Learning Dynamic Boltzmann Distributions for Multiscale Modeling

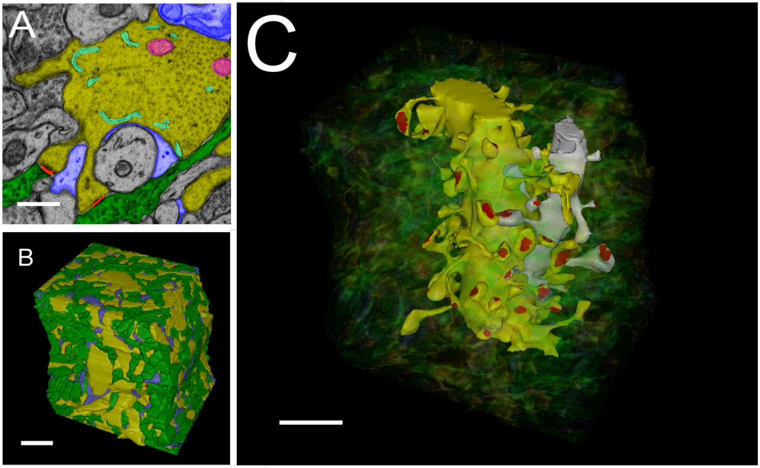
Oliver K. Ernst

CNL, Salk Institute
& UC San Diego Physics

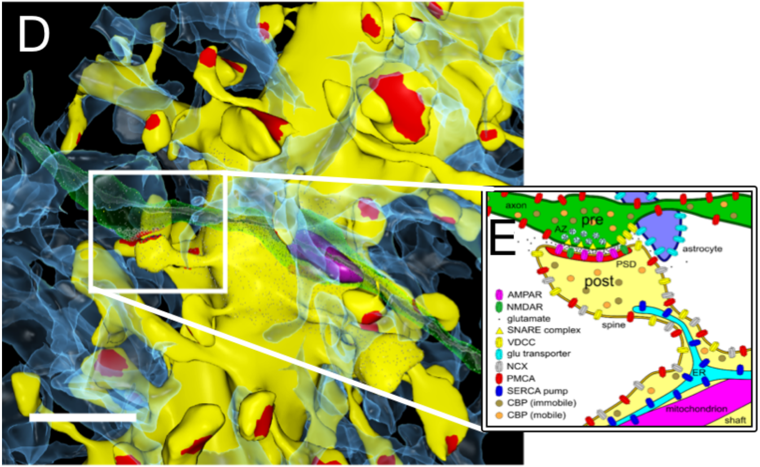
with Tom Bartol (Salk), Terrence Sejnowski (Salk), Eric Mjolsness (UCI)

July 18, 2018

- Abstracting multiscale/multiphysics biochemistry
- Learning problem for diff. eqs.
- Physics + ML

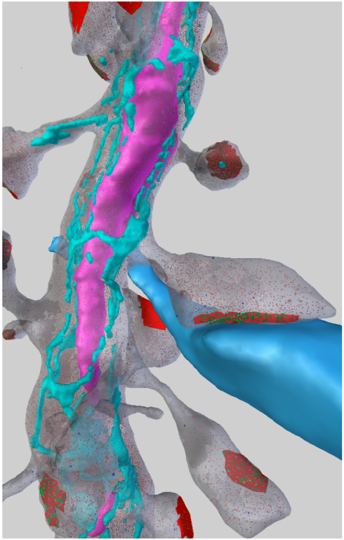


Bartol (2015) *Frontiers in Synaptic Neuroscience*

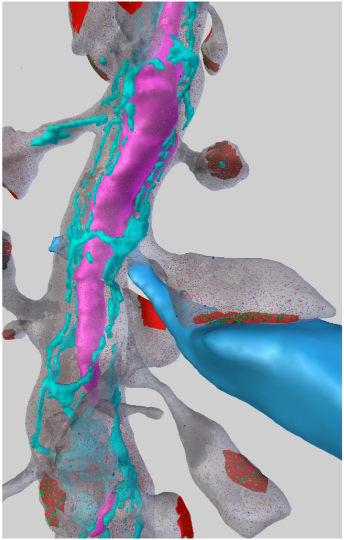


Bartol (2015) *Frontiers in Synaptic Neuroscience*

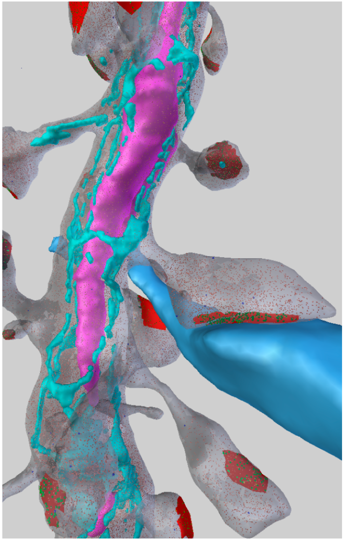
- How correlated is activity between ion channels on different synapses?
- On what timescales do small conductance channels become relevant?
- How is the cytoskeleton regulated (spine morphodynamics)?



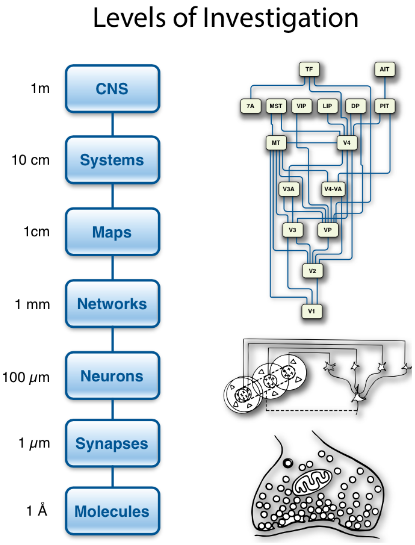
- How correlated is activity between ion channels on different synapses?
- On what timescales do small conductance channels become relevant?
- How is the cytoskeleton regulated (spine morphodynamics)?



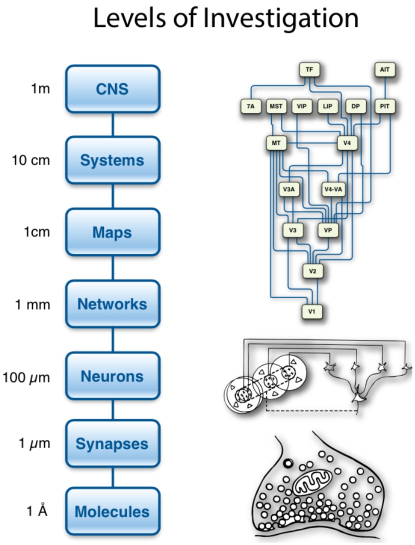
- How correlated is activity between ion channels on different synapses?
- On what timescales do small conductance channels become relevant?
- How is the cytoskeleton regulated (spine morphodynamics)?



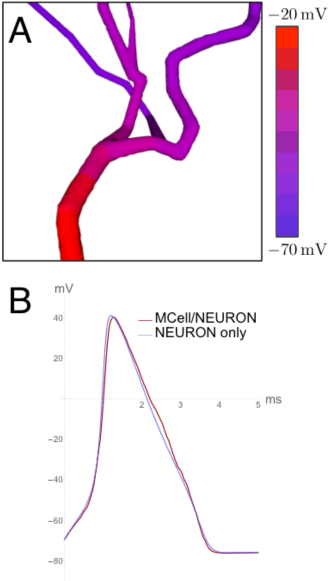
- (Up) How does local synaptic biochemistry regulate whole neuron/network level activity?
- (Down) How does calcium interact with the PSD?



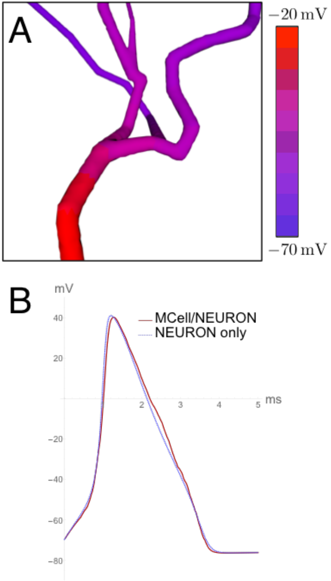
- (Up) How does local synaptic biochemistry regulate whole neuron/network level activity?
- (Down) How does calcium interact with the PSD?



- How does spiking activity affect synaptic biochemistry (e.g. electrodiffusion)?
- Can we integrate reaction-diffusion systems with MD models of ion channel activity?



- How does spiking activity affect synaptic biochemistry (e.g. electrodiffusion)?
- Can we integrate reaction-diffusion systems with MD models of ion channel activity?



- How correlated is activity between ion channels on different synapses?
- On what timescales do small conductance channels become relevant?
- How is the cytoskeleton regulated (spine morphodynamics)?
- (Up) How does local synaptic biochemistry regulate whole neuron/network level activity?
- (Down) How does calcium interact with the PSD?
- How does spiking activity affect synaptic biochemistry (e.g. electrodiffusion)?
- Can we integrate reaction-diffusion systems with MD models of ion channel activity?

Relative Gradient

Low ⟷ High

Particle Number (Low ↕ Infinite)

Low	Gridless SSA (Stochastic Sim. Algorithm)	Particle-Based
	Gridless SSA	Gridded SSA/ Plenum
	Stochastic ODEs	Stochastic PDEs
Infinite	ODEs (Mass action)	PDEs (Finite elements)

1. Motivate yet another learning algorithm (YALA?)
2. Lattice systems
3. Future work

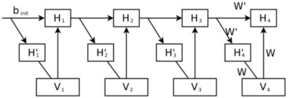
Boltzmann Learning

- Simplest form: Learn a stationary dist.

System is dynamic

ML timeseries

- Learn a conditional dist.
- Recurrent networks



Sutskever (2009) NIPS

- Hidden Markov models
- ...

$\sim$ Discrete time
Discrete space
Blackbox

Machinery available

1. Chemical master eqn.
2. Field theory (Doi-Peliti)

Physics-informed ML?

Well-mixed case

$$\text{true state} = \left| n, t \right\rangle$$
$$\text{reduced state} = \left| \{\nu_k\}_{k=1}^{K}, t \right\rangle$$

k-particle interaction funcs. $\nu_k(t)$

Dynamic Boltzmann distributions

$$\tilde{p}(n,t) = \frac{1}{\mathcal{Z}[\{\nu\}]} \exp\left[-\sum_{k=1}^{K} \binom{n}{k} \nu_k(t) \right] \tag{1}$$

Well-mixed case

Dynamical model

$$\frac{d}{dt}\nu_k(t) = F_k(\{\nu(t)\}) \tag{2}$$

- **Choose F to minimize**:

$$S = \int_0^\infty dt \, \mathcal{D}_{\mathcal{KL}}(p||\tilde{p}) \tag{3}$$

- **Learning rule:** Solve:

$$0 = \frac{\delta S}{\delta F_k(\{\nu\})} =$$

$$\sum_{k'=1}^{K} \int_0^\infty dt' \left(\left\langle \binom{n}{k'} \right\rangle_{p(t')} - \left\langle \binom{n}{k'} \right\rangle_{\tilde{p}(t')} \right) \frac{\delta \nu_{k'}(t')}{\delta F_k(\{\nu\})}$$

PDE-constrained optimization

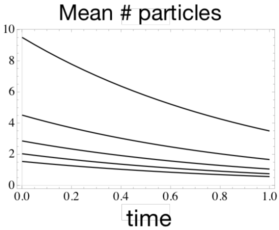
Well-mixed case

Algorithm 1 PDE-constrained optimization

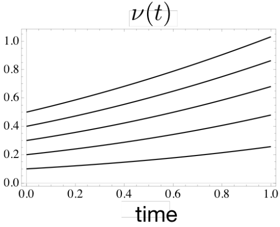
$F_k = 0$
while not converged **do**
  ▷ *Generate trajectory in reduced space $\{\nu\}$:*
  Solve the PDE constraint $d\nu_k/dt = F_k(\{\nu\})$.
  Solve for variational terms $\delta\nu_{k'}(t)/\delta F_k(\{\nu\})$.
  for all times t **do**
    ▷ *Awake phase:*
    $\left\langle \binom{n}{k'} \right\rangle_{p(t)}$ by stoch. sim.
    ▷ *Asleep phase:*
    $\left\langle \binom{n}{k'} \right\rangle_{\tilde{p}(t)}$ by sampling.
  ▷ *Update F_k to decrease objective function:*
  $F_k \to F_k - \lambda \Delta F_k$.

Well-mixed case: Exponential Decay Example

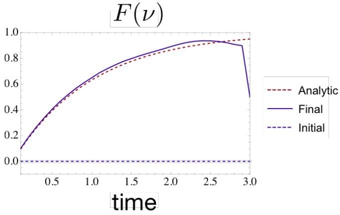
Training Data



Reduced model: $\tilde{p}(n,t) = \frac{1}{\mathcal{Z}} \exp[-n\nu(t)]$

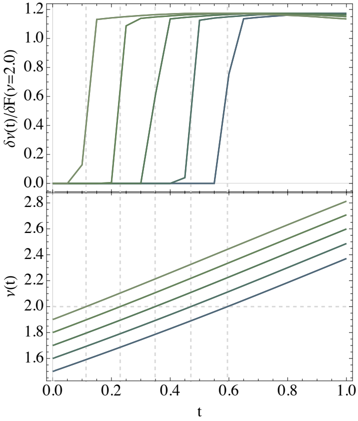


Function learned: $\dot{\nu}(t) = F(\nu(t))$



Well-mixed case

$$\frac{d}{dt'}\left(\frac{\delta\nu_{k'}(t')}{\delta F_k[\{\nu\}]}\right) = \sum_{l=1}^{K} \frac{\partial F_{k'}(\{\nu(t')\})}{\partial \nu_l(t')} \frac{\delta \nu_l(t')}{\delta F_k(\{\nu\})} + \delta_{k',k}\delta(\{\nu\} - \{\nu(t')\}) \quad (4)$$



ANN learning rule deriv:

$$\frac{d}{dt}\frac{\partial f}{\partial y_j} = -\sum_{i=1}^{n+1} \frac{\partial f}{\partial y_i}\frac{\partial g_i}{\partial y_j}$$

Dreyfus, Stuart. The numerical solution of variational problems. *Journal of Mathematical Analysis and Applications* **5**.1 (1962): 30-45.

Spatial systems

$$\text{true state } = \left| n, \boldsymbol{\alpha}, \boldsymbol{x}, t \right\rangle$$
$$\text{reduced state } = \left| \{\nu_k\}_{k=1}^K, t \right\rangle$$

k-particle interaction funcs. $\nu_k(\boldsymbol{\alpha}_{\langle i \rangle_k^n}, \boldsymbol{x}_{\langle i \rangle_k^n}, t)$

Spatial dynamic Boltzmann distributions

$$\tilde{p}(n, \boldsymbol{\alpha}, \boldsymbol{x}, t) = \frac{1}{\mathcal{Z}[\{\nu\}]} \exp\left[-\sum_{k=1}^K \sum_{\langle i \rangle_k^n} \nu_k(\boldsymbol{\alpha}_{\langle i \rangle_k^n}, \boldsymbol{x}_{\langle i \rangle_k^n}, t) \right] \quad (5)$$

Spatial systems

Dynamical model

$$\frac{d}{dt}\nu_k(\boldsymbol{\alpha}_{\langle i \rangle_k^n}, \boldsymbol{x}_{\langle i \rangle_k^n}, t) = \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}] \quad (6)$$

$\mathcal{F}$ is a **functional** $\rightarrow$ PDEs

- **Learning rule:** . . . need to parameterize $\mathcal{F}$!

Proposition

Fix: reaction network, $\{\nu_k\}_{k=1}^K$
Then: linearity of CME in reaction operators $\dot{p} = \sum_r \boldsymbol{W}^{(r)} p$ *extends to the functionals* $\mathcal{F}_k = \sum_r \mathcal{F}_k^{(r)}$

Proposition

Fix: reaction network, $\{\nu_k\}_{k=1}^K$
Then: linearity of CME in reaction operators $\dot{p} = \sum_r \boldsymbol{W}^{(r)} p$ *extends to the functionals* $\mathcal{F}_k = \sum_r \mathcal{F}_k^{(r)}$

Diffusion from point source

$$p(\boldsymbol{x}, t) = \frac{\exp\left[- \sum_{i=1}^n \frac{(x_i - x_0)^2}{4Dt} \right]}{(4\pi Dt)^{n/2}} \sim \frac{\exp[- \sum_{i=1}^n \nu_1(x_i, t)]}{\mathcal{Z}} \tag{7}$$

$$F[\nu_1(y,t)] = D\partial_y^2 \nu_1(y,t) - D\left(\partial_y \nu_1(y,t)\right)^2$$

Proposition

Fix: reaction network, $\{\nu_k\}_{k=1}^K$
Then: linearity of CME in reaction operators $\dot{p} = \sum_r \boldsymbol{W}^{(r)} p$ *extends to the functionals* $\mathcal{F}_k = \sum_r \mathcal{F}_k^{(r)}$

Diffusion from point source

$$p(\boldsymbol{x}, t) = \frac{\exp\left[-\sum_{i=1}^n \frac{(x_i - x_0)^2}{4Dt}\right]}{(4\pi D t)^{n/2}} \sim \frac{\exp[-\sum_{i=1}^n \nu_1(x_i, t)]}{\mathcal{Z}}$$
$$F[\nu_1(y,t)] = D\partial_y^2 \nu_1(y,t) - D\left(\partial_y \nu_1(y,t)\right)^2 \tag{7}$$

Unimolecular rxns

$$p(n, \boldsymbol{x}) = p(n)p(\boldsymbol{x}) = p(n)p(x_1)\dots p(x_n)$$

multiply Boltz. $\rightarrow$ additive energy terms

$$\sum_i \nu_{\text{diff}}(x_i, t) + \sum_{k=1}^K \binom{n}{k} \nu_{k,\text{unimolecular}}(t) \tag{8}$$

Spatial systems

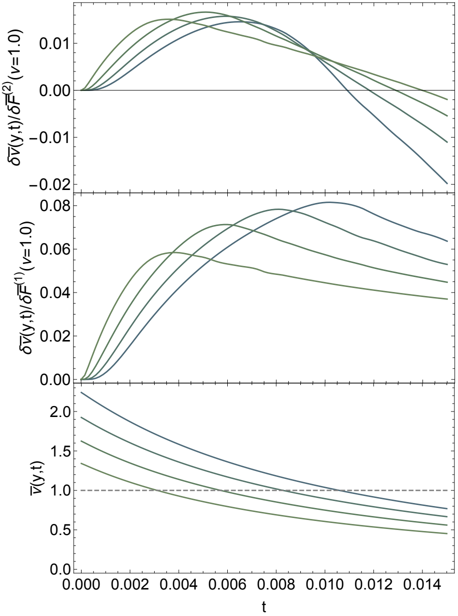
- **Parameterized** dynamical model:

$$\frac{d}{dt}\nu_k(\boldsymbol{\beta},\mathbf{y},t) = F_k^{(0)}(\{\nu(\boldsymbol{\beta},\mathbf{y},t)\}) + \sum_{\lambda=1}^{k}\Bigg(F_k^{(1,\lambda)}(\{\nu(\boldsymbol{\beta},\mathbf{y},t)\})\sum_{\langle i\rangle_\lambda^k}\sum_{m=1}^{\lambda}\Big(\partial_m\nu_\lambda(\boldsymbol{\beta}_{\langle i\rangle_\lambda^k},\mathbf{y}_{\langle i\rangle_\lambda^k},t)\Big)^2 + F_k^{(2,\lambda)}(\{\nu(\boldsymbol{\beta},\mathbf{y},t)\})\sum_{\langle i\rangle_\lambda^k}\sum_{m=1}^{\lambda}\partial_m^2\nu_\lambda(\boldsymbol{\beta}_{\langle i\rangle_\lambda^k},\mathbf{y}_{\langle i\rangle_\lambda^k},t)\Bigg), \tag{9}$$

- **Learning rule:** Solve:

$$0 = \frac{\delta S[\{\nu[\{F\}]\}]}{\delta F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\mathbf{y})\})} = \sum_{k'=1}^{K}\sum_{\boldsymbol{\beta}'}\int d\mathbf{y}'\int dt'\ \left(\mu_{k'}(\boldsymbol{\beta}',\mathbf{y}',t') - \tilde{\mu}_{k'}(\boldsymbol{\beta}',\mathbf{y}',t')\right)\frac{\delta\nu_{k'}(\boldsymbol{\beta}',\mathbf{y}',t')}{\delta F_k^{(\gamma)}(\{\nu(\boldsymbol{\beta},\mathbf{y})\})} \tag{10}$$

Spatial systems



Spatial systems

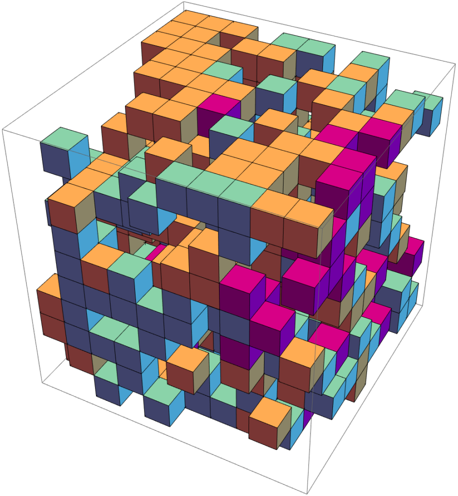
$$\tilde{p}(n, \boldsymbol{\alpha}, \boldsymbol{x}, t) = \frac{1}{\mathcal{Z}[\{\nu\}]} \exp\left[-\sum_{k=1}^{K} \sum_{\langle i \rangle_k^n} \nu_k(\boldsymbol{\alpha}_{\langle i \rangle_k^n}, \boldsymbol{x}_{\langle i \rangle_k^n}, t) \right]$$

$$\frac{d}{dt} \nu_k(\boldsymbol{\alpha}_{\langle i \rangle_k^n}, \boldsymbol{x}_{\langle i \rangle_k^n}, t) = \mathcal{F}_k[\{\nu(\boldsymbol{\alpha}, \boldsymbol{x}, t)\}]$$

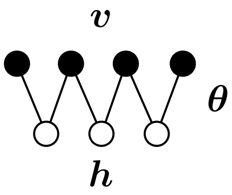
Hierarchy for learning problems

- $\mathcal{F}$ is a function**al**
- function**al** $\mathcal{F}$ $\rightarrow$ ordinary functions F
  - Variational problem $\delta S / \delta F = 0$
- functions F $\rightarrow$ parameters θ
  - Minimization problem $\partial S / \partial \theta = 0$

Lattice systems



Lattice systems



$$\tilde{p}(\boldsymbol{v}, \boldsymbol{h} | \boldsymbol{\theta}(t)) = \frac{\exp[-E(\boldsymbol{v}, \boldsymbol{h}, \boldsymbol{\theta}(t))]}{Z(\boldsymbol{\theta}(t))}$$

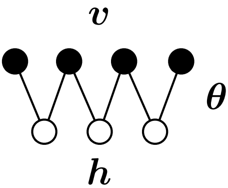
Dynamical model

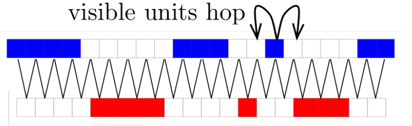
$$\frac{d}{dt}\theta_k(t) = F_k(\boldsymbol{\theta}(t)) \tag{11}$$

- **Learning rule:** Solve:

$$\frac{\delta S}{\delta F_k(\boldsymbol{\theta})} = \sum_{k'=1}^{K} \int_{t_0}^{\infty} dt' \ \left(\mu_{k'}(t') - \tilde{\mu}_{k'}(t')\right) \frac{\delta \theta_{k'}(t')}{\delta F_k(\boldsymbol{\theta})} = 0 \tag{12}$$

Moment closure





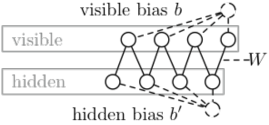
- Example: $A + A \to \varnothing$

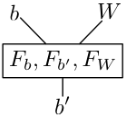
$$\frac{d}{dt}\left\langle \sum_i v_i \right\rangle = -\, 2k_r \left\langle \sum_i v_i v_{i+1} \right\rangle ,$$

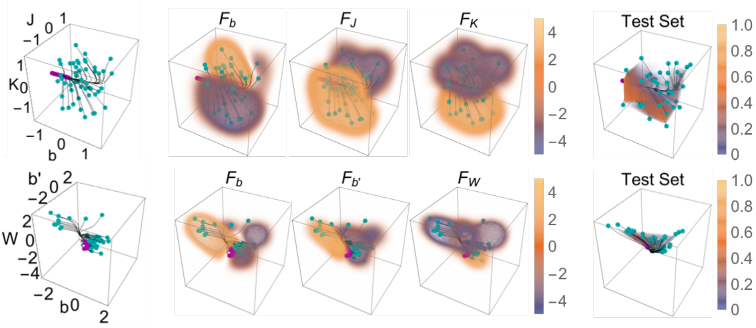
$$\frac{d}{dt}\left\langle \sum_i v_i v_{i+1} \right\rangle = 2D \left\langle \sum_i v_i v_{i+2} \right\rangle - 2k_r \left\langle \sum_i v_i v_{i+1} v_{i+2} \right\rangle + (k_r - 2D) \left\langle \sum_i v_i v_{i+1} \right\rangle , \tag{12}$$

- Hidden layers:
  1. Trained to capture relevant moments
  2. Separate indistinguishable states
  3. Combinatorial explosion
- Infer graph structure from CME

Moment closure

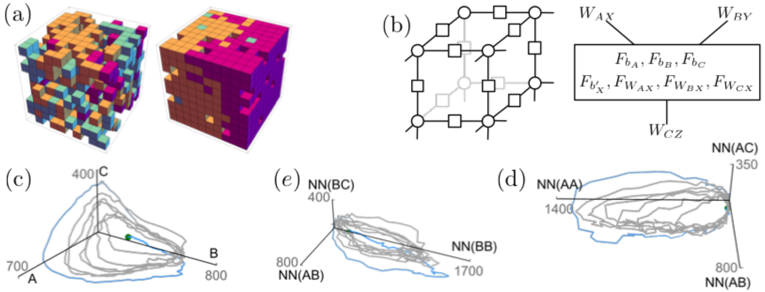




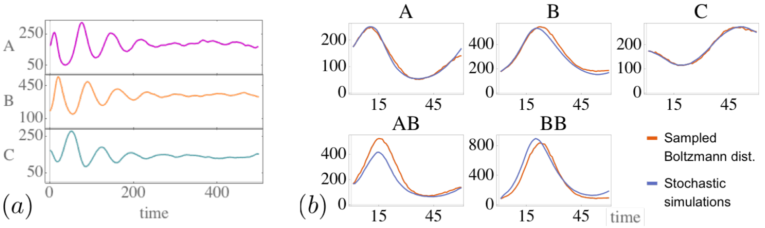


Rössler oscillator

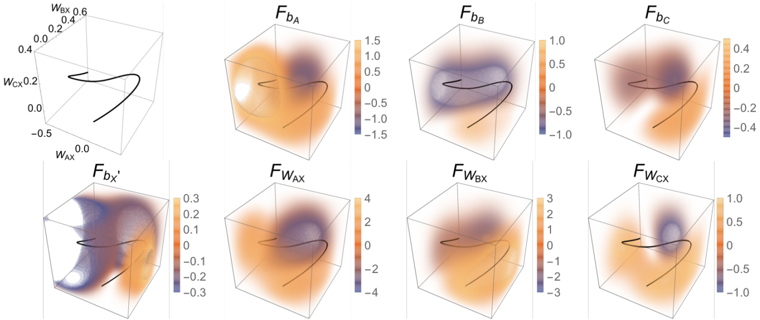
$$A \underset{p_1}{\overset{k_1}{\rightleftharpoons}} 2A \qquad A + B \xrightarrow{p_2} 2B \qquad A + C \xrightarrow{p_3} \varnothing \qquad B \xrightarrow{k_2} \varnothing \qquad C \underset{p_4}{\overset{k_3}{\rightleftharpoons}} 2C$$



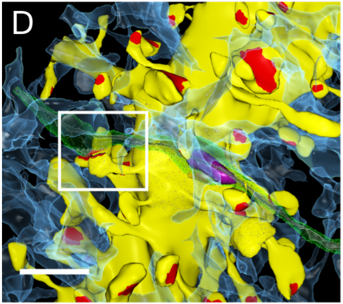
Rössler oscillator

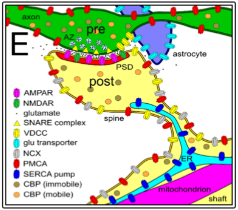


Rössler oscillator



Future



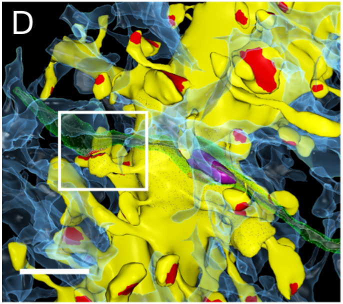


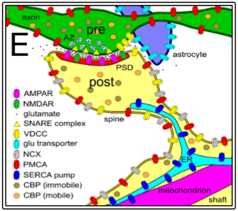
- $\mathcal{F}$ is a functiona**l** diff. eq. model
- Parameterize: functiona**l** $\mathcal{F} \to$ ordinary functions F
- Parameterize: functions $F \to$ parameters θ

What do we gain?

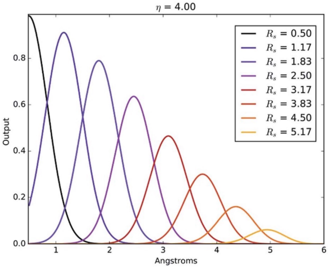
- Introduce physics
- Efficient learning algorithm

Future





- $\mathcal{F}$ is a functional diff. eq. model
- Parameterize: functional $\mathcal{F} \to$ ordinary functions F
- Parameterize: functions $F \to$ parameters θ



Smith (2017) *Chem. Sci.*, **8**, 3192

- Abstracting multiscale/multiphysics biochemistry
- Learning problem for diff. eqs.
- Physics + ML

This work:

Ernst 2018 - Learning Dynamic Boltzmann Distributions as Reduced Models of Spatial Chemical Kinetics. *arXiv:1803.01063*.

Prior work (GCCD):

Johnson 2015 - Model reduction for stochastic CaMKII reaction kinetics in synapses by graph-constrained correlation dynamics. *Phys. Biol.*

A roadmap:

Mjolsness 2018 - Prospects for Declarative Mathematical Modeling of Complex Biological Systems. *arXiv:1804.11044*.

Thanks!

Tom, Terry, Eric Mjolsness (UCI)