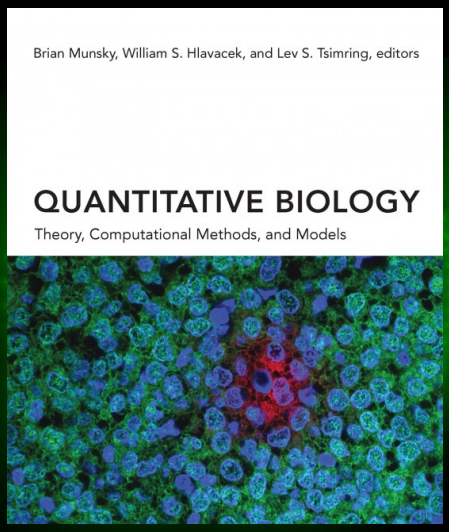
Designing Optimal Microscopy Experiments to Harvest Single-Cell Fluctuation Information while Rejecting Image Distortion Effects

Brian Munsky

Associate Professor,
Chemical and Biological Engineering
School of Biomedical Engineering
Colorado State University

June 29, 2021

munsky@colostate.edu

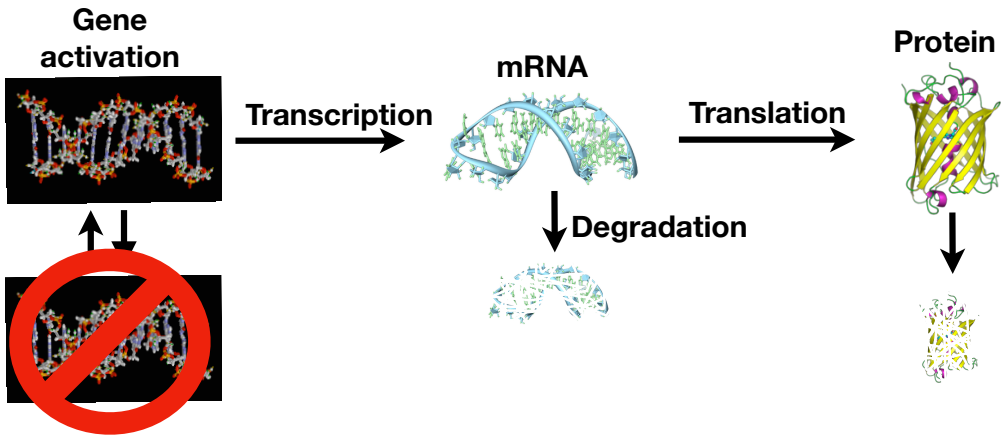


NIGMS
W.M. KECK FOUNDATION
NSF CAREER Award

Outline

1. **Introduction — the origin and importance of single-cell noise.**
2. Motivation — progress toward quantitative measuring and modeling every stage of the central dogma of molecular biology and at single-molecule resolution.
3. Key challenges:
   - ✳ optimal integration of single-cell experiments and stochastic computational models
   - ✳ estimating and reducing uncertainty in stochastic gene regulation models.

Introduction - The Central Dogma of Molecular Biology



Genetically identical cells in identical environments produce **stochastic, spatial, temporal fluctuations**.

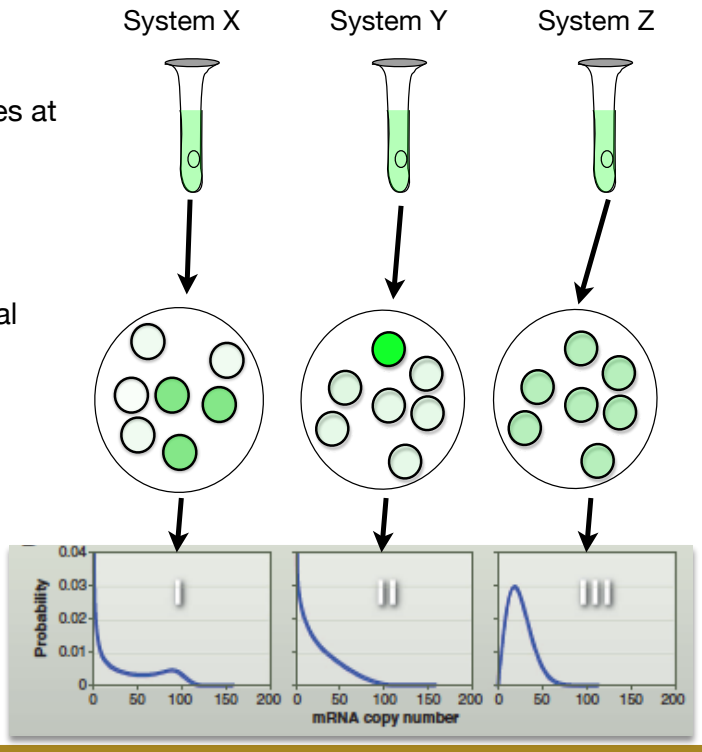
It is possible to **measure** and **predict every** stage of these fluctuations.

Information in fluctuation

Different systems (species, inputs, mechanisms, …) may express genes at equal *average* levels.

Single-cell measurements may reveal hidden response differences.

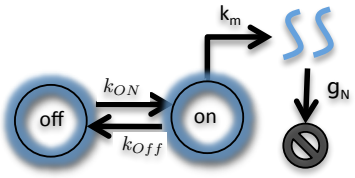
Collective responses can exhibit distinctive “fluctuation fingerprints”.



Munsky, et al, *Science*, 2012

Information in fluctuation

- Consider a model of bursting gene expression:



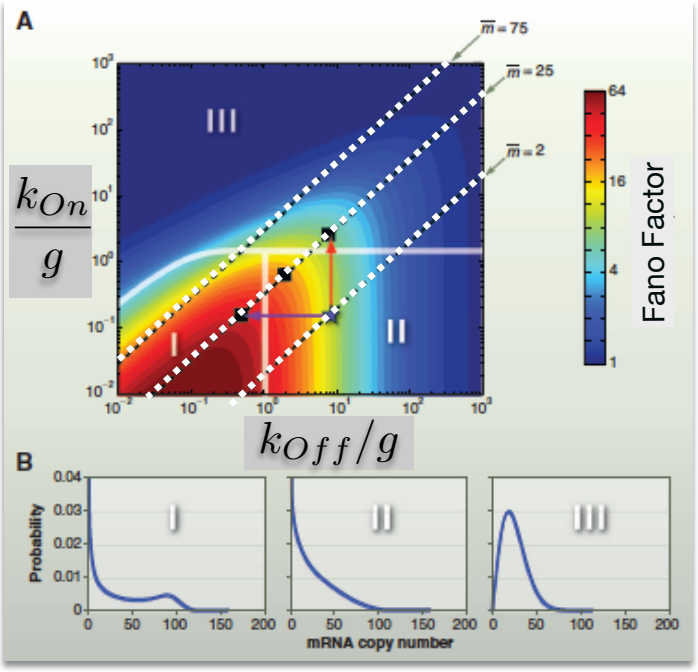
- We can compute the expression mean and variability as functions of all parameters.

$$f_{on} = \frac{k_{ON}}{k_{ON} + k_{OFF}}$$

$$\mu = f_{on}\frac{k_m}{g_m}$$

$$\frac{\sigma^2}{\mu} = 1 + \frac{(1 - f_{on})\, k_m}{k_{ON} + k_{OFF} + g_m}$$

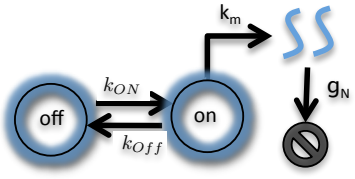
Expression 'Noise' versus parameters



Munsky, et al, *Science*, 2012

Information in fluctuation

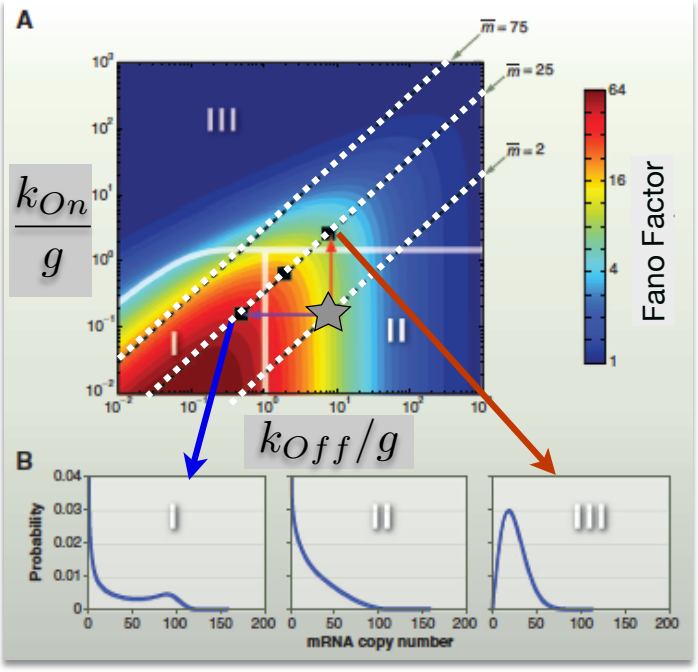
- Consider a model of bursting gene expression:



- We can compute the expression mean and variability as functions of all parameters.

- Tuning k_{Off} or k_{On} can increase expression, but in doing so:
- **Tuning k_{Off} increases variability.**
- **Tuning k_{On} decreases variability.**

Expression 'Noise' versus parameters

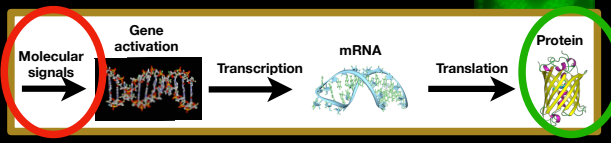


Munsky, et al, *Science*, 2012

Outline

1. Introduction — the origin and importance of single-cell noise.
2. **Motivation — progress toward quantitative measuring and modeling every stage of the central dogma of molecular biology and at single-molecule resolution.**
3. Key challenges:
   - optimal integration of single-cell experiments and stochastic computational models
   - estimating and reducing uncertainty in stochastic gene regulation models.

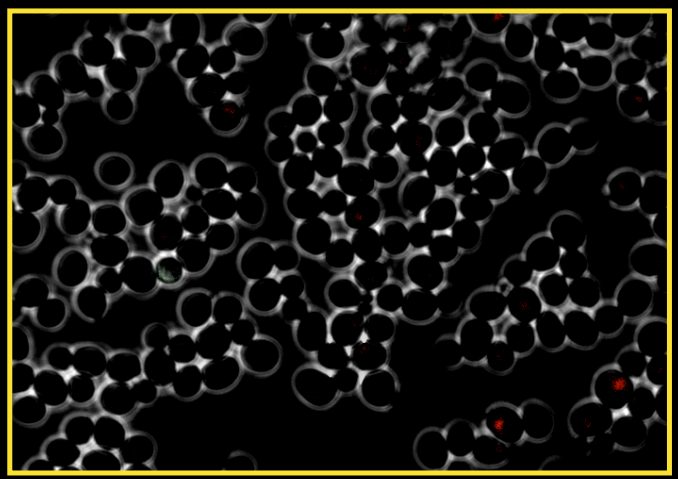
Fluorescent Proteins to Measure Signaling and Responses



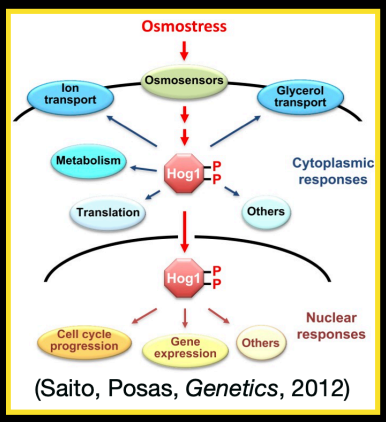
Time lapse fluorescence microscopy measures temporal properties of:

SIGNALS (in this case a mitogen activated protein kinase, MAPK).

RESPONSES (in this case STL1-GFP).



Different perturbations yield different MAPK signals and different downstream responses.



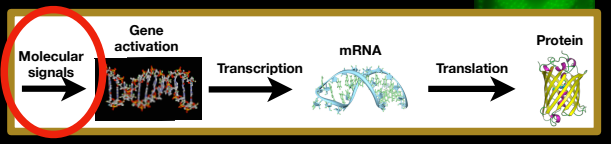
(Saito, Posas, *Genetics*, 2012)

Gregor Neuert
Vanderbilt

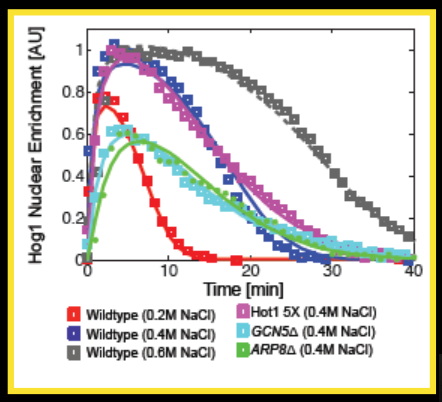
Hossein Jahnsaz
Vanderbilt

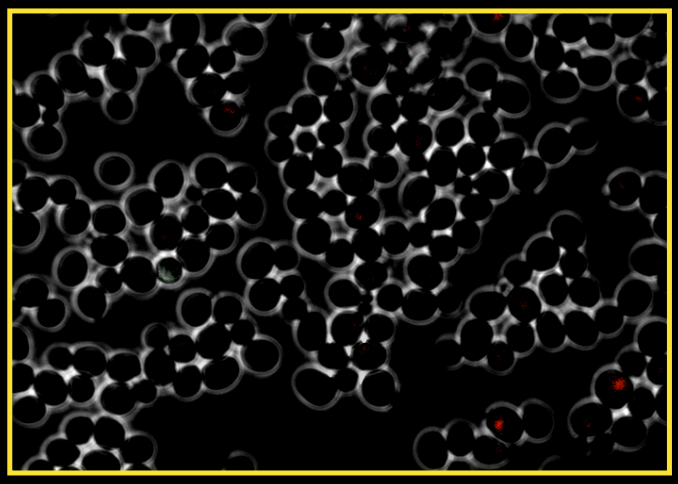
Neuert, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018
Jahnsaz, et al, *iScience*, 2020
Jahnsaz, et al, *Star Protocols*, 2021

Fluorescent Proteins to Measure Signaling and Responses



ODE Models can be parametrized to capture these MAPK dynamics as functions of time and environmental conditions.



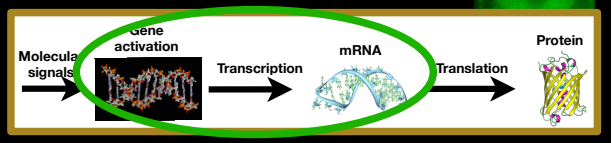


Gregor Neuert
Vanderbilt

Hossein Jahnsaz
Vanderbilt

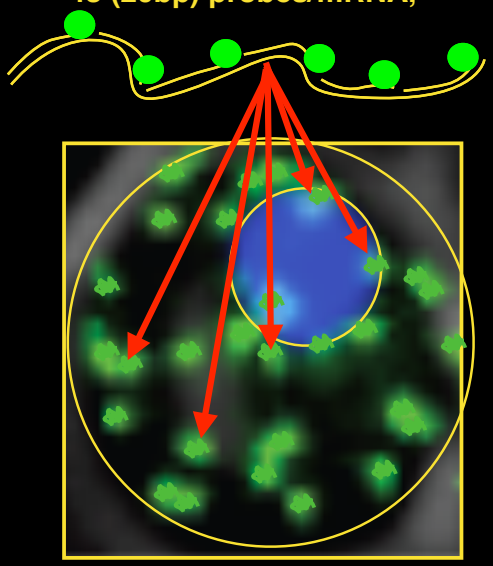
Neuert, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018
Jahnsaz, et al, *iScience*, 2020
Jahnsaz, et al, *Star Protocols*, 2021

Single-Molecule Fluorescence in situ Hybridization (smFISH)



- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,

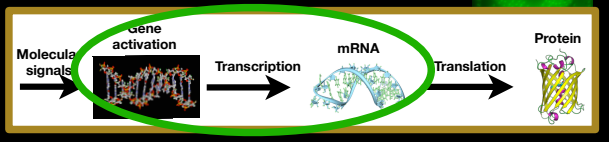
48 (20bp) probes/mRNA,



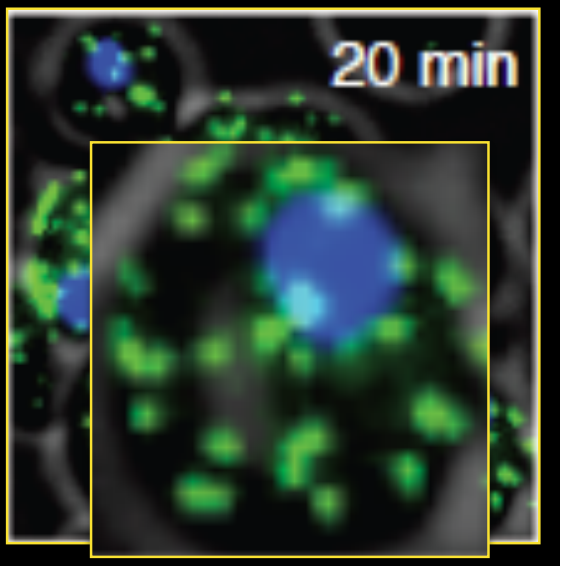
Gregor Neuert,
Vanderbilt

Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

Single-Molecule Fluorescence in situ Hybridization (smFISH)



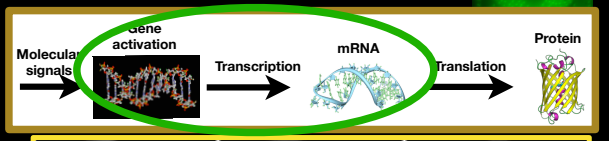
- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,



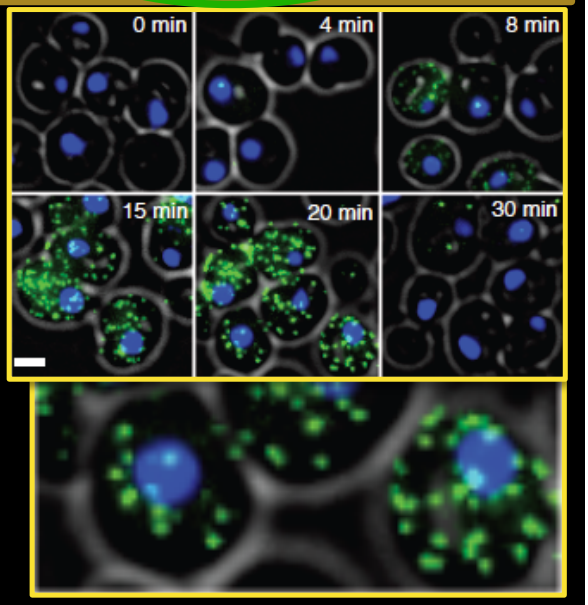
Gregor Neuert, Vanderbilt

Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

Single-Molecule Fluorescence in situ Hybridization (smFISH)



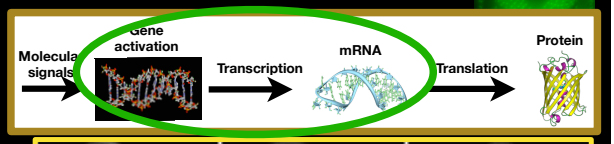
- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,
  - Fast (1-2 minute) time resolution,
  - 100s or 1000s of cells per time point or condition.



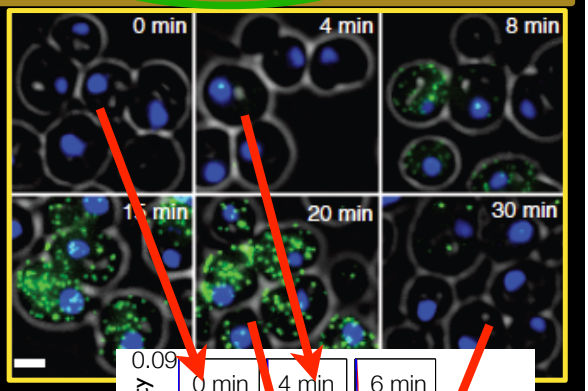
Gregor Neuert, Vanderbilt

Neuert, Munsky, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018

Single-Molecule Fluorescence in situ Hybridization (smFISH)

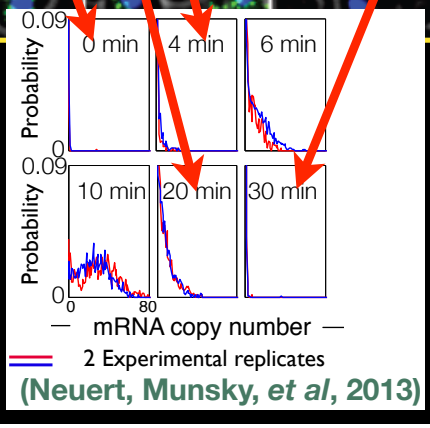


- SM-FISH allows quantification of endogenous transcription response:
  - **Number** of *individual* mRNA per cell,
  - **3D Location** of *individual* mRNA,
  - **DNA transcription site** activity,
  - Fast (1-2 minute) time resolution,
  - 100s or 1000s of cells per time point or condition.



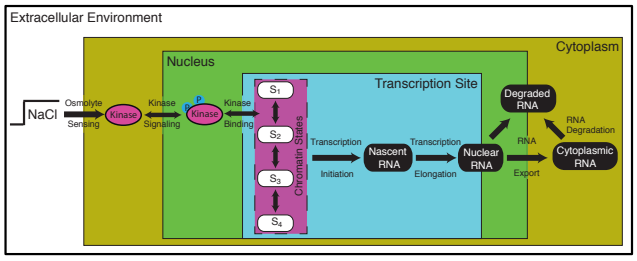
Gregor Neuert, Vanderbilt

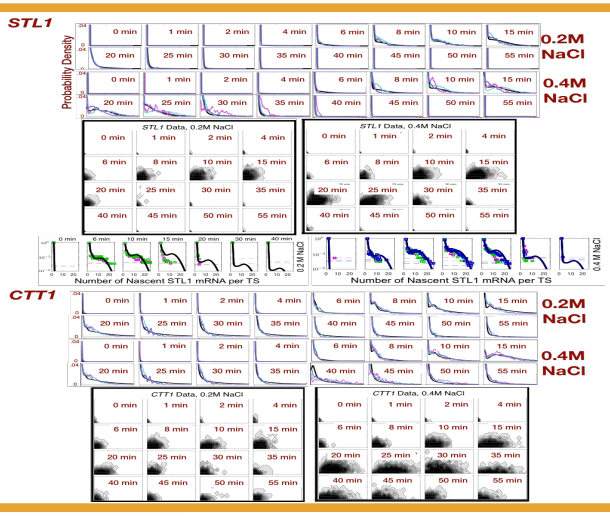
smFISH yields highly **reproducible** & **quantitative** measurements of (noisy) single-cell responses.



(Neuert, Munsky, *et al*, 2013)

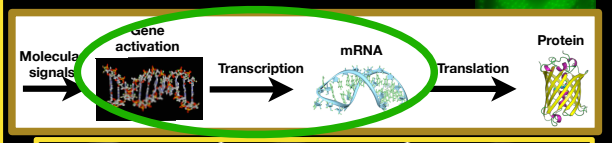
The HOG1-activated distributions of STL1 and CTT1 are well fit and well *predicted* by a 4-state bursting gene expression model with spatial compartments.

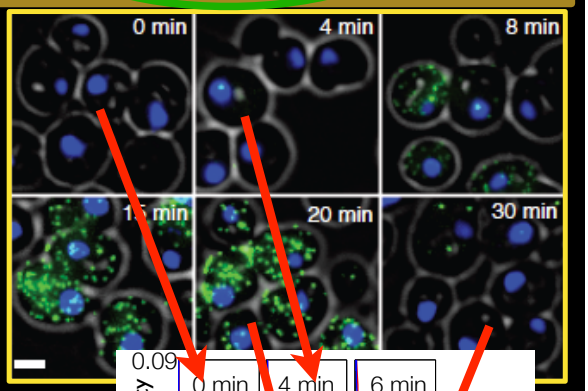


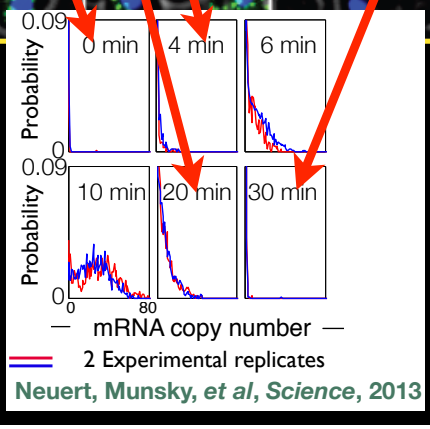


Munsky, *et al*, *PNAS*, 2018

situ Hybridization (smFISH)

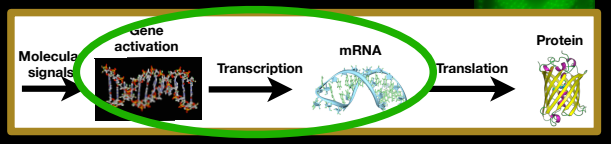




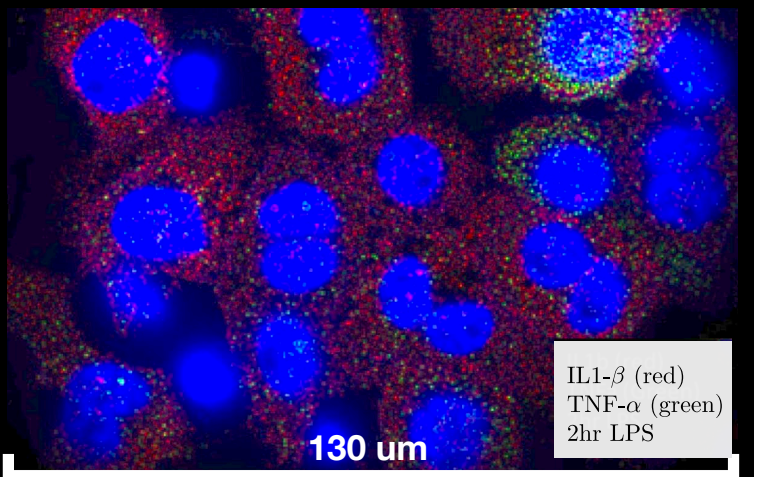


Neuert, Munsky, *et al*, *Science*, 2013

Single-Molecule Fluorescence in situ Hybridization (smFISH)



- smFISH is used to label *individual* mRNA in cells.
- For example, here we examine THP1 cells two hours after induction by bacterial LPS to simulate infection.
- We are interested in the response of two cytokines:
  - red spots — IL1b
  - green spots — TNFa



IL1-β (red)
TNF-α (green)
2hr LPS

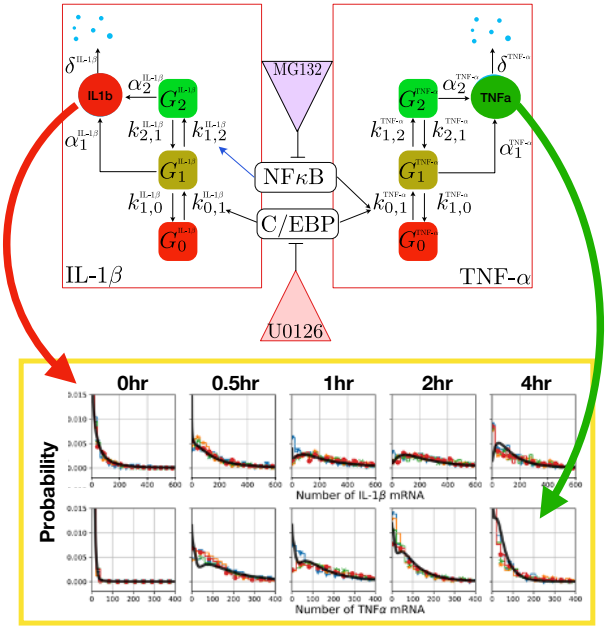

Daniel Kalb LANL


Huy Vo

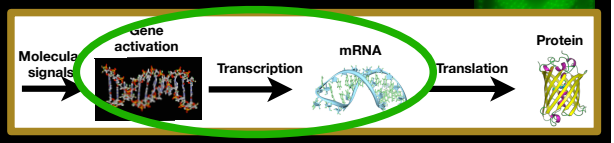

Jim Werner LANL

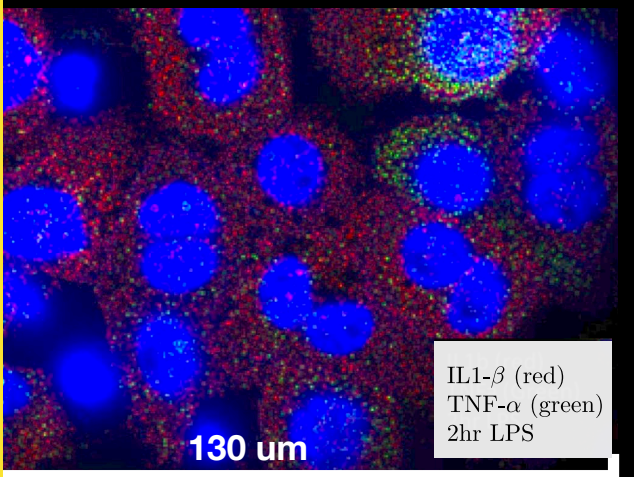
Neuert, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018
Kalb, Vo et al, *Scientific Reports,* 2021

The LPS-activated distributions of IL1b and TNFa are well fit and well *predicted* by an integrated pair of two 3-state bursting gene expression models.



Hybridization (smFISH)





IL1-β (red)
TNF-α (green)
2hr LPS

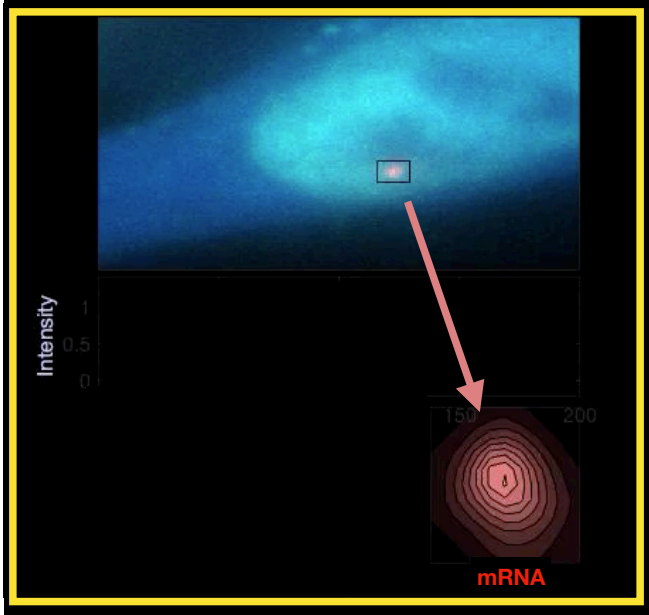

Daniel Kalb LANL

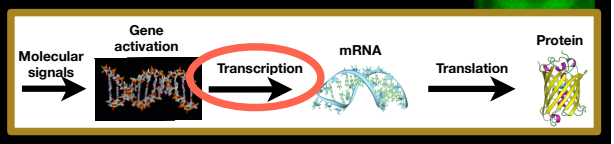

Huy Vo


Jim Werner LANL

Neuert, et al, *Science* 2013
Munsky, et al, *PNAS*, 2018
Kalb, Vo et al, *Scientific Reports,* 2021

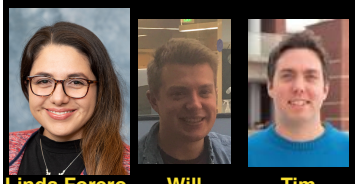
MS2/Fab for Live-cell Nascent TRANSCRIPTION Tracking



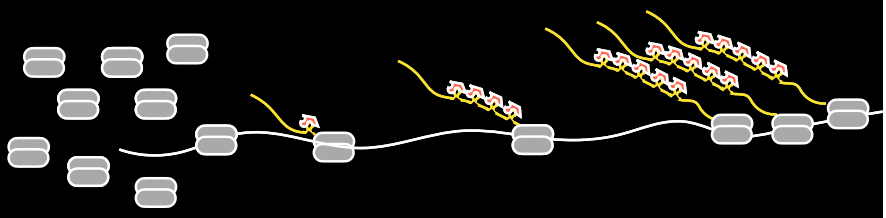


- Using MS2/MCP labeling, we observe *live nascent RNA transcription*.

MCP Label

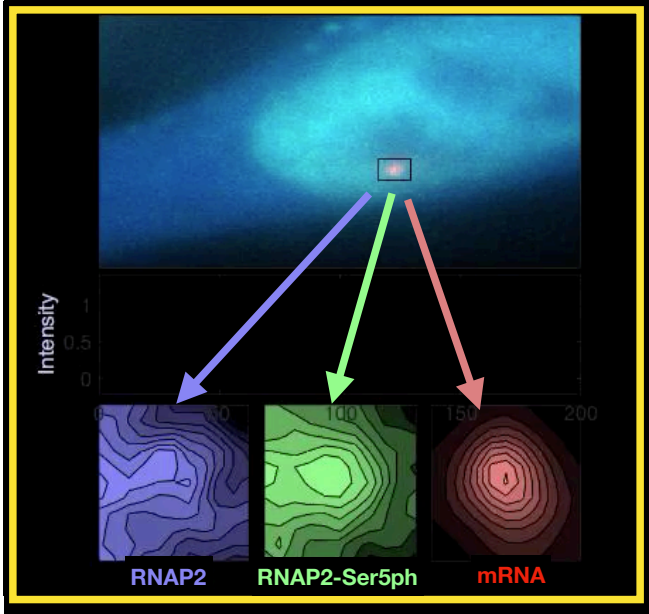


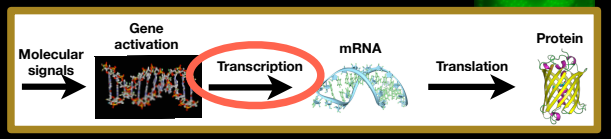
Linda Forero-Quintero | Will Raymond | Tim Stasevich



Forero, Raymond et al, ***Nat. Comms.***, 2021

MS2/Fab for Live-cell Nascent TRANSCRIPTION Tracking



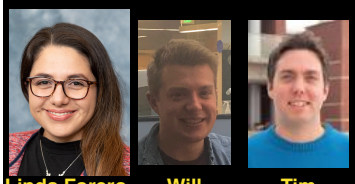


- Using MS2/MCP labeling, we observe *live nascent RNA transcription*.
- Fragmented antibody (Fab) probes allow us to quantify RNA Polymerase II (RNAP2) before (green) and after (green+blue) Ser5 phosphorylation.

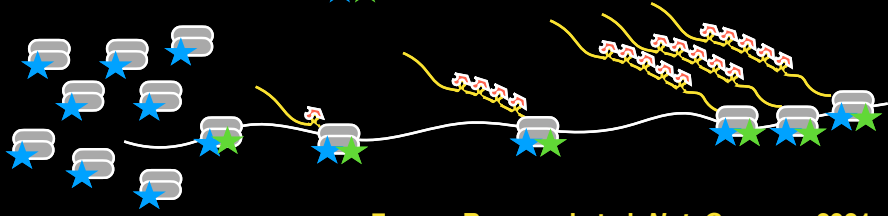
MCP Label

RNAP2-CTD

RNAP2-CTD-Ser5ph



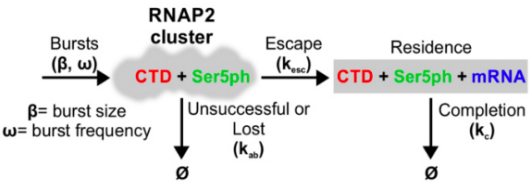
Linda Forero-Quintero | Will Raymond | Tim Stasevich



Forero, Raymond et al, ***Nat. Comms.***, 2021

MS2/Fab for Live-cell Nascent TRANSCRIPTION Tracking

The distributions and temporal correlations of RNAP2 localization, RNAP2 phosphorylation, and nascent transcription are well captured with a 2-state Bursting Transcription Model.

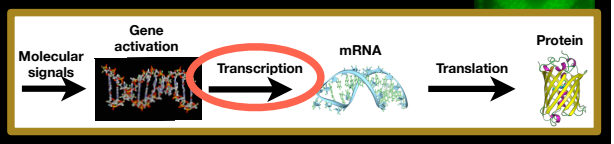


Parameter	Fit
β	15.405
ω (min^{-1})	0.434
k_{ab} (min^{-1})	0.778
k_{esc} (min^{-1})	0.666
k_c (min^{-1})	0.199

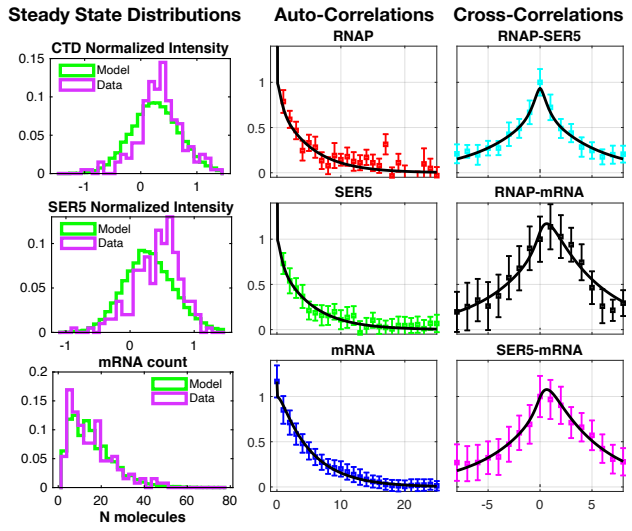

Linda Forero-Quintero


Will Raymond


Tim Stasevich

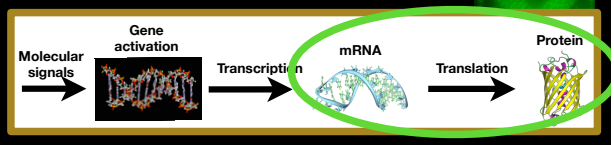


Bursting gene expression models capture the stationary distributions and correlation dynamics of the RNAP2 and transcription dynamics.

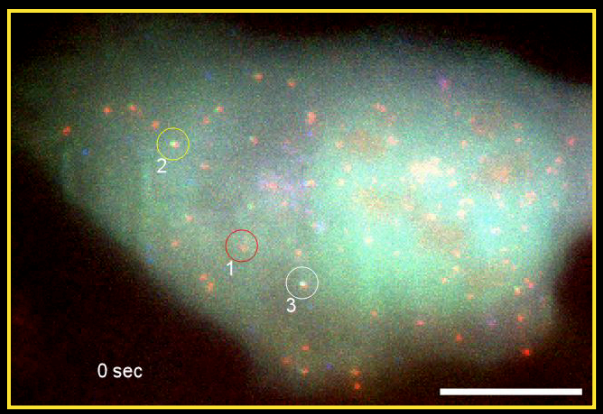


Forero, Raymond et al, *Nat. Comms.*, 2021

MS2/Fab for Live-cell Nascent TRANSLATION Tracking



- Fabs can also be used to quantify *Nascent Protein translation* in living cells.
- Different colors can be used to observe different open reading frames or different ribosomal entry sites.



MCP Label
MS2 Hairpin
FAB Labels
SM Peptide


Kenneth Lyon


Luis Aguilera


Amanda Koch

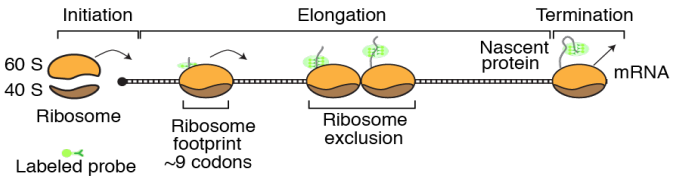

Tim Stasevich


Tatsuya Morisaki

Lyon, Aguilera, et al, *Molecular Cell*, 2019
Aguilera, Raymond, et al, *PLoS Comp Biol*, 2019
Koch, Aguilera et al, *Nat. Struct, Mol. Biol.*, 2020

MS2/Fab for Live-cell Nascent TRANSLATION Tracking

Nascent protein dynamics are captured by a **Totally Asymmetric Simple Exclusion Process** (TASEP) model.



The TASEP Model has two parameters:

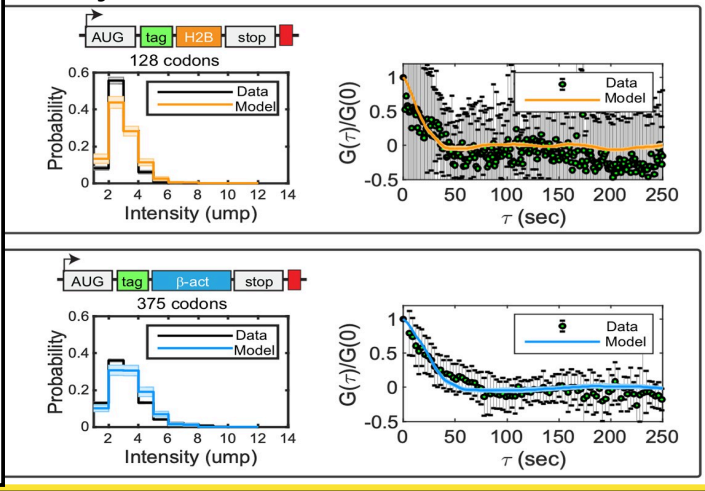
- initiation rate
- average elongation rate*

With just these two parameters, the model captures:

- the distribution of nascent proteins per mRNA in units of mature protein.
- the auto-covariance of the protein translation signal.

**Codon-dependent translation rates are defined by the Codon Adaptation Index.*

Steady-State Distributions **Auto-Covariances**

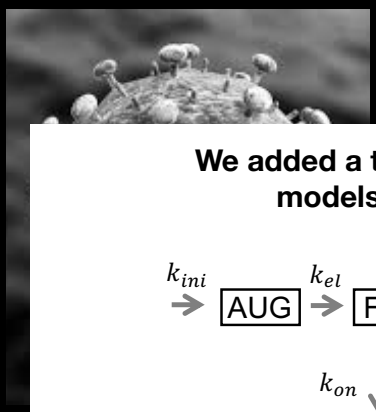



Luis Aguilera


Will Raymond

Lyon, Aguilera, et al, *Molecular Cell*, 2019
Aguilera, Raymond, et al, *PLoS Comp Biol*, 2019
Koch, Aguilera et al, *Nat. Struct, Mol. Biol.*, 2020

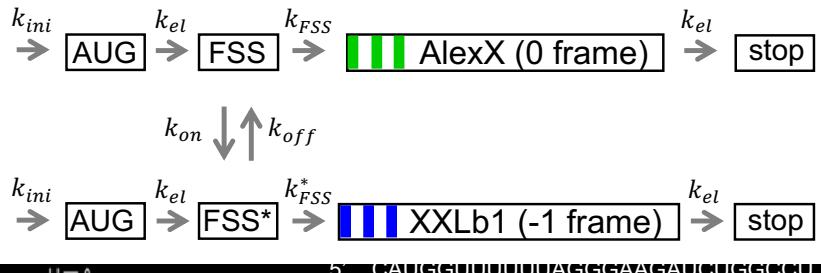
Example 2: Viral Frame-Shifting



Two for the price of one!

HIV and other viruses use frameshift stimulatory

We added a third color in the -1 frame and extended models to allow bursts of frame-shifting.

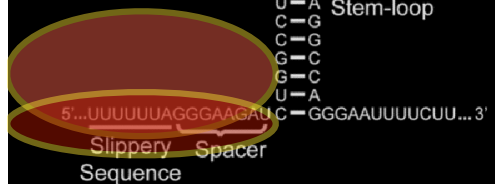


te two

HIV-1

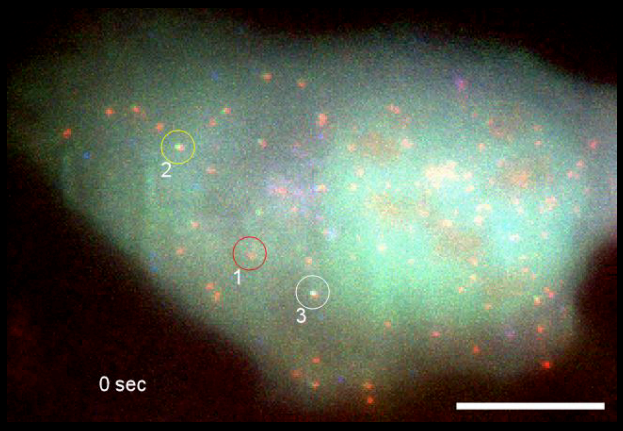
5'... CAUGGUUUUUUAGGGAAGAUCUGGCCU .... 3'

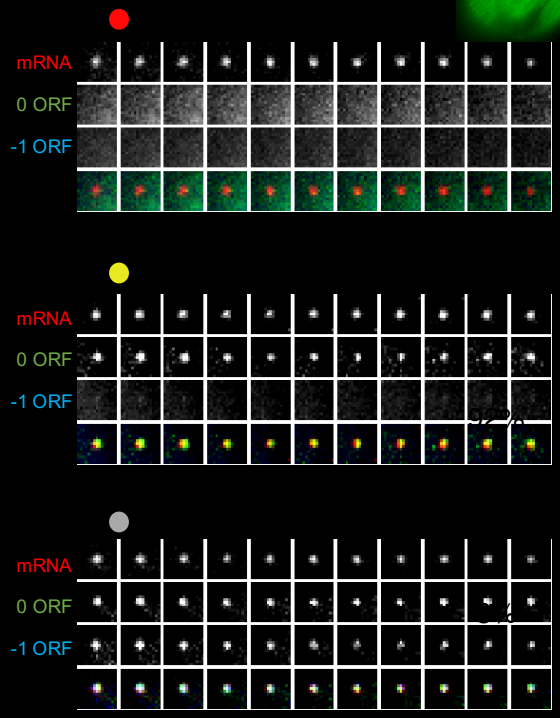
Original Protein



Lyon, Aguilera, et al, *Molecular Cell*, 2019

Watching Frame-Shifting in single-molecule resolution






Kenneth Lyon

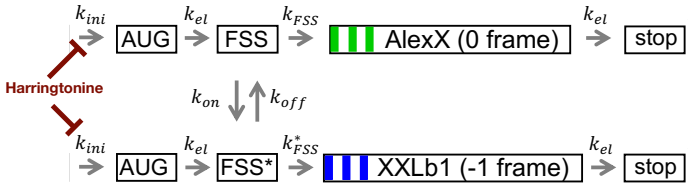

Luis Aguilera

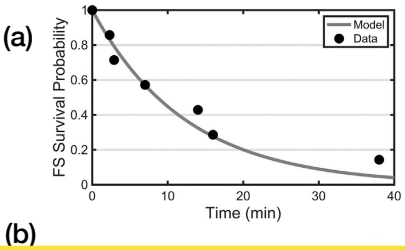

Tim Stasevich


Tatsuya Morisaki

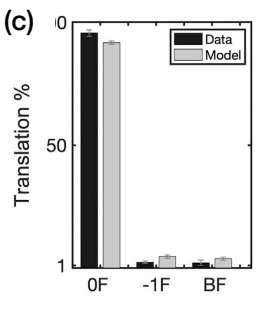
Lyon, Aguilera, et al, *Molecular Cell,* 2019

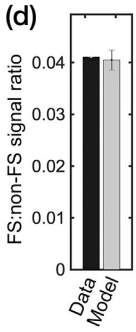
Modeling of Live-Cell Frame-Shifting

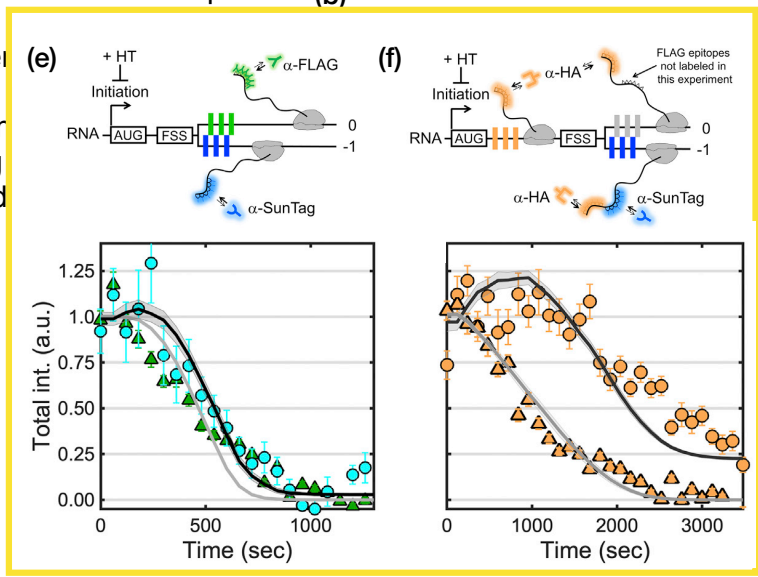




A single **bursting frame-shift TASEP model** captures:
(a) survival time of -1 bursts
(b) the loading of ribosomes per
(c) fraction of mRNA with 0, -1,
(d) Total production ratio of 0 ar
(e) Run-off dynamics after drug
(f) Run-off dynamics for extend



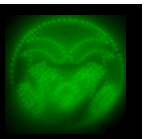




Lyon, Aguilera, et al, *Molecular Cell,* 2019

Outline

1. Introduction — the origin and importance of single-cell noise.
2. **Motivation — progress toward quantitative measuring and modeling every stage of the central dogma of molecular biology and at single-molecule resolution.**
3. **Key challenges:**
   - **optimal integration of single-cell experiments and stochastic computational models**
     - ✳ estimating and reducing uncertainty in stochastic gene regulation models.



✳ **These single-cell experiments are powerful yet expensive**, and there are an vast number of possible experiment designs

Experiment Design Considerations
- **Number of cells**
- **Sampling period**
- **Choice of fluorophore(s)**
- **Number and placement of probes**
- **Inducer/drug concentrations and delivery times**

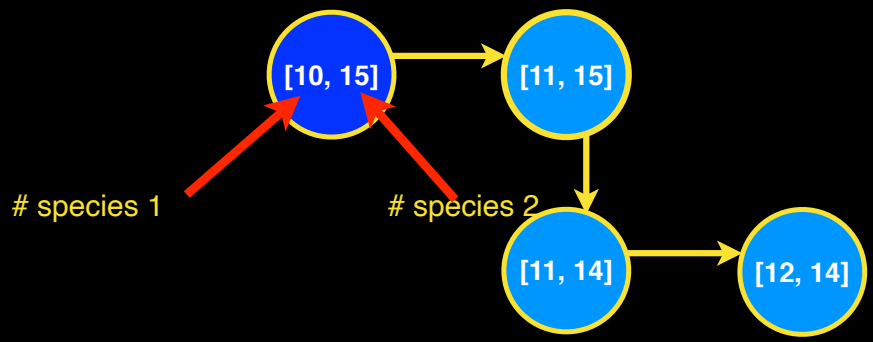
Measurement Error Considerations
- **Microscope resolution**
- **Image processing errors (segmentation, spot detection, track linking)**
- **Photobleaching**
- **Autofluorescence**
- **Delays due to drug diffusion and nuclear import**

✳ We want to get as much insight as possible out of each experiment.

✳ We want to choose experiments that minimize uncertainty about the mechanisms or parameters of interest.

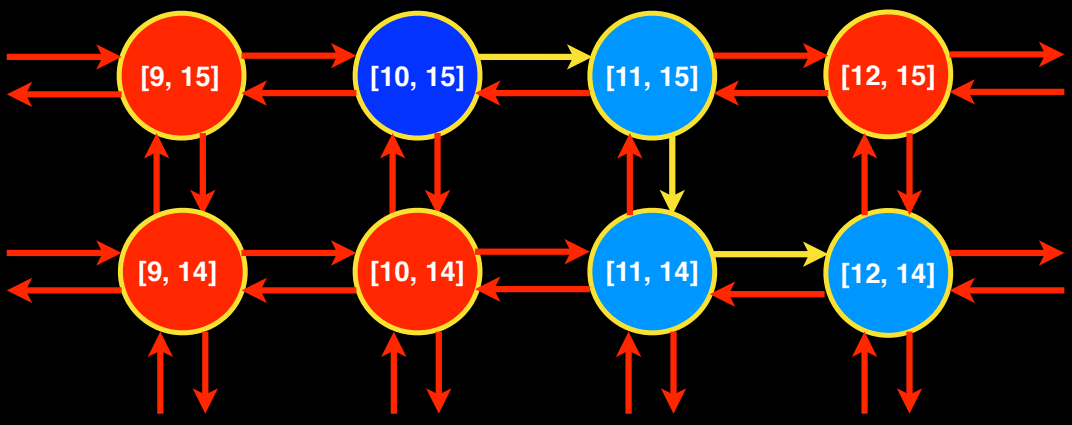
The Markov description of gene expression

- At any time, the state of the system is defined by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another.

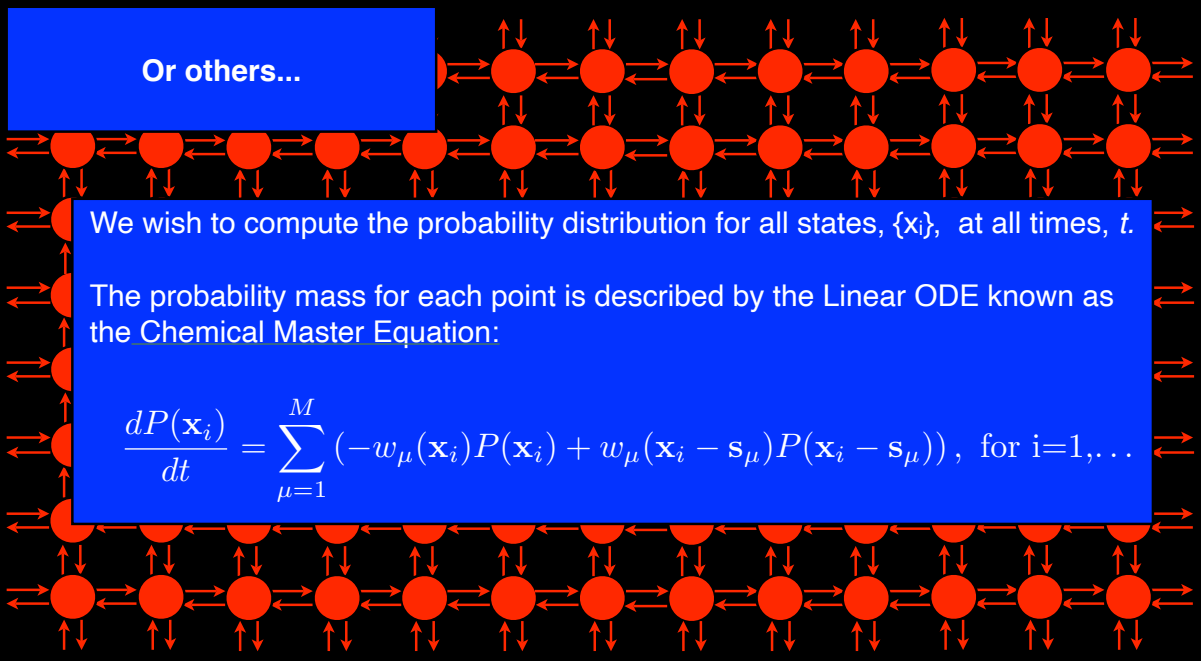


The Markov description of gene expression

- At any time, the state of the system is defined by its integer population vector: $\mathbf{x} \in \mathbb{Z}^N$
- Reactions are transitions from one state to another.
- These reactions are random, others could have occurred:



The Markov description of gene expression



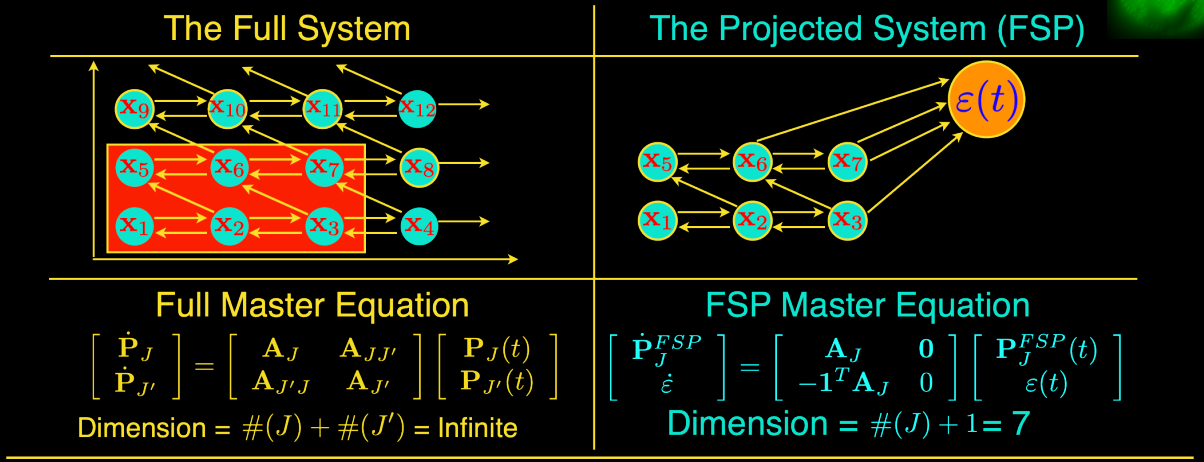
Or others...

We wish to compute the probability distribution for all states, {x_i}, at all times, *t*.

The probability mass for each point is described by the Linear ODE known as the Chemical Master Equation:

$$\frac{dP(\mathbf{x}_i)}{dt} = \sum_{\mu=1}^{M} \left(-w_\mu(\mathbf{x}_i)P(\mathbf{x}_i) + w_\mu(\mathbf{x}_i - \mathbf{s}_\mu)P(\mathbf{x}_i - \mathbf{s}_\mu)\right), \text{ for i=1,...}$$

The finite state projection approach



1) **Strict lower bound** on solution: $\mathbf{P}_J(t) \geq \mathbf{P}_J^{FSP}(t)$

2) **Exact error** of solution: $\left\| \begin{bmatrix} \mathbf{P}_J(t) \\ \mathbf{P}_{J'} \end{bmatrix} - \begin{bmatrix} \mathbf{P}_J^{FSP}(t) \\ \mathbf{0} \end{bmatrix} \right\|_1 = \varepsilon_J(t)$

3) **Monotonic** convergence: $\varepsilon_{J_1}(t) \geq \varepsilon_{J_2}(t)$ for any $J_1 \subseteq J_2$

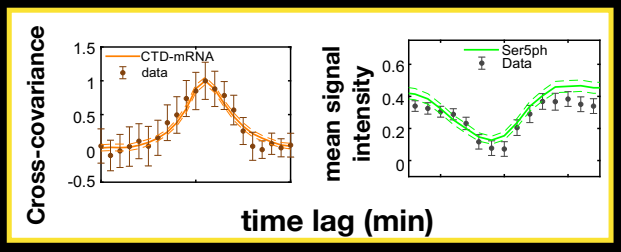
Munsky et al, *JCP* 2006

Inferring parameters from single-cell measurements

Consider some arbitrary set of independent data $\mathcal{D} = [d_1, d_2, \ldots]^{\mathrm{T}}$ and a hypothetical probability distribution to explain those data $p(d; \Lambda)$.

The **LIKELIHOOD** of independent cells' data given our model can be written:

$$L(\mathcal{D}; \Lambda) = \prod_j p(d_j; \Lambda) \qquad \text{or} \qquad \log L(\mathcal{D}; \Lambda) = \sum_j \log p(d_j; \Lambda)$$



For Gaussian distributions with mean μ_k and variance σ^2 (e.g., SEM):

$$p(d_k; \mu_k) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\frac{(-d_k - \mu_k)^2}{2\sigma^2}} \quad \Rightarrow \quad \log L(\mathcal{D}; \Lambda) = C - \frac{1}{2\sigma^2} \sum_k (d_k - \mu_k)^2$$

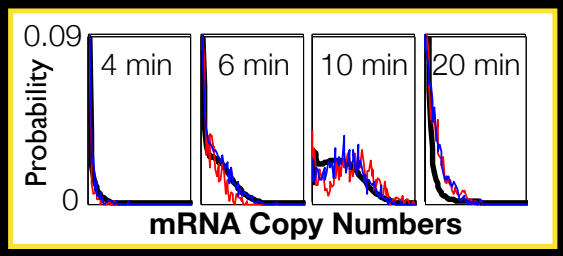
When noise is independent with constant variance, the maximum likelihood estimate (MLE) is the minimum sum of square error (SSE) estimate.

Inferring parameters from single-cell measurements

Consider some arbitrary set of independent data $\mathcal{D} = [d_1, d_2, \ldots]^{\mathrm{T}}$ and a hypothetical probability distribution to explain those data $p(d; \Lambda)$.

The **LIKELIHOOD** of independent cells' data given our model can be written:

$$L(\mathcal{D}; \Lambda) = \prod_j p(d_j; \Lambda) \qquad \text{or} \qquad \log L(\mathcal{D}; \Lambda) = \sum_j \log p(d_j; \Lambda)$$



The FSP provides computable upper and lower bounds on the likelihood of single-cell data given a stochastic model:

$$\sum d_j \log P_j^{FSP}(\Lambda) \leq \log L(D|\Lambda) \leq \max_{\substack{|\mathbf{f}|_1 = 1 \\ f_j \geq 0}} \left(\sum d_j \left(\log P_j^{FSP}(\Lambda) + f_j \varepsilon \right) \right)$$

Zach Fox

Fox et al, *JCP* 2016

Outline

1. Introduction — the origin and importance of single-cell noise.
2. **Motivation — progress toward quantitative measuring and modeling every stage of the central dogma of molecular biology and at single-molecule resolution.**
3. **Key challenges:**
   - optimal integration of single-cell experiments and stochastic computational models
   - **estimating and reducing uncertainty in stochastic gene regulation models.**

Error, Uncertainty (and Bias)

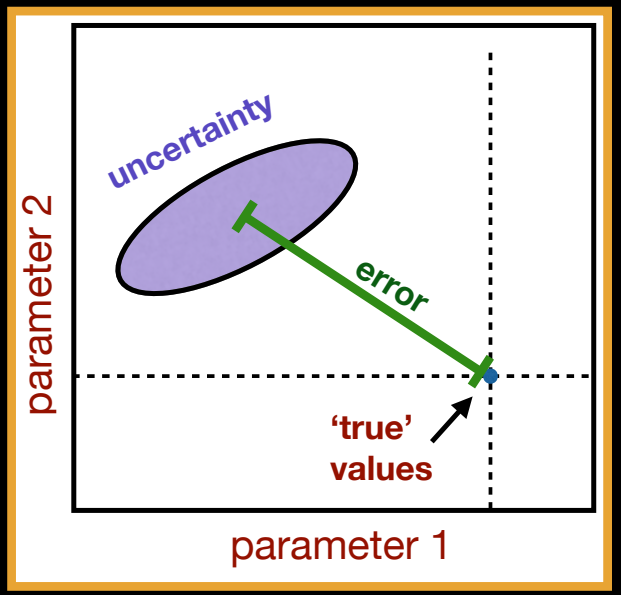
Models that cannot fit data can be invalidated and discarded.

Models that fit to data will still contain uncertainty and errors.

Error is the distance between the best estimate and the actual parameter values.

Uncertainty is the (estimated) variability in the fit parameters for a given data set.

Bias is the average error after fitting to a large (infinite) number of independent experiments.

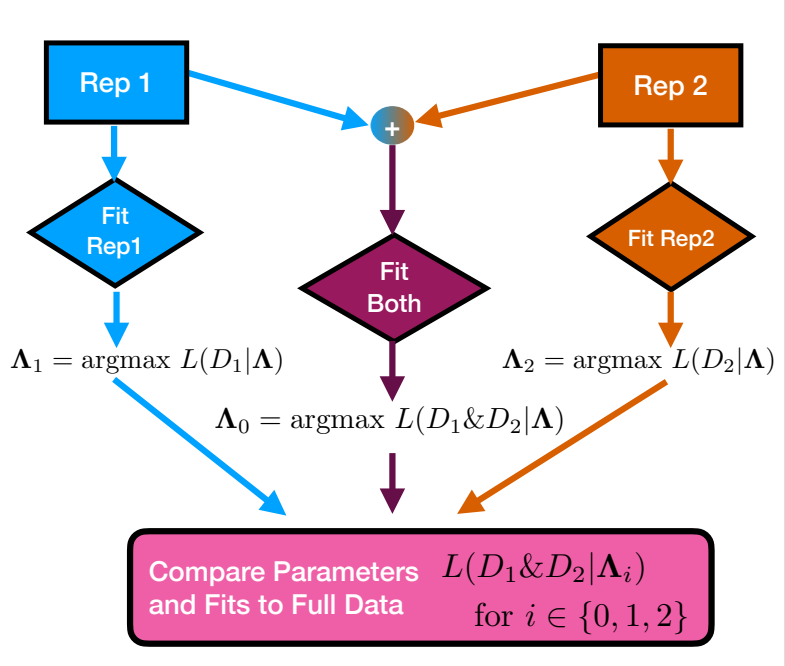


Error, Uncertainty, and Bias are all affected by choice of estimator!

Estimating Uncertainty from DATA (I. Cross-Validation)

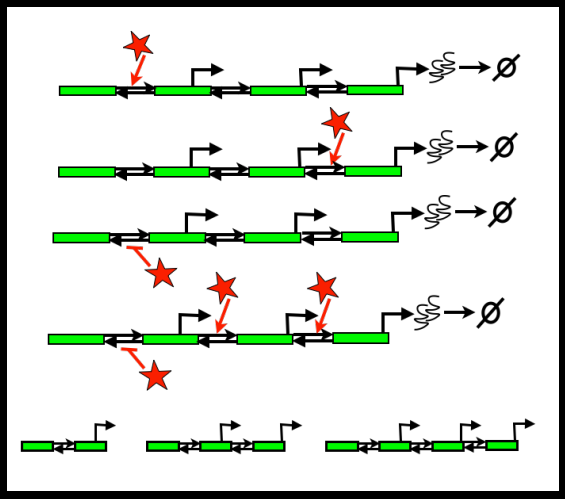
The (Frequentist) **Cross-Validation** approach to estimate model uncertainty:

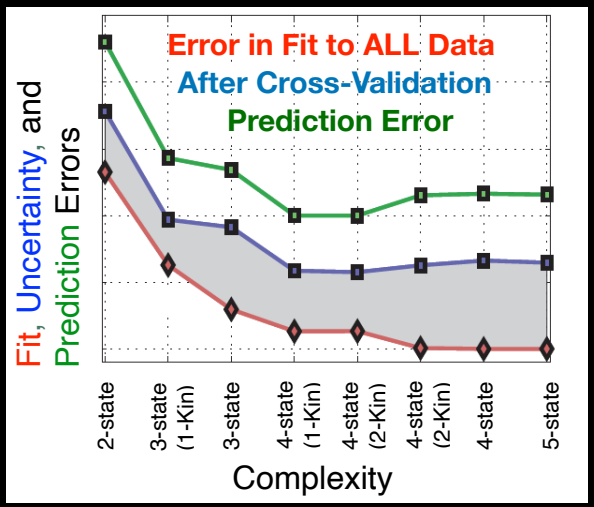
- Conduct multiple replica experiments.
- Fit replicas separately and together.
- Record best fit parameters for individual and lumped datasets.
- Compare parameter and prediction uncertainties.

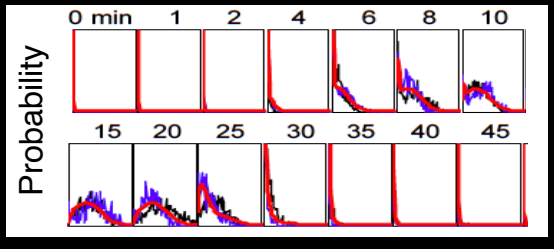


Cross-Validation to Select Model Structure

We performed model cross-validation on the Hog1 -> STL1 data for many different models.





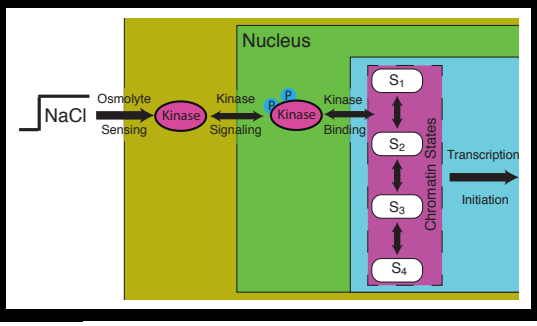


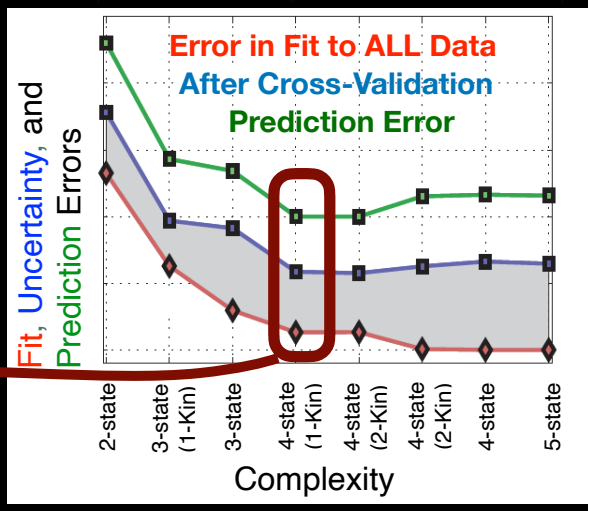
(Neuert, Munsky, et al, *Science* 2013)

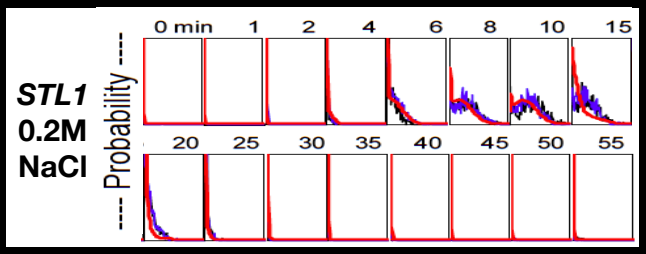
Using Cross-Validation to Select Model Structure

(Neuert, Munsky, et al, *Science* 2013)

We performed model cross-validation on the Hog1 -> STL1 data for many different models.







model predictions (no free parameters)

New experiments verify that we chose the **most predictive** model.

(Neuert, Munsky, et al, *Science* 2013)

Estimating Uncertainty from DATA (II. Bayesian MCMC)

Bayes' rule provides another way to estimate parameter uncertainty:

$$P(\Lambda|D) = \frac{P(D|\Lambda)}{P(D)} P(\Lambda)$$

The 'likelihood' (how the data matches the model)

The 'posterior' (what we want to estimate)

Normalization constant.

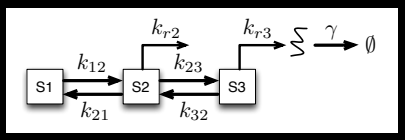
The 'prior' (what new thought we knew before the experiment)

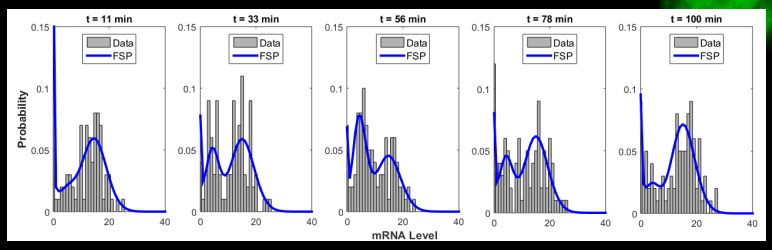
The likelihood $P(D;\Lambda)$ can be computed using known probability functions (e.g., Gaussian) or using the FSP.

We assume a convenient prior (e.g., log normal).

We use Markov Chain Monte Carlo (Gibbs, Metropolis Hastings, Hamiltonian, etc.) to sample the posterior.

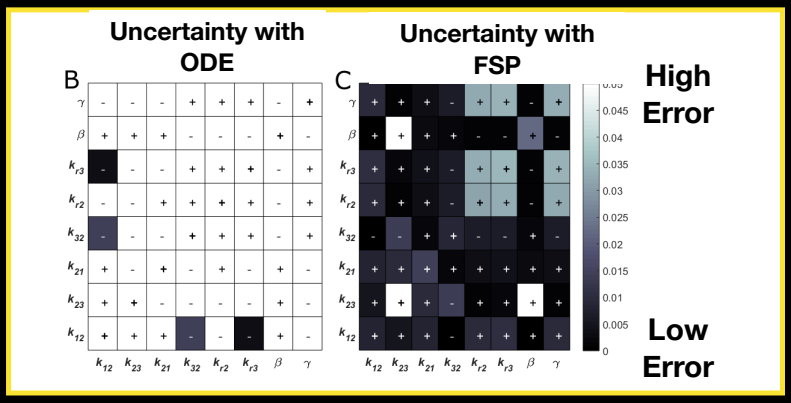
MCMC Uncertainty Quantification for a Bursting Gene Expression Model





We used MCMC quantify uncertainty when a 3-state bursting gene model (above left) was fit to simulated data (above right).

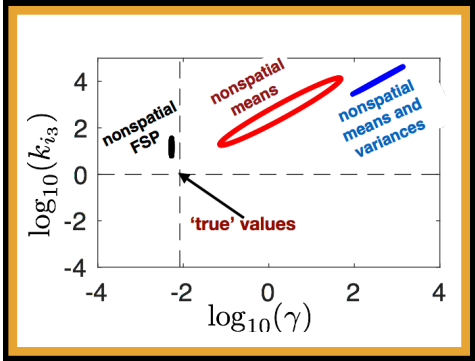
FSP-based likelihood functions resulted in parameter determination that was several orders of magnitude more precise that ODE analyses.

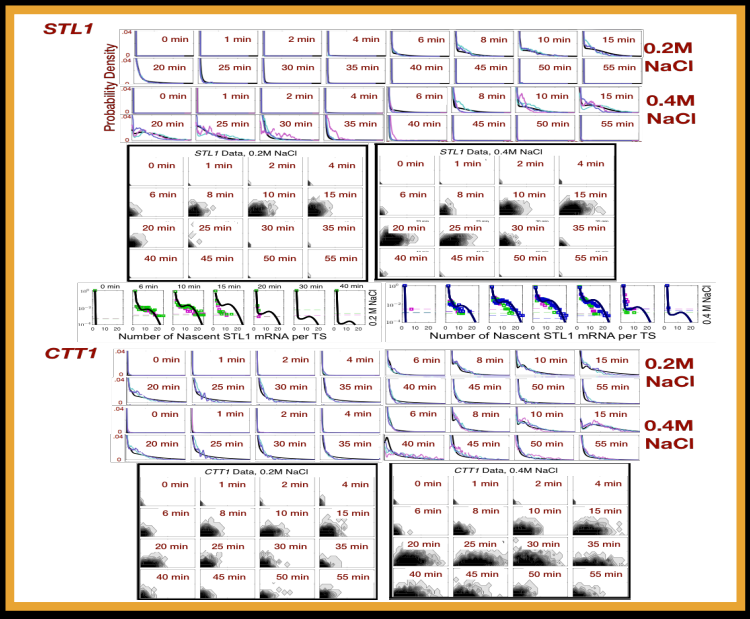


Lisa Weber

(Weber, et al, *Physical Biology,* 2018)

For real smFISH data, the Uncertainty Quantification advantage of the FSP is equally apparent.





Gregor Neuert, Vanderbilt

Zach Fox

Using FSP-MCMC uncertainty quantification, we could fit and predict the **spatio-temporal** dynamics of **nascent** and **mature** mRNA for multiple genes in **multiple conditions**.

Munsky, et al., PNAS, 2018

FSP-Based Fisher Information

The Fisher Information Matrix (FIM) quantifies the information that an observed random variable is expected to have about each model parameter:

$$\mathcal{I}(\theta) = \mathbb{E}_{\mathbf{D}}\left[\Big(\nabla_\theta \log L(\mathbf{D};\theta)\Big)^T \Big(\nabla_\theta \log L(\mathbf{D};\theta)\Big)\right]$$

Using the FSP, we can compute the distributions:

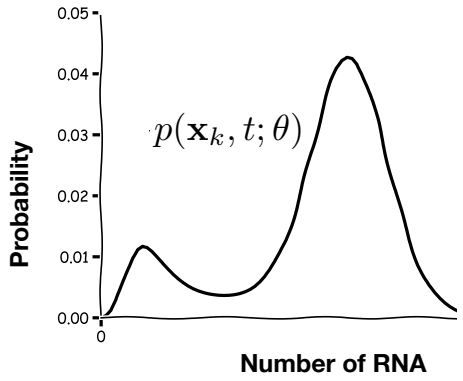
$$p(\mathbf{x}_k, t; \theta)$$

and the CME sensitivities:

$$\mathbf{s}(t)_i^k = \frac{\partial}{\partial \theta_i} p(\mathbf{x}_k, t; \theta)$$

From these, we can derive the FIM:

$$\mathcal{I}_{i,j} = n_{\text{Cells}} \sum_{k=1}^{N} \frac{1}{p(\mathbf{x}_k;\boldsymbol{\theta})} \mathbf{s}_i^k \mathbf{s}_j^k$$



Zach Fox

Fox et al, *PLoS Comp. Biol*, 2019
Fox et al, *Complexity*, 2020

Estimating Expected MLE Uncertainty using Fisher Information

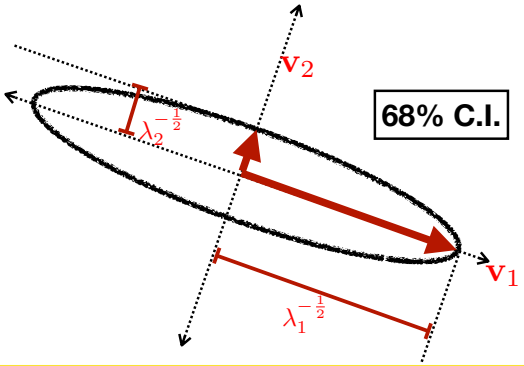
The FIM provides an asymptotic (multivariate Gaussian) estimate for the Maximum Likelihood Estimator.

Asymptotic normality of the MLE:

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{dist} \mathcal{N}(0, I(\boldsymbol{\theta}^*)^{-1})$$

The FIM's eigenvalues $\{\lambda_i\}$ and its eigenvectors $\{\mathbf{v}_i\}$ estimate the magnitudes and directions of uncertainty in MLE parameters (Cramer Rao Lower Bound).

Gaussian approximation of MLE estimate distribution



Zach Fox

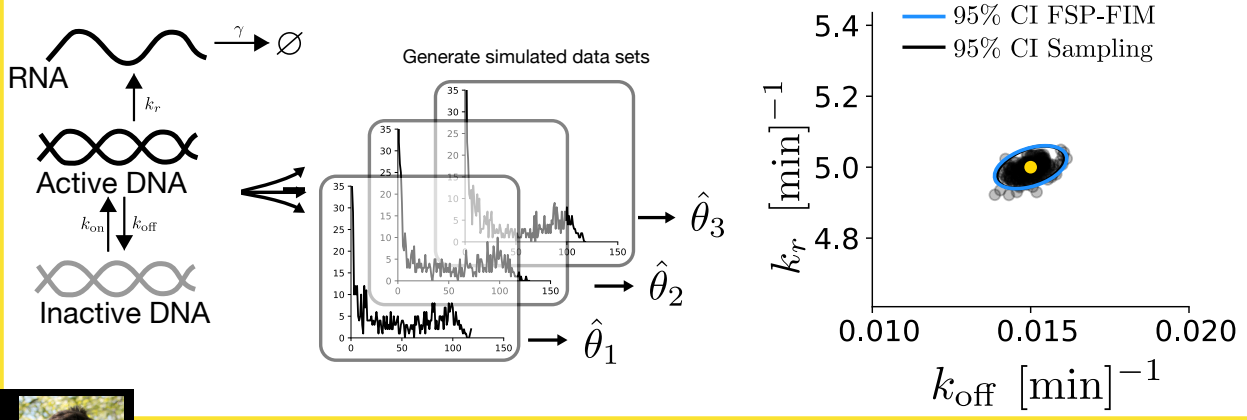
Fox et al, *PLoS Comp. Biol*, 2019
Fox et al, *Complexity*, 2020

FSP-Based Fisher Information

The FSP-FIM correctly predicts the asymptotic normal spread of MLE samples.

Asymptotic normality of the MLE

$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{dist} \mathcal{N}(0, I(\boldsymbol{\theta}^*)^{-1})$$

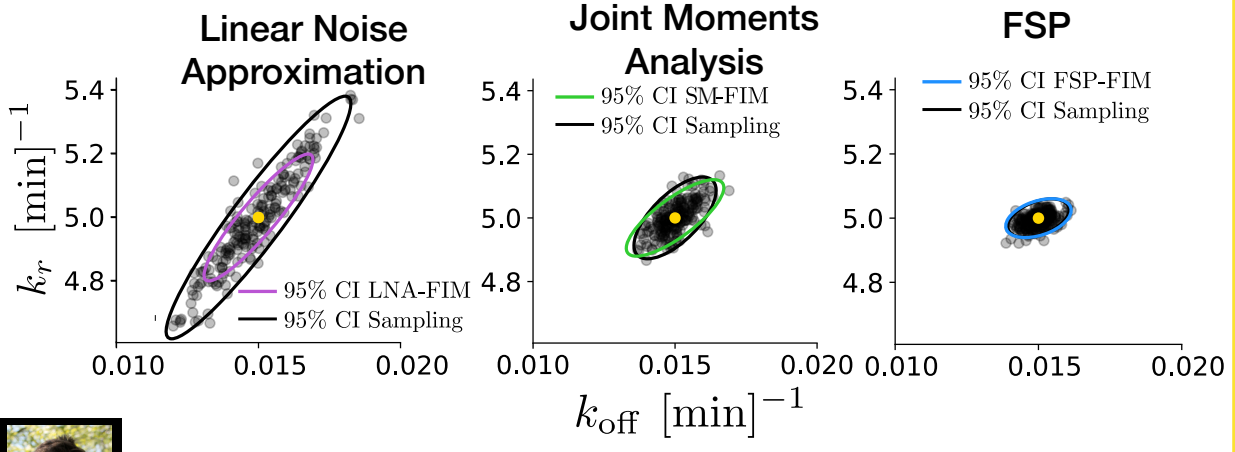


Zach Fox

Fox et al, *PLoS Comp. Biol*, 2019
Fox et al, *Complexity*, 2020

FSP-Based Fisher Information

The FSP-FIM is more accurate and more consistent than moments based approaches.



Asymptotic normality of the Maximum Likelihood Estimator

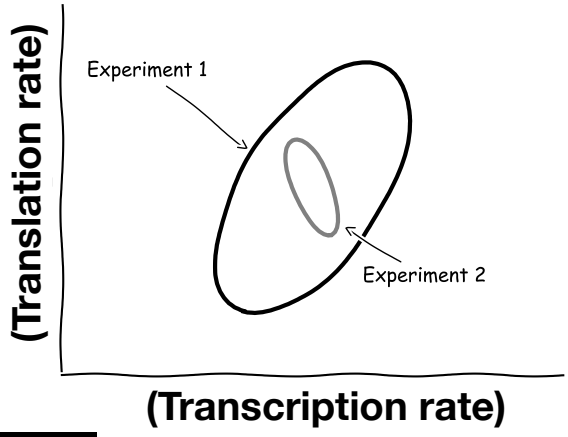
$$\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*) \xrightarrow{dist} \mathcal{N}(0, I(\boldsymbol{\theta}^*)^{-1})$$

Zach Fox

Fox, et al, *PLoS Comp. Biol*, 2019
Fox, et al, *Complexity*, 2020

Using Fisher Information to Design Experiments

Different single-cell experiments reveal different amounts of information about model parameters.



The FIM can estimate which experiments will provide tighter MLE results.

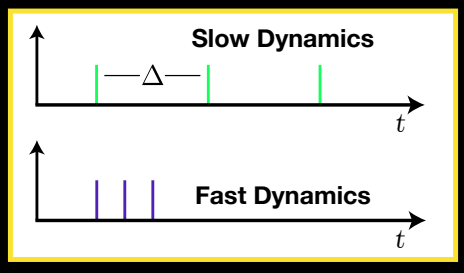
In this case, $|\mathcal{I}_2(\theta)| > |\mathcal{I}_1(\theta)|$

Zach Fox

Fox et al, *PLoS Comp. Biol*, 2019

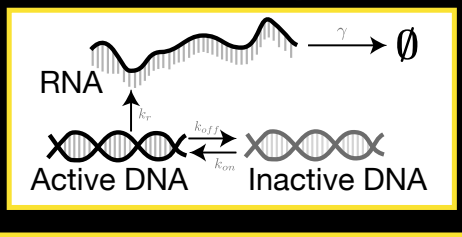
Using Fisher Information to Design Experiments

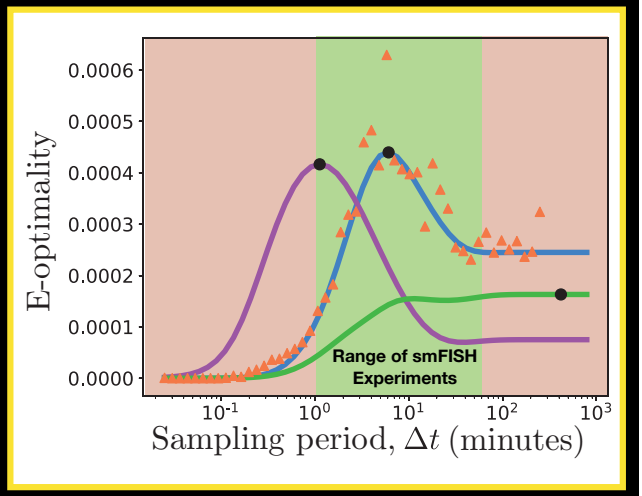
Experiments can be optimized by comparing the FIM for different designs (e.g. sampling periods).



The FSP-FIM (blue) correctly identifies the optimal experiments, whereas the moment based approach (purple, green) do not.

Bursting gene expression

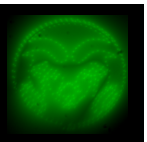




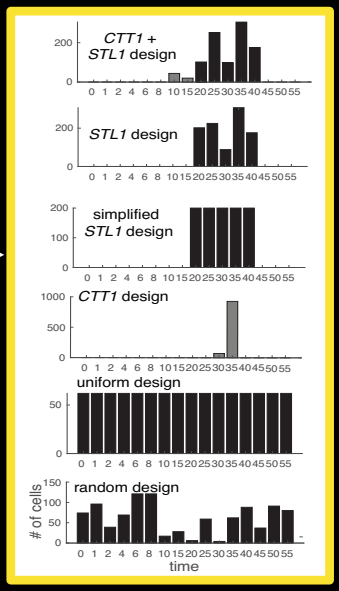
Zach Fox

Fox et al, *PLoS Comp. Biol*, 2019

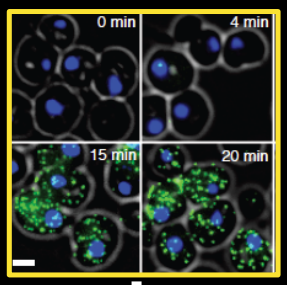
Experimental Validation of FSP-FIM Experiment Design



Real NaCl Concentration

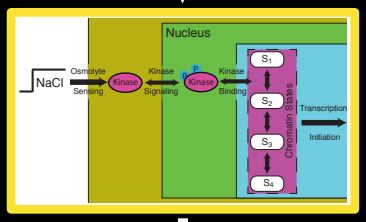


Potential Experiment Designs



smFISH Data:

1000 sets of 75 cells per set per design



Previously identified model

Estimate Uncertainty for:

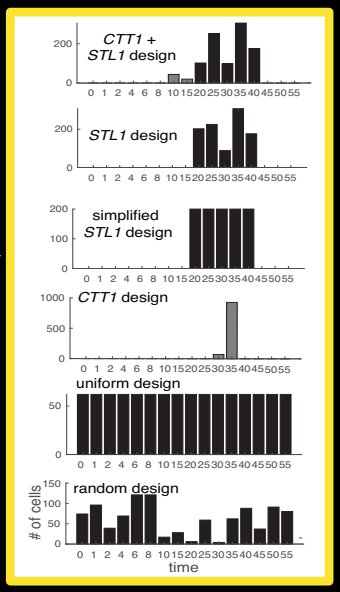
- MAPK Input Signal
- NaCl Concentration

Zach Fox

Munsky et al, *PNAS*, 2018
Fox, et al, *Complexity*, 2020

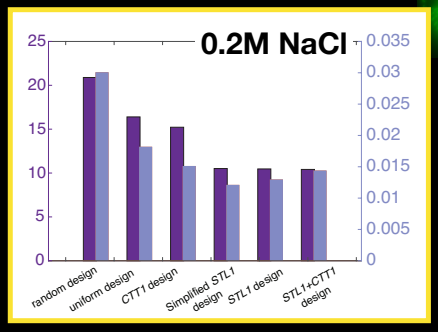
Experimental Validation of FSP-FIM Experiment Design

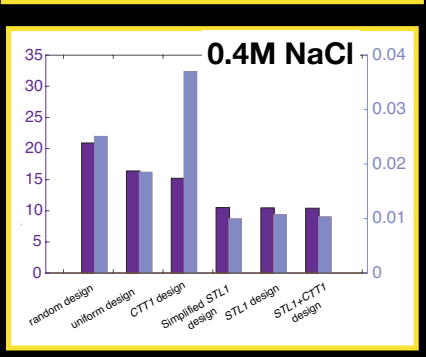
Real NaCl Concentration



Potential Experiment Designs

Signal Deviation Predicted by FIM (s)





Deviation in Fits Estimate of NaCl (M)

Different Experiment Designs

Zach Fox

The FSP-FIM correctly estimates the variance of experiments.

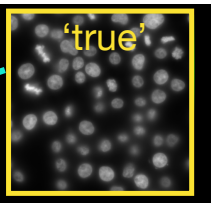
Fox, et al, *Complexity*, 2020

Using FIM to Evaluate Microscope Distortion

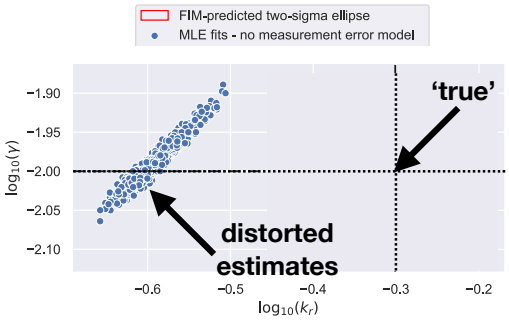
No labeling strategy, microscope, or image processing tool is perfect.

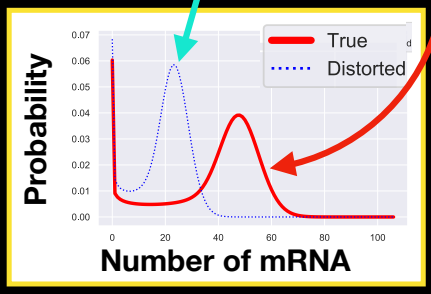
All measurements are noisy.

All data is distorted.



Failure to account for distortion can lead to biased estimates with false confidence.





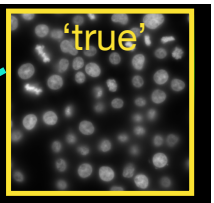
Huy Vo Vo, et al, *bioRxiv*, 2021

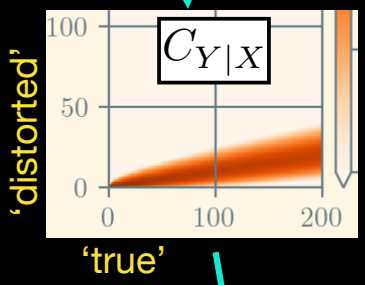
Using FIM to Evaluate Microscope Distortion

No labeling strategy, microscope, or image processing tool is perfect.

All measurements are noisy.

All data is distorted.





Identity Matrix

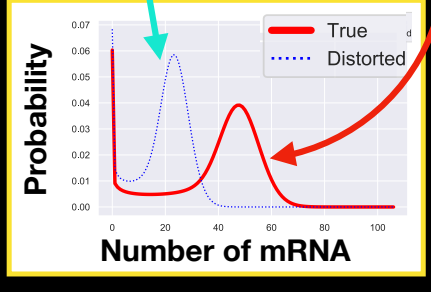
Probabilistic distortion of data can be described by a **Markov Kernel**.

$$P^Y = C_{Y|X} P^X$$

observed distribution

'true distribution'

probabilistic distortion kernel, (derived or empirical)



Huy Vo Vo, et al, *bioRxiv*, 2021

Using FIM to Evaluate Microscope Distortion

The FSP-FIM is easily adapted to consider arbitrary Markov distortion kernels:

Distortion: $P^Y = C_{Y|X} P^X$

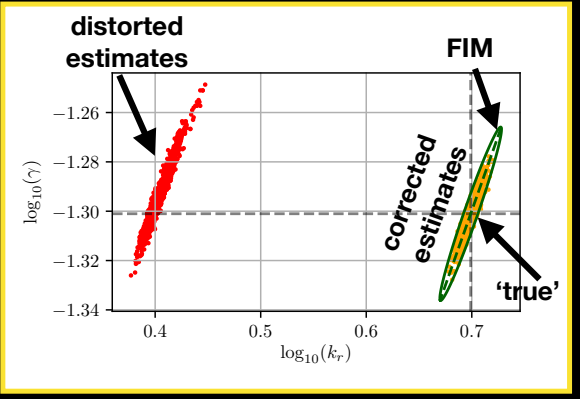
$$\mathbf{s}_i^Y = C_{Y|X} \mathbf{s}_i^X$$

FIM: $\mathcal{I}_{ij}^Y = \mathbb{E}_y \left\{ \partial_i \log P^Y(y) \partial_j \log P^Y(y) \right\}$

$$= \int \frac{\partial_i P^Y(y) \partial_j P^Y(y)}{P^Y(y)} \, \mathrm{dy}$$

$$= \int \frac{s_i^Y(y) s_j^Y(y)}{P^Y(y)} \, \mathrm{dy}$$

Including the distortion kernel corrects estimation errors and improves uncertainty quantification.



Huy Vo Vo, et al, *bioRxiv*, 2021

Distortions Affect Design of Optimal Experiments

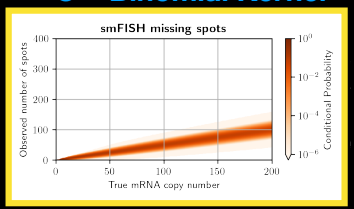
Consider three experiment assays with different distortion kernels:

"Perfect smFISH" (C = Identity Matrix)

1000 cells per time point

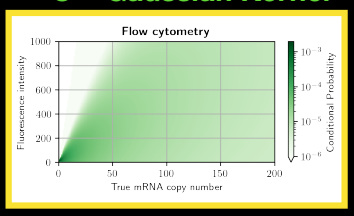
"Lossy smFISH" **C = Binomial Kernel**

1000 cells



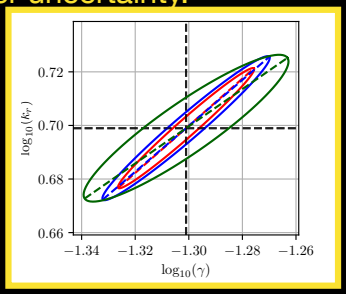
"Flow Cytometry" **C = Gaussian Kernel**

2000 cells

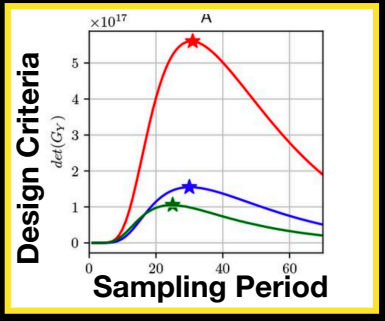


Huy Vo Vo, et al, *bioRxiv*, 2021

The choice of experiment assay changes both **magnitude** and **orientation** of parameter uncertainty.



Effects of distortion can be taken into account during FIM-based experiment design.



Conclusions

- **The Central Dogma is a Noisy Process that can be measured at Single-Molecule resolution.**
  - **Single-cell experiments can quantify, and discrete stochastic models can reproduce, every step of these processes.**
- But ***experiments are expensive;*** there are an infinite number of possible designs; and each choice will affect potential conclusions and uncertainty.
- The **Fisher Information Matrix (FIM)** can estimate expected uncertainties for potential experiment designs.
- **Finite State Projection** allows for computation of the FIM even for arbitrary non-Gaussian processes, and for circumstances when data are subject to unavoidable probabilistic distortions.

Acknowledgments and Collaborators (2019-2021)

Hog Signaling/Transcription Activation
Gregor Neuert, Vanderbilt
Guoliang Li, Vanderbilt
Hossein Jahnsaz, Vanderbilt
Jason Hughes, Vanderbilt
Douglas Shepherd, ASU

Single-mRNA Translation
Tim Stasevich, CSU
Tatsuya Morisaki, CSU
Kenneth Lyon, CSU
Amanda Koch, CSU
Linda Forero Quintero, CSU

Cell Heterogeneity in Algal Biofuels:
Babetta Marrone, LANL
Claire Sanders, LANL
Elaheh Alizadeh, CSU

Cell-free Single-BioMolecule Kinetics:
Sabrina Leslie, McGill University
Shane Scott, McGill University

MultiScale Bayesian FSP tools:
Ania Baetica, CalTech
Thomas Catanach, Sandia

Soil Microbiome Machine Learning
John Dunbar Lab, LANL

Inflammation driven mRNA expressio
James Werner, LANL
Dan Kalb, LANL

q-bio Training and Textbook:
William S Hlavacek, LANL
Lev S. Tsimring, UCSD
Marek Kimmel, Rice University
Former students of the 2010-2017 q-bio Summer Schools

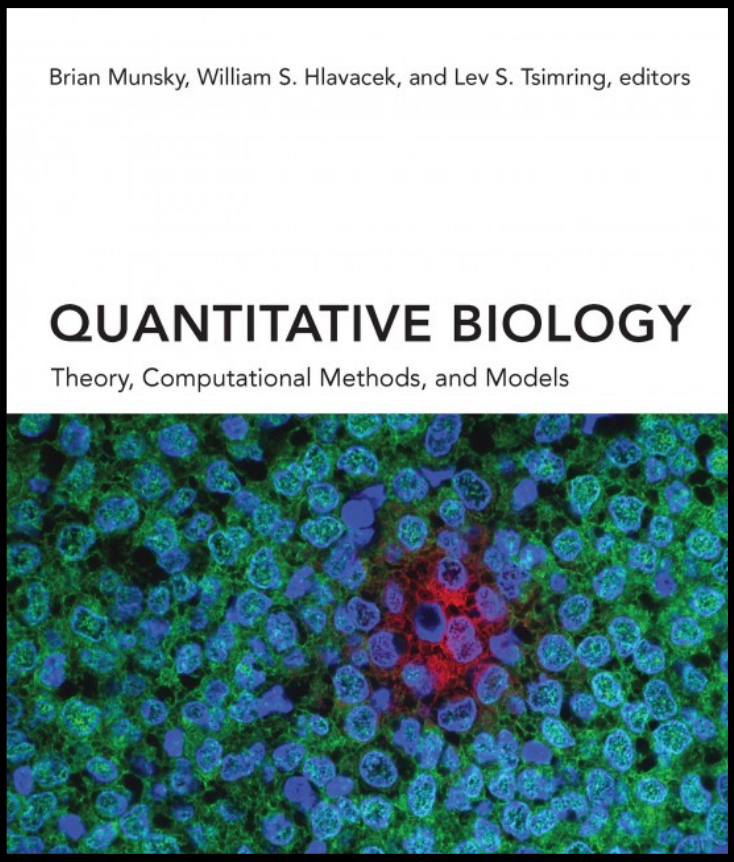
Postdocs:
Luis Aguilera
Linda Forero Quintero
Huy Vo

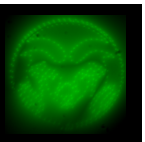
Graduate Students:
Zachary Fox
Lisa Weber
Mohammad Tanhaemami
Michael May
Jaron Thompson
Will Raymond
Eric Ron

Undergrads:
Elliot Djokic
Charis Ellis
Torin Moore
Zach Haigh
Charlotte Mitchell
Joshua Cook
Rachel Keating
Stuart McKnight
Katie Davis

Funding:
WMKF
NIH/NIGMS
DOE/BER
NSF CAREER

munsky@colostate.edu





An introduction to the quantitative modeling of biological processes, presenting modeling approaches, methodology, practical algorithms, software tools, and examples of current research.