Learning PDE model reductions/ moment closures of stochastic reaction-diffusion dynamics

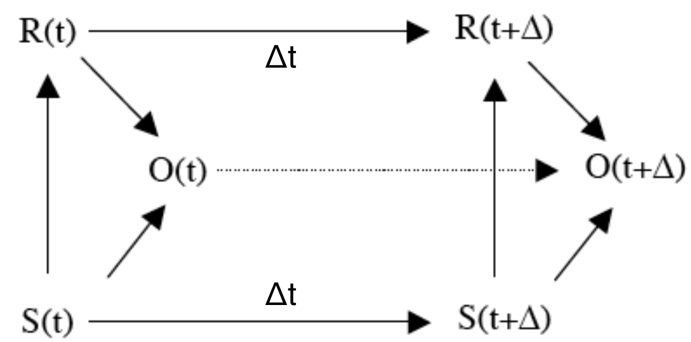
Eric Mjolsness

University of California, Irvine

http://emj.ics.uci.edu

NIH
24 October 2019

Mapping: Model reduction



[Johnson, Bartol, Sejnowski, and Mjolsness. Physical Biology 12:4, July 2015]

$$\Psi \mathscr{R} \simeq \mathscr{R} \Psi$$

$$\frac{dp}{dt} = W \cdot p$$

- Nonspatial: $\hat{p}(R,t) = \exp\left[-\sum_{\alpha} \mu_\alpha(t) V_\alpha(R)\right] / \hat{Z}(\mu(t))$

[Johnson, Bartol, Sejnowski, and Mjolsness. Physical Biology 12:4, July 2015]

  - Graph-Constrained Correlation Dynamics
  - warmup case for …
- Spatial generalization: $\tilde{p}(n, \boldsymbol{x}, \boldsymbol{\alpha}, t) = \frac{1}{Z} \exp\left[-\sum_{k=1}^{K} \sum_{\langle j \rangle} \nu_k(\boldsymbol{x}_{\langle j \rangle}, \boldsymbol{\alpha}_{\langle j \rangle}, t)\right],$
  - Dynamic Boltzmann distributions

MaxEnt Problem

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

$$\text{w/ } \mathcal{D}_{\mathcal{KL}}(p||\tilde{p}) = \sum_{n=0}^{\infty} \int d\boldsymbol{x}\; p \ln \frac{p}{\tilde{p}}$$

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

Variational problem

$$\frac{\delta S}{\delta F_k[\{\nu_k(\mathbf{x})\}_{k=1}^K]} = 0 \text{ for } k = 1, \dots, K \text{ at all } \boldsymbol{x} \tag{12}$$

where the variation is with respect to a set of **functionals**

$$\boxed{\dot{\nu}_k(\mathbf{x}) = F_k[\{\nu_k\}_{k=1}^K]} \tag{13}$$

… Higher-order calculus!

Slide: Oliver Ernst, Salk

O. Ernst, T. Bartol, T. Sejnowksi, and E. Mjolsness , J of Chem Phys 149, 034107, July 2018. Also arXiv 1803.01063

Variational Problem: Spatial systems

$$\frac{\delta S}{\delta F_k[\boldsymbol{\nu}(\boldsymbol{x})]} = \sum_{k'=1}^{K} \int d\boldsymbol{x}' \int dt \, \frac{\delta S}{\delta \nu_{k'}(\boldsymbol{x}', t)} \frac{\delta \nu_{k'}(\boldsymbol{x}', t)}{\delta F_k[\boldsymbol{\nu}(\boldsymbol{x})]} = 0 \qquad (19)$$

1

$$\frac{\delta S}{\delta \nu_{k'}(\boldsymbol{x}', t)} = \left\langle \sum_{\langle i \rangle_{k'}^{n}} \delta(\boldsymbol{x}' - \boldsymbol{x}_{\langle i \rangle_{k'}^{n}}) \right\rangle_{p} - \left\langle \sum_{\langle i \rangle_{k'}^{n}} \delta(\boldsymbol{x}' - \boldsymbol{x}_{\langle i \rangle_{k'}^{n}}) \right\rangle_{\tilde{p}} \qquad (20)$$

e.g. $k' = 1 : \left\langle \sum_{i=1}^{n} \delta(x_i - x') \right\rangle$ for all x'

$k' = 2 : \left\langle \sum_{i=1}^{n} \sum_{j>i} \delta(x_i - x_1')\delta(x_j - x_2') \right\rangle$ for all x_1', x_2'

2

Need to choose a parametrization for functional!

Slide: Oliver Ernst, Salk

O. Ernst, T. Bartol, T. Sejnowksi, and E. Mjolsness , J of Chem Phys 149, 034107, July 2018. Also arXiv 1803.01063

$\mathscr{R}$

Diffusion-inspired parametrization

$$p(x) \sim \exp\left[-\frac{(x-x_0)^2}{4Dt}\right] \to \exp[-\nu_1(x)]$$

$$\text{satisfies: } \frac{\partial \nu_1}{\partial t} = D\nabla^2\nu_1(x) - D(\nabla\nu_1(x))^2$$

$$\therefore F_k[\boldsymbol{\nu}(\boldsymbol{x})] = F_k^{(0)} + \sum_{\lambda=1}^{k} F_{k\lambda}^{(1)} (\nabla\nu_\lambda)^2 + \sum_{\lambda=1}^{k} F_{k\lambda}^{(2)} (\nabla^2\nu_\lambda) \qquad (20)$$

where: $F =$ some funcs of ν on LHS

$$\frac{\delta S}{\delta F_k^{(0)}} = 0, \frac{\delta S}{\delta F_{k\lambda}^{(1)}} = 0, \frac{\delta S}{\delta F_{k\lambda}^{(2)}} = 0$$

PDE-constrained Optimization Problem

$$\text{Minimize} \sum_{k'=1}^{K} \int_0^\infty dt \left(\left\langle \sum_{\langle i\rangle_{k'}^n} \delta(\boldsymbol{x}' - \boldsymbol{x}_{\langle i\rangle_{k'}^n}) \right\rangle_p - \left\langle \sum_{\langle i\rangle_{k'}^n} \delta(\boldsymbol{x}' - \boldsymbol{x}_{\langle i\rangle_{k'}^n}) \right\rangle_{\tilde{p}} \right) \frac{\delta \nu_{k'}(t)}{\delta F} \quad (23)$$

subject to PDE constraints for $\delta\nu_{k'}(t)/\delta F$.

Slide: Oliver Ernst, Salk

O. Ernst, T. Bartol, T. Sejnowksi, and E. Mjolsness , J of Chem Phys 149, 034107, July 2018. Also arXiv 1803.01063

Adjoint method BMLA-like learning algorithm

Algorithm 1 Stochastic Gradient Descent for Learning Restricted Boltzmann Machine Dynamics

---

Initialize
    Parameters $\boldsymbol{u}_k$ controlling the functions $F_k(\boldsymbol{\theta}; \boldsymbol{u}_k)$ for all $k = 1, \dots, K$.
    Time interval $[t_0, t_f]$, a formula for the learning rate λ.
while not converged **do**
    Initialize $\Delta F_{k,i} = 0$ for all $k = 1, \dots, K$ and parameters $i = 1, \dots, M_k$.
    for sample in batch **do**
        ▷ *Generate trajectory in reduced space* $\boldsymbol{\theta}$*:*
        Solve the PDE constraint (27) for $\theta_k(t)$ with a given IC $\theta_{k,0}$ over $t_0 \le t \le t_f$, for all k.

$$\frac{d}{dt}\theta_k(t) = F_k(\boldsymbol{\theta}(t); \boldsymbol{u}_k)$$

        ▷ *Wake phase:*
        Evaluate moments $\mu_k(t)$ of the data for all k, t.
        ▷ *Sleep phase:*
        Evaluate moments $\tilde{\mu}_k(t)$ of the Boltzmann distribution.
        ▷ *Solve the adjoint system:*
        Solve the adjoint system (31) for $\phi_k(t)$ for all k, t.

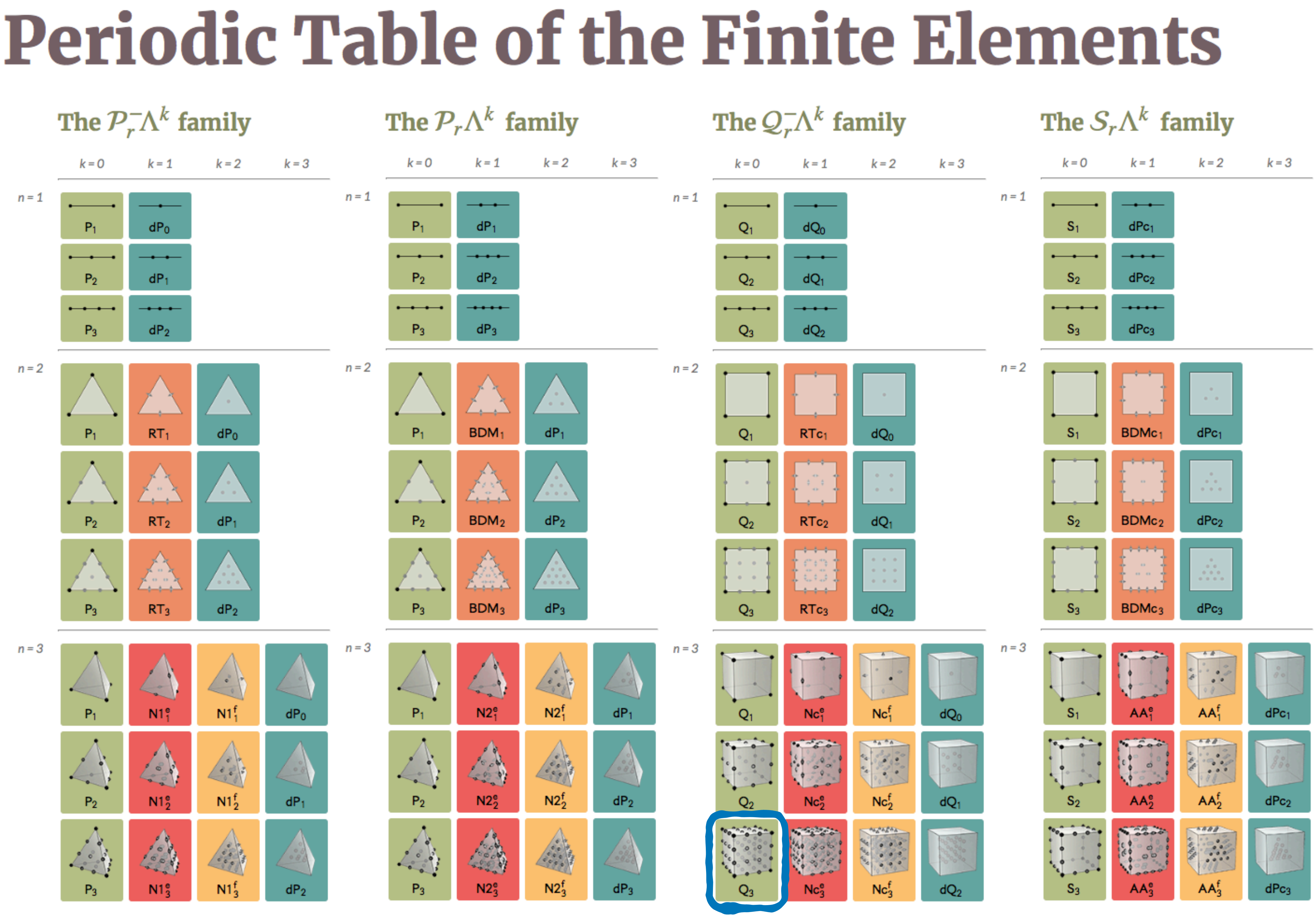
$$\frac{d}{dt}\phi_k(t) = \tilde{\mu}_k(t) - \mu_k(t) - \sum_{l=1}^{K} \frac{\partial F_l(\boldsymbol{\theta}(t); \boldsymbol{u}_l)}{\partial \theta_k(t)} \phi_l(t),$$

        ▷ *Evaluate the objective function:*
        Update $\Delta F_{k,i}$ as the cumulative moving average of the sensitivity equation (30) over the batch.

$$\frac{dS}{du_{k,i}} = -\int_{t_0}^{t_f} dt \, \frac{\partial F_k(\boldsymbol{\theta}(t); \boldsymbol{u}_k)}{\partial u_{k,i}} \phi_k(t).$$

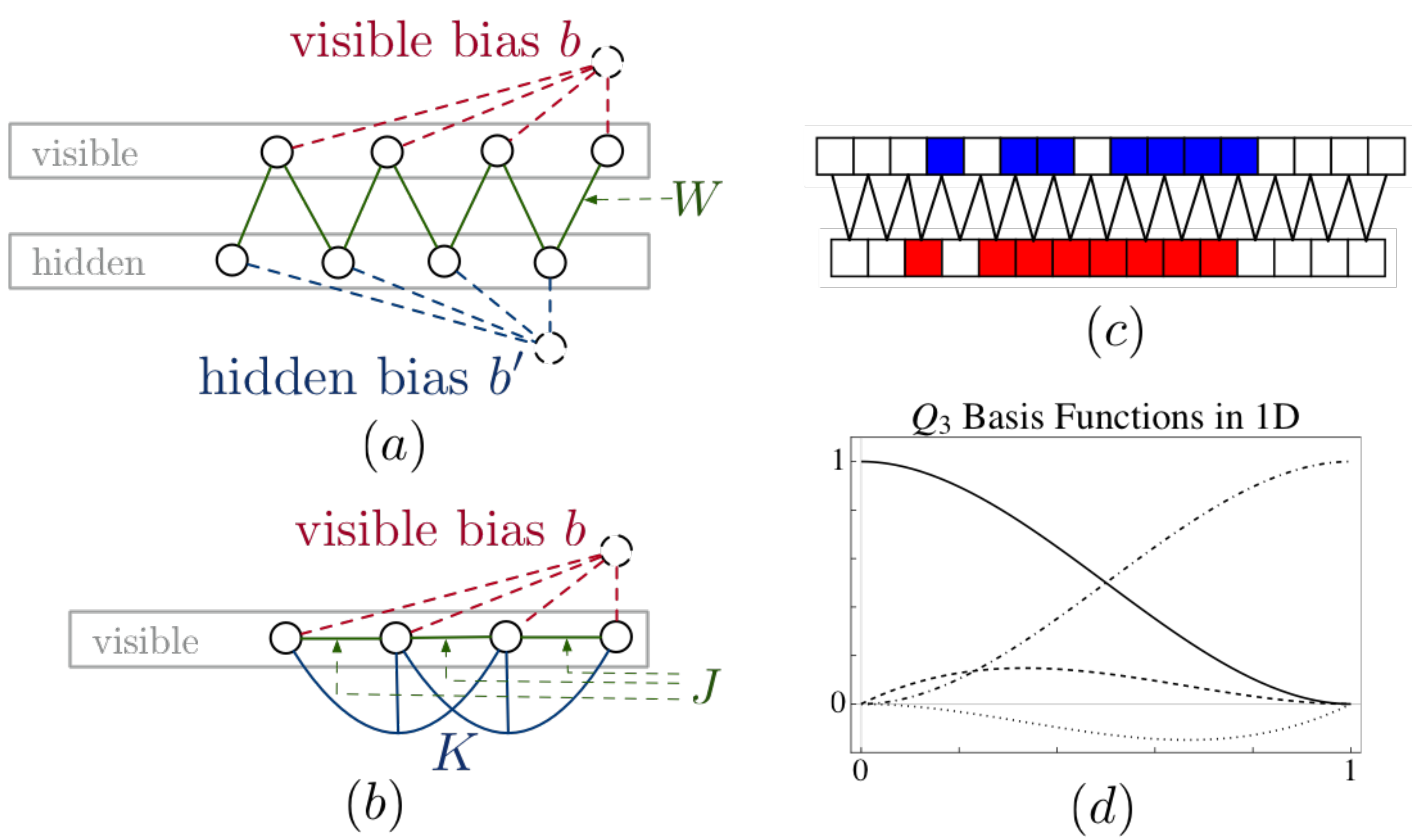
    ▷ *Update to decrease objective function:*
    $u_{k,i} \to u_{k,i} - \lambda \Delta F_{k,i}$ for all k, i.

Interpolation of F on LHS



Benefit of Hidden Units

Network: fratricide + lattice diffusion



$$E(\boldsymbol{v}, \boldsymbol{h}, b(t), W(t), b'(t)) = -b(t)\sum_{i=1}^{N} v_i - b'(t)\sum_{j=1}^{N-1} h_j - W(t)\sum_{i=1}^{N}\sum_{j=i-1,i} v_i h_j,$$

$$\frac{d}{dt}\gamma = F_\gamma(b, b', W; \boldsymbol{u}_\gamma) \quad \text{for} \quad \gamma = b, b', W.$$

$$E(\boldsymbol{v}, b(t), J(t), K(t)) = -b(t)\sum_{i=1}^{N} v_i - J(t)\sum_{i=1}^{N-1} v_i v_{i+1} - K(t)\sum_{i=1}^{N-2} v_i v_{i+1} v_{i+2},$$

$$\frac{d}{dt}\gamma = F_\gamma(b, J, K; \boldsymbol{u}_\gamma) \quad \text{for} \quad \gamma = b, J, K.$$

[Ernst, Bartol, Sejnowski, Mjolsness, arXiv:1808.08630 v2 April 2019]

Benefit of Hidden Units $\mathscr{R}$

Network: fratricide + lattice diffusion

- Learned DBD ODE RHS, without and with hidden units

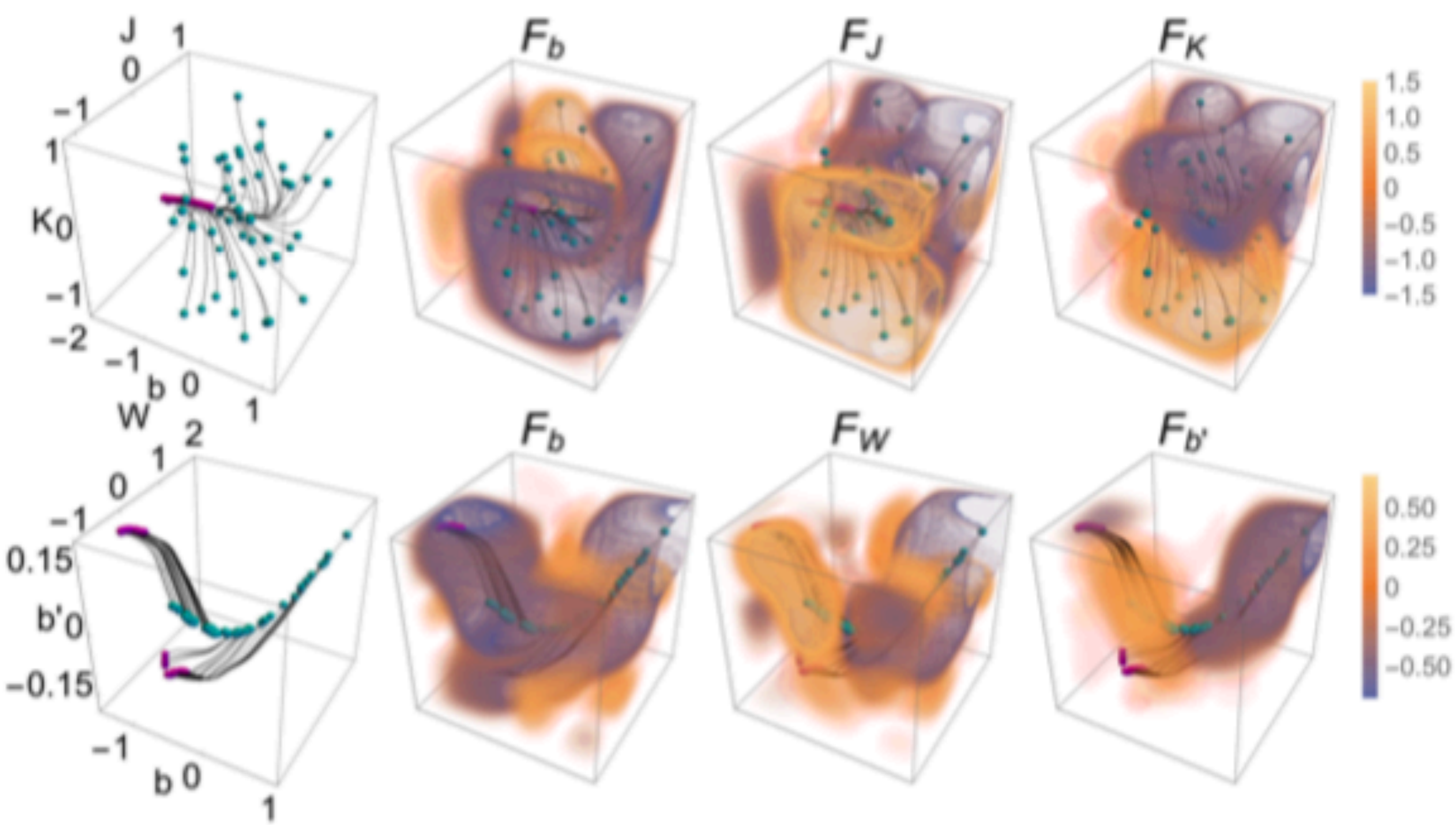
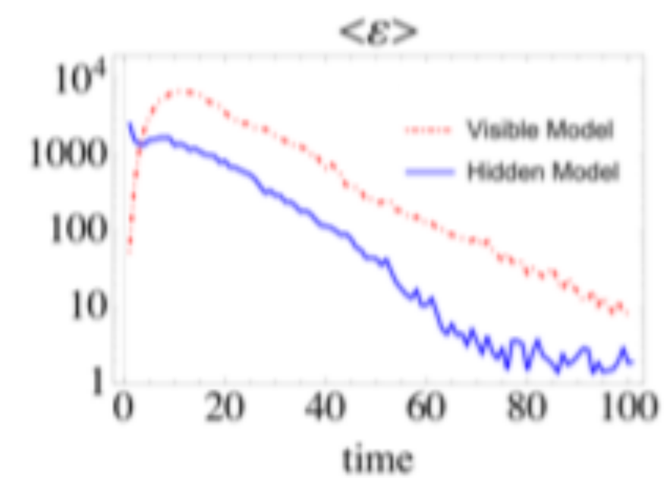


FIG. 2. *Top row:* Learned time-evolution functions for the fully visible model (19), using the Q_3, C_1 finite element parameterization (21) with $5 \times 5 \times 5$ evenly spaced cubic cells. Left: Training set of initial points (b, J, K) (cyan) sampled evenly in $[-1, 1]$. Stochastic simulations for each initial point are used as training data (learned trajectories shown in black, endpoints in magenta). Other panels: the time evolution functions learned. *Bottom row:* Hidden layer model (20) and parameterization (21) with the same number of cells as the visible model. Initial points are generated by BM learning the points of the visible model.

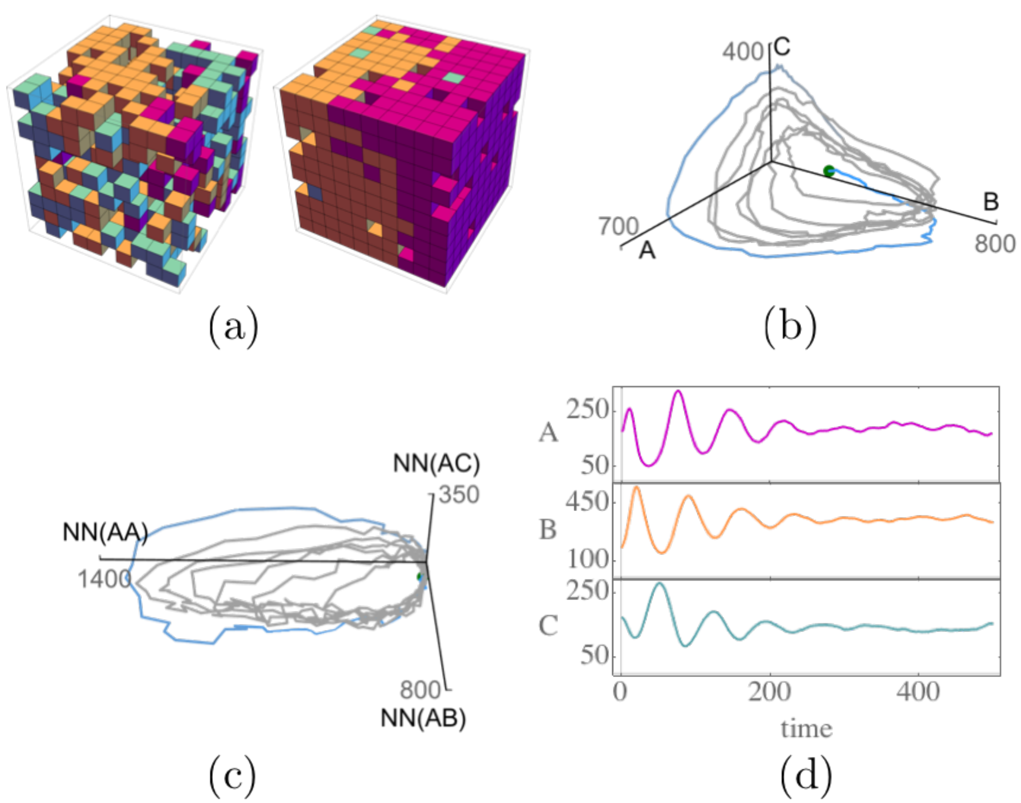


MSE of 4th order stats

[Ernst, Bartol, Sejnowski, Mjolsness, Phys Rev E 99 063315, 2019]

Rössler Oscillator in 3D

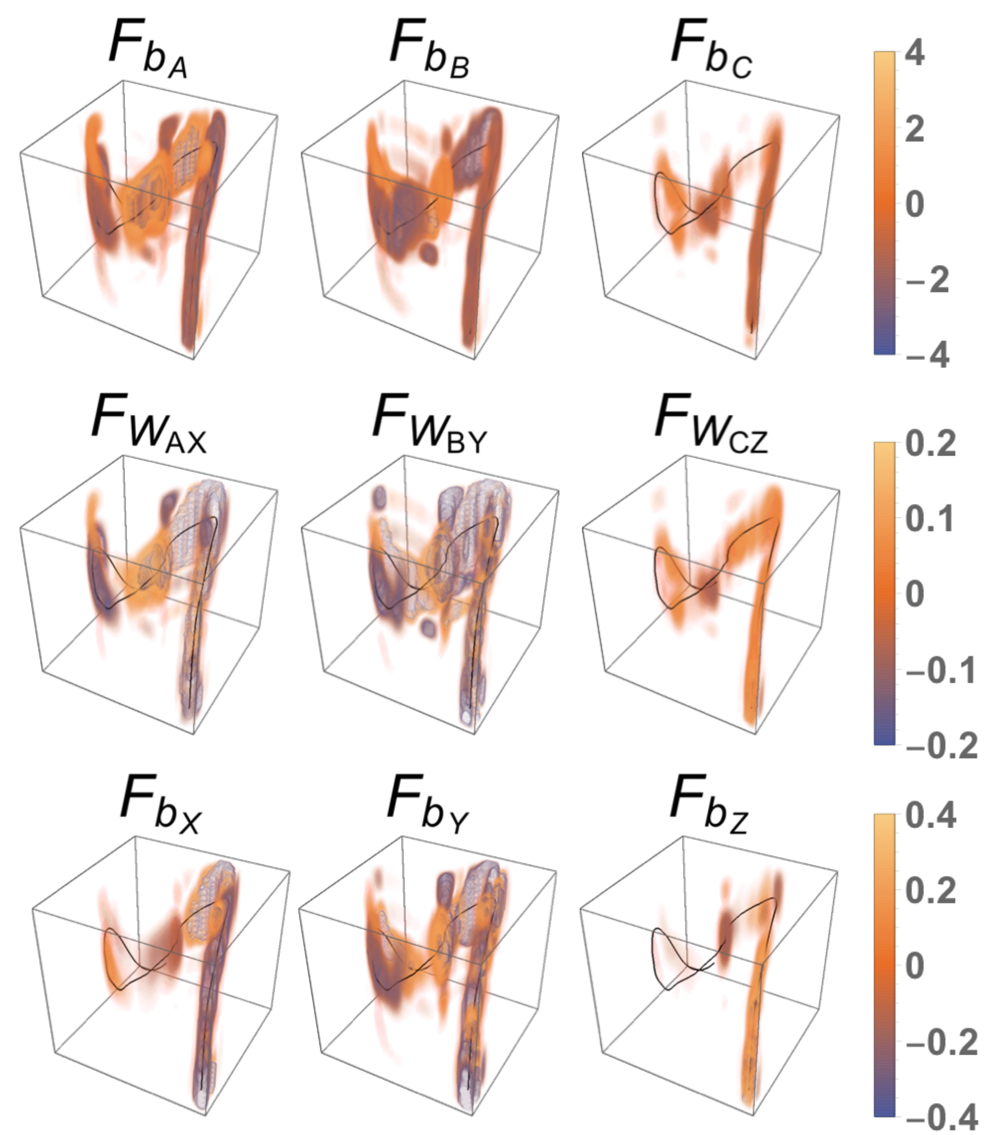
- Function:



(a) (b)

(c) (d)

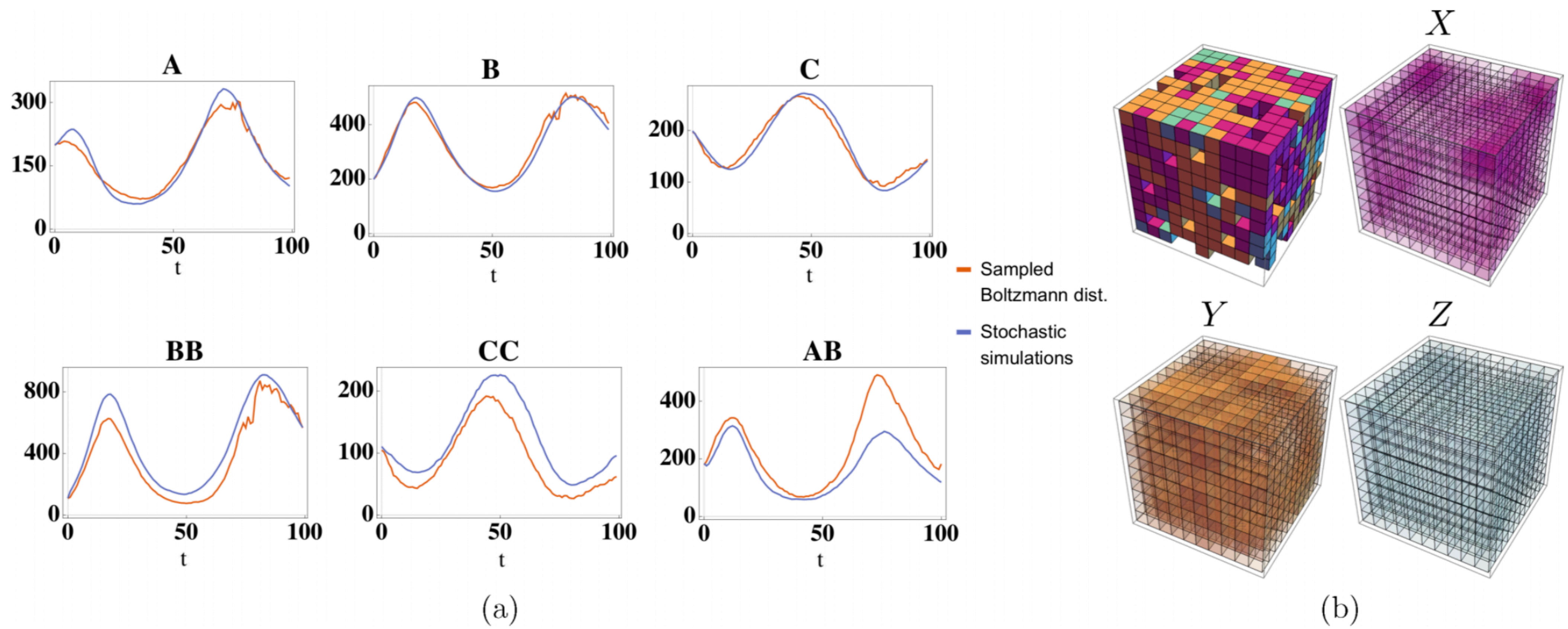
- Learned DBD ODE RHS:



[Ernst, Bartol, Sejnowski, Mjolsness, Phys Rev E 99 063315, 2019]

Rössler Oscillator in 3D

- Learned correlations:
- Learned Configuration



[Ernst, Bartol, Sejnowski, Mjolsness, Phys Rev E 99 063315, 2019]

$\Psi, \mathscr{I}$

Fields to Structures

$$\frac{dp}{dt} = W \cdot p$$

- Fields: PDE differential operator dynamics in W
- Software dichotomy: particle vs. PDE solvers
  - Usual integration: operator splitting. Deeper unification?
  - Regardless, a third category is necessary: dynamic structures.
- Dynamical Graph Grammars (DGGs):
  - operator addition of reactions, GGs, ODEs;
  - but what about PDEs?
- Approximately eliminate fields by:
  - Cell complexes in PDE (adaptive) meshing / FEMs, FVMs

MT fiber
Stochastic Parametrized Graph Grammar

$$(\bullet_1)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1)\rangle\!\rangle \longrightarrow (\circ_1 \rightarrow \bullet_2)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1),(\boldsymbol{x}_2,\boldsymbol{u}_2)\rangle\!\rangle$$

$$\textbf{with}\quad \hat{\rho}_{\text{grow}}([\text{tubulin}])\mathcal{N}(\boldsymbol{x}_1-\boldsymbol{x}_2;L\boldsymbol{u}_1,\sigma)\mathcal{N}(\boldsymbol{u}_2;\boldsymbol{u}_1/(|\boldsymbol{u}_1|+\epsilon),\epsilon),$$

$$(\blacksquare_1 \rightarrow \circ_2)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1),(\boldsymbol{x}_2,\boldsymbol{u}_2)\rangle\!\rangle \longrightarrow (\blacksquare_2)\,\langle\!\langle(\boldsymbol{x}_2,\boldsymbol{u}_2)\rangle\!\rangle$$

$$\textbf{with}\quad \hat{\rho}_{\text{retract}}$$

$$\left(\begin{tikzcd} \circ_1 \arrow[r] & \circ_2 \arrow[r] & \circ_3 \\ & \bullet_4 & \end{tikzcd}\right)\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1),(\boldsymbol{x}_2,\boldsymbol{u}_2),(\boldsymbol{x}_3,\boldsymbol{u}_3),(\boldsymbol{x}_4,\boldsymbol{u}_4)\rangle\!\rangle$$

$$\longrightarrow \left(\begin{tikzcd} \circ_1 \arrow[r] & \blacktriangle_2 \arrow[r] & \circ_3 \\ \circ_4 \arrow[ur] & & \end{tikzcd}\right)\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1),(\boldsymbol{x}_2,\boldsymbol{u}_2),(\boldsymbol{x}_3,\boldsymbol{u}_3),(\boldsymbol{x}_4,\boldsymbol{u}_4)\rangle\!\rangle$$

$$\textbf{with}\quad \hat{\rho}_{\text{bundle}}(|\boldsymbol{u}_2\cdot\boldsymbol{u}_4|/|\cos\theta_{\text{crit}}|)\exp(-|\boldsymbol{x}_2-\boldsymbol{x}_4|^2/2L^2)$$

$$(\blacksquare_1 \rightarrow \bullet_2)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1),(\boldsymbol{x}_2,\boldsymbol{u}_2)\rangle\!\rangle \longleftrightarrow \varnothing \qquad \textbf{with}\ \Big(\hat{\rho}_{\text{retract}},$$

$$\hat{\rho}_{\text{nucleate}}([\text{tubulin}])\mathcal{N}(\boldsymbol{x};\mathbf{0},\sigma_{\text{broad}})\delta_{\text{Dirac}}(|\boldsymbol{u}_1|-1)\delta_{\text{Dirac}}(\boldsymbol{u}_1-\boldsymbol{u}_2)\Big)$$

$$(\bullet_1)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1)\rangle\!\rangle \longleftrightarrow (\blacksquare_1)\,\langle\!\langle(\boldsymbol{x}_1,\boldsymbol{u}_1)\rangle\!\rangle$$

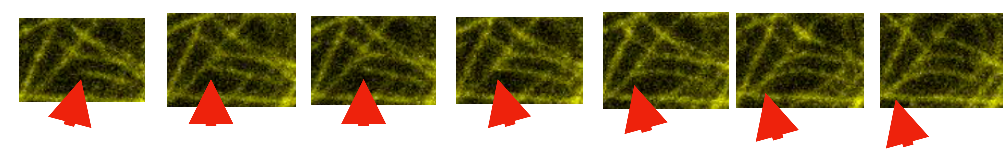
$$\textbf{with}\quad (\hat{\rho}_{\text{retract}\leftarrow\text{growth}},\hat{\rho}_{\text{growth}\leftarrow\text{retract}})$$

[EM, Bull. Math Biol. 81:8 Aug 2019
+arXiv:1804.11044]

MT fiber
Stochastic Parametrized Graph Grammar

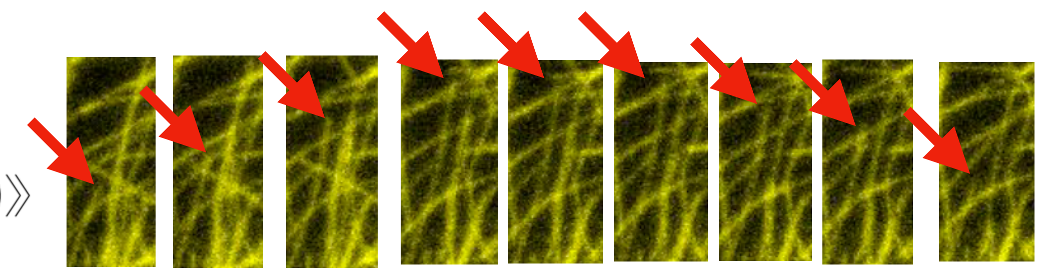
// (continued)

// Fiber collision, with several alternative discrete outcomes:



$$\left(\begin{array}{c} \quad \star_1 \text{ — } \bigcirc_2 \text{ — } \star_3 \\ \bigcirc_4 \text{ — } \bullet_5 \end{array}\right) \langle\!\langle (x_1,u_1),(l_2,u_2),(x_3,u_3),(l_4,u_4),(x_5,u_5) \rangle\!\rangle$$

$$\longrightarrow \left(\begin{array}{c} \star_1 \text{ — } \bigcirc_6 \text{ — } \blacktriangle_2 \text{ — } \bigcirc_7 \text{ — } \star_3 \\ \diagup \\ \bigcirc_4 \end{array}\right) \langle\!\langle (x_1,u_1),(x_2,u_2),(x_3,u_3),(l_4,u_4),\varnothing,(\alpha l_4,u_2),((1-\alpha)l_4,u_2$$



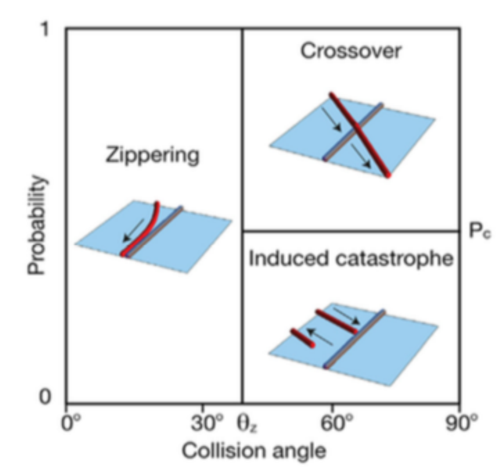
$$\textbf{with } \hat{\rho}_{\text{bundle}}(|u_2 \cdot u_4|/|\cos\theta_{\text{crit}}|)\exp\left(-\gamma^2/2\epsilon^2\right)\Theta(\epsilon \leqslant \alpha \leqslant 1-\epsilon)$$

$$\longrightarrow \left(\begin{array}{c} \quad \star_1 \text{ — } \bigcirc_2 \text{ — } \star_3 \\ \bigcirc_4 \text{ — } \blacksquare_5 \end{array}\right) \langle\!\langle (x_1,u_1),(l_2,u_2),(x_3,u_3),(l_4,u_4),(x_5,u_5) \rangle\!\rangle$$

$$\textbf{with } \hat{\rho}'_{\text{bundle}}(|u_2 \cdot u_4|/|\cos\theta_{\text{crit}}|)\exp\left(-\gamma^2/2\epsilon^2\right)\Theta(\epsilon \leqslant \alpha \leqslant 1-\epsilon)$$

$$\longrightarrow \left(\begin{array}{c} \bullet_9 \\ \diagup \\ \bigcirc_8 \\ \diagup \\ \star_1 \text{ — } \bigcirc_6 \text{ — } \blacklozenge_2 \text{ — } \bigcirc_7 \text{ — } \star_3 \\ \diagup \\ \bigcirc_4 \end{array}\right) \left\langle\!\!\left\langle \begin{array}{l} (x_1,u_1),((1-\alpha)x_1+\alpha x_3,u_2),(x_3,u_3),(l_4,u_4),\varnothing, \\ (\alpha l_2,u_2),((1-\alpha)l_2,u_2),(\epsilon l_4,u_4),(x_2+\epsilon l_4 u_4),u_4) \end{array} \right\rangle\!\!\right\rangle$$

$$\textbf{with } \hat{\rho}''_{\text{bundle}}(|u_2 \cdot u_4|/|\cos\theta_{\text{crit}}|)\exp\left(-\gamma^2/2\epsilon^2\right)\Theta(\epsilon \leqslant \alpha \leqslant 1-\epsilon)$$



[Chakrabortty et al. Current Biology

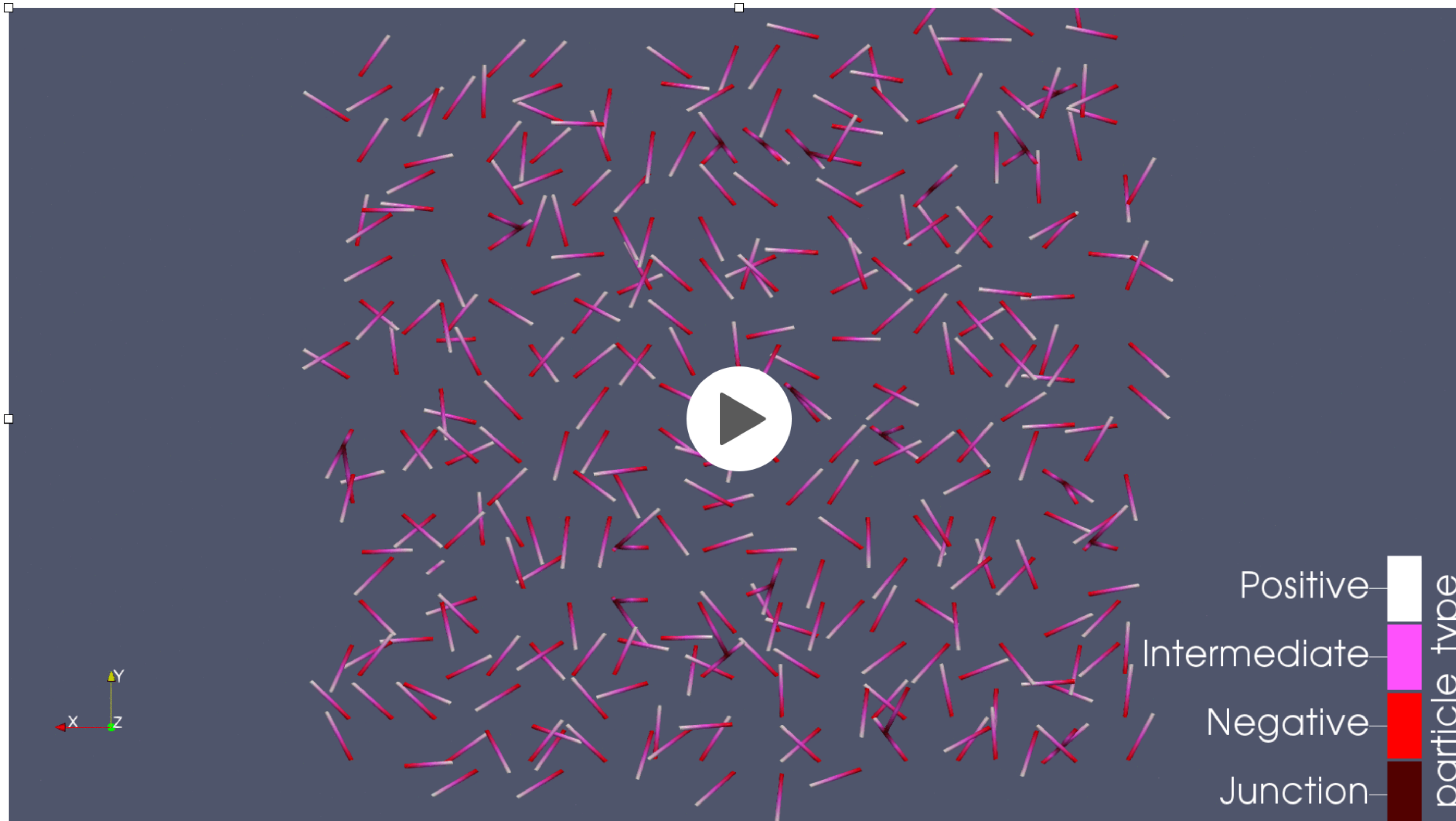
$$\textbf{where} \quad \begin{aligned} \gamma &= -[(x_3-x_1)\times(x_1-x_5)]_z/[(x_3-x_1)\times u_5]_z && \text{// rel. distance to intersection along } u_5 \\ \alpha &= -[(x_1-x_5)\times u_5]_z/[(x_3-x_1)\times u_5]_z && \text{// fractional location of intersection along } u_2 \end{aligned}$$

[EM, Bull. Math Biol. 81:8 Aug 2019
+arXiv:1804.11044]

Cajete MT: First Light

Eric Medwedeff, UCI; Bob Bird, LANL

Cajete MT: First Light



Eric Medwedeff, UCI; Bob Bird, LANL

$\mathcal{J}$

Why operator algebra yields algorithms

- Baker Campbell Hausdorff theorem
  - => operator splitting algorithms e.g. Trotter Product Formula …

$$\lim_{n\to\infty} \left[e^{(t/n)\,H_0}\ e^{(t/n)\,H_1}\right]^n$$

- Time-ordered product expansions => Stochastic Simulation Algorithm (SSA)
  - – [EMj, Phys Bio 2013]

$$\boxed{\begin{aligned}\exp(t\,(W_0+W_1)) &= \exp(t\,W_0)\left(\exp\left(\int_0^t \exp(-\tau\,W_0)\,W_1\,\exp(\tau\,W_0)\,d\tau\right)\right)_+ \\ &\equiv \exp(t\,W_0)\left(\exp\left(\int_0^t W_1\,(\tau)\,d\tau\right)\right)_+\end{aligned}}$$

  - – weighted SSA (wSSA) possible too

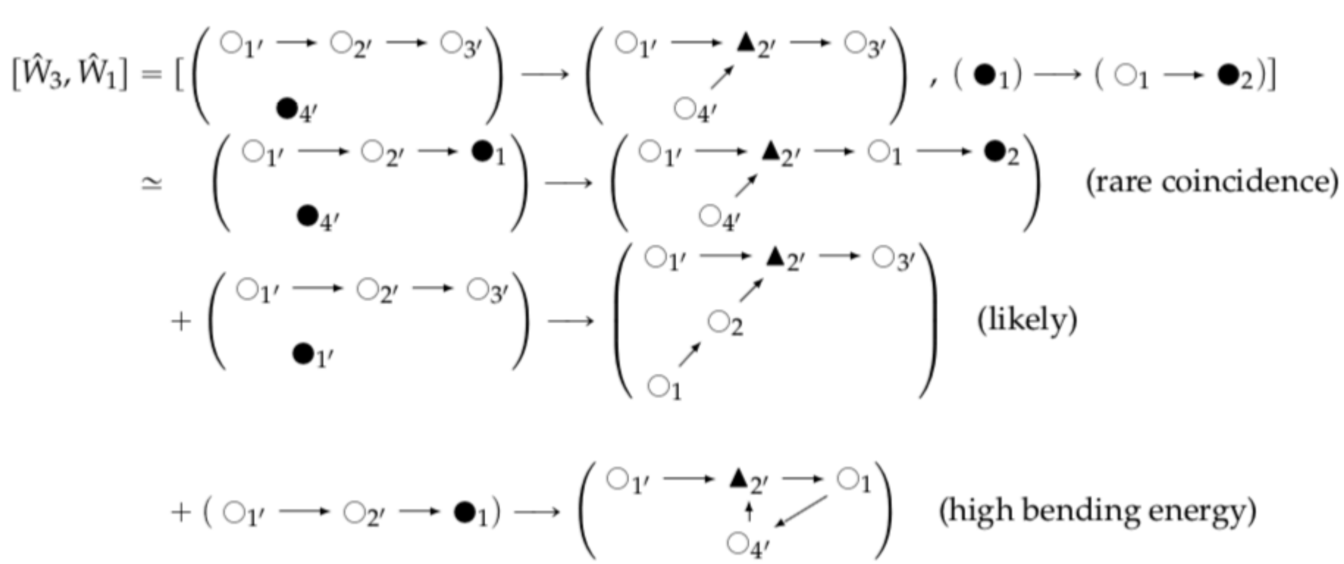
Ψ

Algebra of Labelled-Graph Rewrite Rules

$$\hat W_{G^{r_2\,\mathrm{in}}\to G^{r_2\,\mathrm{out}}}\hat W_{G^{r_1\,\mathrm{in}}\to G^{r_1\,\mathrm{out}}} \simeq \sum_{\substack{H\subseteq G^{r_1\,\mathrm{out}}\simeq \tilde H\subseteq G^{r_2\,\mathrm{in}}\\ |\ \text{edge-maximal}}} \;\sum_{h:H \xhookrightarrow{1\text{-}1}\tilde H} \hat W_{G^{r_1\,\mathrm{in}}\cup(G^{r_2\,\mathrm{in}}\setminus\tilde H)\underset{h}{\to} G^{r_2\,\mathrm{out}}\cup(G^{r_1\,\mathrm{out}}\setminus H)}$$

E.g.:

$$\begin{aligned}[\hat W_2,\hat W_1] &= [(\blacksquare_{1'}\longrightarrow \circ_{2'})\longrightarrow(\blacksquare_{2'})\ ,\ (\bullet_1)\longrightarrow(\circ_1\longrightarrow\bullet_2)]\\ &\simeq\ (\blacksquare_{1'}\longrightarrow\bullet_1)\longrightarrow(\blacksquare_1\longrightarrow\bullet_2)\end{aligned}$$

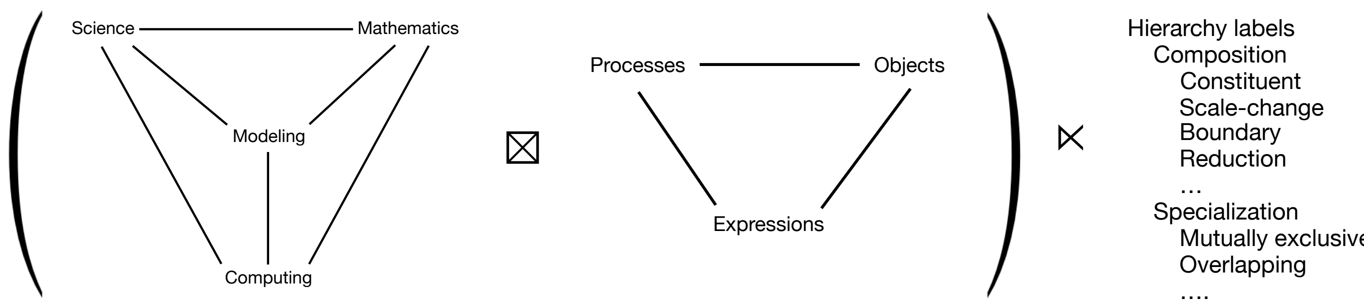


[EM, http://arxiv.org/abs/1909.04118]

+ (3 terms whose LHS rely on MT syntax violations - omitted)

“Tchicoma” Architecture for Mathematical Modeling

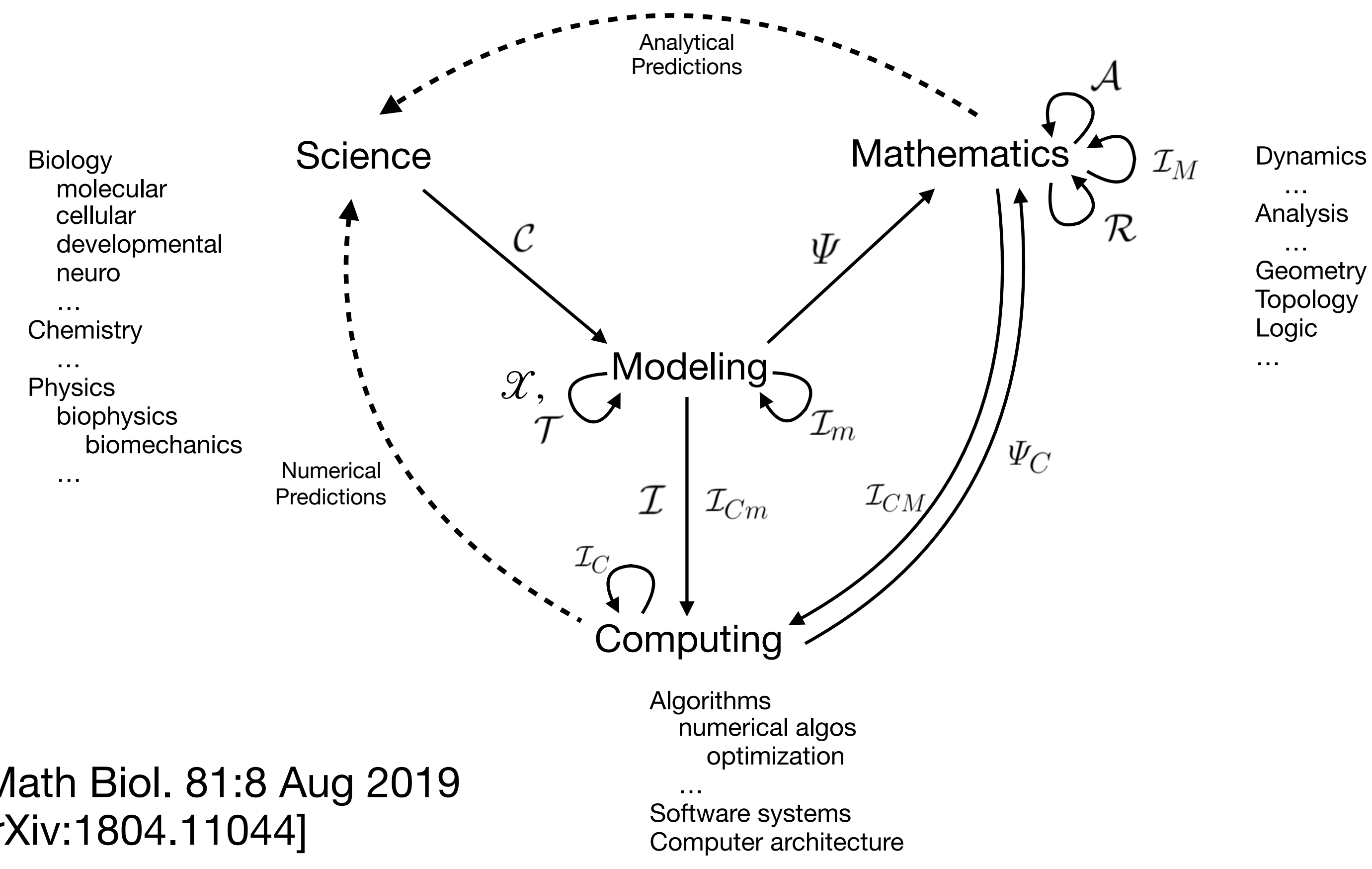
- Language meta-hierarchy: *(a DAG with edge labels in a tree)*



- Mappings therein:

respecting compositional structure

Features:
Enables problem-solving
via chaining, theorem-proving
Foments abstraction
via commutation
Decoupled, yet can be efficient



[EM, Bull. Math Biol. 81:8 Aug 2019
+arXiv:1804.11044]

Conclusions

- Model reduction can be achieved by machine learning, in spatial stochastic models.
  - Coarse scale: Discretized PDEs for multi-particle Boltzmann potential energies.
  - Fine scale: Stochastic reaction/diffusion examples.
- Morpho-dynamic spatial structures can be modeled by dynamical graph grammars with operator algebra semantics.
  - Bio-universal (includes particles and *DEs).
  - Scalability is in progress.
  - MT examples.
- Model stacks are the key data structure for understanding complex bio systems.
  - They require model reduction and bio-universal modeling languages (perhaps as above).
  - They can intersect productively, and could be curated in a proposed conceptual architecture "Tchicoma".
- In these ways, both symbolic and numeric AI can (and should!) be brought to bear on understanding complex biological systems.