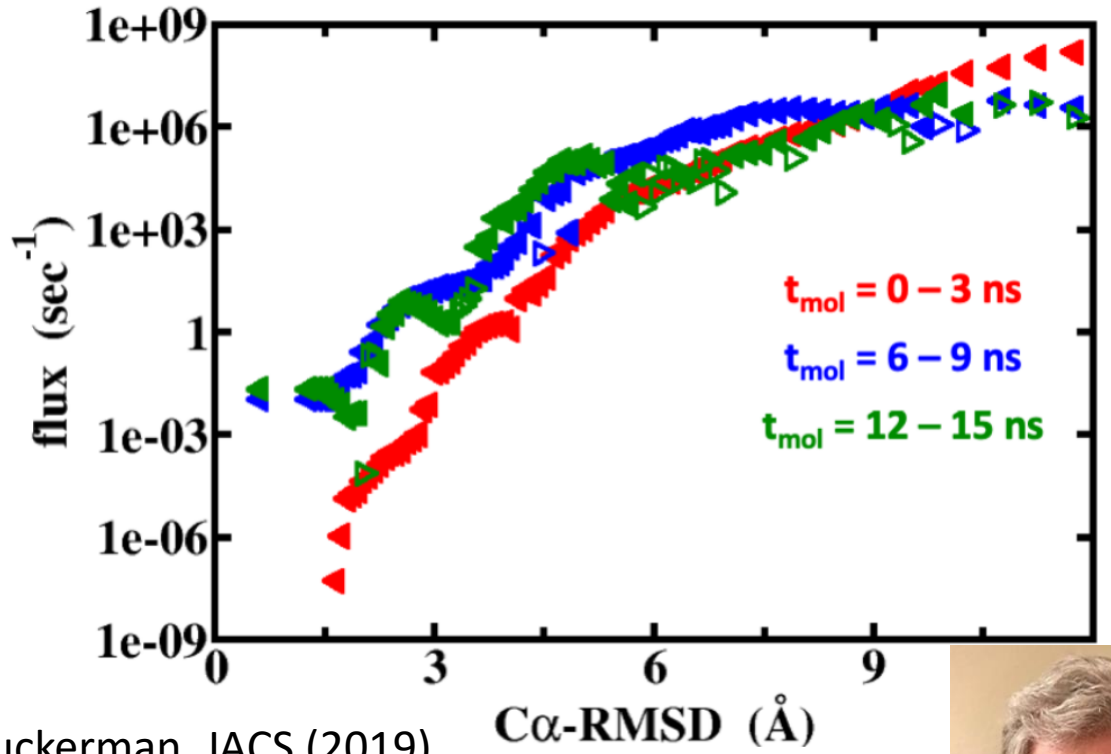
When is a trajectory ensemble converged?

being stuck looks a lot like convergence/equilibration

flux vs. time looks flat…



flux profile not converged

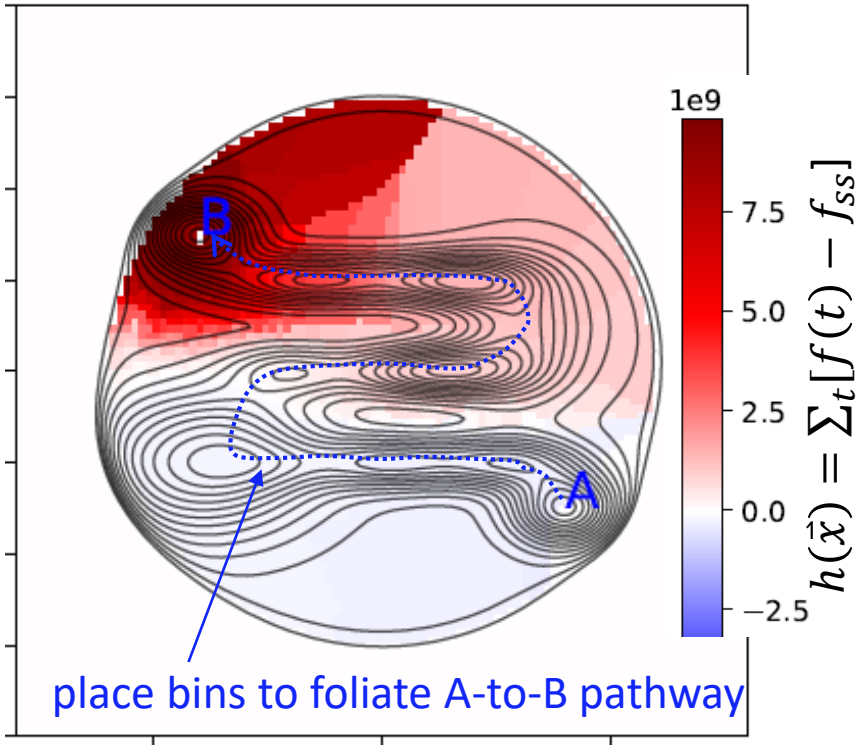


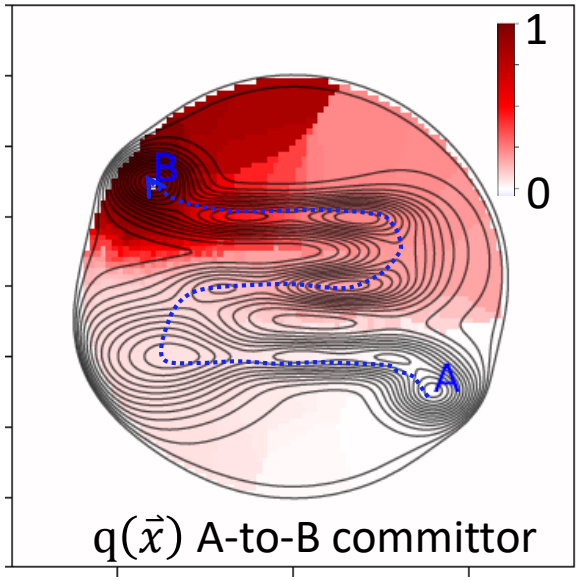
Adhikari, Mostofian, Copperman, Subramanian, Petersen, and Zuckerman. JACS (2019).

Lets tackle convergence in MD– need more absolute metrics (beyond leveling off or self-consistency) which can say if a trajectory ensemble is converged

Controlling variance in A-to-B trajectory ensembles

- Optimal reaction coordinate $h(\vec{x})$ for controlling error is committor-like foliating progress from A-to-B



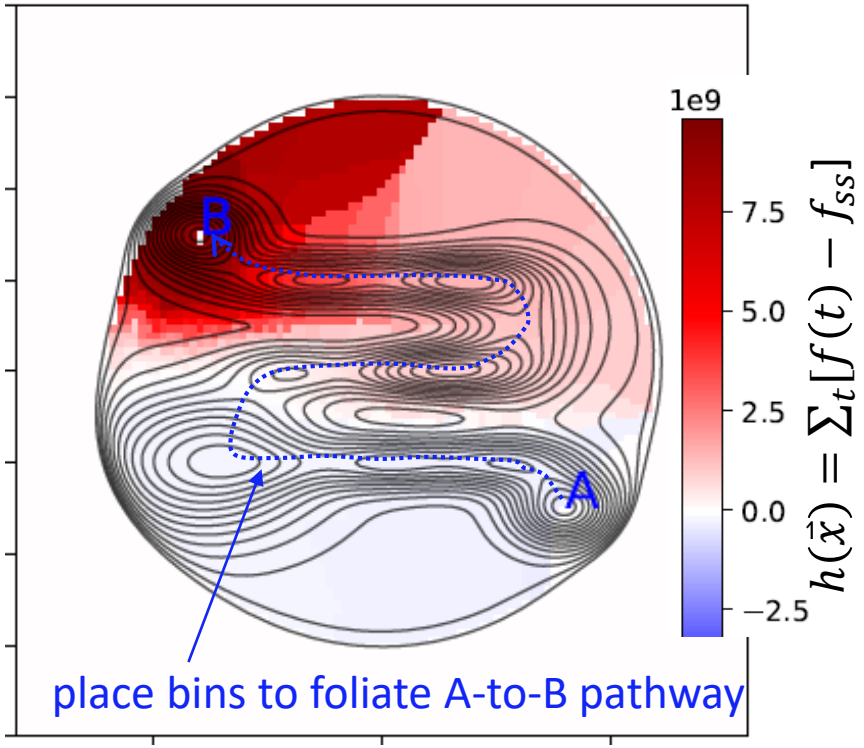


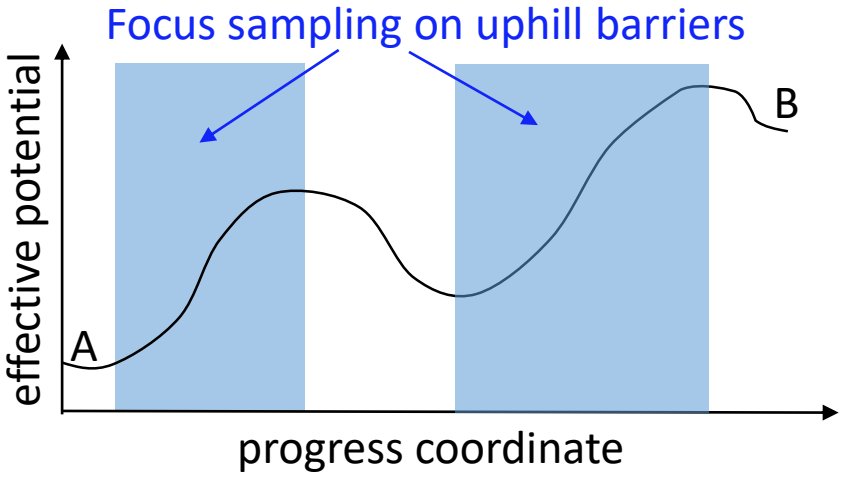
$K \equiv$ Koopman

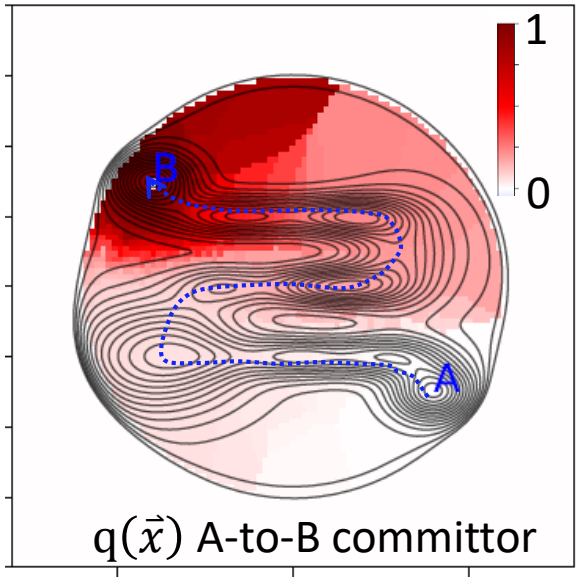
Aristoff and Zuckerman, SIAM (2020). Aristoff, Copperman, Simpson, Webber, and Zuckerman, man. in prep.

Controlling variance in A-to-B trajectory ensembles

- Optimal reaction coordinate $h(\vec{x})$ for controlling error is committor-like foliating progress from A-to-B
- Optimal allocation focuses sampling where it is most needed (where h variance $v^2 = Kh^2 - (Kh)^2$ due to sampling is high)
- in the low temperature limit this is the uphill side of barriers





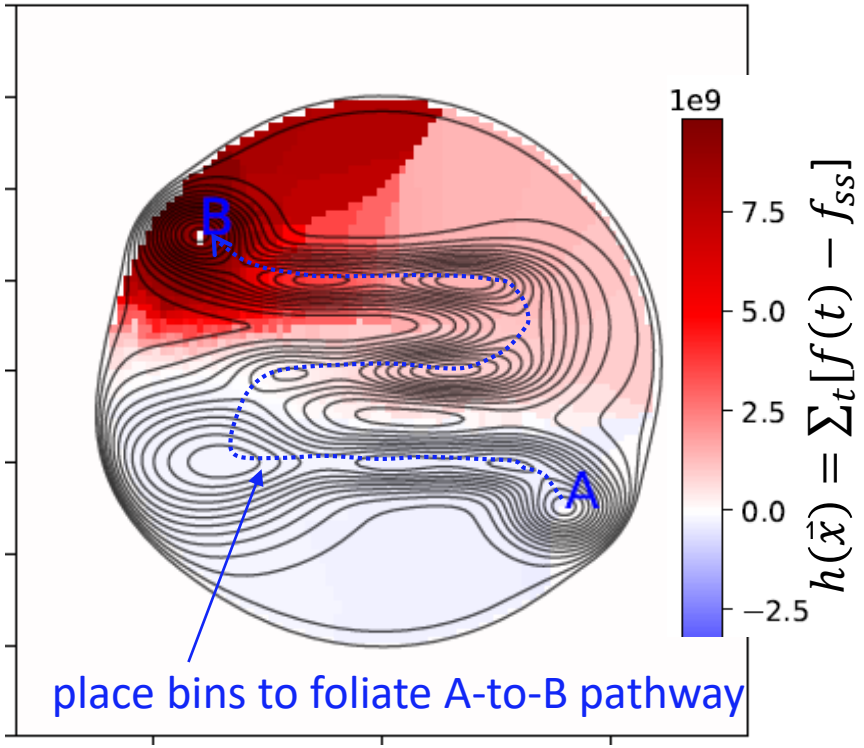


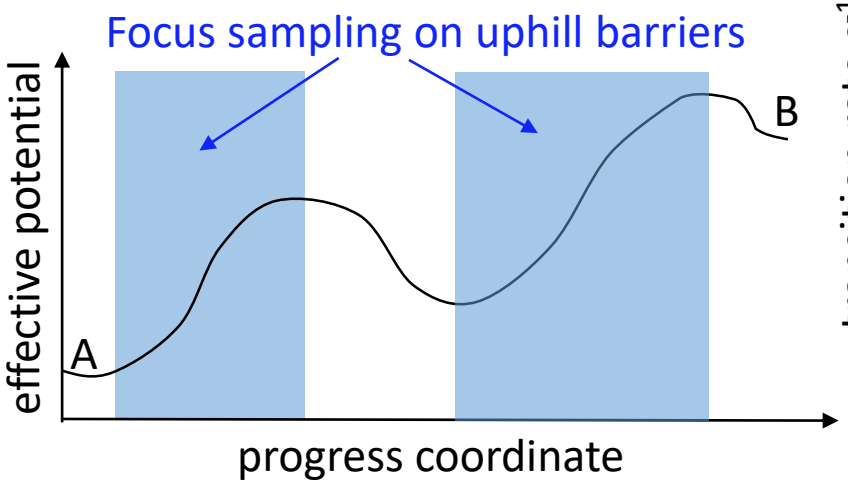
$K \equiv$ Koopman

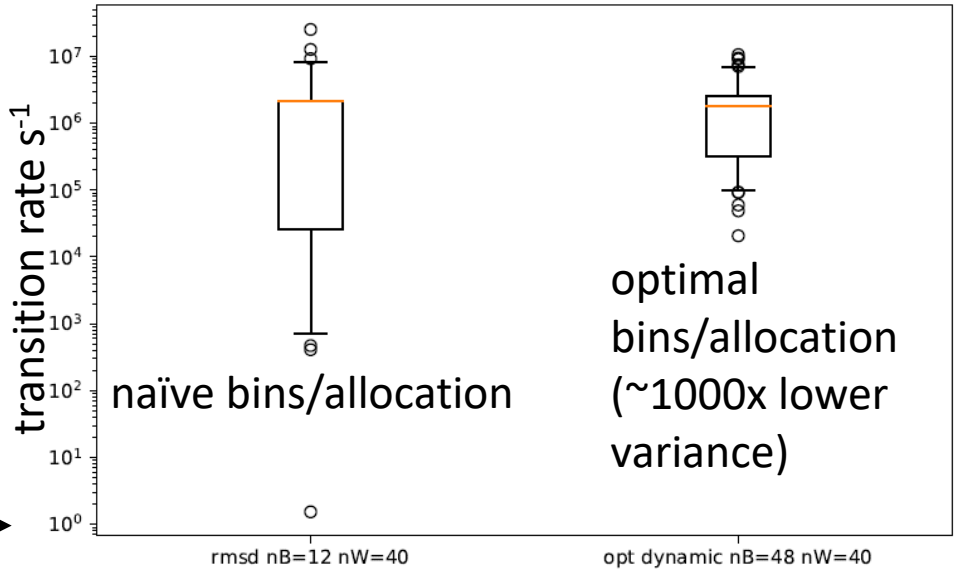
Aristoff and Zuckerman, SIAM (2020). Aristoff, Copperman, Simpson, Webber, and Zuckerman, man. in prep.

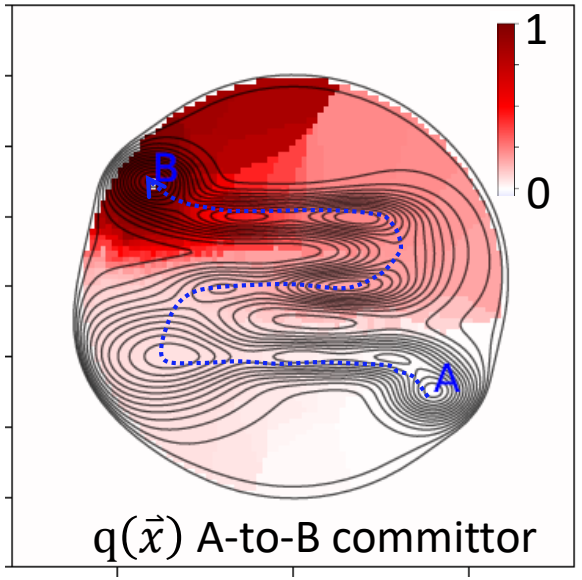
Controlling variance in A-to-B trajectory ensembles

- Optimal reaction coordinate $h(\vec{x})$ for controlling error is committor-like foliating progress from A-to-B
- Optimal allocation focuses sampling where it is most needed (where h variance $v^2 = Kh^2 - (Kh)^2$ due to sampling is high)
- in the low temperature limit this is the uphill side of barriers







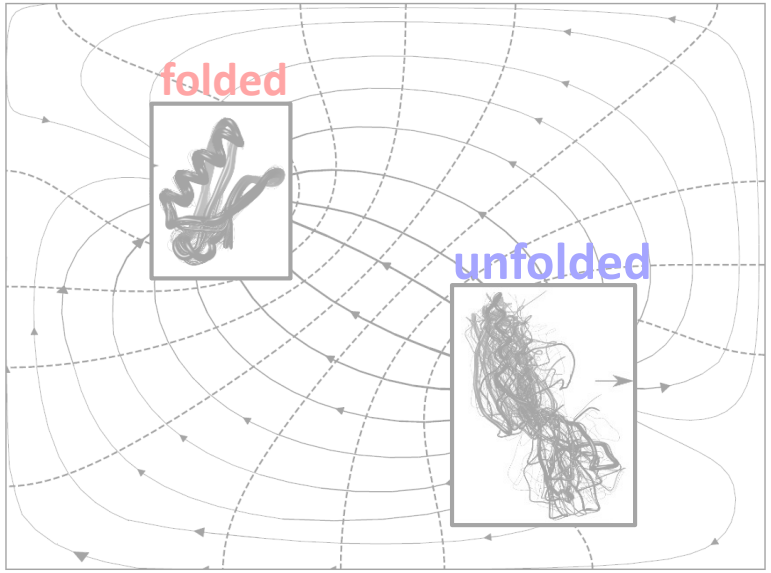


$K \equiv$ Koopman

Aristoff and Zuckerman, SIAM (2020). Aristoff, Copperman, Simpson, Webber, and Zuckerman, man. in prep.

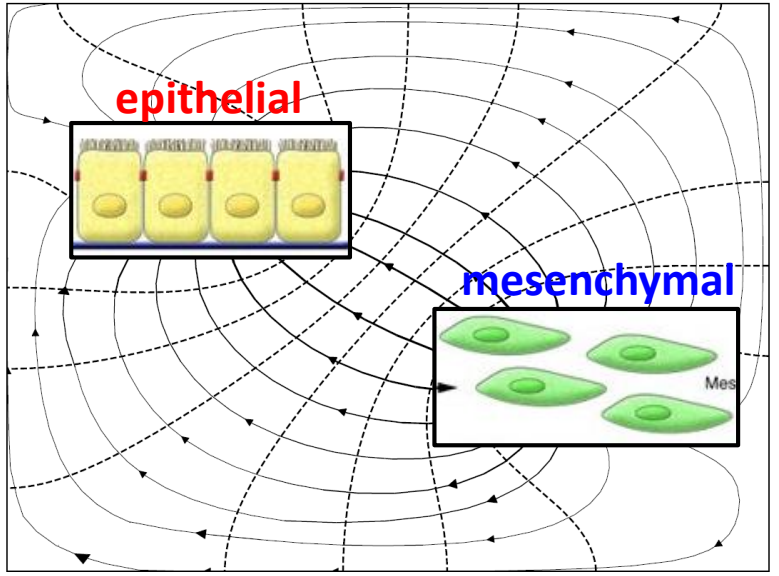
1. Strategies for rate estimation using weighted ensemble: history-augmented Markov State Models (haMSMs) and optimal binning

   with John Russo, David Aristoff, Gideon Simpson, and Daniel Zuckerman

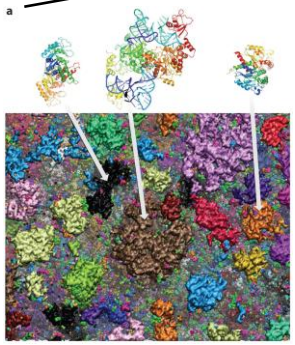


2. Do cells have transition states which can be leveraged to control cell state?

   with Young Hwan Chang, Laura Heiser, and Daniel Zuckerman

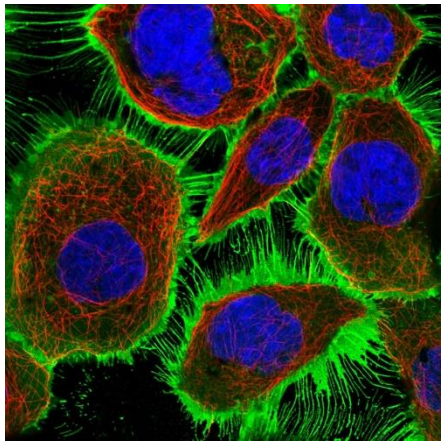


In the face of massive multiscale complexity...



Feig, Ann. Rev., (2019).

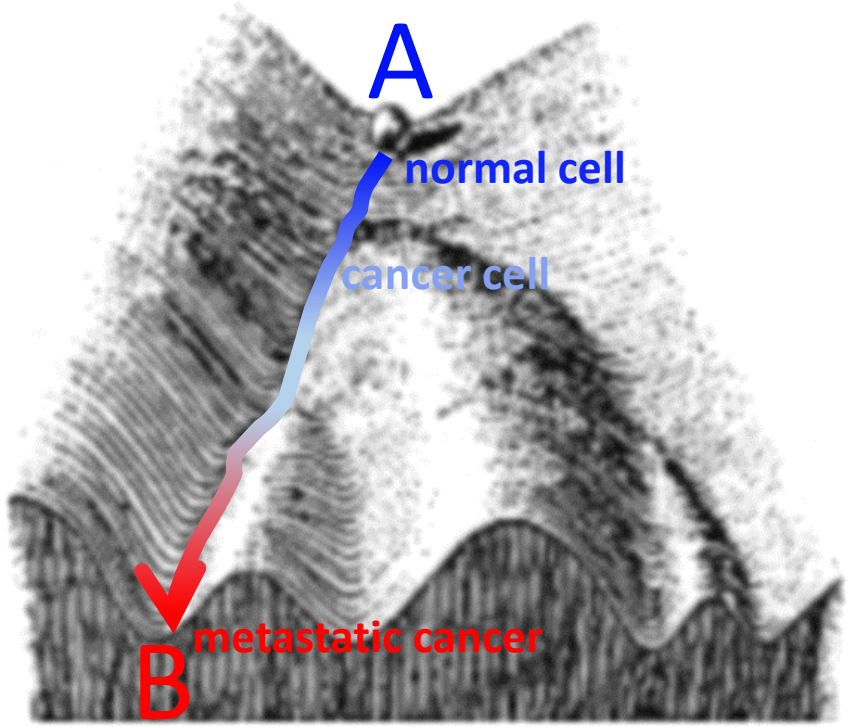
Atoms…



Chalmers University of Technology

are to cells…

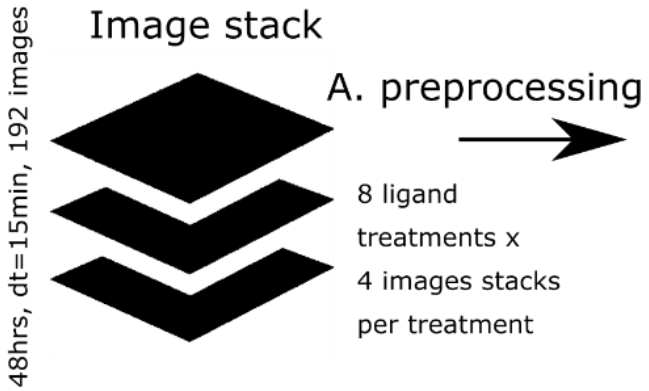
as cells are to humans

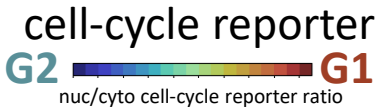


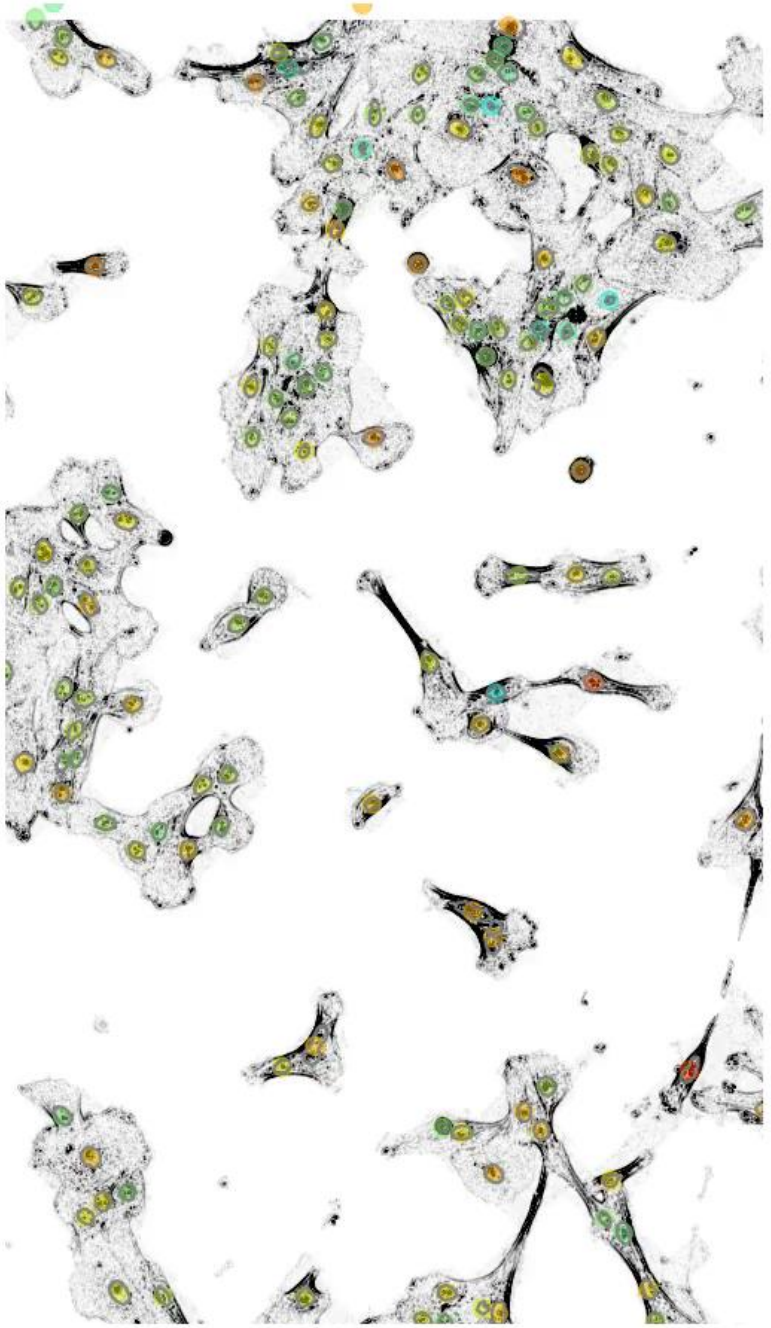
modified Waddington's landscape as an A-to-B ensemble

Can trajectory ensembles provide insight into the control of dysregulated disease cell states?

Live-cell imaging provides single-cell trajectories

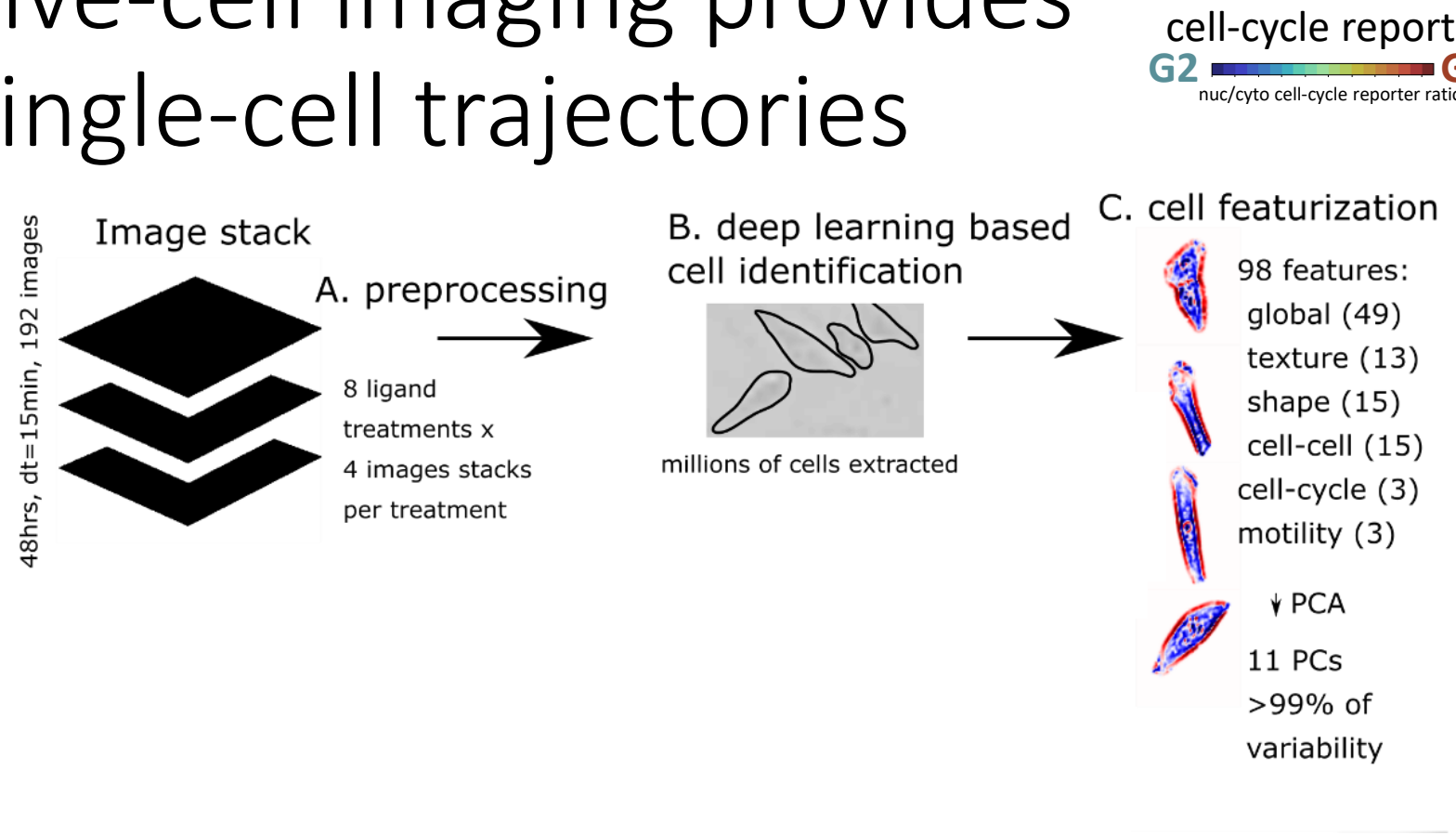




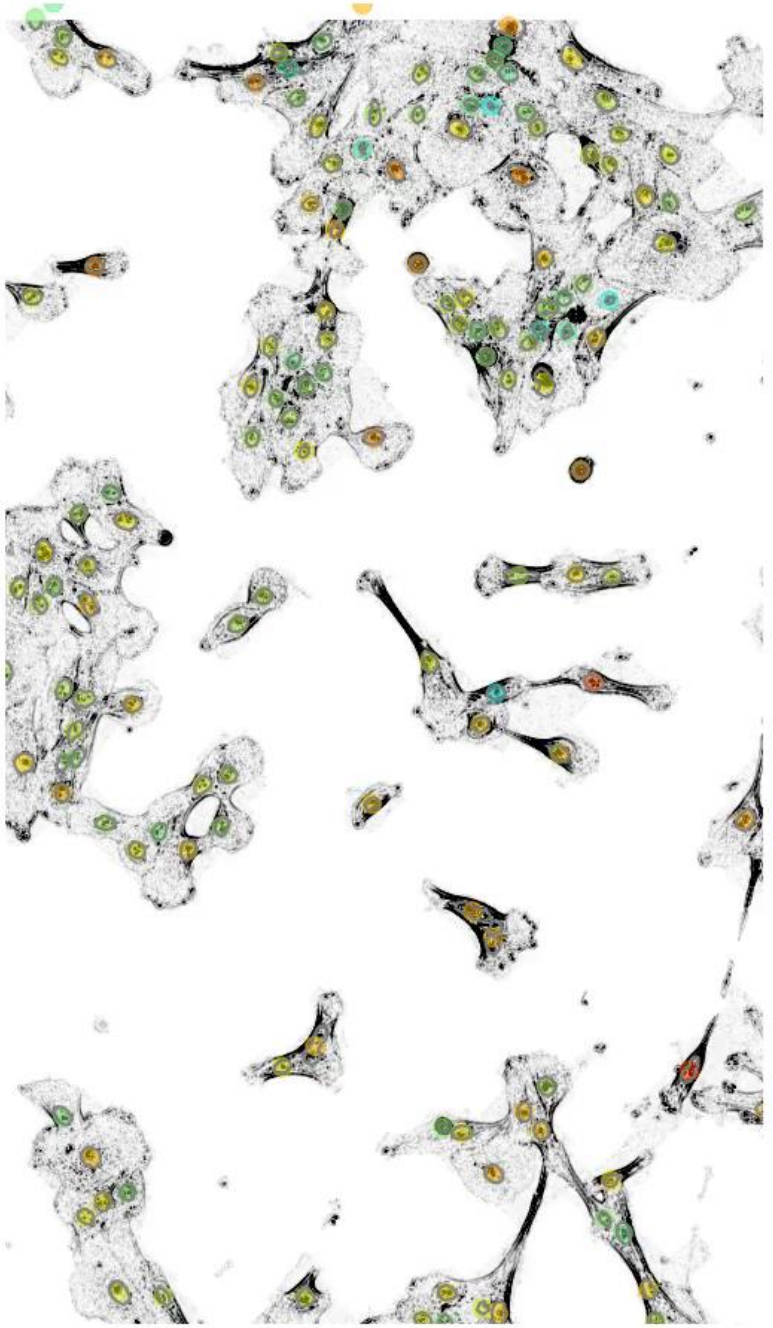


MCF10A cells in 2D culture, live-cell imaging with cell-cycle reporter, 15-minutes / frame, 48 hours

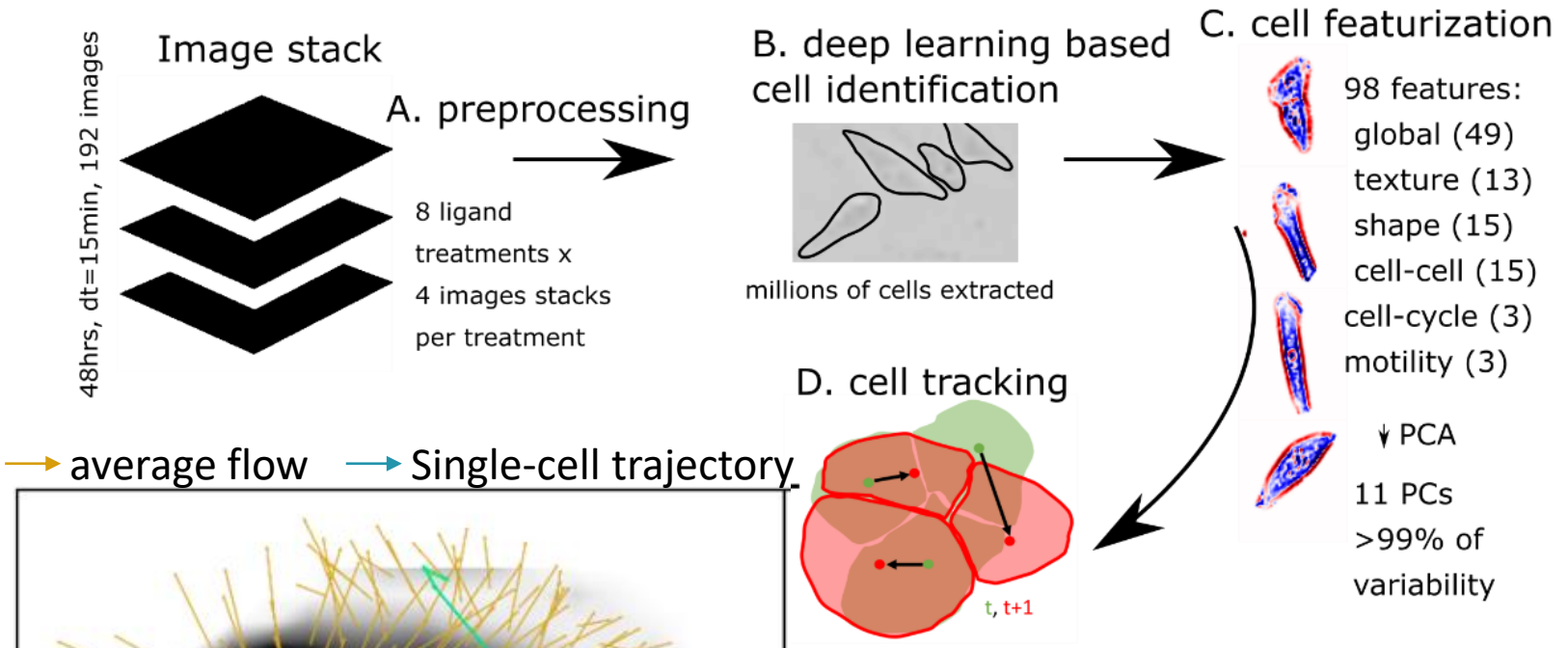
Live-cell imaging provides single-cell trajectories

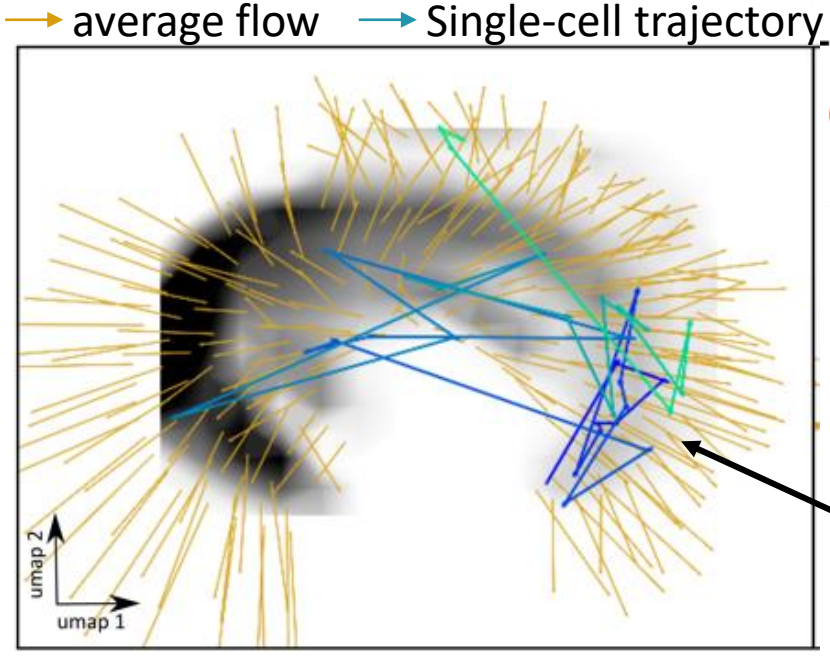


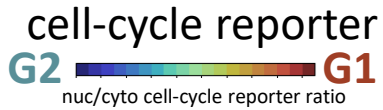
MCF10A cells in 2D culture, live-cell imaging with cell-cycle reporter, 15-minutes / frame, 48 hours

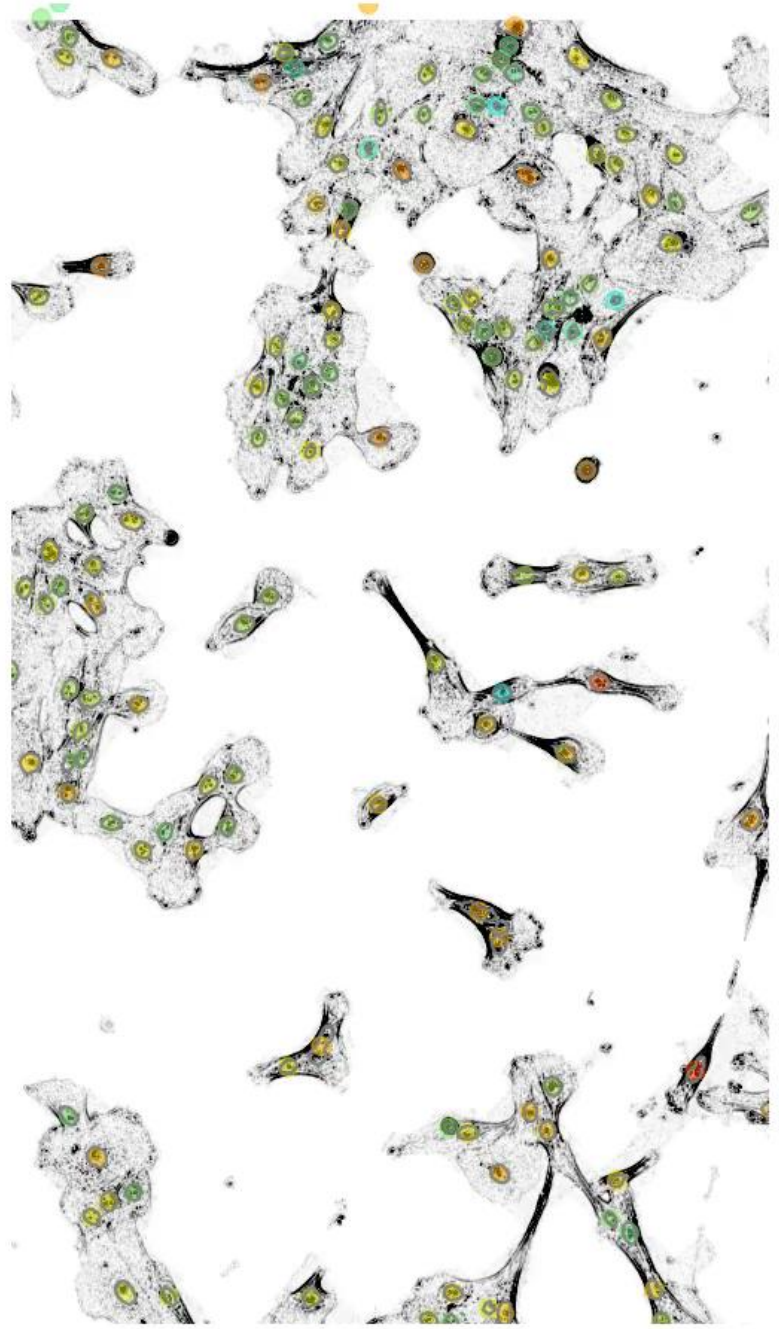


Live-cell imaging provides single-cell trajectories









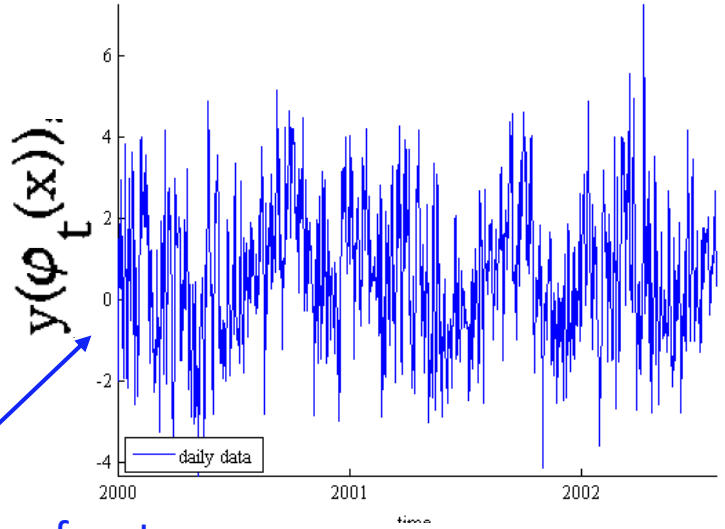
MCF10A cells in 2D culture, live-cell imaging with cell-cycle reporter, 15-minutes / frame, 48 hours

- morphological and motility feature trajectories appear highly stochastic

An observable is a smooth function $y:M \to \mathbb{R}$. The first problem is this : if, for some dynamical system with time evolution φ_t, we know the functions $t \mapsto y(\varphi_t(x))$, $x \in M$, then how can we obtain information about the original dynamical system (and manifold) from this. The next three theorems deal with this problem. (After the

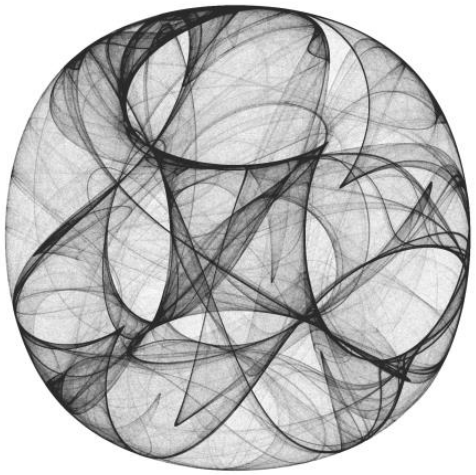
Theorem 1. Let M be a compact manifold of dimension m. For pairs (φ, y), $\varphi:M \to M$ a smooth diffeomorphism and $y:M \to \mathbb{R}$ a smooth function, it is a generic property that the map $\Phi_{(\varphi, y)}:M \to \mathbb{R}^{2m+1}$, defined by

$$\Phi_{(\varphi, y)}(x) = (y(x), y(\varphi(x)), \ldots, y(\varphi^{2m}(x)))$$



Incomplete observations of system

Trajectory embedding (delay embedding) from instantaneous snapshots to trajectory chunks of observable(s)



Takens: diffeomorphism between trajectory embedding of observable(s) and full dynamical manifold

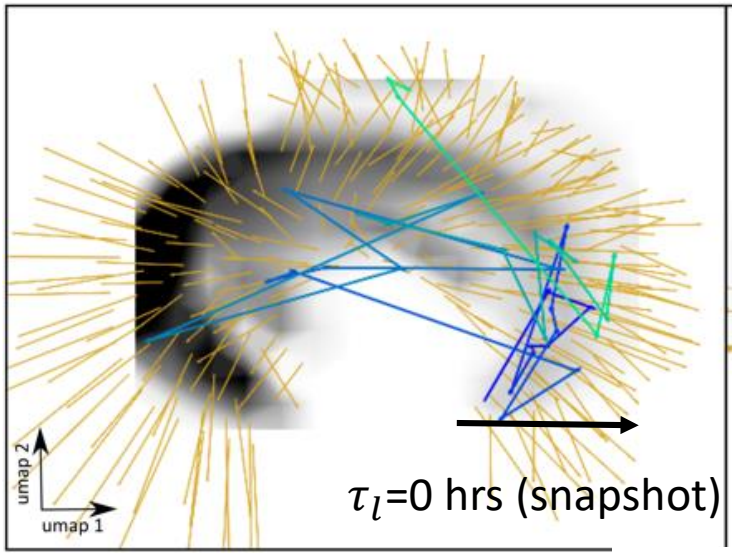
Corollary 5. Let M be a
consisting of a vector field
For generic such (X, y, p, α)
conditions depending on X a
the set of limit points of th

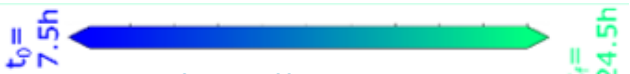
Detecting strange attractors in turbulence.

Floris Takens.

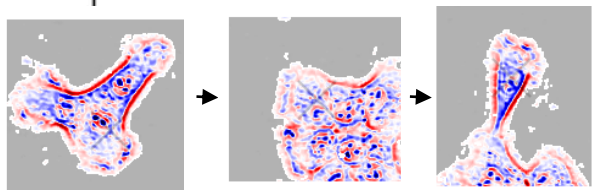
Single timepoint morphological features

stochastic trajectory in morphological snapshot space





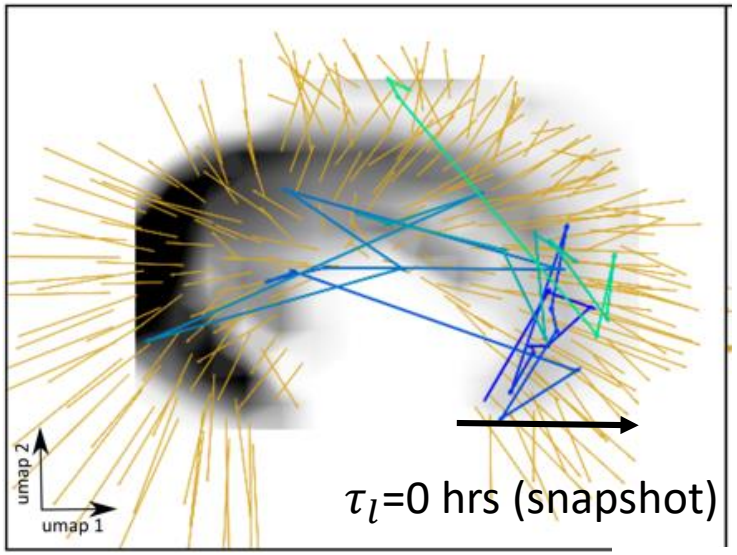
single-cell trajectory

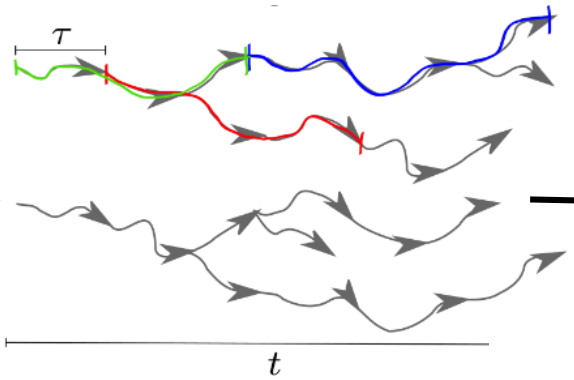


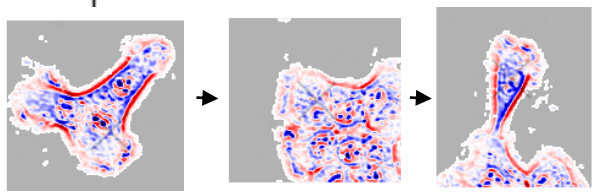
average flow Single-cell trajectory

Single timepoint morphological features

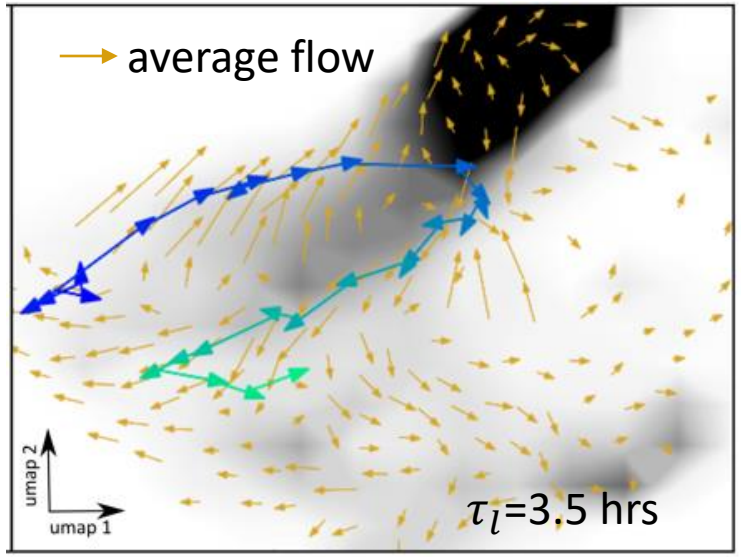
stochastic trajectory in morphological snapshot space







systematic trajectory in morphodynamical trajectory space

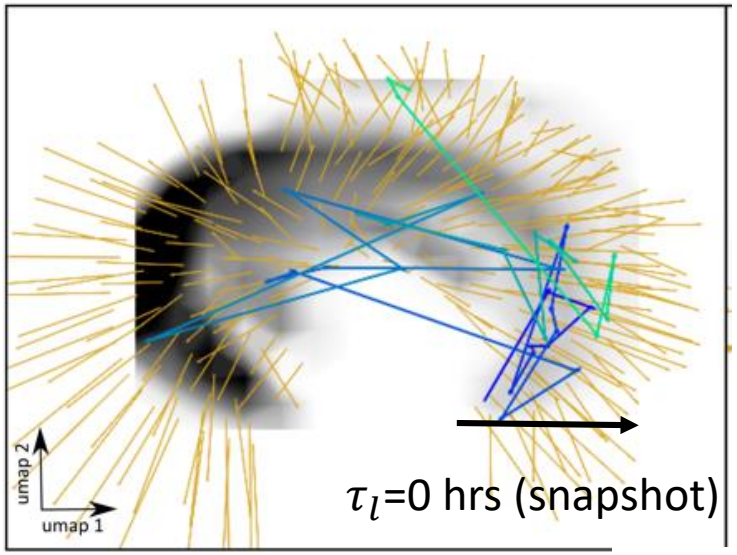


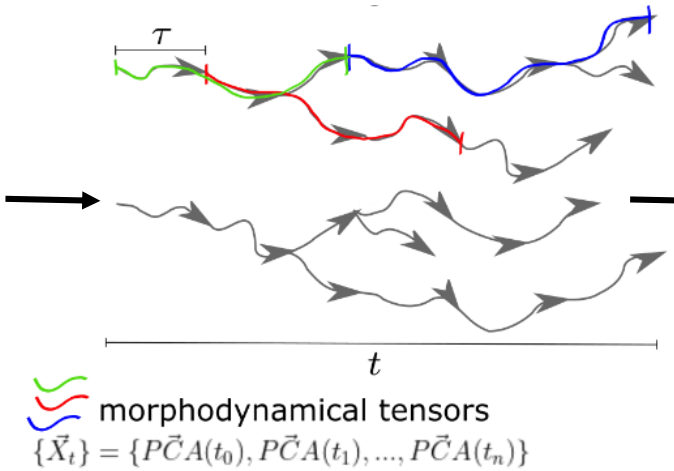
Morphodynamical Trajectory Embedding

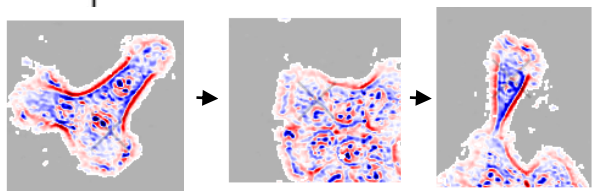
average flow Single-cell trajectory

Single timepoint morphological features

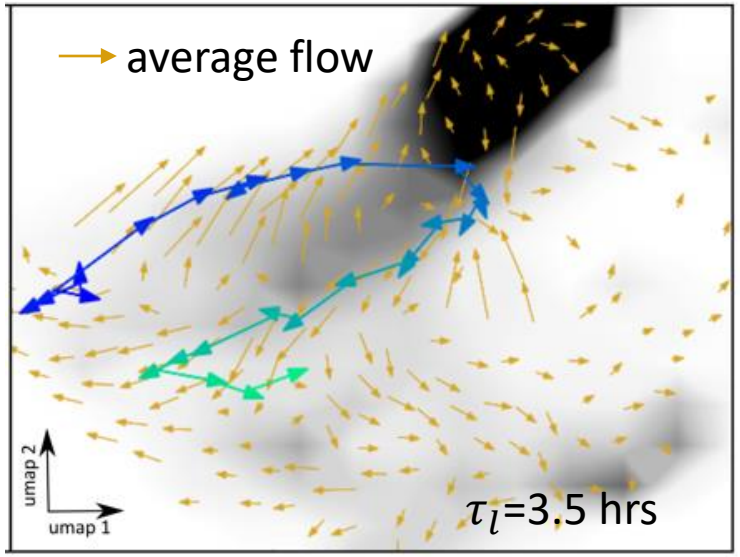
stochastic trajectory in morphological snapshot space





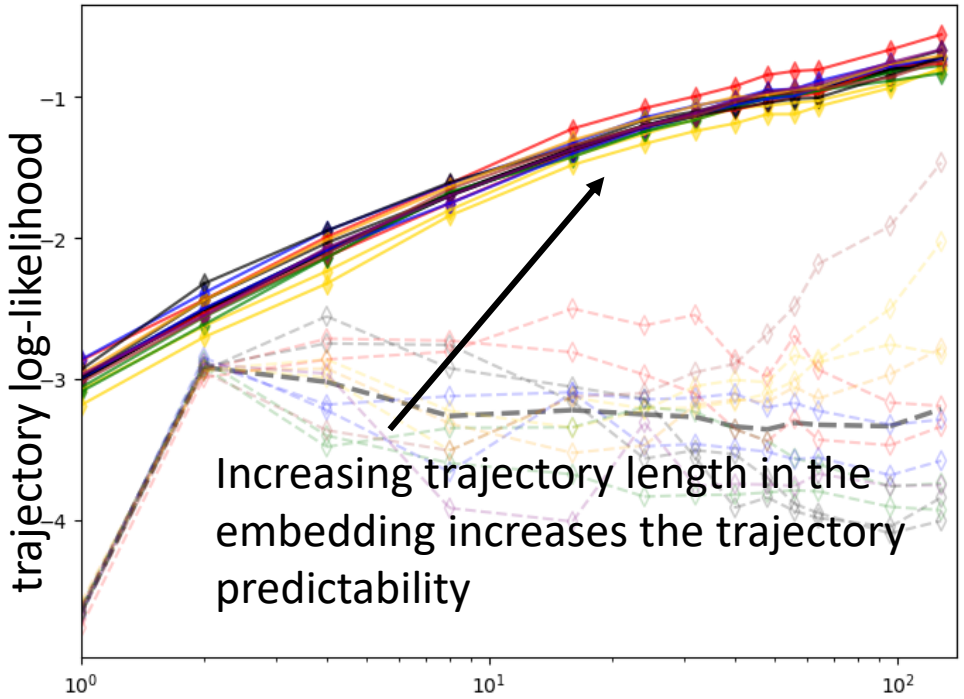


systematic trajectory in morphodynamical trajectory space



Morphodynamical Trajectory Embedding

Improved cell-state representation via trajectory embedding of single-timepoint cell features

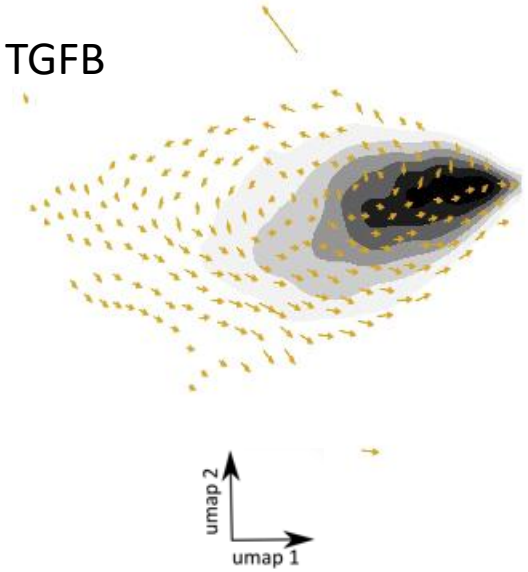


Copperman, Gross, Chang, Heiser, and Zuckerman. *bioRxiv* (2021).

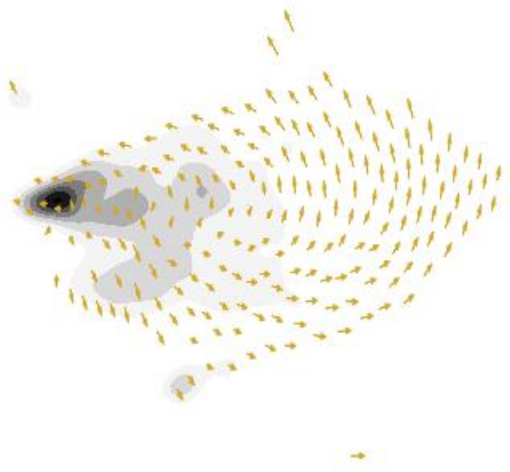
Coupled cell clustering and cell cycle dynamics

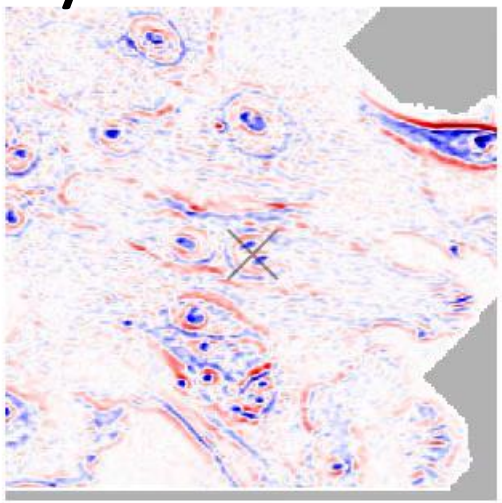
2D UMAP of trajectory embedding of multiple ligand conditions, τ_l=10 hrs

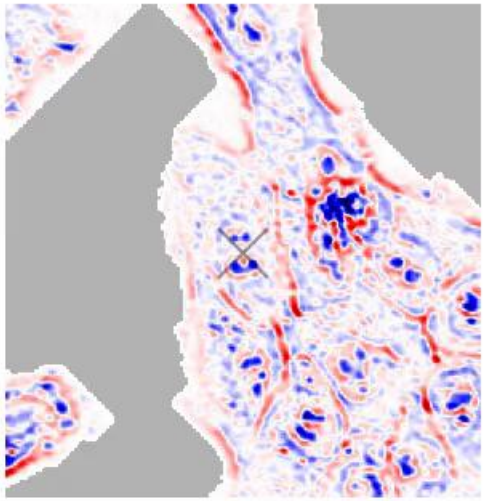
EGF + TGFB



HGF



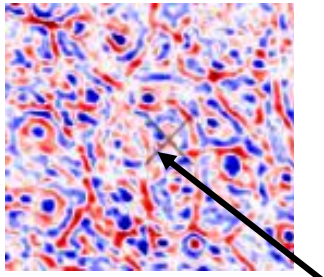


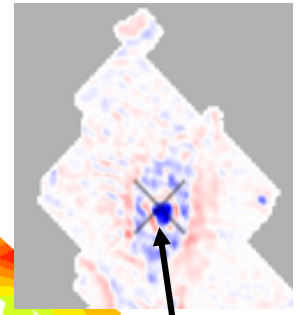


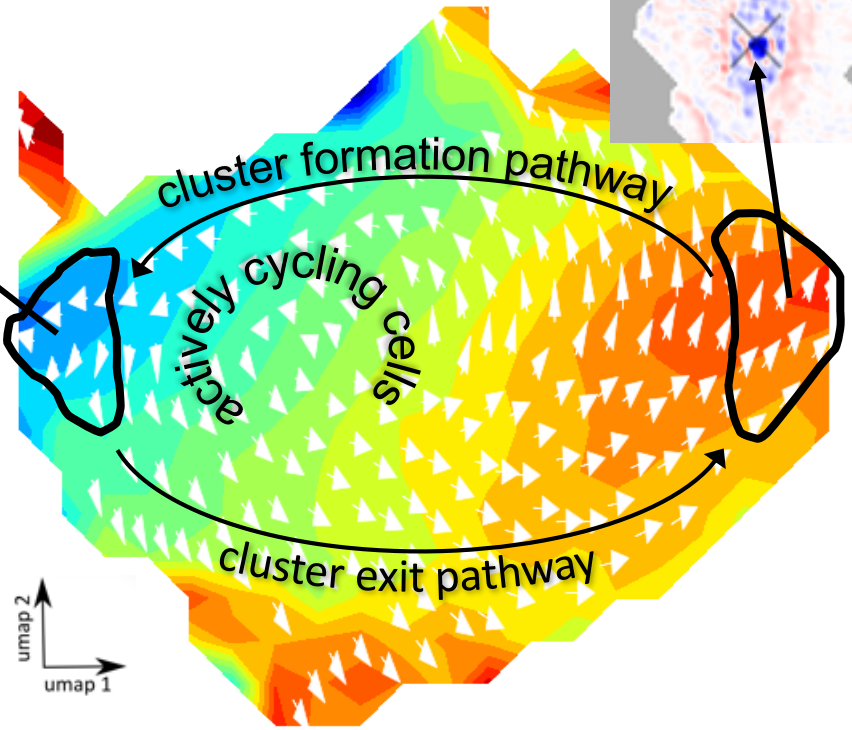
epithelial-like G2 – associated cell clusters

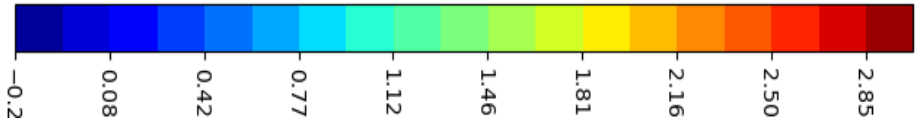
collective motility

G1-associated mesenchymal-like +lamellopodia +individual motility



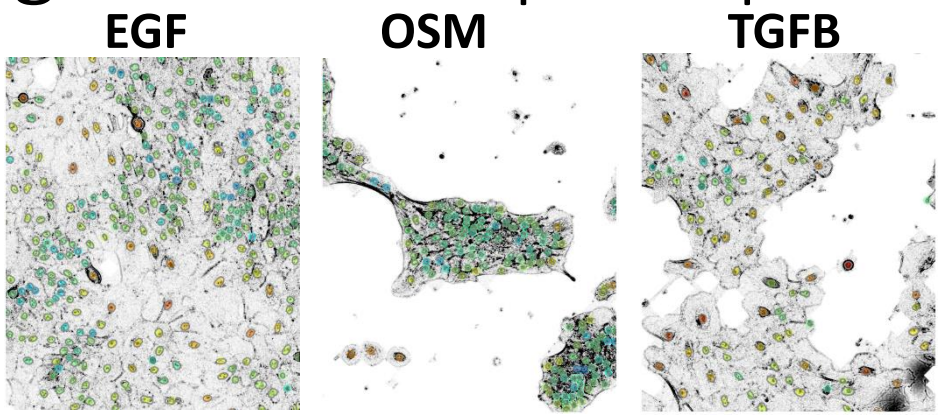






log2(nuc/cyto cell-cycle reporter ratio)

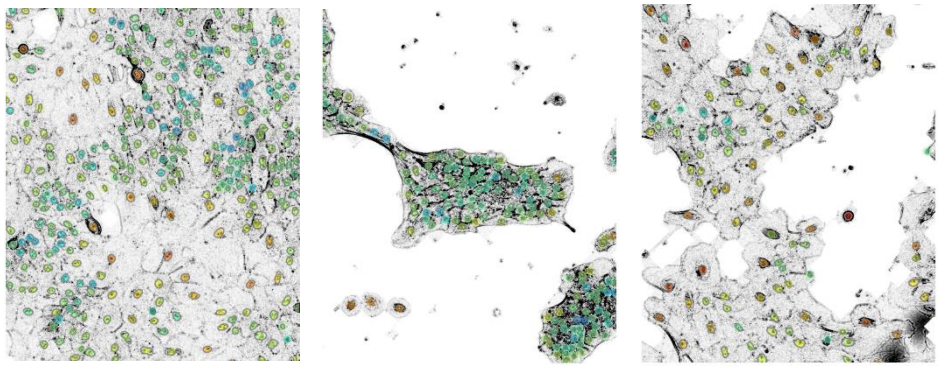
Using live-cell trajectories to define cell states and state-specific gene transcription profiles



Paired live-cell imaging and bulk RNA sequencing in 11 ligand conditions

Using live-cell trajectories to define cell states and state-specific gene transcription profiles

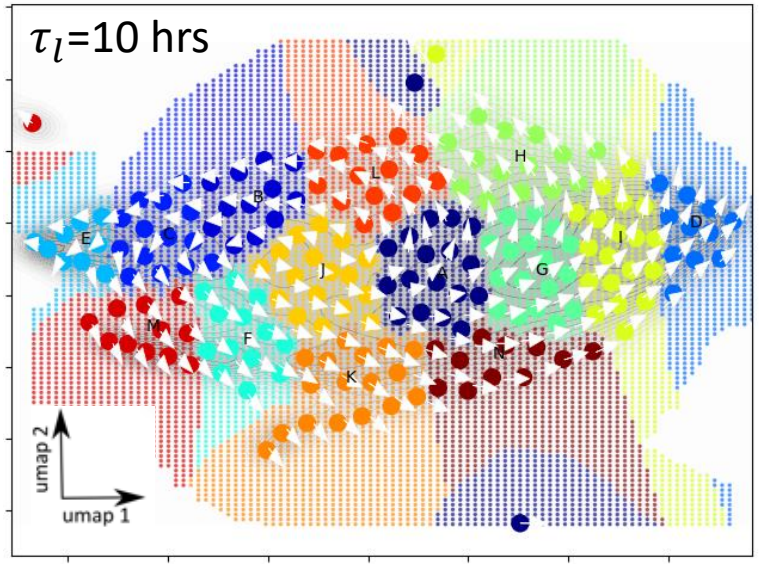
EGF **OSM** **TGFB**



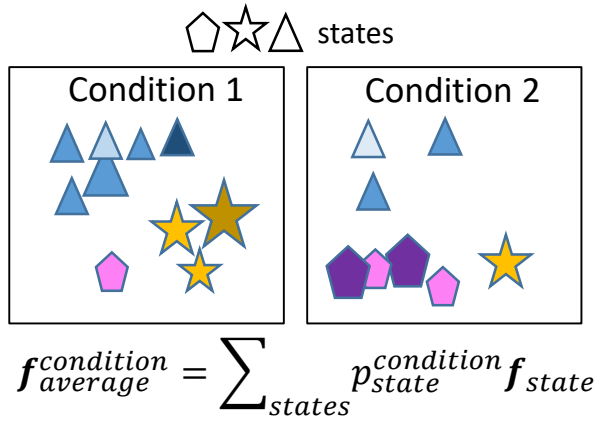
Paired live-cell imaging and bulk RNA sequencing in 11 ligand conditions

trajectory embedding

use single-cell trajectory ensemble to perform dynamical clustering



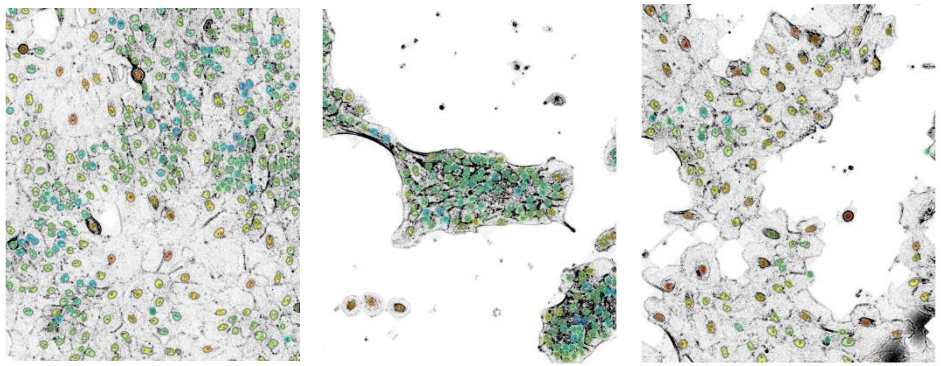
decompose bulk RNAseq into state-specific profiles using condition-specific cell-state populations



$$\boldsymbol{f}_{average}^{condition} = \sum_{states} p_{state}^{condition} \boldsymbol{f}_{state}$$

Using live-cell trajectories to define cell states and state-specific gene transcription profiles

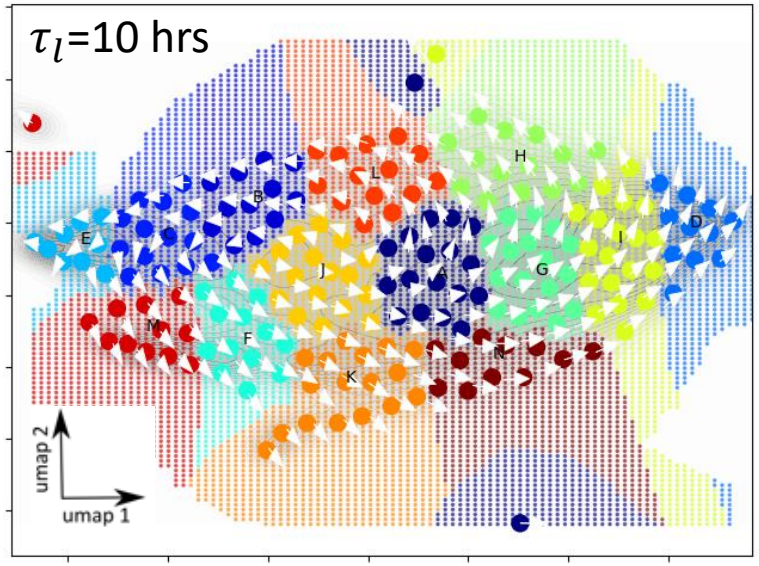
EGF **OSM** **TGFB**



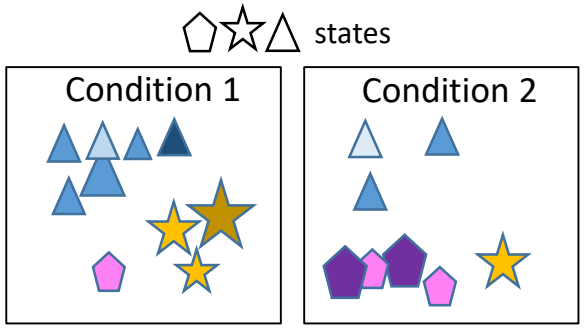
Paired live-cell imaging and bulk RNA sequencing in 11 ligand conditions

trajectory embedding

use single-cell trajectory ensemble to perform dynamical clustering

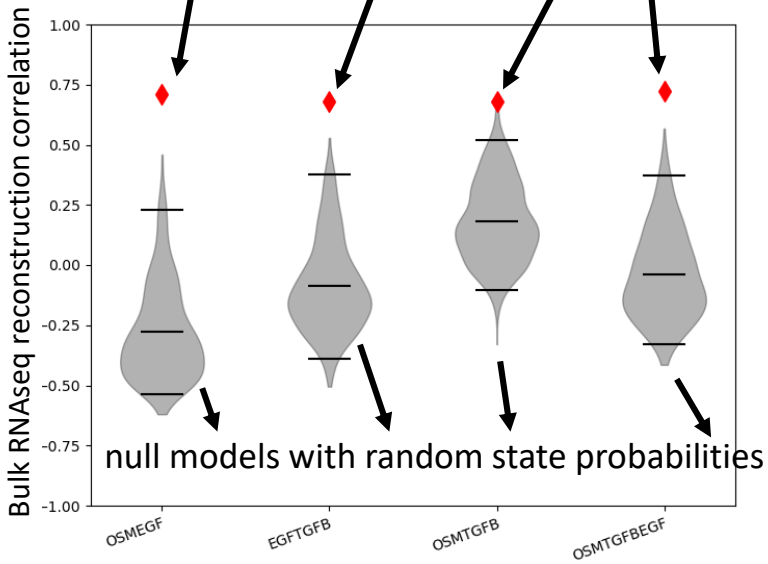


decompose bulk RNAseq into state-specific profiles using condition-specific cell-state populations

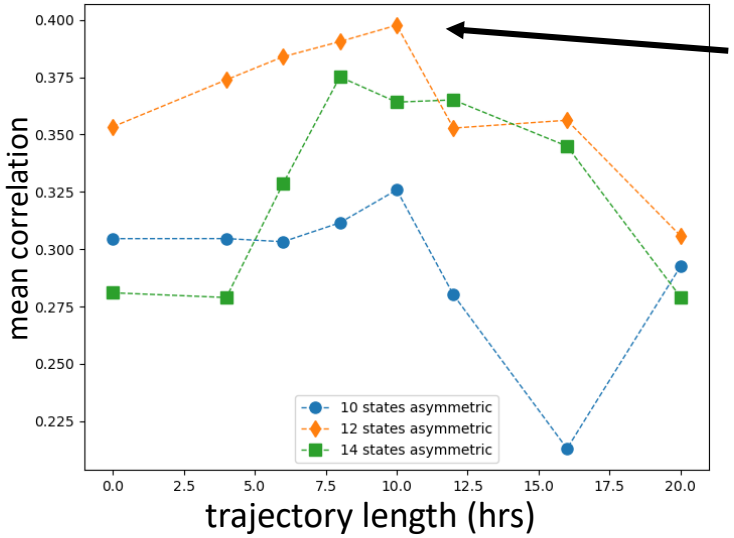


$$f_{average}^{condition} = \sum_{states} p_{state}^{condition} f_{state}$$

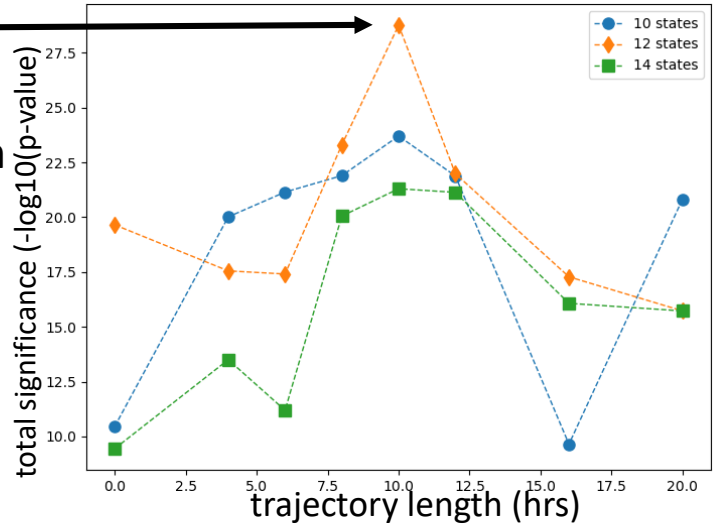
live-cell imaging morphodynamical state populations



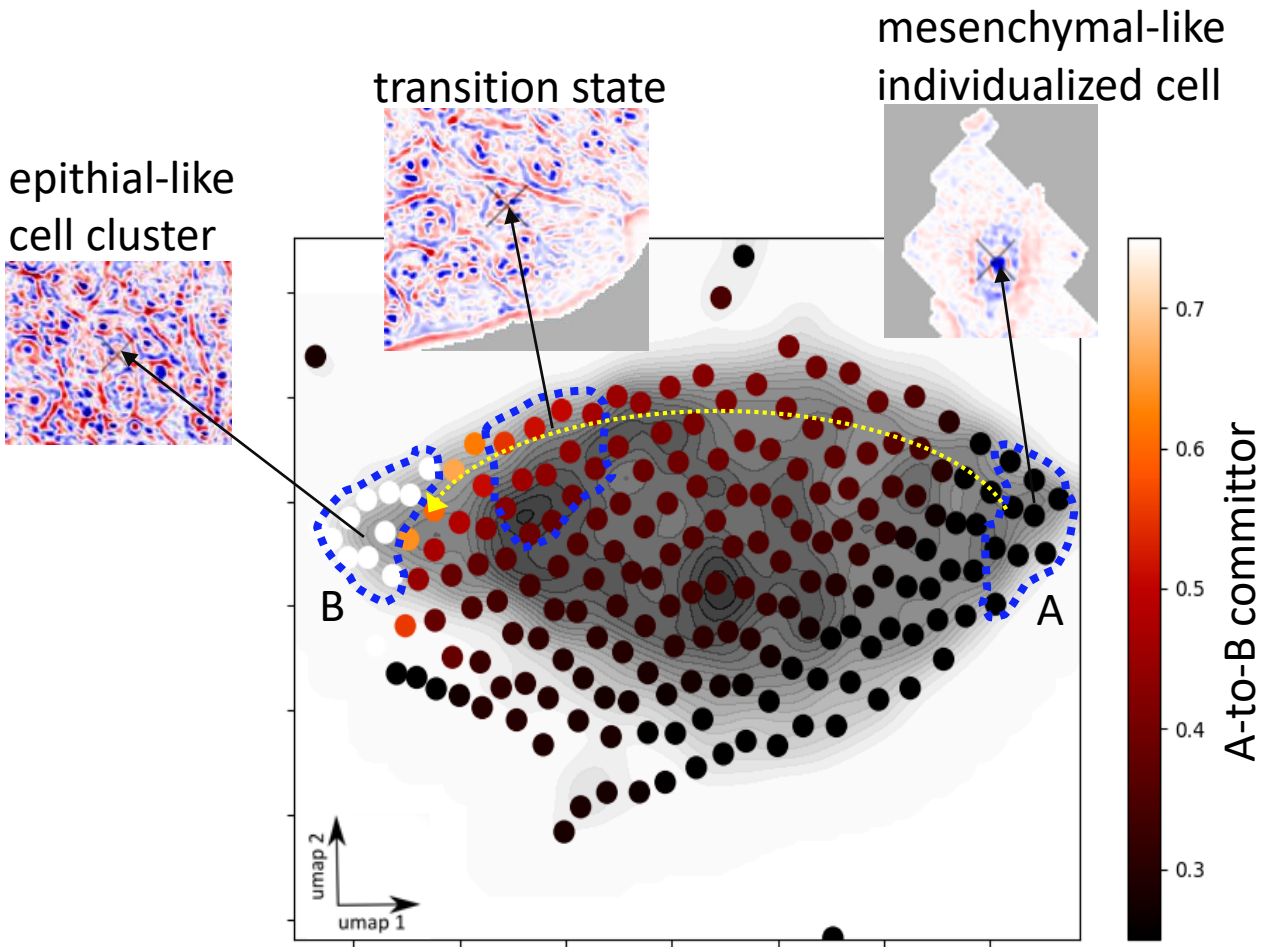
null models with random state probabilities



Improved cell state representation via trajectory embedding

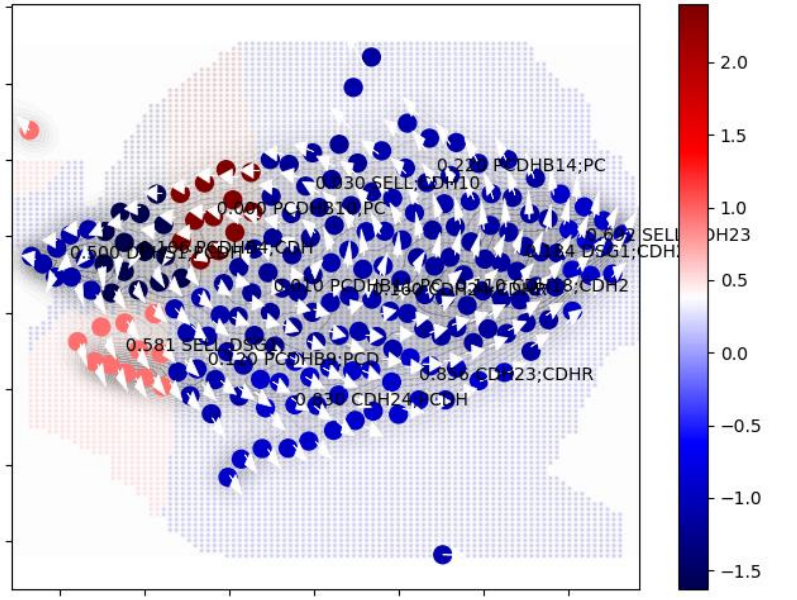


Cluster formation transition state

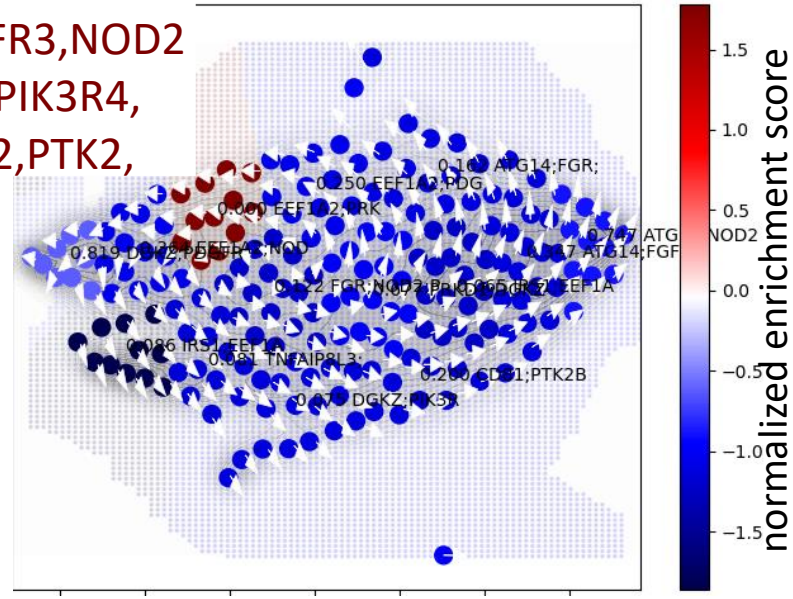


state-specific gene set enrichment analysis

calcium-dependent cell-cell adhesion: PCDHB10,11,14,13,9,16,CDH24; PCDHB3,5,4,2,6,DCHS1;DSG1

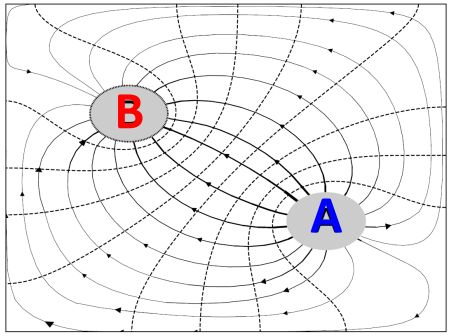


Positive regulation of lipid kinase activity: EEF1A2,PRKD1,FGFR3,NOD2,FGR,DGKZ,ATG14,PIK3R4,IRS1,AMBRA1,FGF2,PTK2,CD81,PDGFRB



Can cell transition states be directly targeted to control specific live-cell behaviors? WIP

A-to-B trajectory ensembles…



- … can be efficiently sampled using feedback
- … may have slow steady-state convergence but can be accelerated using haMSM reweighting
- … define optimal reaction coordinates and sampling allocation for minimizing variance in rate estimation
- … can define the mechanism and control of complex dynamical processes across scales
- … may provide insight into novel molecular targets and the specific control of observed live-cell behaviors

Acknowledgements

- Daniel Zuckerman
- Laura Heiser
- Young Hwan Chang
- Joe Gray
- Sean Gross
- Heiser Lab
  - Ian Mclean
  - Mark Dane
  - Nicholas Calistri
- Chang Lab
  - Luke Ternes
- Zuckerman Lab
  - John Russo
  - August George
  - Harry Ryu
  - Shelby Santos

OHSU School of Medicine Biomedical Engineering

Advanced Computing Center

NIH

KNIGHT CANCER Institute

DAMON RUNYON CANCER RESEARCH FOUNDATION