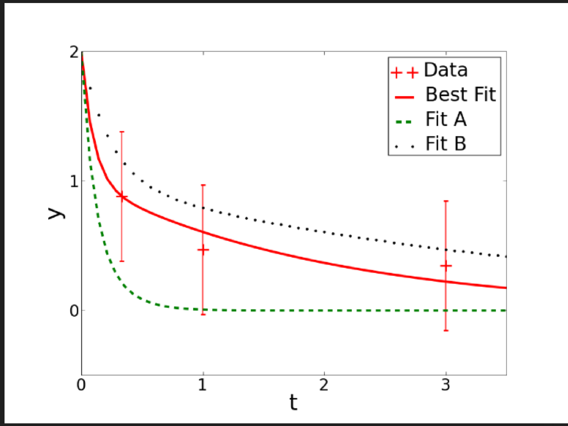
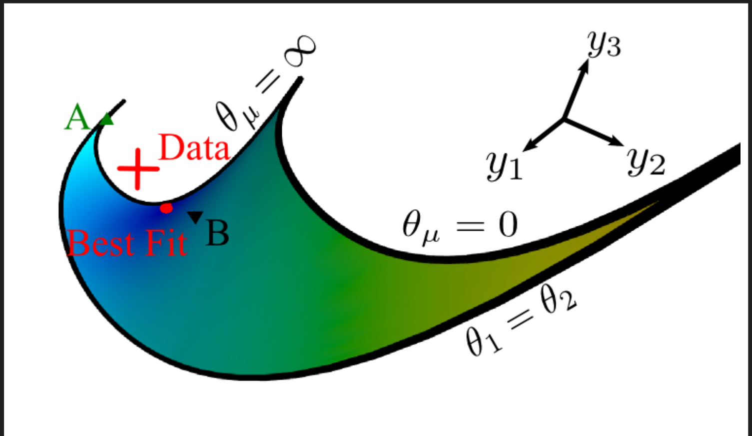
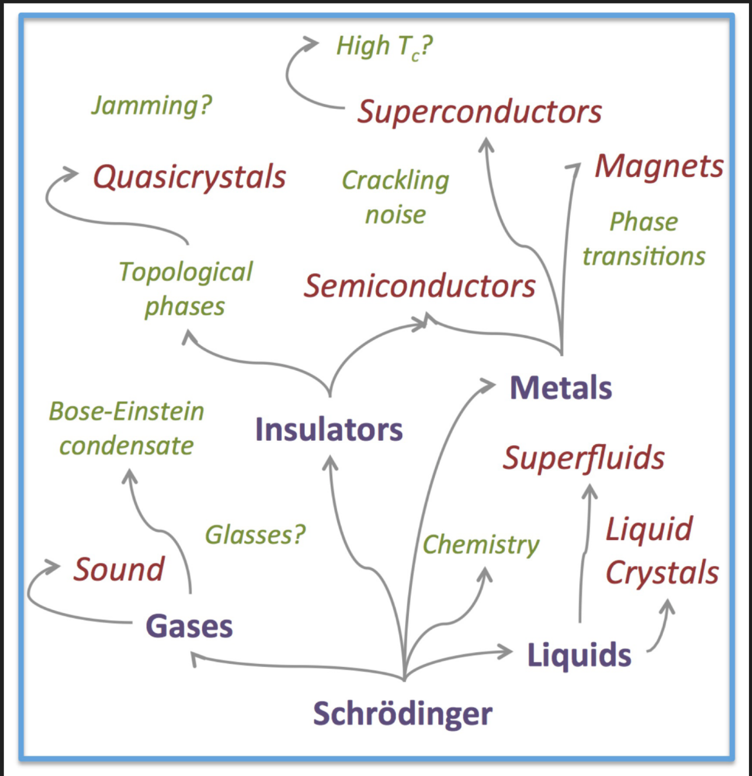
# SLOPPY MODELS AND EFFECTIVE THEORIES IN PHYSICS, BIOLOGY, AND BEYOND





MARK K. TRANSTRUM


IPAM WORKSHOP: UNCERTAINTY QUANTIFICATION FOR STOCHASTIC SYSTEMS AND APPLICATIONS

NOVEMBER 17, 2017

# OUTLINE

1. UQ and Effective Theories in Physics
2. Sloppy Models
3. Information Geometry
4. Manifold Boundary Approximation Method (MBAM)
5. Information Topology

# UNCERTAINTY QUANTIFICATION

## TYPES OF UNCERTAINTY

- Parametric Uncertainty
- Structural Uncertainty
- Experimental Uncertainty
- Algorithmic Uncertainty

## MODELING QUESTIONS

- Model Selection/Model Reduction
- Model Complexity
  - Accuracy
  - Computability

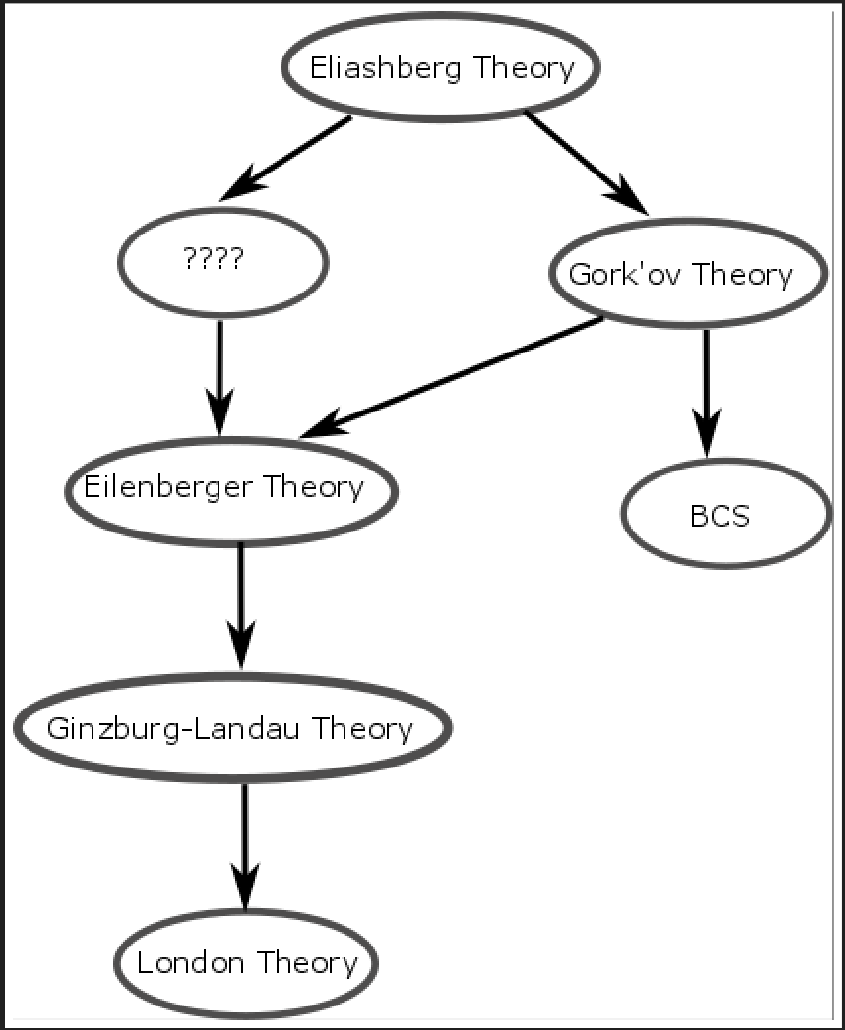
# EXAMPLE: SUPERCONDUCTIVITY

## CENTER FOR BRIGHT BEAMS

cbb.cornell.edu

Goal: Make material-specific predictions to guide development of the next generation of particle accelerators.

- Hierarchy of potential models
- Simple models are special cases of more complicated theories
- Partially ordered set (Hasse Diagram)



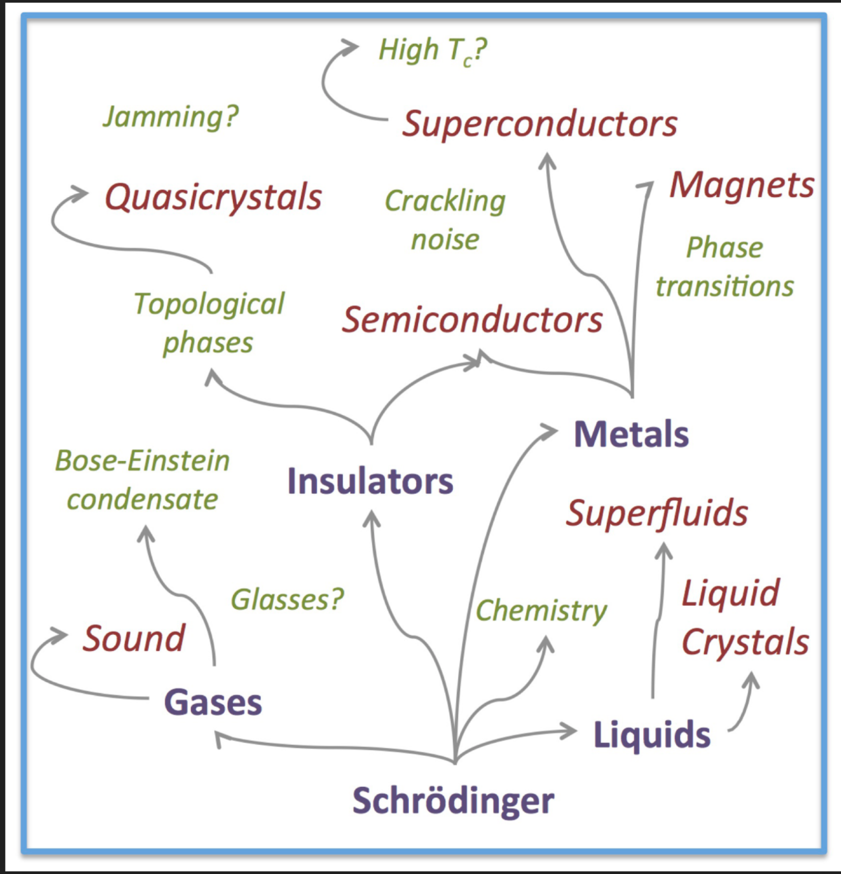
# HIERARCHY OF SCIENTIFIC THEORIES

## MANY PRACTICAL QUESTIONS:

- Parameter estimation
- Numerical methods
- Experimental design
- Large scale models demand new computational and theoretical tools (UQ).

More broadly, how do we organize our knowledge about the world? And why?

- Reductionism
- Emergence



# PHILOSOPHY OF SCIENCE

- Wigner: Mathematics [modeling] has been unreasonably successful at explaining physical phenomena

- Anderson: More is Different
  - Reductionism vs. Constructionism
  - Simple theories are organized hierarchically
  - Very different systems can be described by the same theory
  - Equivalence classes of physical systems described by the same model (Universality)

- Goldenfeld & Kadanoff: Don't model bulldozers with quarks.
  - Use the right level of description to catch the phenomena of interest.

- David Pines: Protectorates
  - A collective behavior whose "properties are determined by higher organizing principles and nothing else"

# AN INFORMTION THEORY PERSPECTIVE: (EMERGENT) PHYSICAL LAWS ARE EFFICIENT DATA COMPRESSIONS SCHEMES.

Shannon Information: $H(X) = -\langle \log(P(X)) \rangle$

- Absolute bound on lossless compression

Fisher Information: $I(\theta) = -\langle \nabla_\theta^2 \log P \rangle = \langle \nabla_\theta \log P \nabla_\theta \log P \rangle$

- Information carried by random variable about parameters $\theta$

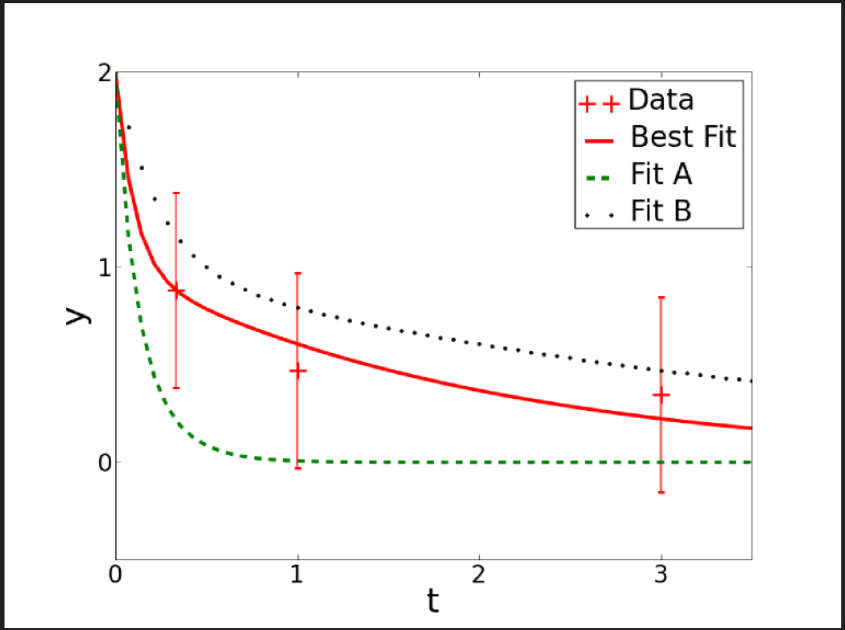
Algorithmic Complexity

- Kolmogorov: shortest string (program) that will reproduce some data
- Human-Interpretable
- Floating Point Operations
- Memory
- Internal Precision

# OUR FRAMEWORK

- Model: Mapping from parameters ($\theta \in \mathbb{R}^N$) to predictions (probability distribution for a quantity of interest).
- The model mapping implies a computational algorithm for reproducing an observed distribution
- Goal: Find a computationally simpler, approximate distribution.
- Kullback-Leibler Divergence: Cost of using the wrong distribution:
  - $D_{KL}(p||q) = -\langle \log(q/p) \rangle_p$
  - How many extra bits required for encoding data using distribuion $q$ that was drawn from $p$.
  - Not a proper metric.
  - Infintesimal separation: $q = p(\theta + d\theta) \to D_{KL}(p||q) \propto d\theta^T I d\theta$
- Fisher Information ($I$):
  - Riemannian Metric on the space of probability distributions $I \to g$.
  - Information Geometry
- Strategy: Use computational differential geometry to explore the space equivalent models.

# NONLINEAR REGRESSION (NLLS)

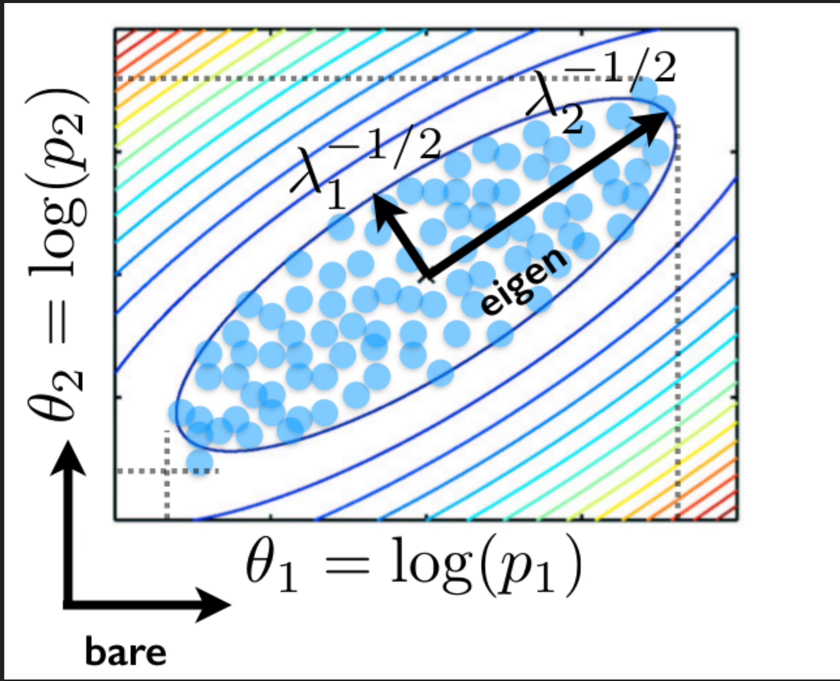



$$y(t,\theta) = e^{-\theta_1 t} + e^{-\theta_2 t}$$

$$C(\theta) = \frac{1}{2}\sum_m \left(\frac{d_m - y(t_m,\theta)}{\sigma_m}\right)^2 = \frac{1}{2}\sum_m r_m^2$$

# COST SURFACE

Which parameter values are consistent with the data?




- Asymptotic Inference
- Quadratic Approximation at the best fit
- Hessian Matrix (Fisher Information Matrix)

$$\partial_\mu \equiv \frac{\partial}{\partial \theta_\mu}$$

$$H_{\mu\nu} = \partial_\mu \partial_\nu C = \frac{1}{2} \partial_\mu \partial_\nu \sum_m r_m^2$$

$$H_{\mu\nu} = \sum_m \partial_\mu r_m \partial_\nu r_m + \sum_m r_m \partial_\mu \partial_\nu r_m$$

$$J_{m\mu} = \partial_\mu r_m$$

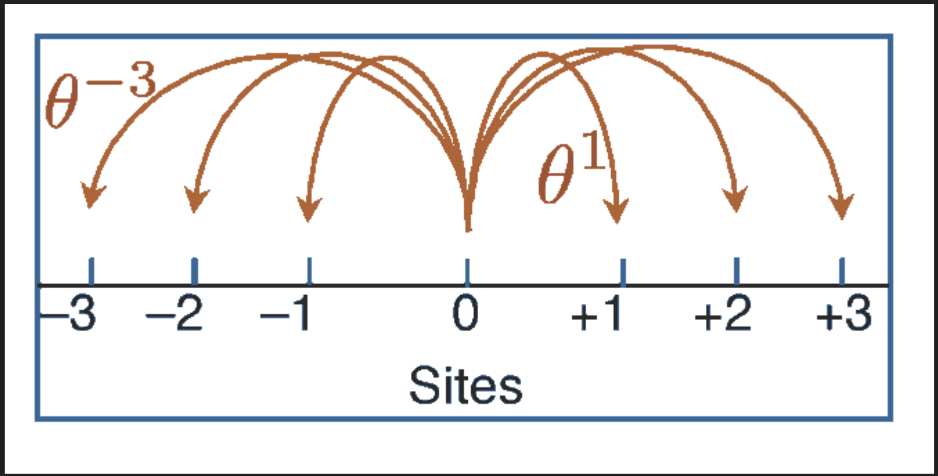
$$H_{\mu\nu} \approx I_{\mu\nu} = (J^T J)_{\mu\nu}$$

$I_{\mu\nu}$ is intrinsic to the model, independent of the data.

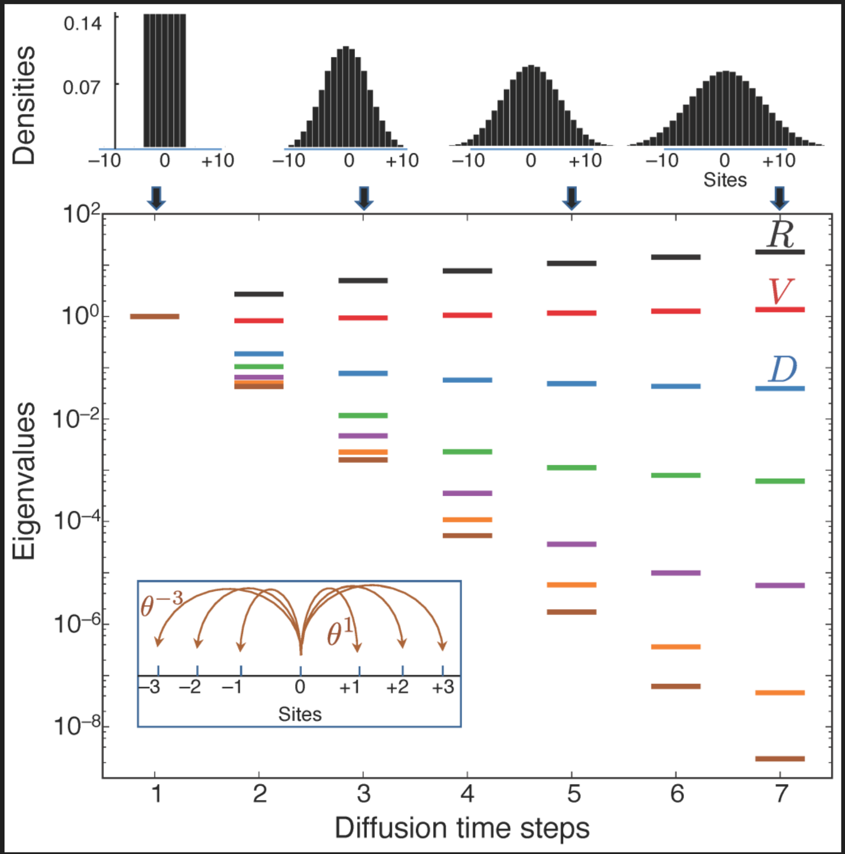
Use it to study a model class rather than a specific fitting problem.

# EXAMPLE: MICROSCOPIC DIFFUSION

- Particles "hopping" on a lattice
- Parameters: probability of hopping a given number of sites away
- Initial density: All particles at the origin
- Data: Observed density after several steps
- Goal: Infer the parameters from observations of density at later times



# EXAMPLE: MICROSCOPIC DIFFUSION




- Short time scales, parameter inference is easy
- More difficult after observations are coarsened
- Only a few parameter combinations need to be carefully tuned to explain data
- Most parameter combinations are irrelevant: can take on *ANY* values.
- Small parameter: Ratio of scales

Machta, Benjamin B., et al. "Parameter space compression underlies emergent theories and predictive models." *Science* **342**.6158 (2013): 604-607.

# DIFFUSION: MICROSCOPIC TO MACROSCOPIC

- Continuum limit: Gradient expansion
  - Formally justified by the small parameter (ratio of scales)

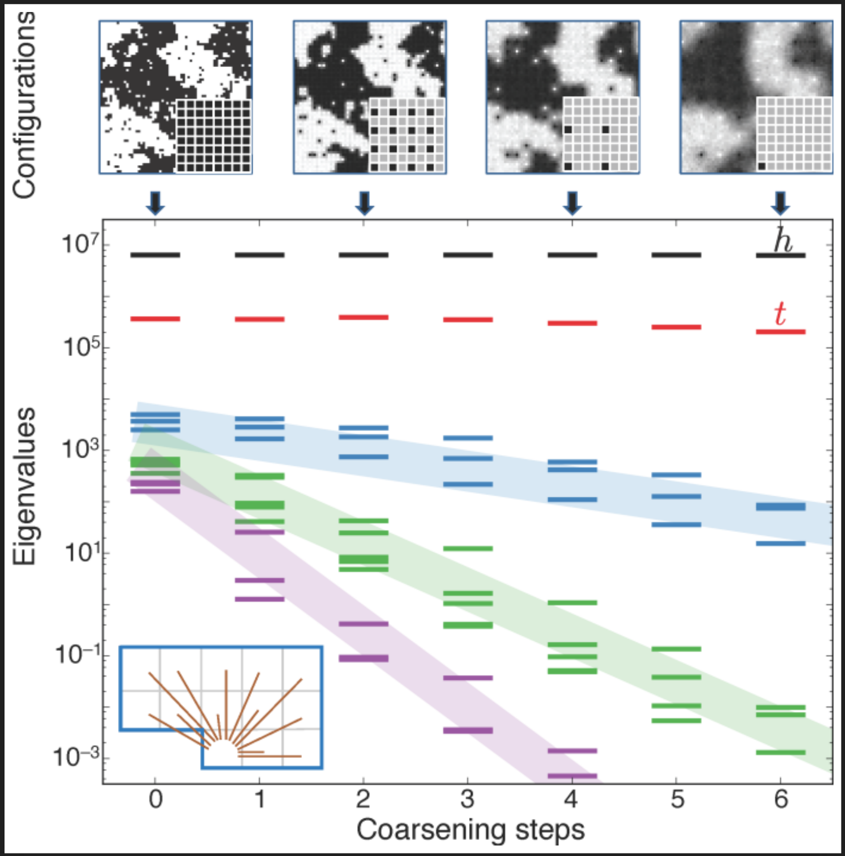
$$\frac{\partial \rho}{\partial t} = R\,\rho + \mathbf{v}\cdot\nabla\rho + D\nabla^2\rho + \ldots$$

- In most cases, first two terms are zero because of symmetry:

$$\frac{\partial \rho}{\partial t} = D\nabla^2\rho$$

- Fisher Information justifies the use of a simple theory
- Intuitively, are these models equivalent?
  - One model has many parameters
  - The other model has only one parameter
  - How can a low-dimensional object approximate a higher-dimensional object?

# EXAMPLE: CRITICAL PHENOMENON




Ising Model:

$$\mathcal{H} = -\sum_\mu \sum_{d(i,j)=\mu} \theta_\mu s_i s_j$$

Small parameter: (inverse system size)

Ising universality class:

- Magnets
- Liquid-Gas
- Biological Membranes

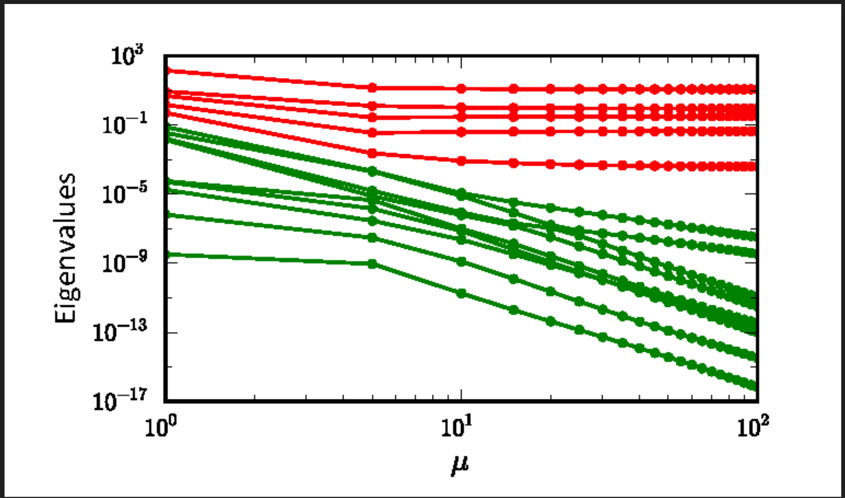
Machta, Benjamin B., et al. "Parameter space compression underlies emergent theories and predictive models." *Science* **342**.6158 (2013): 604-607.

# EXAMPLE: SLOW MANIFOLDS

Van der pol oscillator:

$$\mu^{-2}\dot{x} = x - \frac{x^3}{3} - y + \sum_{m,n} a_{m,n}\left(x - \frac{x^3}{3} - y\right)^m x^n$$

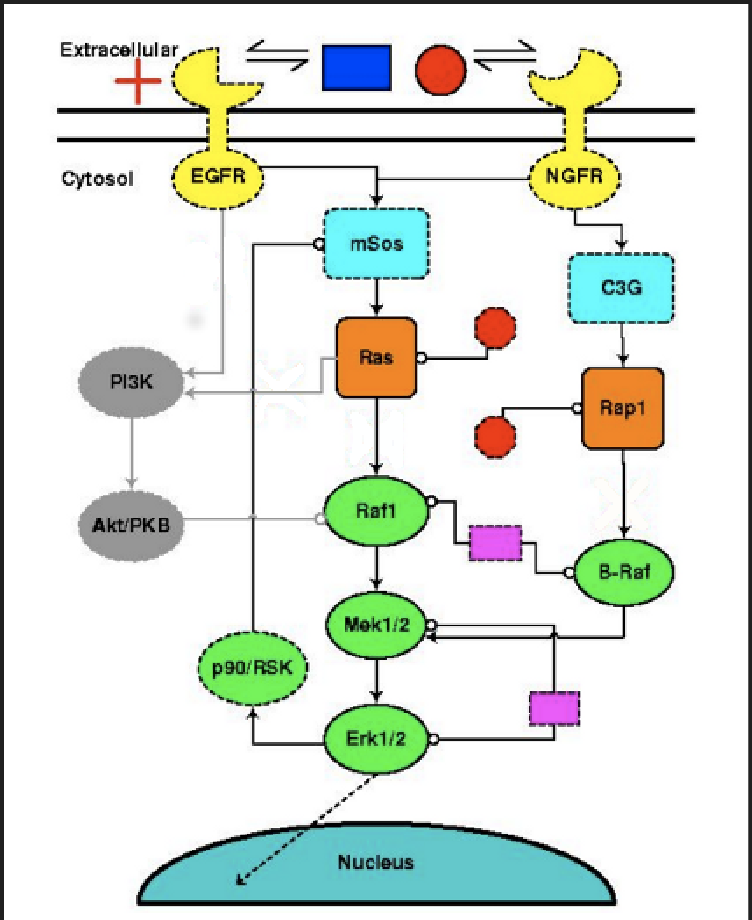
$$\dot{y} = x$$

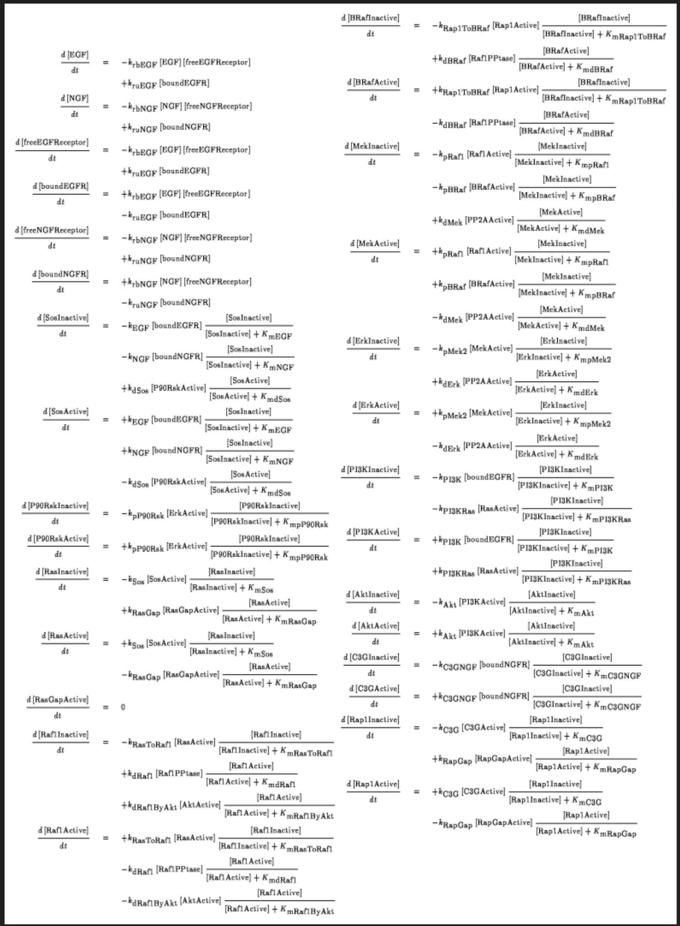



Small parameter: Ratio of time scales

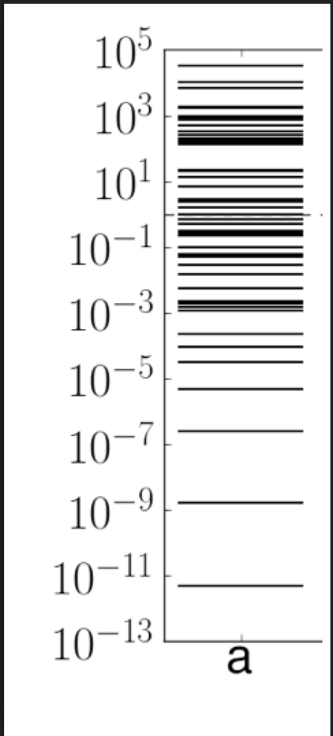
Chachra, Ricky, Mark K. Transtrum, and James P. Sethna. "Structural susceptibility and separation of time scales in the van der Pol oscillator." *Physical Review E* **86**.2 (2012): 026712.

# EXAMPLE: SYSTEMS BIOLOGY



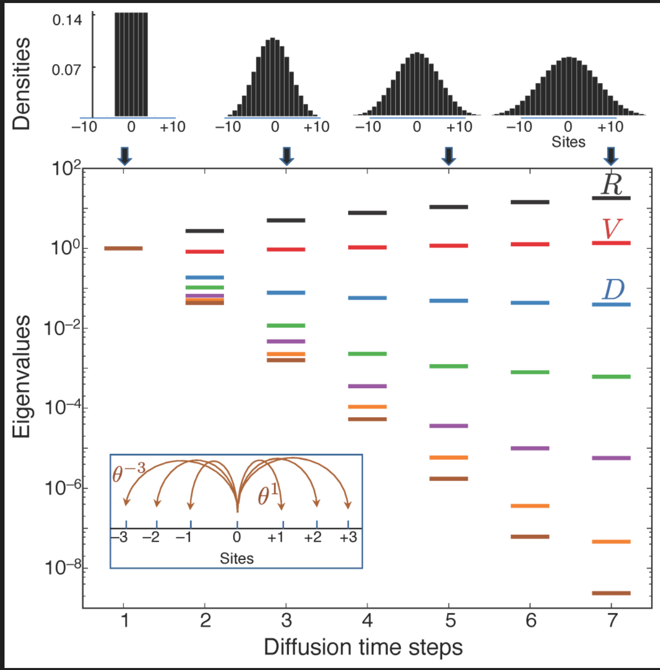



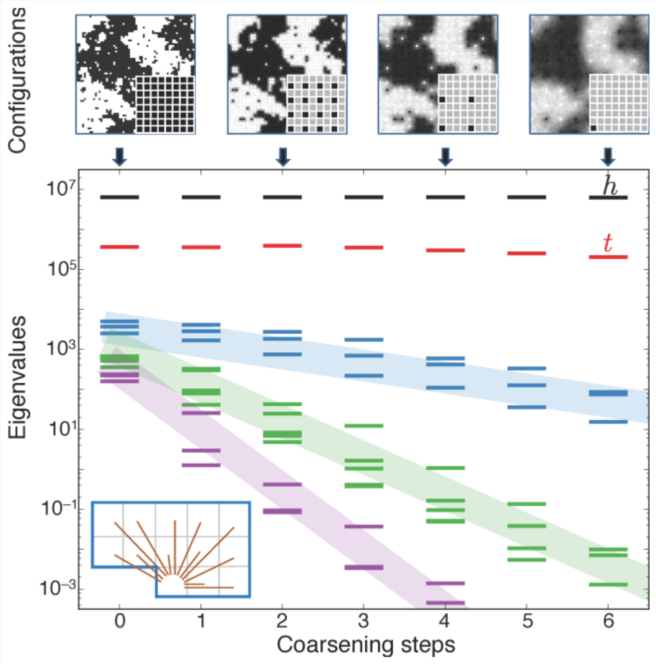
29 Coupled ODEs
48 Parameters

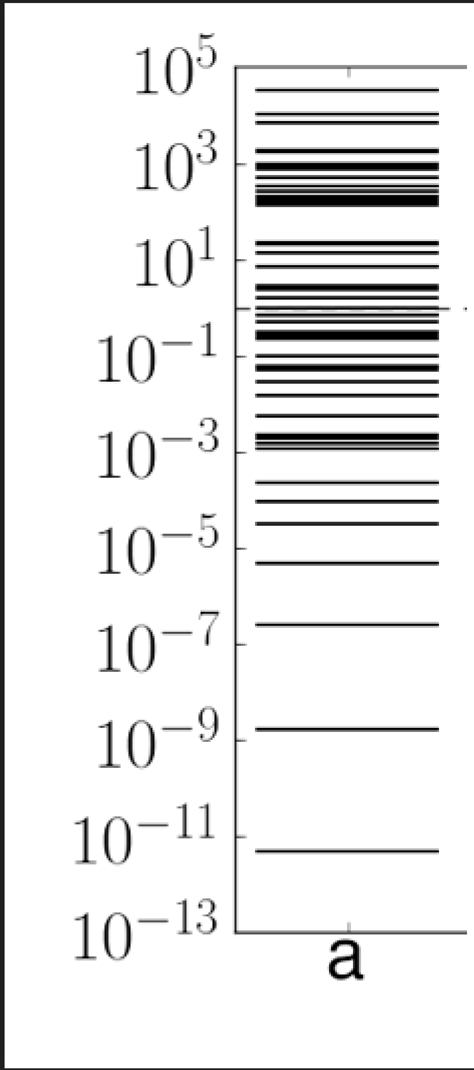


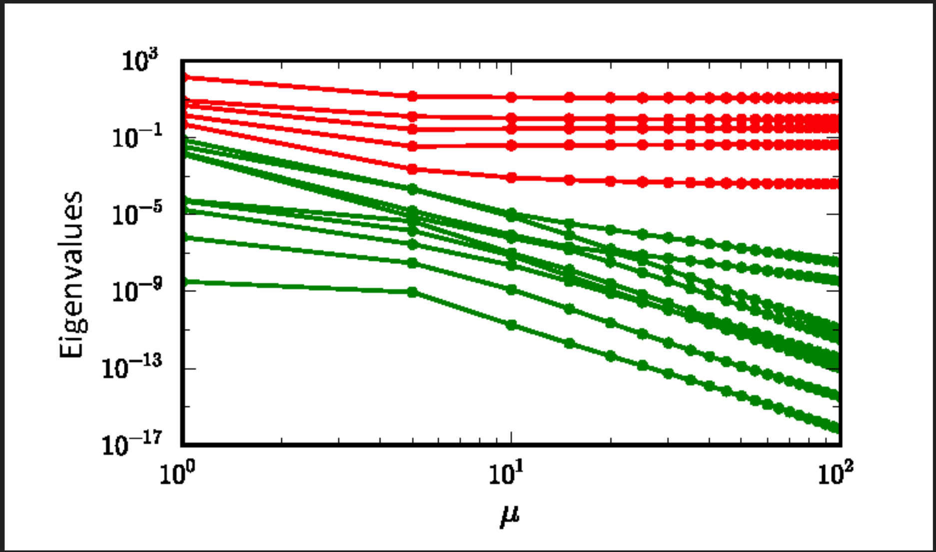

No obvious small parameter? Simple theory?

Brown, Kevin S., et al. "The statistical mechanics of complex signaling networks: nerve growth factor signaling." *Physical biology* **1**.3 (2004): 184.

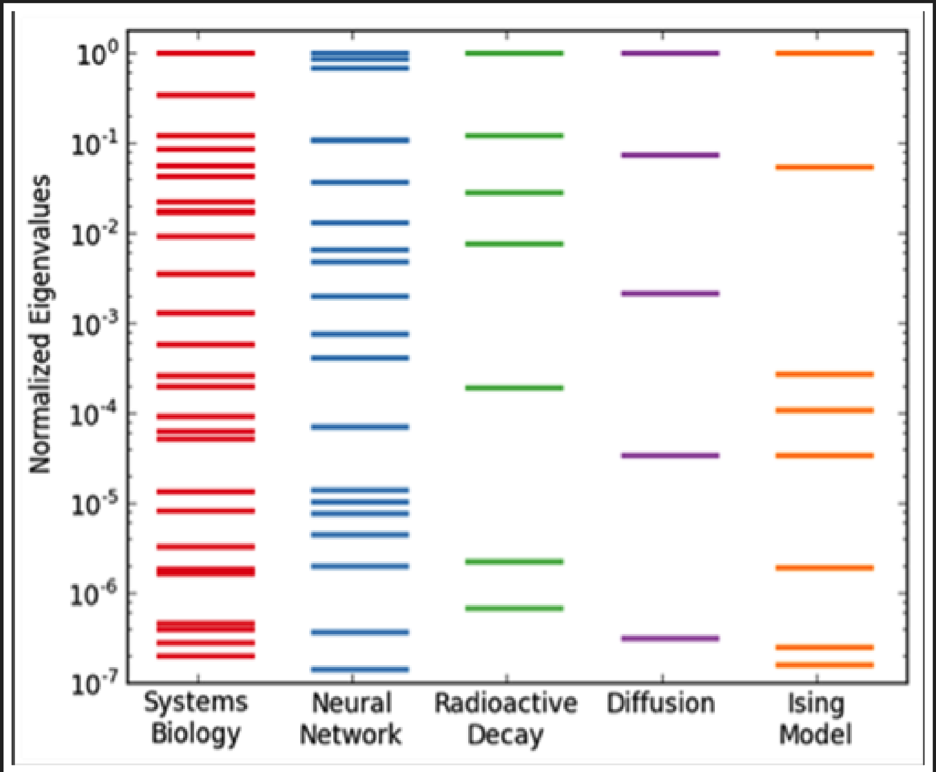

Densities
0.14
0.07
−10
0
+10
Sites
Eigenvalues
R
V
D
θ−3
θ1
−3
−2
−1
0
+1
+2
+3
Sites
1
2
3
4
5
6
7
Diffusion time steps


Configurations
Eigenvalues
h
t
0
1
2
3
4
5
6
Coarsening steps


10^5
10^3
10^1
10^−1
10^−3
10^−5
10^−7
10^−9
10^−11
10^−13
a


Eigenvalues
10^3
10^−1
10^−5
10^−9
10^−13
10^−17
10^0
10^1
10^2
μ

# SLOPPY MODELS




- Exponential Distribution of FIM eigenvalues
- Parameter inference is hard
- Lots of equivalent microscopic models
- Irrelevant (sloppy) parameter combinations can take on ***ANY*** values.
- Emergent "theory" hidden inside the complexity
- Low effective dimensionality

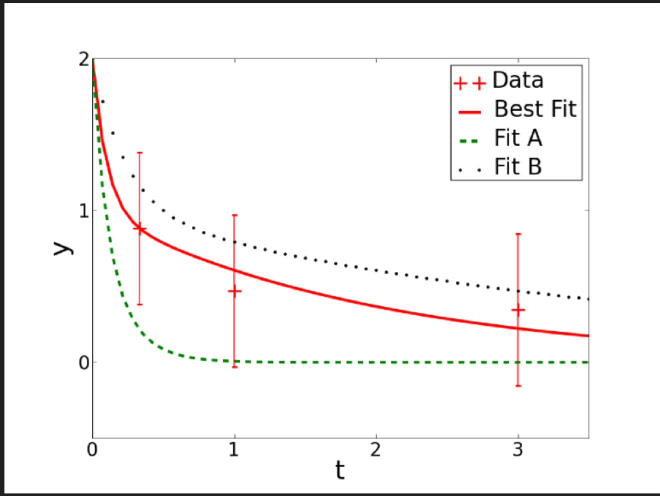

Machta, Benjamin B., et al. "Parameter space compression underlies emergent theories and predictive models." *Science* **342**.6158 (2013): 604-607.

# SLOPPINESS AND FISHER INFORMATION

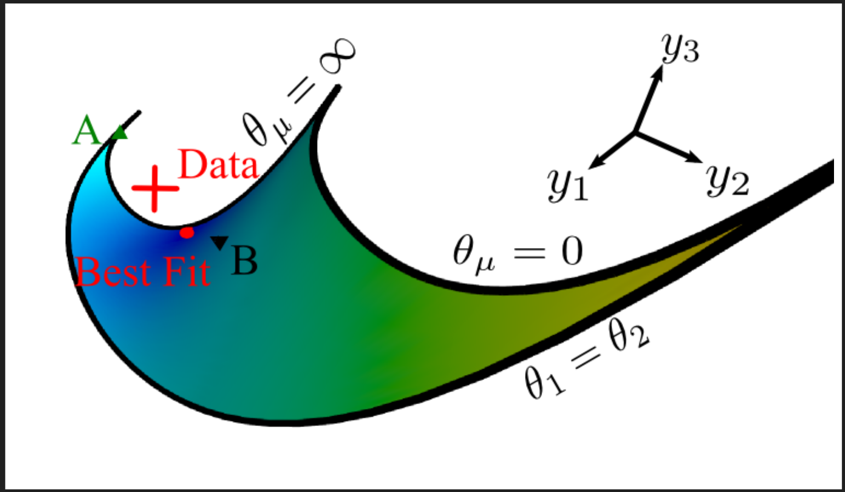
- FIM eigenvalues suggest a low effective dimensionality
  - Can we make this idea rigorous?
- What happens if we reparameterize the model?
  - Change units?
  - Rates vs. time constants?
  - Other combinations of parameters
- FIM eigenvalues can be transformed into any positive values by reparameterizing the model.
- Is there a more intrinsic measure of sloppiness?
- Answer: Information Geometry

# THE MODEL MANIFOLD

- *Model Manifold*: Mapping from parameters to data
- All possible model predictions embed a manifold in the space of data with ***parameters as coordinates.***
- Riemannian Metric (induced from data space) = Fisher Information Matrix
- Sloppy models are ***bounded manifolds***




$$y(t, \theta) = e^{-\theta_1 t} + e^{-\theta_2 t}$$




Transtrum, Mark K., Benjamin B. Machta, and James P. Sethna. "Why are nonlinear fits to data so challenging?." *Physical review letters* **104**.6 (2010): 060201.

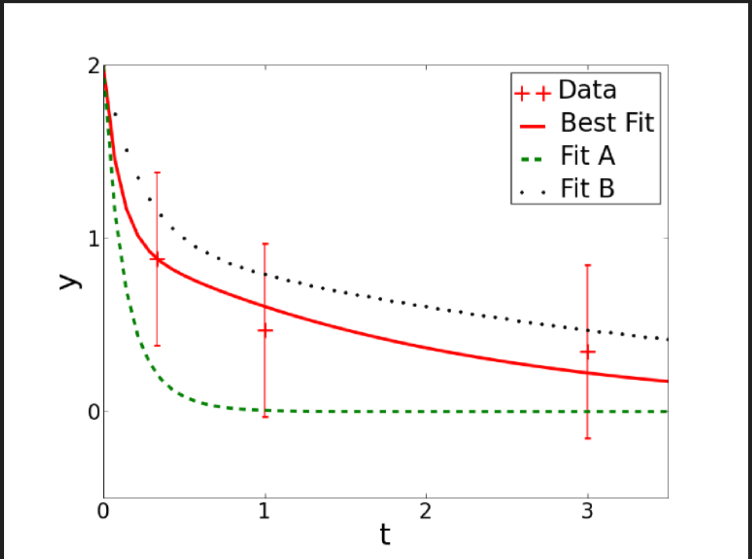
# GALLERY OF MODEL MANIFOLDS

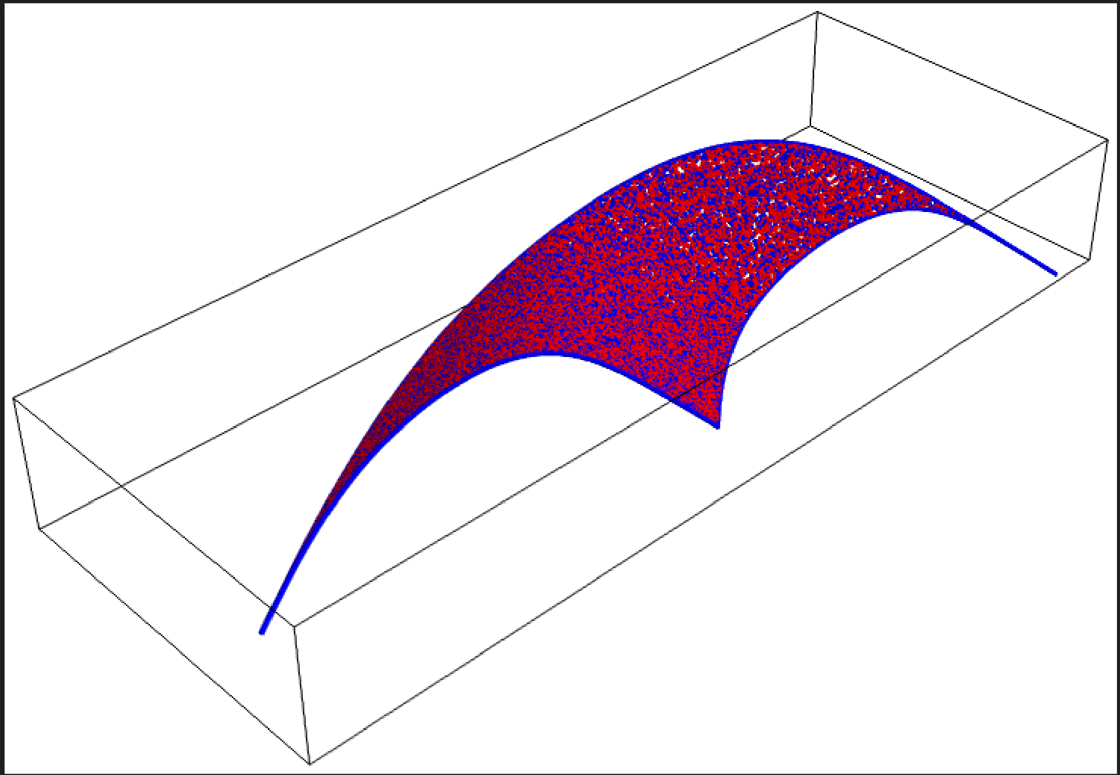
# TWO EXPONENTIALS

$$y(t, \theta) = e^{-\theta_1 t} + e^{-\theta_2 t}$$

2 Parameters

3 Observations



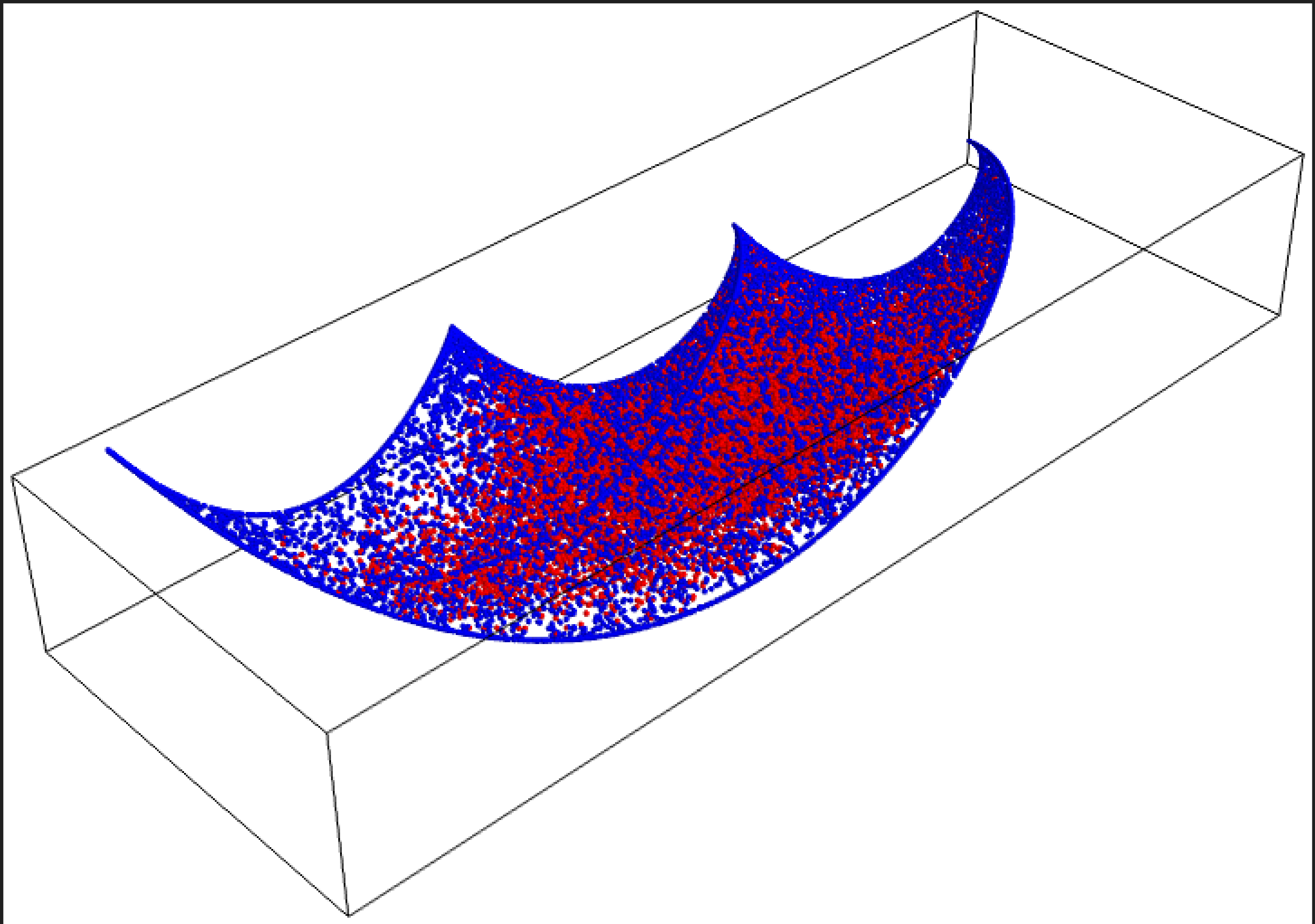



# THREE EXPONENTIALS

$$y(t, \theta) = e^{-\theta_1 t} + e^{-\theta_2 t} + e^{-\theta_3 t}$$

3 Parameters

5 Observations

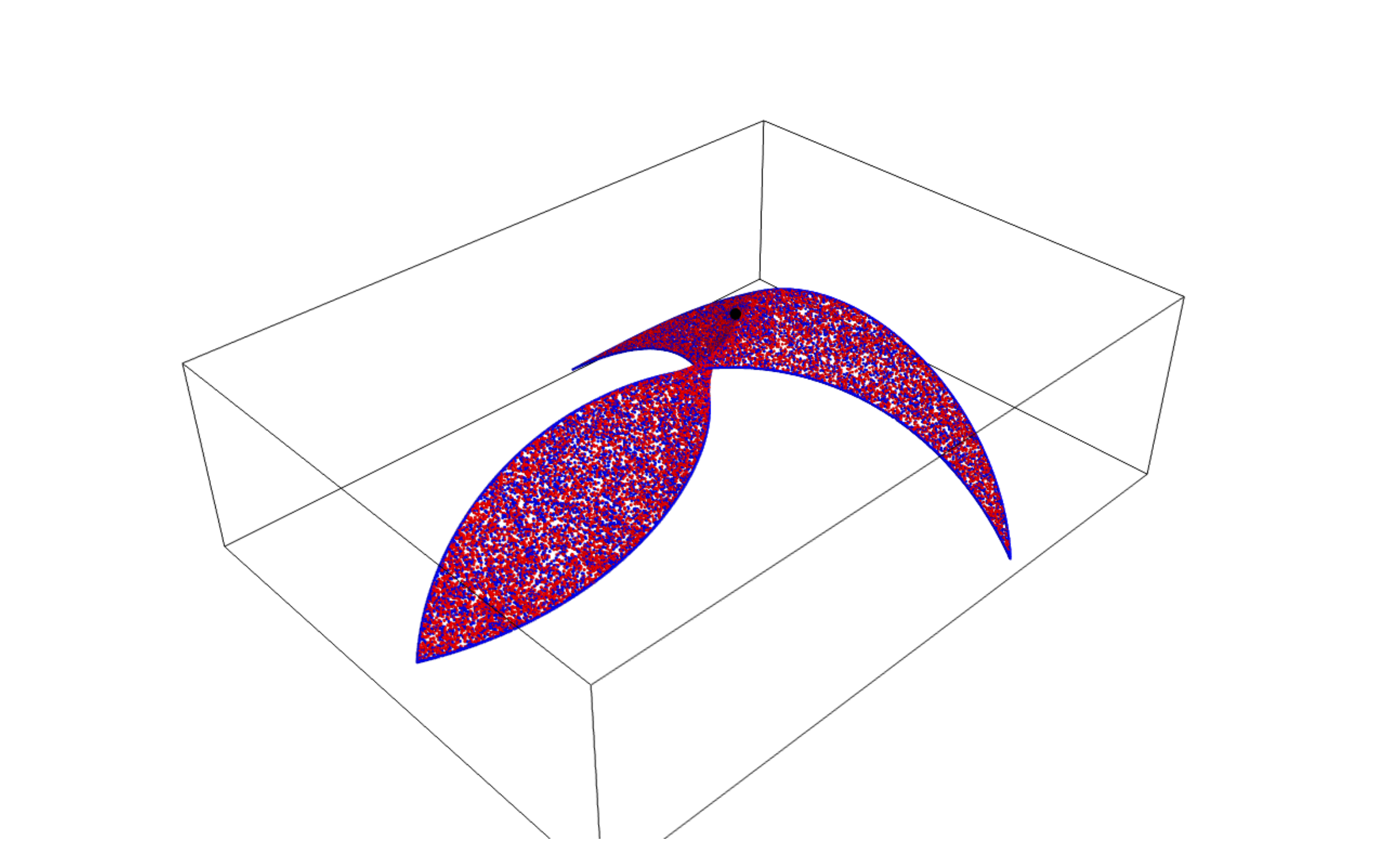


# ENZYME CATALYZED REACTION

$$y(x, \theta) = \frac{\theta_1(x^2+\theta_2 x)}{(x^2+\theta_3 x+\theta_4)}$$

4 (2) Parameters

11 Observations

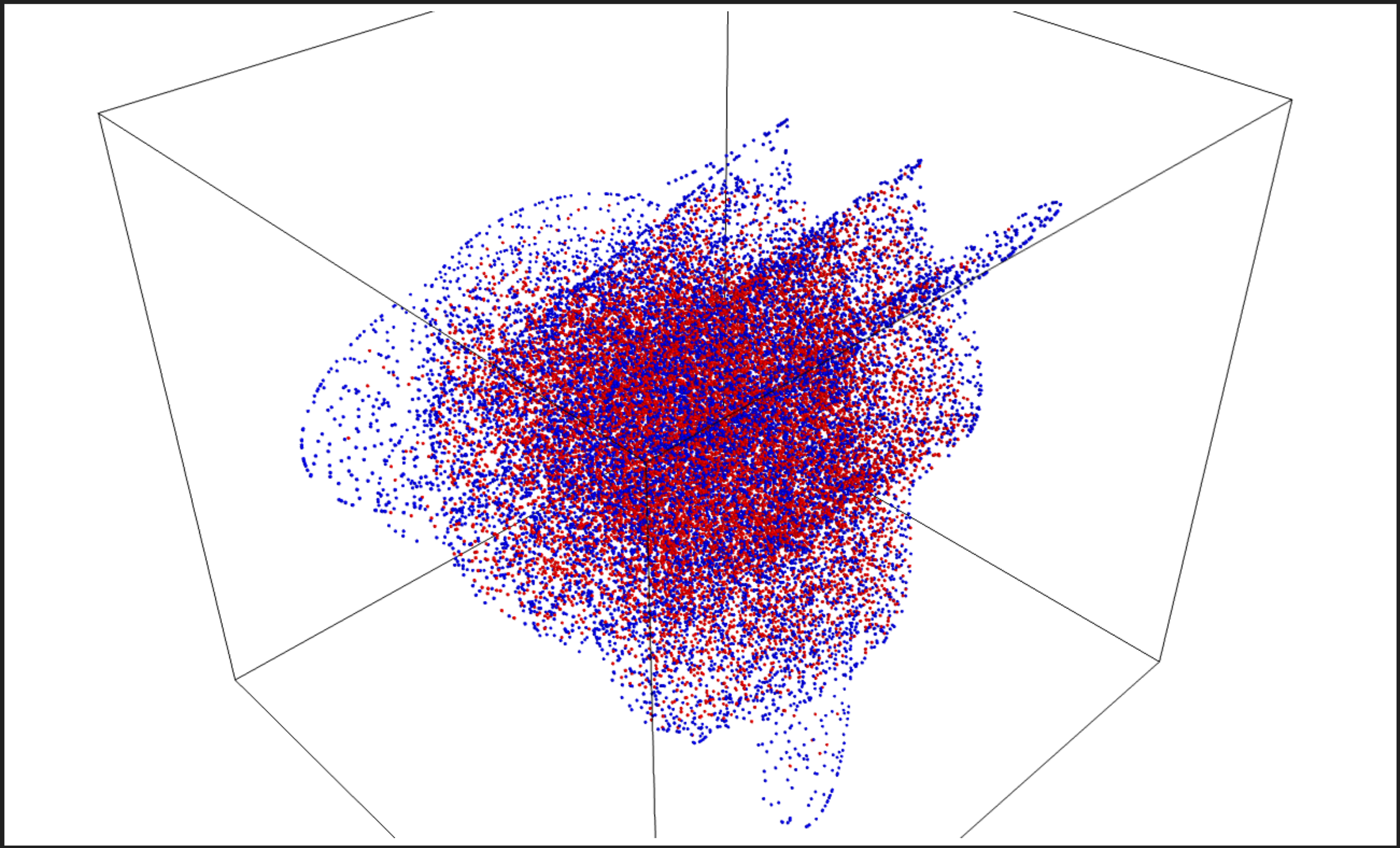


# CHEBYSHEV QUADRATURE

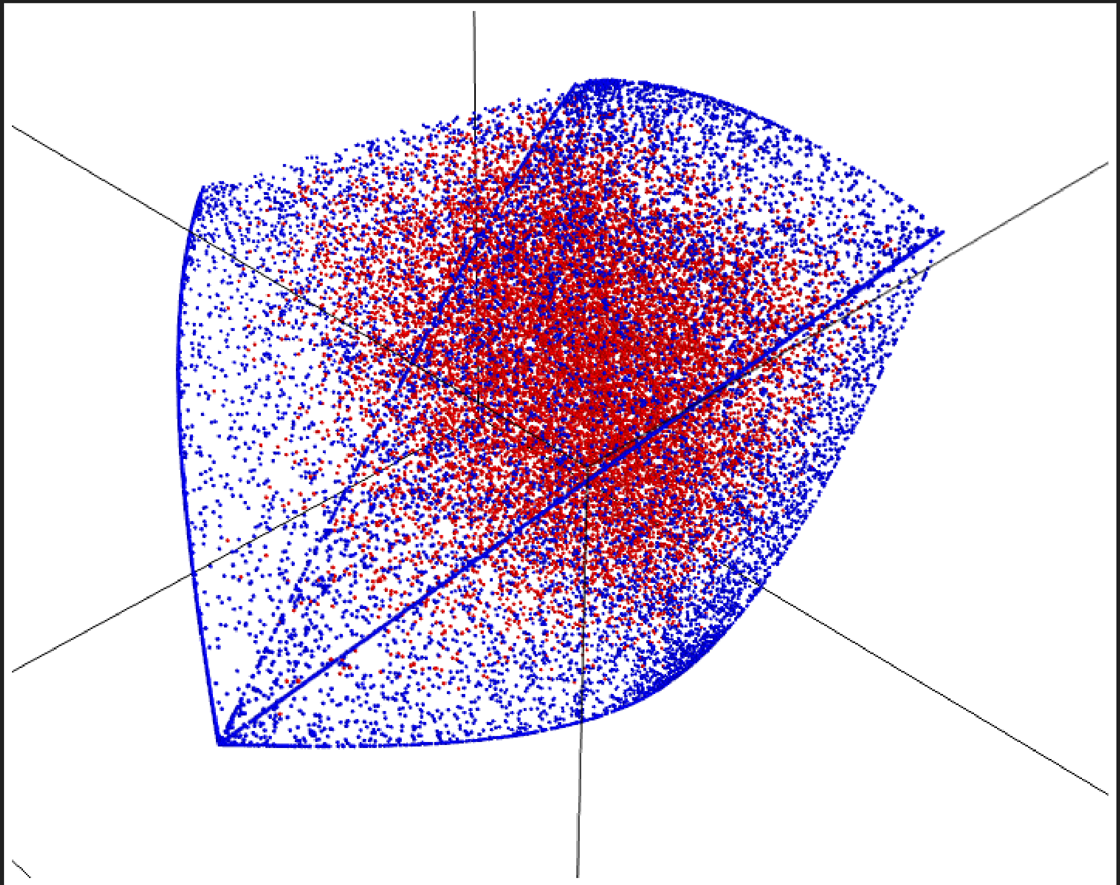
$$y_i = \frac{1}{M} \sum_j T_i(\theta_j)$$

3 Parameters

5 Observations



# ISOMERIZATION OF $\alpha$-PINENE



$$\dot{y}_1 = -(\theta_1 + \theta_2)y_1$$
$$\dot{y}_2 = \theta_1 y_1$$
$$\dot{y}_3 = \theta_2 y_1 - (\theta_3 + \theta_4)y_3 + \theta$$
$$\dot{y}_4 = \theta_3 y_3$$
$$\dot{y}_5 = \theta_4 y_3 - \theta_5 y_5$$

5 Parameters

40 Observations

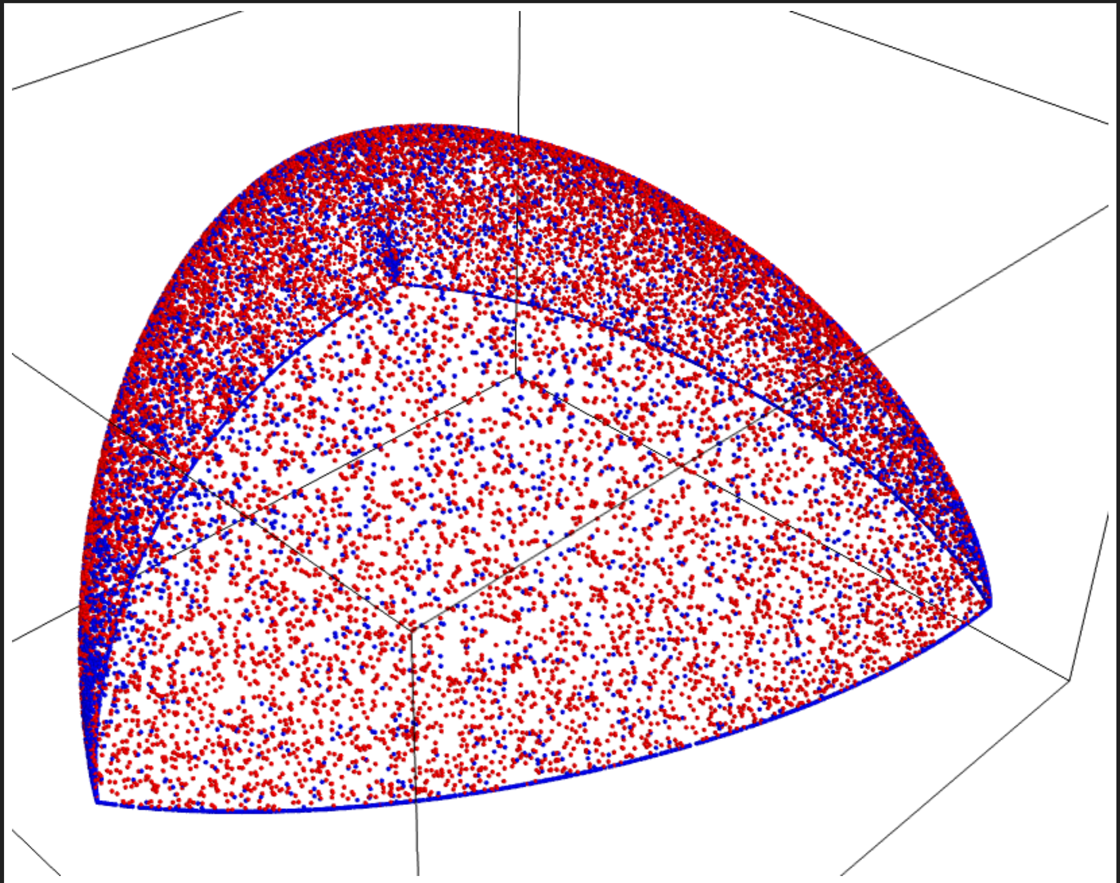
# ISING MODEL

$$\mathcal{H} = -J_1 \sum_i s_i s_{i+1} - J_2 \sum_i s_i s_{i+2}$$

$$P(s) \propto e^{-\mathcal{H}}$$

2 Parameters

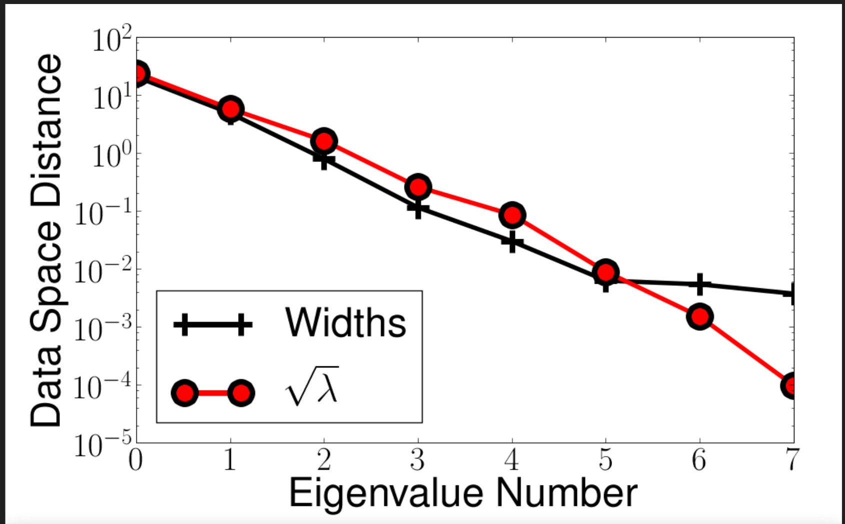
16 Observations (4 spins)

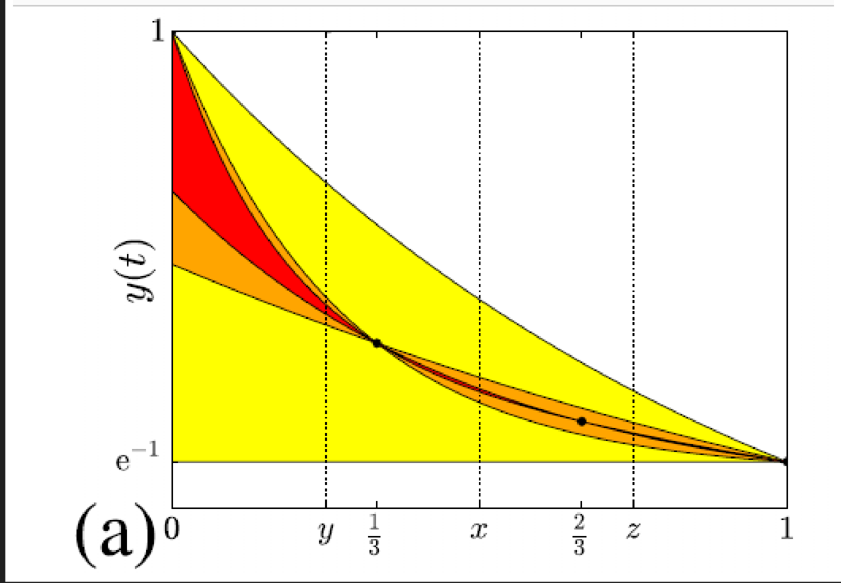


# LOW EFFECTIVE DIMENSIONALITY

# MANIFOLD WIDTHS

- "Measure" the manifold with geodesics.
- Widths are exponentially narrow: ***Hyper-ribbon***
- Interpolation
- Parameterization independent measure of sloppiness
- Widths are approximated by the FIM eigenvalues
- Intrinsic measure of the effective dimensionality of the model

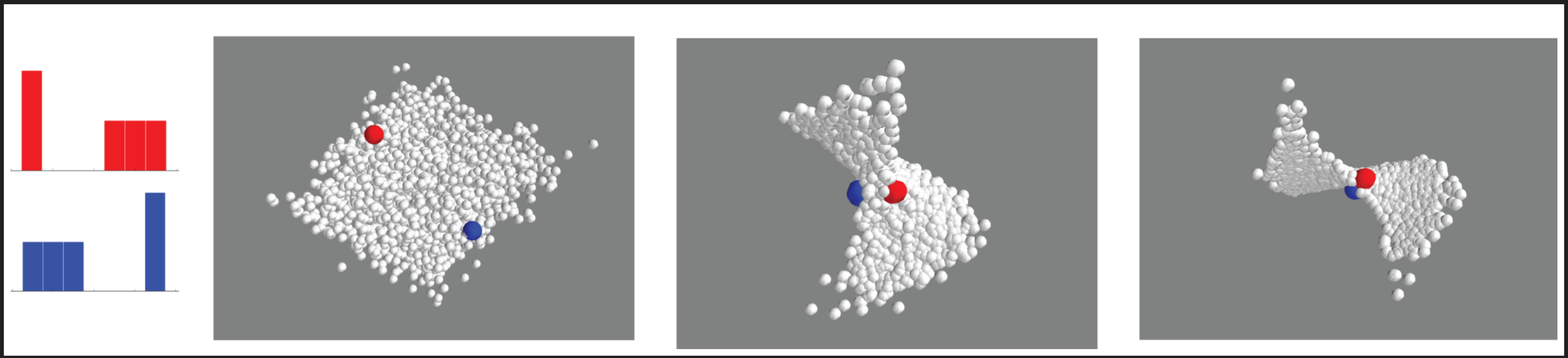



Transtrum, Mark K., et al. "Perspective: Sloppiness and emergent theories in physics, biology, and beyond." *The Journal of chemical physics* **143**.1 (2015): 07B201_1.

# DIFFUSION REVISITED

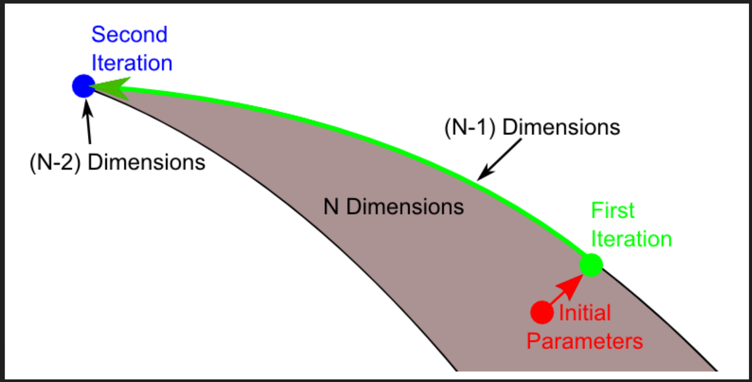
As observations are coarsened, the high-dimensional manifold is systematically compressed into something of lower effective dimensionality.

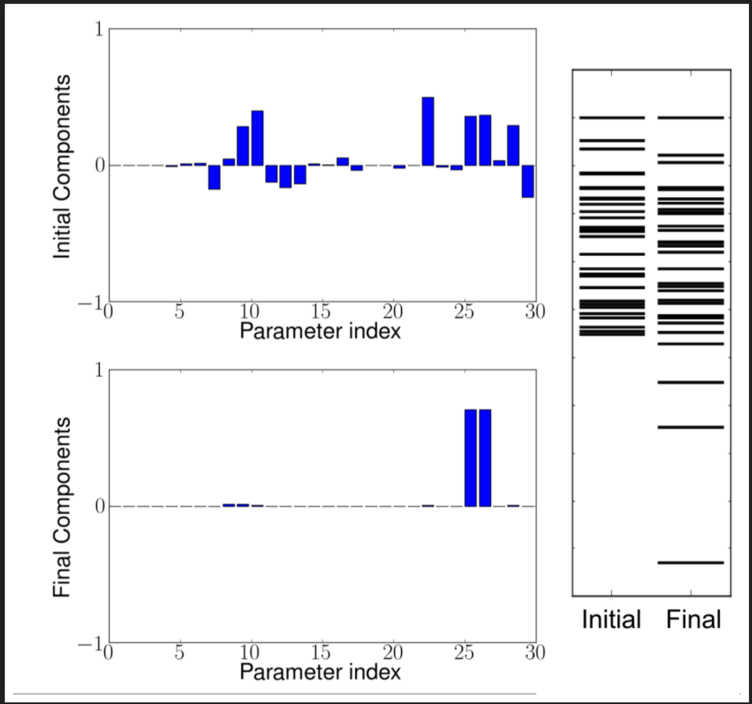


# PARAMETER REDUCTION

# MANIFOLD BOUNDARY APPROXIMATION METHOD (MBAM)

- Can we remove sloppy parameters?
- Find the emergent theory hidden in the model?
- Geodesics along sloppines direction to nearby boundary.
- Eigendirections simplify at the boundary to physically reasonable models
- Reparameterize model such that one parameter goes to zero at boundary (identifying the small parameter)




Transtrum, Mark K., and Peng Qiu. "Model reduction by manifold boundaries." *Physical review letters* **113**.9 (2014): 098701.

# MANIFOLDS (WITH BOUNDARY, CORNER)

## MANIFOLD

- A topological space that locally looks like $\mathbb{R}^N$.
- There exists an invertible map $f : \mathcal{M} \to \mathbb{R}^N$.

## MANIFOLD WITH BOUNDARY

- A topological space that locally looks like $\mathbb{R}^{N-1} \times \mathbb{R}_+$
- Like a manifold, but one coordinate is never negative.
- This coordinate is zero at the boundary

## MANIFOLD WITH CORNER

- A topological space that locally looks like $\mathbb{R}^{N-p} \times \mathbb{R}_+^p$
- Like a manifold with boundary, but $p$ coordinates are never negative.
- Each coordinate is individually zero on a boundary "cell".
- These cells meet at a corner, where all $p$ coordinates are zero.

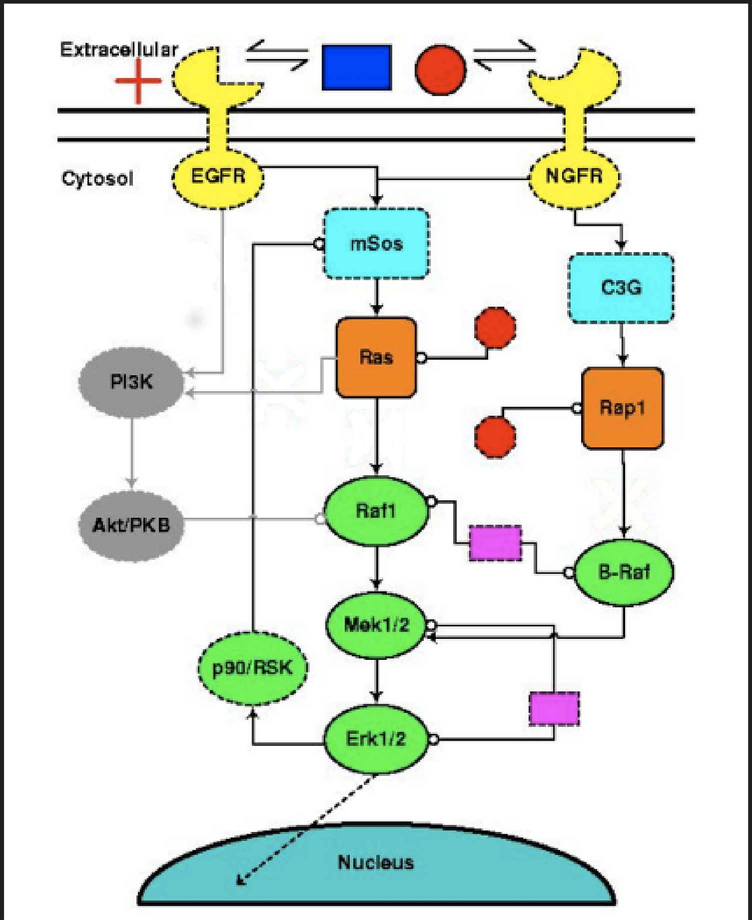
# MBAM EXAMPLE

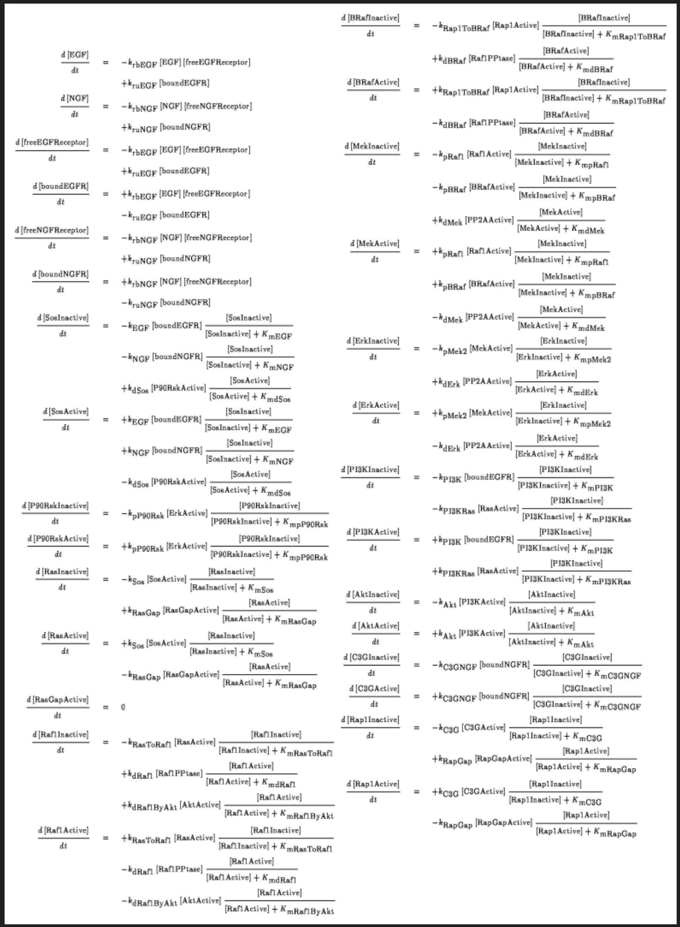
- A Michaelis-Menten reaction $[\dot{P}] = -[\dot{S}] = \frac{k[E][S]}{K+[S]}$ with parameters $k$ and $K$.
- The geodesic reveals that at the boundary both $k$ and $K$ become infinite but that the ratio $k/K$ remains finite.
- Reparameterize: $\tilde{k} = k/K, \epsilon = 1/K$ (These are boundary coordinates) So that

$$[\dot{P}] = \frac{k[E][S]}{K + [S]} = \frac{k}{K} \frac{[E][S]}{1 + [S]/K}$$
$$= \frac{\tilde{k}[E][S]}{1 + \epsilon[S]}$$

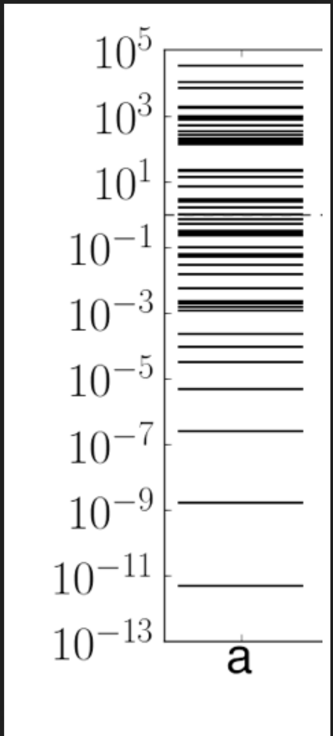
- Taking the limit $\epsilon \to 0$ gives the boundary model: $[\dot{P}] = \tilde{k}[E][S]$
- But can it be done for nontrivial cases?

# EGFR SIGNALING MODEL

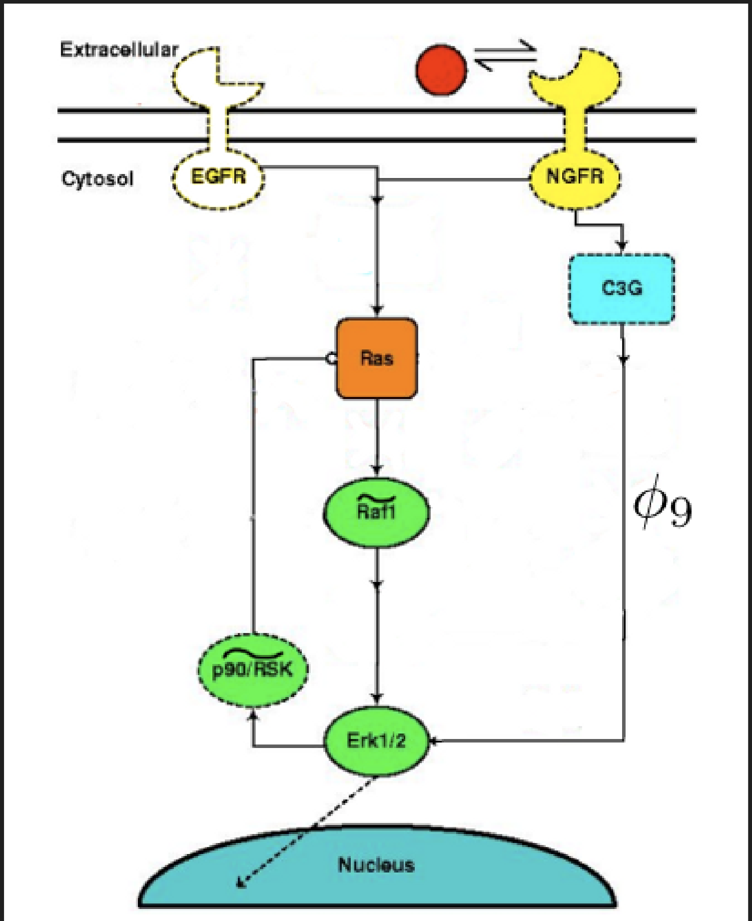


29 Coupled ODEs
48 Parameters



# REDUCED EGFR SIGNALING MODEL




$$[\mathrm{bEGFR}] = \begin{cases} 1 & \text{EGF Present} \\ 0 & \text{Otherwise} \end{cases}$$

$$\frac{d}{dt}[\mathrm{bNGFR}] = \theta_1[\mathrm{NGF}][\mathrm{fNGFR}]$$

$$\frac{d}{dt}[\mathrm{NGF}] = -\theta_1[\mathrm{NGF}][\mathrm{fNGFR}]$$

$$\frac{d}{dt}[\mathrm{RasA}] = -[\mathrm{RasA}][\widetilde{\mathrm{P90RskA}}] + \theta_2[\mathrm{bEGFR}] + \theta_3[\mathrm{bNGFR}]$$

$$\frac{d}{dt}[\widetilde{\mathrm{Raf1A}}] = \theta_4[\mathrm{RasA}] - \theta_5[\widetilde{\mathrm{Raf1A}}]/([\widetilde{\mathrm{Raf1A}}] + \theta_6)$$

$$\frac{d}{dt}[\mathrm{C3GA}] = \theta_7[\mathrm{bNGFR}][\mathrm{C3GI}]$$

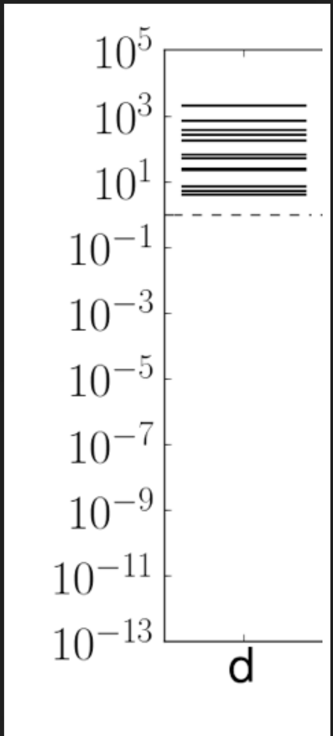
$$[\mathrm{Rap1A}] = \theta_8[\mathrm{C3GA}]$$

$$[\mathrm{MekA}] = [\widetilde{\mathrm{Raf1A}}][\mathrm{MekI}] + \theta_9[\mathrm{Rap1A}]$$

$$\frac{d}{dt}[\mathrm{Erk}] = -\theta_{10}[\mathrm{ErkA}] + \theta_{11}[\mathrm{MekA}][\mathrm{ErkI}]$$

$$\frac{d}{dt}[\widetilde{\mathrm{P90RskA}}] = \theta_{12}[\mathrm{ErkA}]$$

6 Coupled ODEs
12 Parameters




Transtrum, Mark K., and Peng Qiu. "Model reduction by manifold boundaries." *Physical review letters* **113**.9 (2014): 098701.

# INTERPRETING THE REDUCED MODEL

- Effective "Renormalized" Parameters:

$$\phi_9 = \frac{[\mathrm{BRafI}](\mathrm{kRap1ToBRaf})(\mathrm{KmdBRaf})(\mathrm{kpBRaf})(\mathrm{KmdMek})}{[\mathrm{PP2AA}][\mathrm{Raf1PPtase}](\mathrm{kdBRaf})(\mathrm{KmRap1ToBRaf})(\mathrm{kdMek})}$$

  - Interpretation: Rate of Information Flow through channel
  - Emergent Control Knob
  - No Black Box
  - Bridges the mechanistic and phenomenological descriptions
- Dynamical Variables
  - Biologically Functional Modules
- Proteins $\rightarrow$ Signaling
- Chemistry $\rightarrow$ Biology
- Not specific to systems biology

# WHERE WE HAVE APPLIED IT:

- Chemical/Biochemical kinetics (Conservation of mass)
- Compartment models (Conservation of mass)
- Power Systems Transients (Singular Perturbation)
- Stable Linear Time Invariant Systems (Balanced Truncation/Singular Perturbation)
- Composition of elementary functions (exponential, rational polynomials etc.)
- Bayesian Networks/Makov Chains/Markov Random Fields (Conservation of Probability)
- Molecular Dynamics (Conservation of Energy)
- Multilayer Perceptrons, i.e., Neural Networks
- General Exponential Families (e.g., Ising Models)
- Models with discrete symmetries (Orbifolds)
- Hodgkin-Huxley Neurons

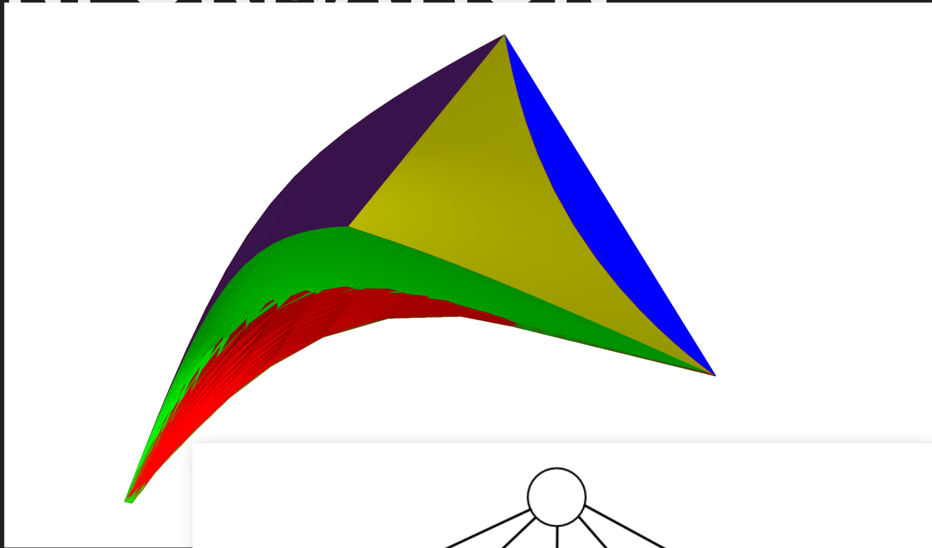
# RELATION TO OTHER REDUCTION METHODS

For an appropriately parameterized model, MBAM is equivalent to several other model reduction methods:

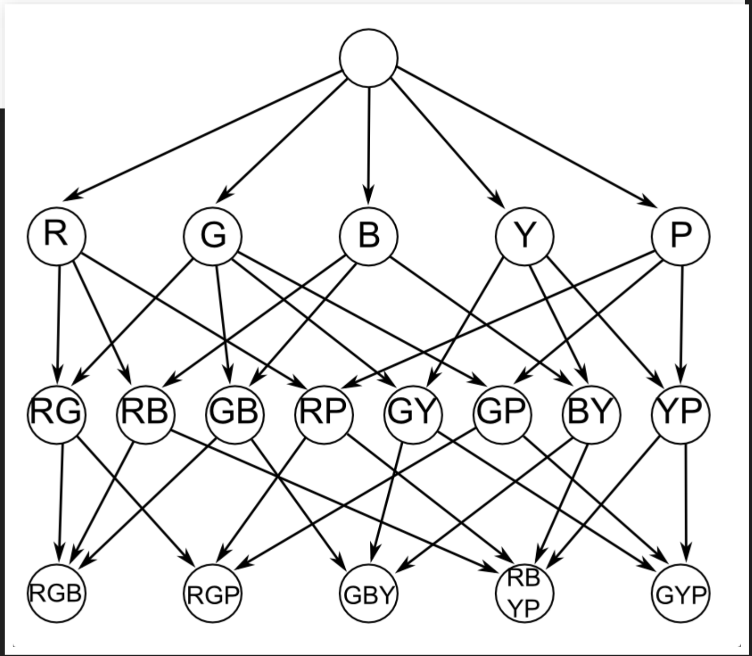
- Real-space/Momentum-space Renormalization Group
- Mean-field Approximations
- Singular Perturations
- Balanced Truncation
- Steady-state/Equilibrium Approximations
- Thermodynamic/Continuum/Classical limits

# A META-THEORY OF MODELING

# BEYOND GEOMETRY: INFORMATION TOPOLOGY



- Manifold Boundaries forms a hierarchical cell-complex
- Each boundary cell is a simplified model
- The boundary complex is a topological property, not a geometric property
- Model manifolds are realizations of abstract polytopes
- Adjacency relationships induce a partial order




Transtrum, Mark K., Gus Hart, and Peng Qiu. "Information topology identifies emergent model classes." arXiv preprint arXiv:1409.6203 (2014).

# EXAMPLE: ENZYME REACTION

$$E + S \leftrightharpoons C \rightarrow E + P$$

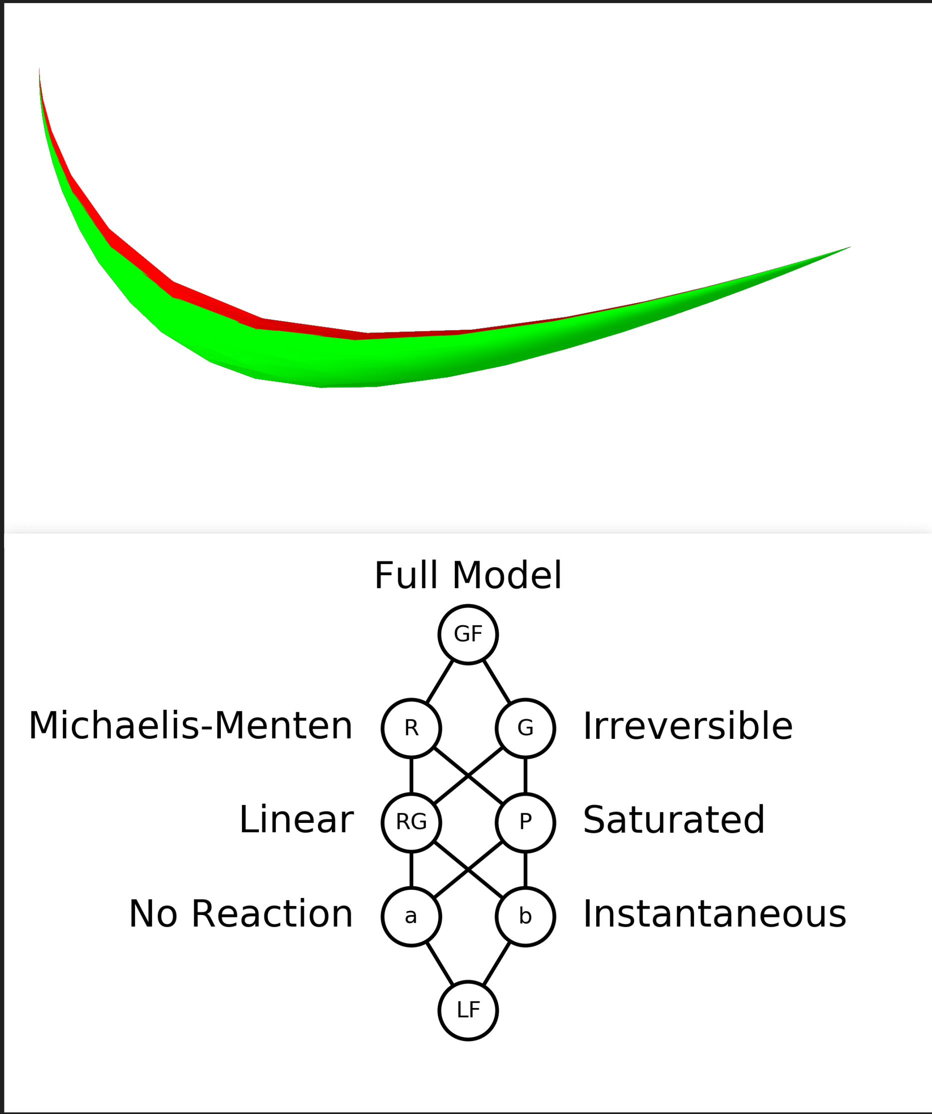
3 Parameters $k_f, k_r, k_c$

$$\frac{d}{dt}[E] = -k_f[E][S] + k_r[C] + k_c[C]$$

$$\frac{d}{dt}[S] = -k_f[E][S] + k_r[C]$$

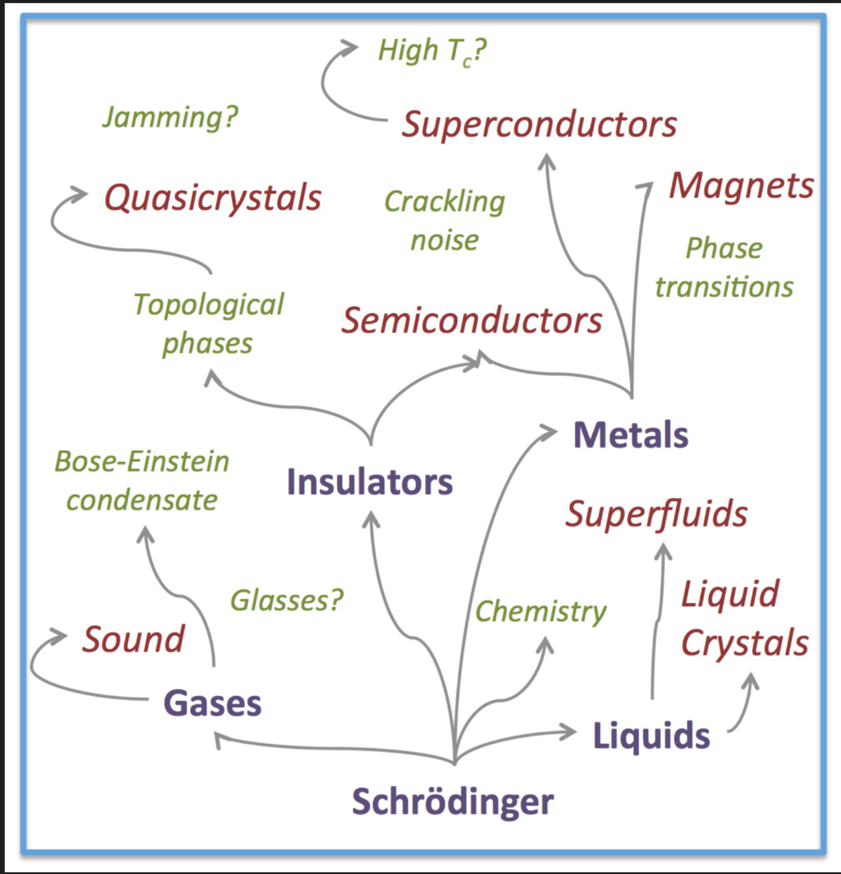
$$\frac{d}{dt}[C] = k_f[E][S] - k_r[C] - k_c[C]$$

$$\frac{d}{dt}[P] = k_c[C]$$



# HASSE DIAGRAMS AND APPROXIMATE MODELS

- Each boundary cell captures a different idealized behavior of the system.
- Depending on the context, these idealized behaviors could represent:
  - "Healthy" vs. "Sick"
  - Phases of matter
  - Fast/Slow Dynamics
  - Developmental Stages?




The Hasse diagram is a road map from the intricate, fully-parameterized description through various types of approximations to distinct behavioral regimes.

# THOUGHTS ABOUT UQ

- Parametric Uncertainty
  - Models with lots of unknown parameters don't always lead to large prediction uncertainty or overfitting
  - The goal is to identify the relevant degrees of freedom (e.g., Cooper pairs)
  - There is not always a sharp distinction between relevant/irrelevant.
- Structural Uncertainty:
  - Consider an effective model to be a submanifold of a more "complete" model.
  - Structural uncertainty is bound by the widths of the complete model manifold.
  - Sloppy models have a exponential distribution of widths.
- Model selection problem:
  - Accuracy vs. Complexity
  - Model Reduction

# SUMMARY AND CONCLUSIONS

- Parameter Reduction & Effective Theories
  - Two Cultures of Mathematical Modeling
  - Fisher Information reflects low-effective dimensionality
  - Physically insightful models often emerge because of a "small parameter"
- Information Geometry
  - Models are equivalent to manifolds
  - Hierarchy of widths
  - Complex Models have low effective dimensionality
  - Parameter Reduction = Manifold Hyper-corners (MBAM)
- Systematically derive new, effective theories from first principles
  - Information Topology: A meta theory for organizing our information about the world

By characterizing the geometric and topological structures underlying scientific models, we can connect bottom-up descriptions of complex processes with top-down inferences drawn from data, paving the way for emergent theories in physics, biology, and beyond.

# ACKNOWLEDGEMENTS

## COLLABORATORS

- James Sethna (Cornell)
- Peng Qiu (Georgia Tech)
- Chris Myers (Cornell)
- Kevin Brown (Connecticut)
- Ben Machta (Princeton, Yale)
- Bryan Daniels (Arizona State)

## FUNDING SOURCES

NSF

U.S.ARMY

Center for BRIGHT BEAMS
A National Science Foundation Science & Technology Center

# REFERENCES


- Brown, Kevin S., et al. "The statistical mechanics of complex signaling networks: nerve growth factor signaling." *Physical biology* **1**.3 (2004): 184.
- Transtrum, Mark K., Benjamin B. Machta, and James P. Sethna. "Why are nonlinear fits to data so challenging?." *Physical review letters* **104**.6 (2010): 060201.
- Chachra, Ricky, Mark K. Transtrum, and James P. Sethna. "Structural susceptibility and separation of time scales in the van der Pol oscillator." *Physical Review E* **86**.2 (2012): 026712.
- Machta, Benjamin B., et al. "Parameter space compression underlies emergent theories and predictive models." *Science* **342**.6158 (2013): 604-607.
- Transtrum, Mark K., and Peng Qiu. "Model reduction by manifold boundaries." *Physical review letters* **113**.9 (2014): 098701.
- Transtrum, Mark K., Gus Hart, and Peng Qiu. "Information topology identifies emergent model classes." arXiv preprint arXiv:1409.6203 (2014).
- Transtrum, Mark K., et al. "Perspective: Sloppiness and emergent theories in physics, biology, and beyond." *The Journal of chemical physics* **143**.1 (2015): 07B201_1.
- Transtrum, Mark K. "Manifold boundaries give" gray-box" approximations of complex models." arXiv preprint arXiv:1605.08705 (2016).