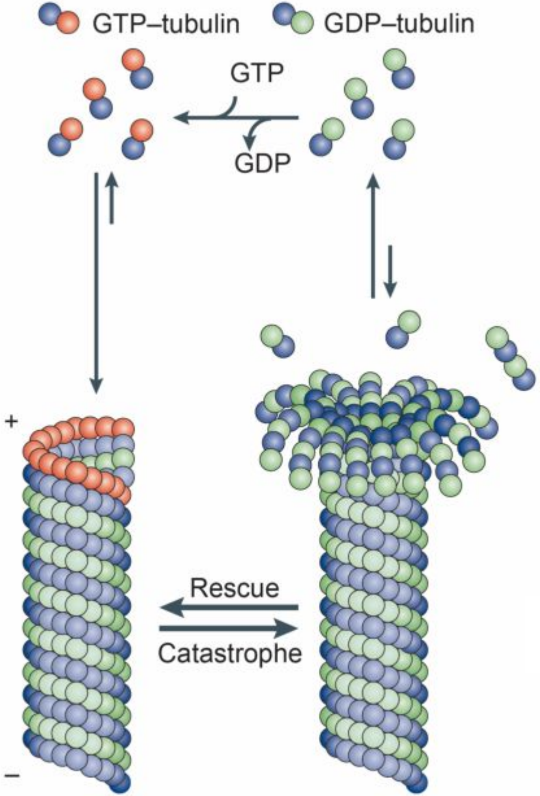
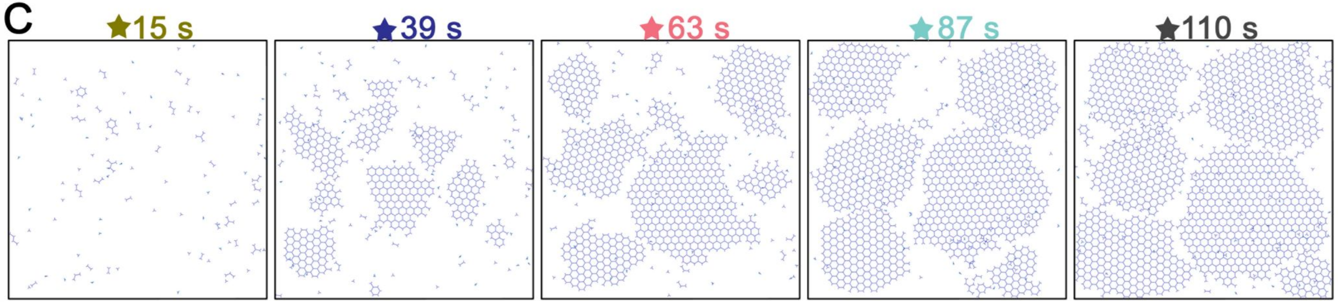
Autodiff-based estimation of kinetic trap rate constants

David Bass
Margaret Johnson lab
26 January 2023

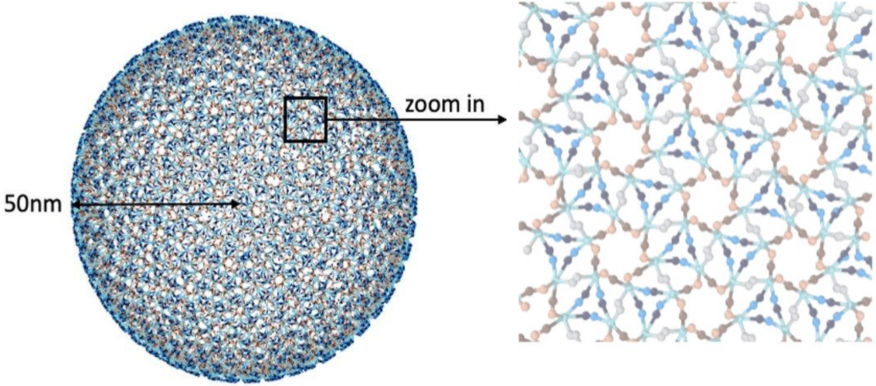
Molecular self-assembly is ubiquitous in cell biology



Cheeseman and Desai 2008,
Nat. Rev. Mol. Cell Biol.

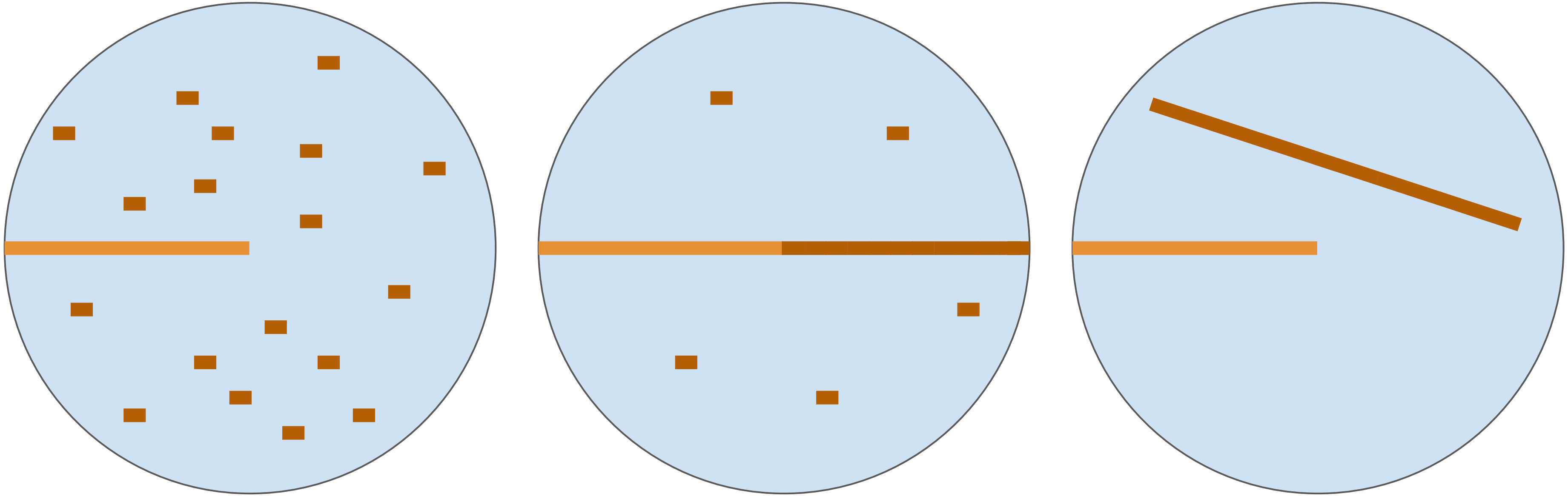


Guo et al. 2022, *PLoS Comput. Biol.*



Qian and Evans et al. 2023, *Biophys. J.*

Kinetic trapping inhibits molecular self-assembly

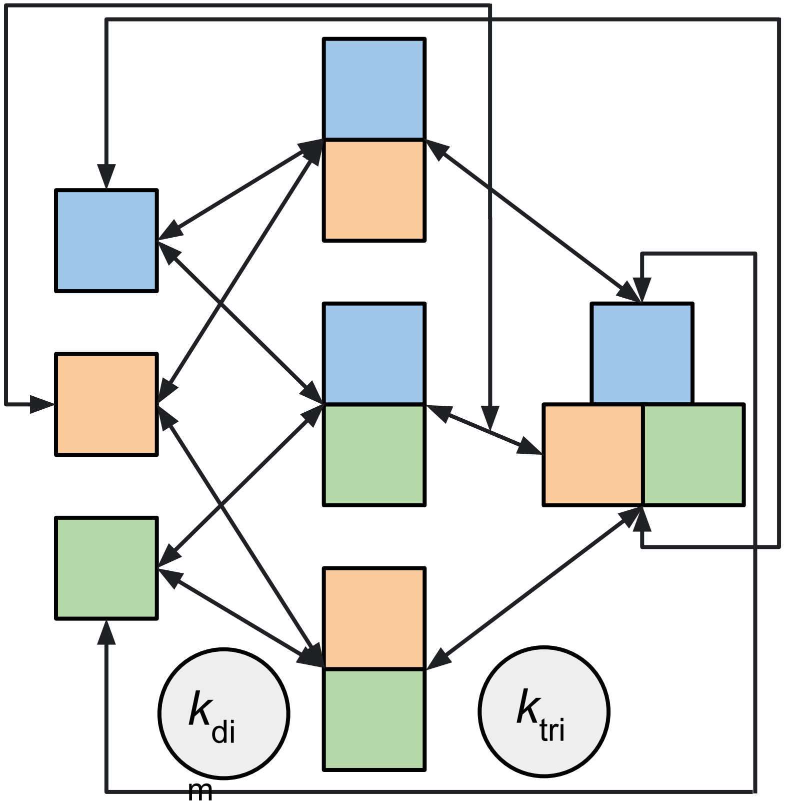


Cell seeks to span self with microtubule

Tubulin subunits bind to existing microtubule: No kinetic trapping

Tubulin subunits form new microtubule: Kinetic trapping

Fully connected hetero-*n*-mers provide a model system



$$\mathrm{X} + \mathrm{Y} \underset{k_{dim\_off}}{\overset{k_{dim\_on}}{\rightleftharpoons}} \mathrm{XY}$$

$$\mathrm{X} + \mathrm{Z} \underset{k_{dim\_off}}{\overset{k_{dim\_on}}{\rightleftharpoons}} \mathrm{XZ}$$

$$\mathrm{Y} + \mathrm{Z} \underset{k_{dim\_off}}{\overset{k_{dim\_on}}{\rightleftharpoons}} \mathrm{YZ}$$

$$\mathrm{XY} + \mathrm{Z} \underset{k_{tri\_off}}{\overset{k_{tri\_on}}{\rightleftharpoons}} \mathrm{XYZ}$$

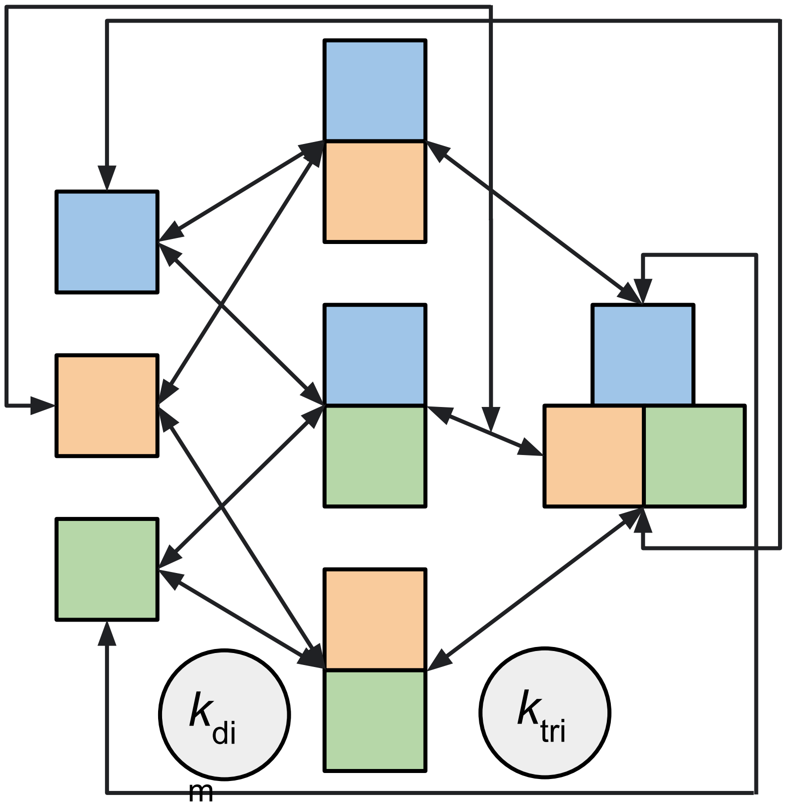
$$\mathrm{XZ} + \mathrm{Y} \underset{k_{tri\_off}}{\overset{k_{tri\_on}}{\rightleftharpoons}} \mathrm{XYZ}$$

$$\mathrm{YZ} + \mathrm{X} \underset{k_{tri\_off}}{\overset{k_{tri\_on}}{\rightleftharpoons}} \mathrm{XYZ}$$

Fully connected:
Every reactants binds to every other reactant

Rate-growth:
Every reaction at a given level of assembly has the same association rate, which determines a single dissociation rate

Fully connected hetero-n-mers provide a model system



Catalyst.jl

$$\frac{\mathrm{d}X(t)}{\mathrm{d}t} = k_{dim\_off}\mathrm{XY}(t) + k_{dim\_off}\mathrm{XZ}(t) + k_{tri\_off}\mathrm{XYZ}(t) - k_{dim\_on}Y(t)X(t) - k_{dim\_on}X(t)Z(t) - k_{tri\_on}X(t)\mathrm{YZ}(t) \tag{1}$$

$$\frac{\mathrm{d}Y(t)}{\mathrm{d}t} = k_{dim\_off}\mathrm{XY}(t) + k_{dim\_off}\mathrm{YZ}(t) + k_{tri\_off}\mathrm{XYZ}(t) - k_{dim\_on}Y(t)X(t) - k_{dim\_on}Y(t)Z(t) - k_{tri\_on}Y(t)\mathrm{XZ}(t) \tag{2}$$

$$\frac{\mathrm{dXY}(t)}{\mathrm{d}t} = -k_{dim\_off}\mathrm{XY}(t) + k_{tri\_off}\mathrm{XYZ}(t) + k_{dim\_on}Y(t)X(t) - k_{tri\_on}\mathrm{XY}(t)Z(t) \tag{3}$$

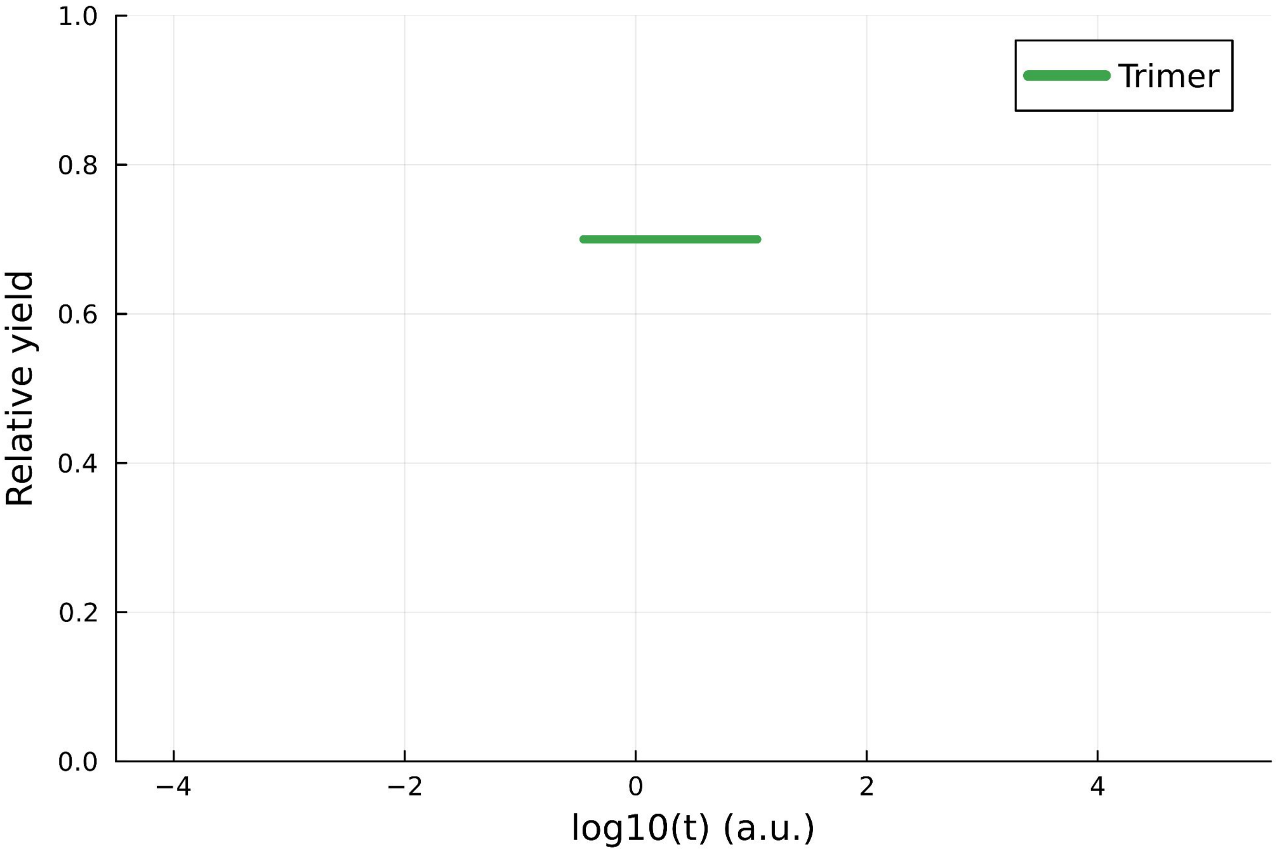
$$\frac{\mathrm{d}Z(t)}{\mathrm{d}t} = k_{dim\_off}\mathrm{XZ}(t) + k_{dim\_off}\mathrm{YZ}(t) + k_{tri\_off}\mathrm{XYZ}(t) - k_{dim\_on}Y(t)Z(t) - k_{dim\_on}X(t)Z(t) - k_{tri\_on}\mathrm{XY}(t)Z(t) \tag{4}$$

$$\frac{\mathrm{dXZ}(t)}{\mathrm{d}t} = -k_{dim\_off}\mathrm{XZ}(t) + k_{tri\_off}\mathrm{XYZ}(t) + k_{dim\_on}X(t)Z(t) - k_{tri\_on}Y(t)\mathrm{XZ}(t) \tag{5}$$

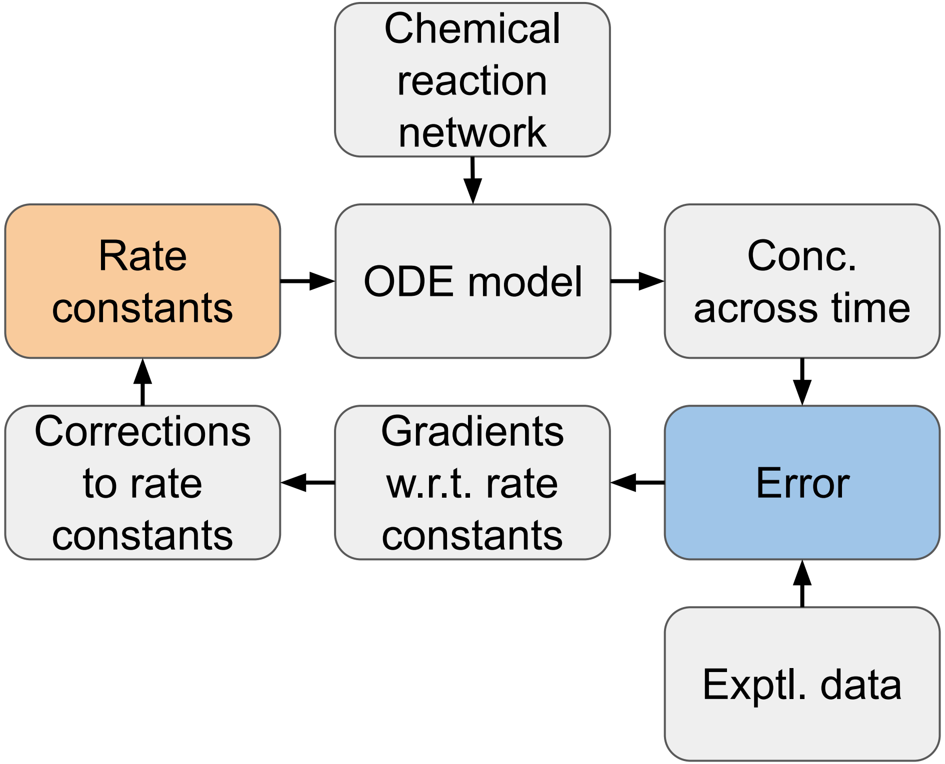
$$\frac{\mathrm{dYZ}(t)}{\mathrm{d}t} = -k_{dim\_off}\mathrm{YZ}(t) + k_{tri\_off}\mathrm{XYZ}(t) + k_{dim\_on}Y(t)Z(t) - k_{tri\_on}X(t)\mathrm{YZ}(t) \tag{6}$$

$$\frac{\mathrm{dXYZ}(t)}{\mathrm{d}t} = -3k_{tri\_off}\mathrm{XYZ}(t) + k_{tri\_on}Y(t)\mathrm{XZ}(t) + k_{tri\_on}\mathrm{XY}(t)Z(t) + k_{tri\_on}X(t)\mathrm{YZ}(t) \tag{7}$$

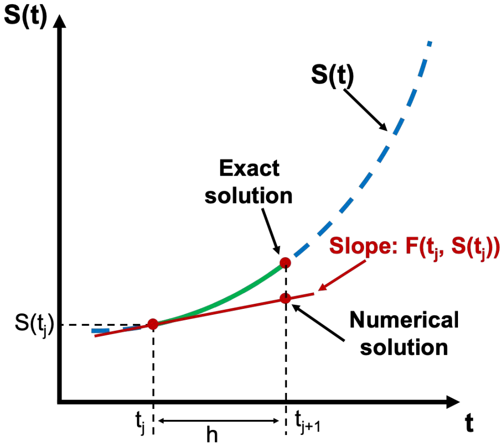
Kinetic trapping may encode information about reactions



Estimating rate constants requires expensive computation



$$\nabla \mathrm{Error} = \left(\frac{\partial \mathrm{Error}}{\partial k_{dim}}, \frac{\partial \mathrm{Error}}{\partial k_{tri}}, \dots \right)$$

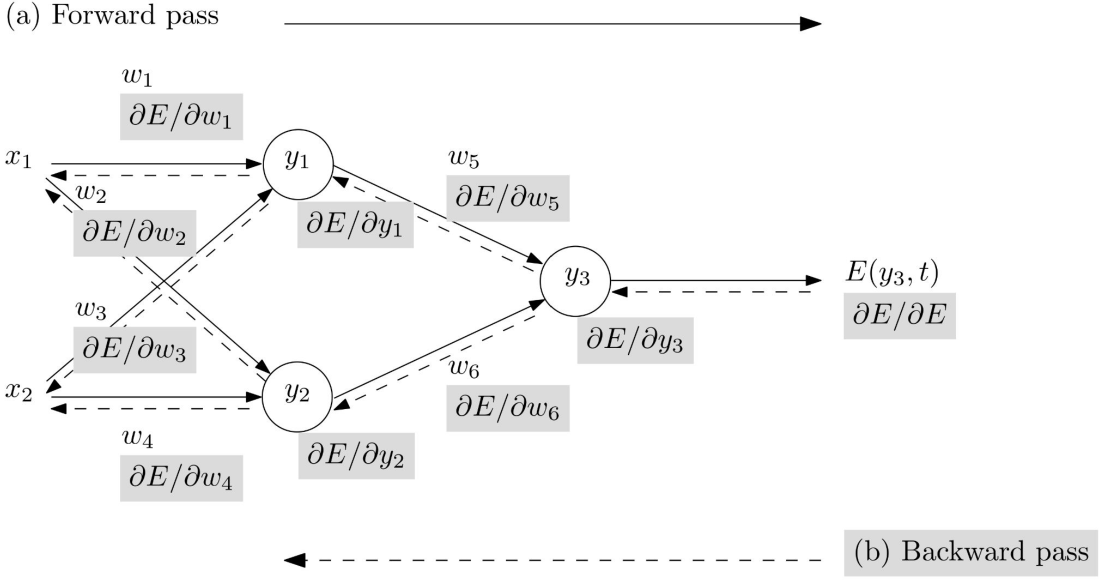


$$\frac{\partial [X]_n}{\partial t} = f([X]_n, [Y]_n, [Z]_n, \dots)$$

Kong et al. 2020, *Academic Press*

$$[X]_{n+1} \approx f([X]_n, [Y]_n, [Z]_n, \dots) \times (t_{n+1} - t_n)$$

Automatic differentiation accelerates gradient calculation



Baydin et al. 2018, *arXiv*

Reverse-mode autodiff excels when the ratio of inputs to outputs is high

Forward-mode autodiff requires less memory

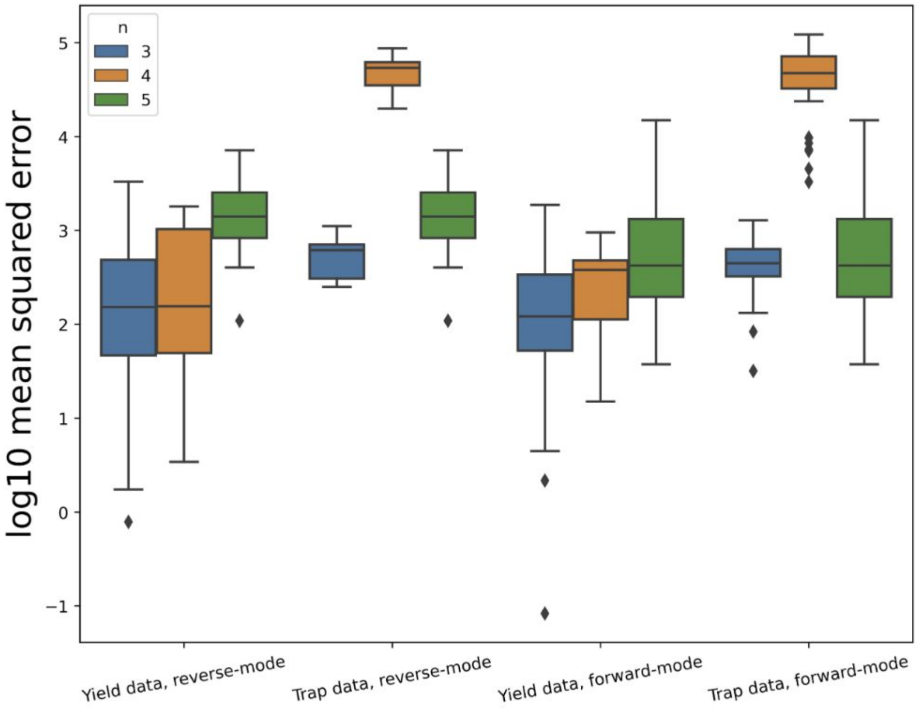
Zygote

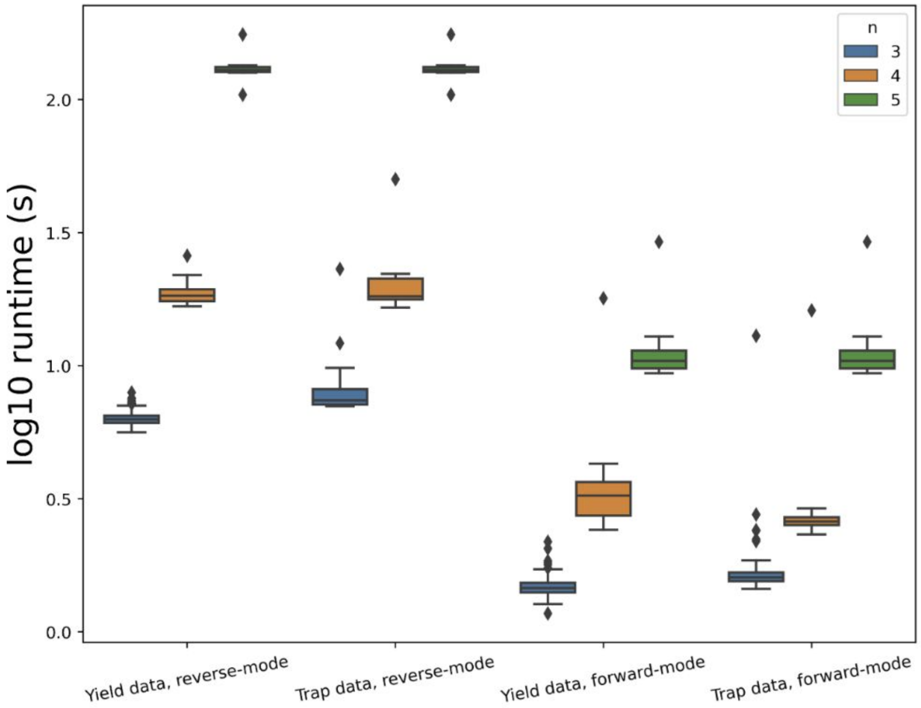
ForwardDiff.jl

What information does kinetic trapping encode about rate constants?

Can autodiff help us to extract that data machine learning?

Forward-mode is faster, yield data provides more accuracy





Next steps

1. Perform non-autodiff control experiments
2. Examine memory consumption of forward- and reverse-mode autodiff
3. Examine more realistic chemical reaction networks and data

Acknowledgements

Johnson lab

Dr. Margaret Johnson

Adip Jhaveri

Jonathan Fischer

Dr. Sam Foley

Spencer Loggia (undergrad. alum.)

Mankun Sang

Moon Ying

Dr. Sikao Guo

JHU symBIOsis

Error regression

```
Call:
lm(formula = error ~ autodiff_mode + exp_type + n + autodiff_mode:exp_type,
    data = df)

Residuals:
   Min     1Q Median     3Q    Max
-30533 -13018    848   2682  99715

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -10004       4709  -2.124   0.0341 *
autodiff_moderev                 -2286       3055  -0.748   0.4546
exp_typeyield                   -17008       2237  -7.603 1.45e-13 ***
n                                 8115       1273   6.377 4.14e-10 ***
autodiff_moderev:exp_typeyield    4192       3748   1.118   0.2639
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 17330 on 498 degrees of freedom
Multiple R-squared:  0.2557,	Adjusted R-squared:  0.2497
F-statistic: 42.77 on 4 and 498 DF,  p-value: < 2.2e-16
```

Time regression

```
Call:
lm(formula = time ~ autodiff_mode + exp_type + n + autodiff_mode:exp_type,
    data = df)

Residuals:
    Min      1Q  Median      3Q     Max
-31.729 -11.304  -3.051   7.536 118.899

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                     -80.031      4.908 -16.305  < 2e-16 ***
autodiff_moderev                 32.209      3.184  10.115  < 2e-16 ***
exp_typeyield                     3.320      2.332   1.424    0.155
n                                23.560      1.326  17.762  < 2e-16 ***
autodiff_moderev:exp_typeyield -16.696      3.906  -4.274  2.3e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 18.07 on 498 degrees of freedom
Multiple R-squared:  0.4973,    Adjusted R-squared:  0.4933
F-statistic: 123.2 on 4 and 498 DF,  p-value: < 2.2e-16
```